nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-12958-0 OPEN

# Multi-ancestry sleep-by-SNP interaction analysis in 126,926 individuals reveals lipid loci stratified by sleep duration

Raymond Noordam et al.#

Both short and long sleep are associated with an adverse lipid profile, likely through different biological pathways. To elucidate the biology of sleep-associated adverse lipid profile, we conduct multi-ancestry genome-wide sleep-SNP interaction analyses on three lipid traits (HDL-c, LDL-c and triglycerides). In the total study sample (discovery + replication) of 126,926 individuals from 5 different ancestry groups, when considering either long or short total sleep time interactions in joint analyses, we identify 49 previously unreported lipid loci, and 10 additional previously unreported lipid loci in a restricted sample of European-ancestry cohorts. In addition, we identify new gene-sleep interactions for known lipid loci such as *LPL* and *PCSK9*. The previously unreported lipid loci have a modest explained variance in lipid levels: most notable, gene-short-sleep interactions explain 4.25% of the variance in triglyceride level. Collectively, these findings contribute to our understanding of the biological mechanisms involved in sleep-associated adverse lipid profiles.

#A full list of authors and their affiliations appears at the end of the paper.

Dyslipidemia is defined as abnormalities in one or more types of lipids, such as high blood LDL-cholesterol (LDL-c) and triglyceride (TG) concentrations and a low HDL-cholesterol (HDL-c) concentration. High LDL-c and TG are well-established modifiable causal risk factors for cardiovascular disease[1–3], and therefore are a primary focus for preventive and therapeutic interventions. Over 300 genetic loci are identified to be associated with blood lipid concentrations[4–10]. Recent studies showed that only 12.3% of the total variance in lipid concentration is explained by common single-nucleotide polymorphisms (SNPs), suggesting additional lipid loci could be uncovered[10]. Some of the unexplained heritability may be due to the presence of gene–environment and gene–gene interactions. Recently, high levels of physical activity were shown to modify the effects of four genetic loci on lipid levels[11], an additional 18 previously unreported lipid loci were identified when considering interactions with high alcohol consumption[12], and 13 previously unreported lipid loci were identified when considering interaction with smoking status[13], suggesting that behavioural factors may interact with genetic loci to influence lipid levels.

Sleep is increasingly recognised as a fundamental behaviour that influences a wide range of physiological processes[14]. A large volume of epidemiological research implicates disturbed sleep in the pathogenesis of atherosclerosis[15], and specifically, both a long and short sleep duration are associated with an adverse blood lipid profile[16–26]. However, it is unknown whether sleep duration modifies genetic risk factors for adverse blood lipid profiles. We hypothesise that short and long habitual sleep duration may modify genetic associations with blood lipid levels. The identification of SNPs involved in such interactions will facilitate our understanding of the biological background of sleep-associated adverse lipid profiles.

We investigate gene–sleep duration interaction effects on blood lipid levels as part of the Gene-Lifestyle Interactions Working Group within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium[27,28]. To permit the detection of both such sleep-duration–SNP interactions and lipid–SNP associations accounting for total sleep duration, a two degree of freedom (2df) test that jointly tests the SNP-main and SNP-interaction effect was applied[29]. Given that there are differences among ancestry groups in sleep behaviours and lipid levels, analysis of data from cohorts of varying ancestries facilitate the discovery of robust interactions between genetic loci and sleep traits. We focus on short total sleep time (STST; defined as the lower 20% of age- and sex-adjusted sleep duration residuals) and long total sleep time (LTST; defined as the upper 20% of age- and sex-adjusted sleep duration residuals) as exposures compared with the remaining individuals in the study population, given that each extreme sleep trait are associated with multiple adverse metabolic and health outcomes[15–26,30–34]. Within this study, we report multi-ancestry sleep-by-SNP interaction analyses for blood lipid levels that successfully identified several previously unreported loci for blood lipid traits.

## Results

**Study population**. Discovery analyses were performed in up to 62,457 individuals (40,041 European-ancestry, 14,908 African-ancestry, 4460 Hispanic-ancestry, 2379 Asian-ancestry and 669 Brazilian/mixed-ancestry individuals) from 21 studies spanning five different ancestry groups (Supplementary Tables 1 and 2; Supplementary Data 1). Of the total discovery analysis, 13,046 (20.9%) individuals were classified as short sleepers and 12,317 (19.7%) individuals as long sleepers. Replication analyses were performed in up to 64,469 individuals (47,612 European-ancestry, 12,578 Hispanic-ancestry, 3133 Asian-ancestry and 1146 African-ancestry individuals) from 19 studies spanning four different ancestry groups (Supplementary Tables 3 and 4; Supplementary Data 2). Of the total replication analysis, 12,952 (20.1%) individuals were classified as short sleepers and 12,834 (19.9%) individuals as long sleepers.

**Genome-wide SNP–sleep interaction analyses**. An overview of the multi-ancestry analyses process for both STST and LTST is presented in Fig. 1. QQ plots of the combined multi-ancestry and European meta-analysis of the discovery and replication analysis are presented in Supplementary Figs. 1 and 2. Lambda values ranged between 1.023 and 1.055 (trans-ancestry meta-analysis) before the second genomic control and were all 1 after second genomic control correction. In the combined discovery and replication meta-analyses comprising all contributing ancestry groups, we found that many SNPs replicated for the lipid traits ($P_{joint}$ in replication < 0.05 with similar directions of effect as in the discovery analyses and $P_{joint}$ in combined discovery and replication analysis $< 5 \times 10^{-8}$). Notably, we replicated 2395 and 2576 SNPs for HDL-c, 2012 and 2074 SNPs for LDL-c, and 2643 and 2734 SNPs for TG in the joint model with LTST and STST, respectively.

Most of the replicated SNPs were mapped to known loci (Supplementary Data 3 and 4). We looked at the 427 known lipid SNPs (Supplementary Data 5), but these did not reveal significant 1df interactions with either LTST or STST. In addition, we identified lead SNPs mapping to previously unreported regions when considering the joint model with potential interaction for either STST or LTST (>1 Mb distance from known locus). Ultimately, in the multi-ancestry analysis, we identified 14 previously unreported loci for HDL-c, 12 for LDL-c and 23 ci for TG ($R^2 < 0.1$; Fig. 2). Of these, seven loci for HDL-c, four loci for LDL-c and seven loci for TG were identified after considering an interaction with LTST (Supplementary Data 6). Furthermore, 7 loci for HDL-c, 8 loci for LDL-c and 16 loci for TG were identified when considering an interaction with STST (Supplementary Data 7). Importantly, none of these loci for the three lipid traits identified through LTST were identified in the analyses with STST, and vice versa. Furthermore, these lipid loci were specific to a single-lipid trait. Regional plots of the previously unreported loci from the multi-ancestry analyses are presented in Supplementary Figs. 3–8. Some of the previously unreported SNPs identified through modelling a short or long sleep duration interaction (1df) also showed suggestive evidence of association with lipid levels in the joint model (2df interaction test). However, this pattern suggested a main effect that appeared once sleep duration was adjusted for rather than an effect due to an interaction between sleep and the SNP (Supplementary Data 6, 7).

Using the R-based VarExp package[35], we calculated the explained variance based on the summary statistics of the combined discovery and replication analysis. Collectively, previously unreported lead lipid SNPs identified with LTST explained 0.97% of the total HDL-c variation, 0.13% of the total LDL-c variation and 1.51% of the total TG variation. In addition, the previously unreported SNPs identified with STST explained 1.00% of the total HDL-c variation, 0.38% of the total LDL-c variation and 4.25% of the total TG variation.

In the analyses restricted to European-ancestry individuals (overview Supplementary Fig. 9), we identified ten additional previously unreported loci (seven with LTST and three with STST; Supplementary Fig. 10), which were not identified in the multi-ancestry analyses. Of these, we identified four loci for HDL-c, two loci for LDL-c and one locus for TG with LTST (Supplementary Data 8). In addition, we identified one locus for HDL-c and two for TG with STST (Supplementary Data 9).

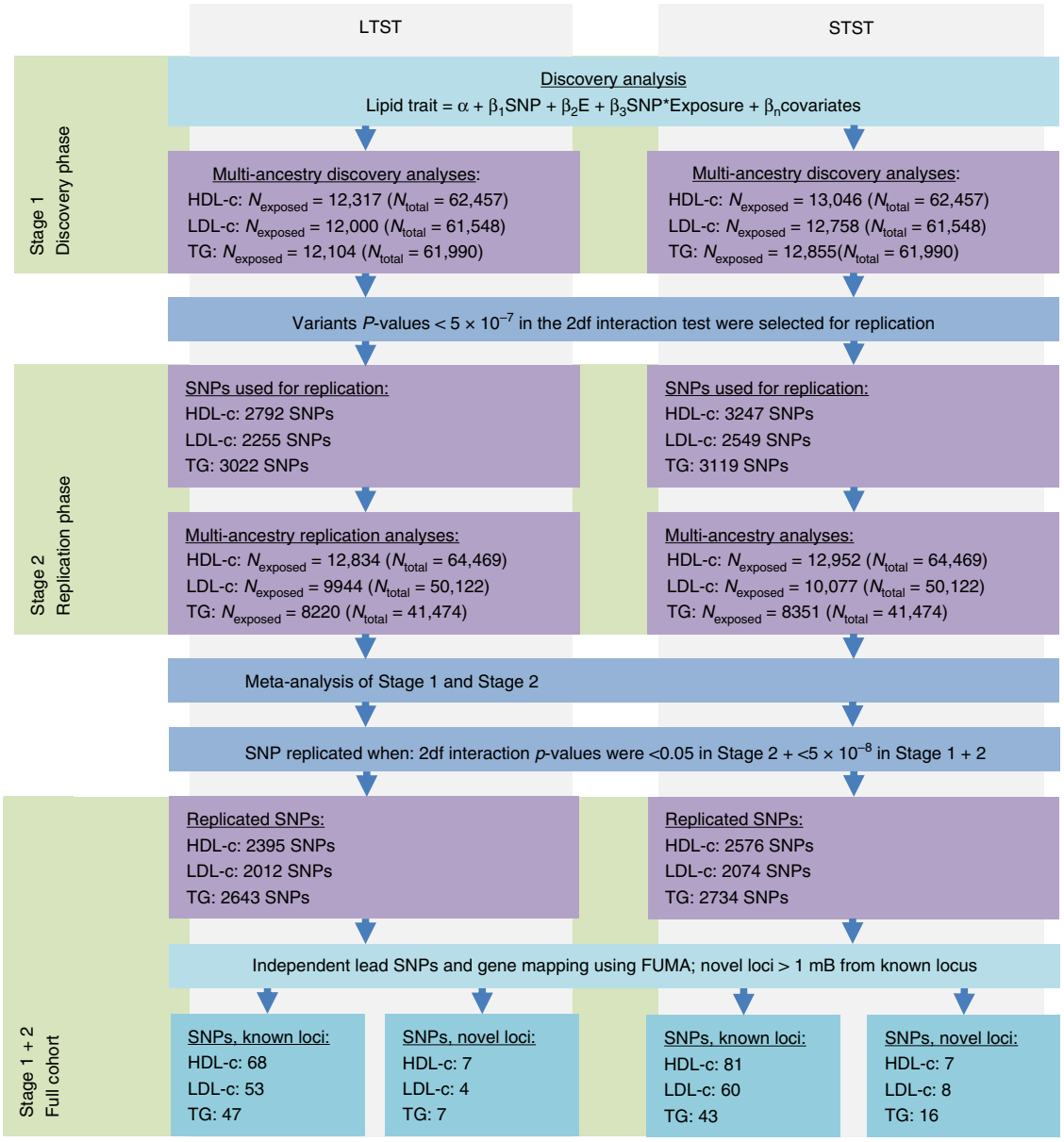



**Fig. 1** Project overview and SNP selection in the multi-ancestry analyses. Project overview of the multi-ancestry analyses of how the new lipid loci were identified in the present project. Replicated variants had to have 2df interaction test *P*-values of Stage 1 < $5 \times 10^{-7}$, Stage 2 < 0.05 with a similar direction of effect as in the discovery meta-analysis, and Stage 1 + 2 < $5 \times 10^{-8}$

Again, we observed no overlapping findings between the two sleep exposures and the three lipid traits. Regional plots of the previously unreported loci were presented in Supplementary Figs. 11–15.

**Gene mapping of known and previously unreported loci**. Based on a total of 402 lead SNPs in known and previously unreported regions for both exposures and the three lipid traits that were identified using the joint test in the combined sample of discovery and replication studies, we subsequently explored the extent the effects were driven by 1df interaction with the sleep exposure trait being tested[29]. We corrected the 1df interaction *P*-value for multiple testing using the false discovery rate[36] considering all 402 lead SNPs for the present investigation, which was equivalent in our study to a 1df interaction *P*-value < $5 \times 10^{-4}$. Overall, in the multi-ancestry meta-analyses, the previously unreported lipid loci show clearly stronger interaction with either LTST or STST than the loci defined as known (Fig. 3). The majority of these identified lead variants were generally common, with minor allele frequencies (MAF) mostly > 0.2, and SNP × sleep interaction effects were not specifically identified in lower frequency SNPs (e.g., MAF < 0.05).

Out of the seven previously unreported HDL-c loci identified in the joint model with LTST, six had a 1df interaction *P*-value$_{FDR}$ < 0.05, notably lead SNPs mapped to *ATP6V1H*, *ARTN2*, *ATP6V0A4*, *KIAA0195*, *MIR331* and *MIR4280*. Based on exposure-stratified analyses in the meta-analysis of the discovery cohorts, we further explored the effect sizes per exposure group. The lead SNPs that showed significant sleep × SNP interaction also showed effect estimates that modestly differed between LTST exposure groups (Supplementary Data 10). Interestingly, two lead SNPs near known HDL-c loci showed a 1df interaction *P*-value$_{FDR}$ < 0.05, including SNPs near *CETP* and *LIPC* (Supplementary Data 4). Out of the seven previously unreported HDL-c loci identified in the joint model with STST, we found six loci with a 1df interaction *P*-value$_{FDR}$ < 0.05, notably

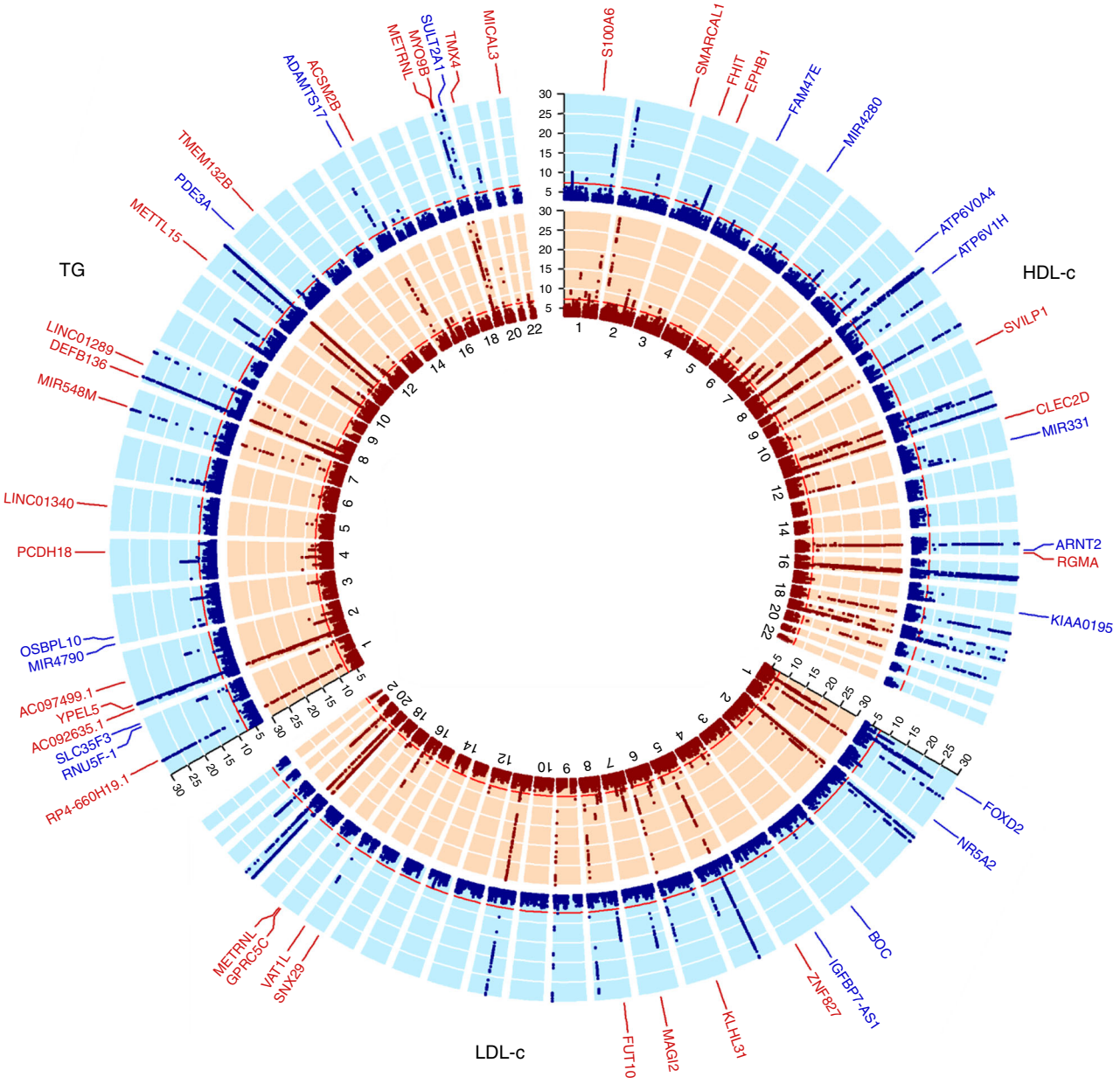



**Fig. 2** log(*P*-value of 2df interaction analyses) plots of the multi-ancestry analyses. Plot visualises the −log(*P*-values in the 2df interaction test) for HDL-c, LDL-c and TG per chromosome. In red (inner circle) are the −log(*P*-value) plots for the analyses taking into account potential interaction with short total sleep time. In blue (outer circle) are the −log(*P*-value plots for the analyses taking into account potential interaction with long total sleep time. Loci defined as novel and replicated are labelled. Replicated variants had to have 2df interaction test *P*-values of Stage $1 < 5 \times 10^{-7}$, Stage 2 < 0.05 with a similar direction of effect as in the discovery meta-analysis and Stage $1 + 2 < 5 \times 10^{-8}$. Labelled gene names in red were identified in the STST analysis; labelled gene names in blue were identified in the LTST analysis. All −log(*P*-value in the 2df interaction test) > 30 were truncated to 30 for visualisation purposes only. The unlabelled regions with $P < 5 \times 10^{-8}$ in the 2df interaction test were in known loci. Figure prepared using the R package circlize[104]

lead SNPs mapped to *S1000A6*, *SMARCAL1, RGMA*, *EPHB1*, *FHIT* and *CLEC2D*. Again, their effect estimates differed between the exposure groups in the discovery multi-ancestry meta-analysis (Supplementary Data 11; Fig. 4). Some lead SNPs near known HDL-c loci showed evidence of a 1df interaction with STST (e.g., *MADD* and *LPL; P*-value$_{FDR} < 0.05$).

For all four lead SNPs in previously unreported regions associated with LDL-c when considering LTST, we observed a 1df interaction *P*-value$_{FDR} < 0.05$; notably, lead SNPs mapped to *IGFBP7-AS1*, *FOXD2*, *NR5A2* and *BOC*. One locus that mapped within a 1 Mb physical distance from known LDL-c locus (*PCSK9*) showed 1df interaction with LTST (Supplementary Data 4). Similarly, all eight lead SNPs in previously unreported regions associated with LDL-c when considering STST, had a 1df interaction *P*-value$_{FDR} < 0.05$; notably, lead SNPs mapped to *MAGI2*, *METRNL*, *VAT1L*, *FUT10*, *SNX29*, *ZNF827*, *GPRC5C* and *KLHL31*. In addition, of the known LDL-c loci, lead SNPs mapped within a physical distance of 1 Mb of *APOB* and *SLC22A1* showed a 1df interaction *P*-value$_{FDR} < 0.05$ (Supplementary Data 5). For both analyses, we observed that effect estimates differed between the LTST and STST exposure groups in the multi-ancestry discovery analysis (Supplementary Data 10 and 11; Fig. 4).

All seven SNPs in previously unreported regions associated with TG when considering LTST, had a 1df interaction *P*-value$_{FDR} < 0.05$; notably, lead SNPs mapped to *RNU5F*-1,

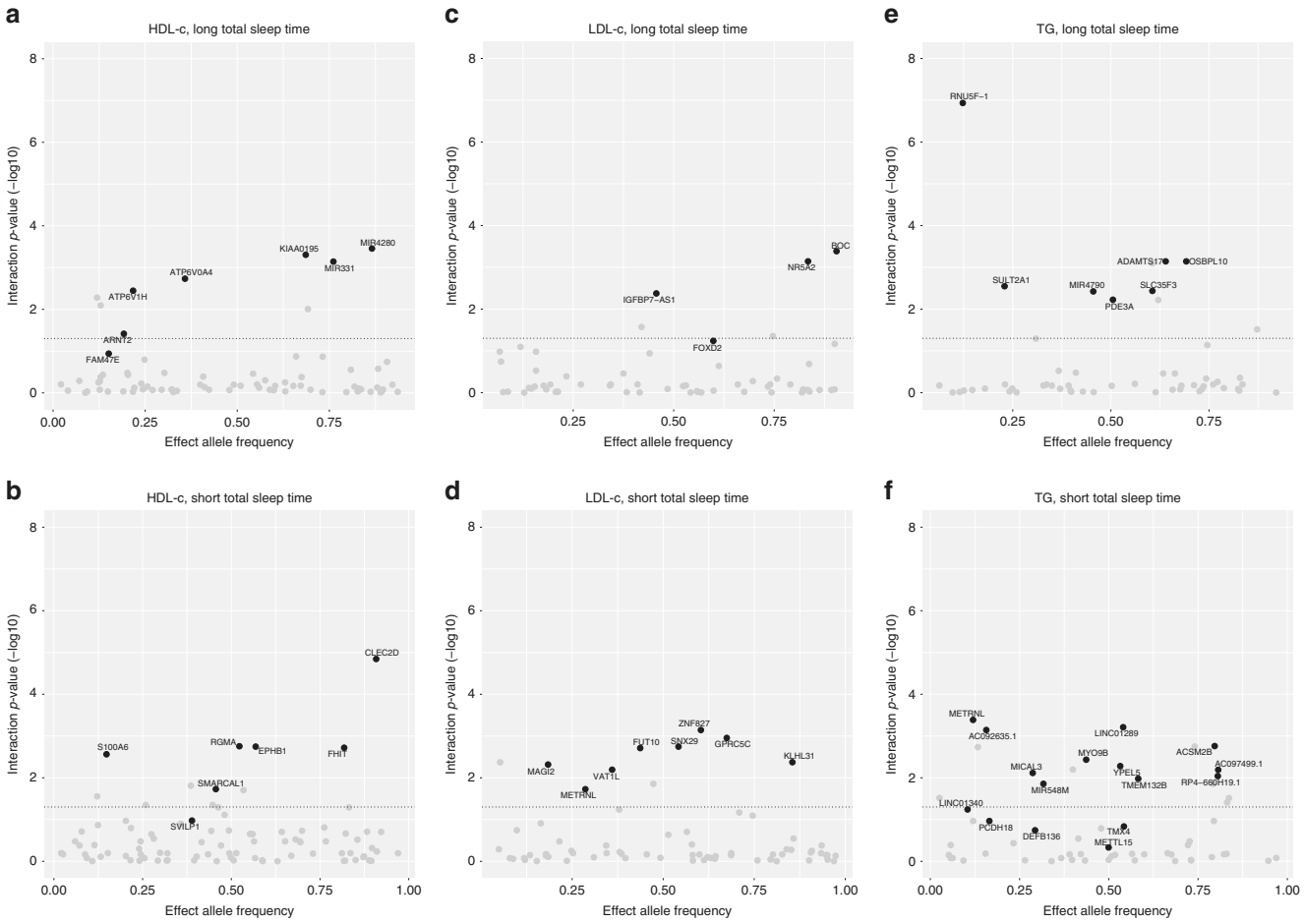



**Fig. 3** Sleep-interactions in known and previously unreported regions. Plot displaying the -log(*P*-value) of the 1df interaction between the SNP and either LTST or STST on the lipid trait after correction for multiple testing using false discovery rate against the allele frequency of the effect allele. Dotted horizontal line resembles the cut-off for the 1df interaction $P\text{-value}_{FDR} < 0.05$ after correction for multiple testing using false discovery rate. In black are the novel loci for lipid traits; in grey are the identified lead SNPs mapped within a 1 -Mb physical distance from a known lipid locus. Visualisation of the plots was performed using the R package ggplot2[105]. **a** HDL-c, long total sleep time; **b** HDL-c, short total sleep time; **c** LDL-c, long total sleep time; **d** LDL-c, short total sleep time; **e** triglycerides, long total sleep time; **f** triglycerides, short total sleep time

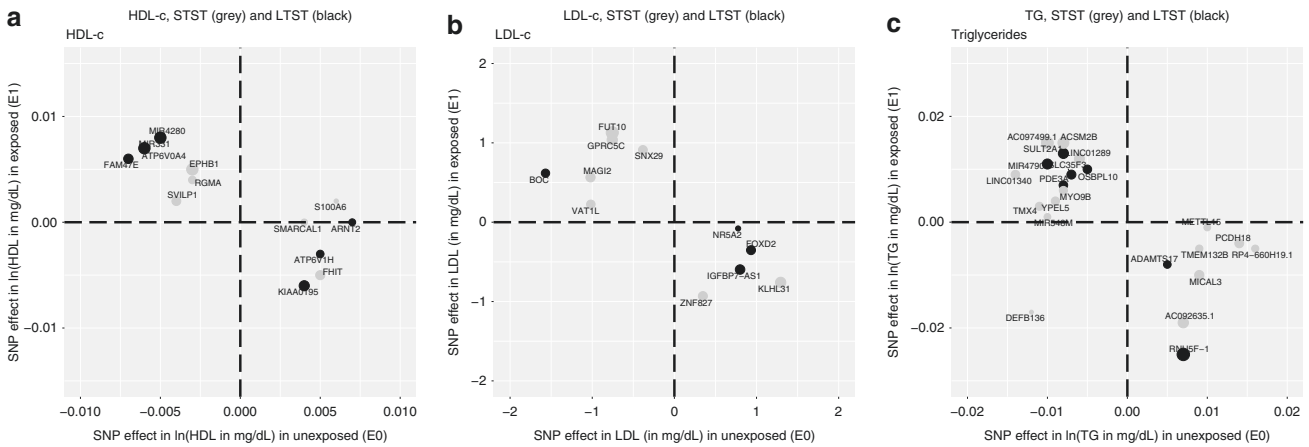



**Fig. 4** Comparison of SNP-main effects stratified by exposure. *X*-axis displays the effect sizes of the novel lead SNPs as observed in the meta-analyses of the unexposed individuals (LTST = '0', STST = '0'). *Y*-axis displays the effect sizes of the novel lead SNPs as observed in the meta-analyses of the exposed individuals (LTST = '1', STST = '1'). In black are the novel lead SNPs identified with LTST; in grey are the novel lead SNPs identified with STST. Sizes of the dots were weighted to the difference observed between exposed and unexposed. Visualisation of the plots was performed using the R package ggplot2[105]. **a** HDL-c; **b** LDL-c; **c** triglycerides

*SULT2A1*, *MIR4790*, *PDE3A*, *SLC35F3*, *ADAMTS17* and *OSBPL10*. In addition, we found some evidence for long sleep–SNP interaction in lead SNPs near known TG loci, including lead SNPs near *AKR1C4* and *NAT2* (Supplementary Data 4). Of the 16 lead SNPs in previously unreported regions associated with TG when considering STST, we observed 12 lead SNPs with a 1df interaction $P$-value $< 5 \times 10^{-4}$ ($P$-value$_{FDR} <$ 0.05), including lead SNPs mapped to *LINC0140*, *METRNL*, *AC092635.1*, *MICAL3*, *MIR548M*, *MYO9B*, *YPEL5*, *LINC01289*, *TMEM132B*, *ACSM2B*, *AC097499.1* and *RP4–660H19.1*. In addition, we observed some lead SNPs within 1 Mb physical distance from known TG loci, such as *MMP3* and *NECTIN2* (Supplementary Data 5). For both LTST and STST analyses, we again observed differing effects dependent on the exposure group in the discovery meta-analyses (Supplementary Data 10 and 11; Fig. 4).

**Look-ups and bioinformatics analyses**. Based on the lead SNPs mapped to the previously unreported loci, we conducted a look-up in GWAS summary statistics data on different questionnaire-based sleep phenotypes from up to 337,074 European-ancestry individuals of the UK Biobank (Supplementary Data 12). We only observed the TG-identified rs7924896 (*METTL15*) to be associated with snoring ($P$-value $= 1e^{-5}$) after correction for a total of 343 explored SNP–sleep associations (seven sleep phenotypes × 49 genes; ten SNPs were unavailable; threshold for significance = $1.46e^{-4}$). Furthermore, we did not observe that any of these identified SNPs was associated with accelerometer-based sleep traits (Supplementary Data 13). In general, we did not find substantial evidence that the identified lead SNPs in previously unreported regions were associated with coronary artery disease in the CARGIoGRAMplusC4D consortium (Supplementary Table 5).

Identified lipid loci from previously unreported regions were further explored in the GWAS catalogue (Supplementary Data 14). Several of the mapped genes of these lead SNPs have previously been identified with multiple other traits, such as body mass index (*FHIT*, *KLH31*, *ADAMTS17*, and *MAGI2*), mental health (*FHIT* [autism/schizophrenia, depression], *SNX13* [cognition]), gamma-glutamyltransferase (*ZNF827*, *MICAL3*), and inflammatory processes (*ZNF827*, *NR5A2*).

We additionally investigated differential expression of these lead SNPs using data from multiple tissues from the GTEx consortium[37,38] (Supplementary Data 15). Lead SNPs were frequently associated with mRNA expression levels of the mapped gene and with trans-eQTLs. For example, rs429921 (mapped to *VAT1L*) was associated with differential mRNA expression levels of *CLEC3A* and *WWOX*, which are located more upstream on chromosome 16 (Supplementary Fig. 6). rs3826692 (mapped to *MYO9B*) was specifically associated with differential expression of the nearby *USE1* gene. Identified SNPs were frequently associated with differential expression in the arteries. For example, rs6501801 (*KIAA0195*) was associated with differential expression in arteries at different locations. Several of the other identified SNPs showed differential expression in multiple tissues, including the gastrointestinal tract, (subcutaneous/visceral) adipose tissue, brain, heart, muscle, lung, liver, nervous system, skin, spleen, testis, thyroid and whole blood.

## Discussion

We investigated SNP–sleep interactions in a large, multi-ancestry, meta-analysis of blood lipid levels. Given the growing evidence that sleep influences metabolism[39–44], at least in part through effects on gene expression, we hypothesised that short/long habitual sleep duration may modify the effects of genetic loci on lipid levels. In a total study population of 126,926 individuals from five different ancestry groups, we identified 49 loci previously unreported in relation to lipid traits when considering either long or short total sleep time in the analyses. An additional ten previously unreported lipid loci were identified in analyses in Europeans only. Of these identified loci, most loci at least in part were driven by differing effects in short/long sleepers compared with the rest of the study population. Multiple of the genes identified from previously unreported regions for lipid traits have been previously identified in relation to adiposity, hepatic function, inflammation or psychosocial traits, collectively contributing to potential biological mechanisms involved in sleep-associated adverse lipid profile.

In addition to the over 300 genetic loci that already have been identified in relation to blood lipid concentrations in different efforts[4–10], we identified 49 additional loci associated with either HDL-c, LDL-c or TG in our multi-ancestry analysis. While for some of the SNPs had no neighbouring SNPs in high LD (e.g., rs7799249; mapped to *ATP6V0A4*), our applied filters (e.g., imputation quality > 0.5) would suggest that the chance of invalidity of the findings is negligible. Furthermore, in the case of rs7799249, no SNPs in high LD are known in individuals from different ancestries[45]. Considering the previously unreported TG loci identified by considering interactions with total sleep duration explain an additional 4.25% and 1.51% of the total variation in TG concentrations, for STST and LTST, respectively. Whilst the additionally explained variance for LDL-c (0.38% and 0.13%) and HDL-c (1.00% and 0.97%) was low/modest, the lead SNPs from previously unreported regions for LDL-c levels map to genes that are known to be associated with adiposity, inflammatory disorders, cognition, and liver function, thus identifying pathways by which sleep disturbances may influence lipid biology.

Across multiple populations, both short and long sleep duration have been associated with cardiovascular disease and diabetes[46]. There are numerous likely mechanisms for these associations. Experimental sleep loss results in inflammation, cellular stress in brain and peripheral tissues, and altered expression of genes associated with oxidative stress[47,48]. The impact of long sleep on metabolism is less well understood than the effect of short sleep, and multiple of the associations seem to overlap with short sleep as well. Long sleep duration is associated with decreased energy expenditure, increased sedentary time, depressed mood and obesity-related factors associated with inflammation and a pro-thrombotic state[49], as well as with higher C-reactive protein and interleukin-6 concentrations[50]. However, studies that adjusted for multiple confounders, including obesity, depression and physical activity, showed that long sleep remained a significant predictor of adverse cardiovascular outcomes[46,51]. Therefore, the adverse effects of long sleep also may partly reflect altered sleep–wake rhythms and chronodisruption resulting from misalignment between the internal biological clock with timing of sleep and other behaviours that track with sleep, such as timing of food intake, activity and light exposure[52]. Altered sleep–wake and circadian rhythms influence glucocorticoid signalling and autonomic nervous system excitation patterns across the day[41], which can influence the phase of gene expression. These inputs appear to be particularly relevant for genes controlling lipid biosynthesis, absorption and degradation, many of which are rhythmically regulated and under circadian control[53]. Moreover, the molecular circadian clock acts as a rate-limiting step in cholesterol and bile synthesis, supporting the potential importance of circadian disruption in lipid biology[54]. Collectively, these data suggest different biological mechanisms involved in short and long sleep-associated adverse lipid profiles.

Consistent with different hypothesised physiological effects of short and long sleep, we observed no overlap in the previously

unreported loci that were identified by modelling interactions with short or long sleep duration. The lipid loci that were identified after considering STST include *FHIT*, *MAGI2* and *KLH3*, which have been previously associated with body mass index (BMI)[55–61]. Interestingly, although not genome-wide significant, variation in *MAGI2* has been associated with sleep duration[62], however, we did not find evidence for an association with rs10244093 in *MAGI2* with any sleep phenotype in the UK Biobank sample. Variants in *MICAL3* and *ZNF827*, that were also identified after considering STST, have been associated with serum liver enzymes gamma-glutamyltransferase measurement and/or aspartate aminotransferase levels[63,64], which have been implicated in cardiometabolic disturbances[65–68] and associated with prolonged work hours (which often results in short or irregular sleep)[69]. Other loci identified through interactions with STST were in genes previously associated with neurocognitive and neuropsychiatric conditions, possibly reflecting associations mediated by heightened levels of cortisol and sympathetic activity that frequently accompany short sleep.

In relation to LTST, the previously unreported lipid genes have been previously related to inflammation-driven diseases of the intestine, blood pressure and blood count measurements, including traits influenced by circadian rhythms[70,71]. However, none of these loci with LTST directly interacted with genes involved in the central circadian clock (e.g., *PER2*, *CRY2* and *CLOCK*) in the KEGG pathways database[72]. The *NR5A2* and *SLC35F3* loci have been associated with inflammation-driven diseases of the intestine[73,74]. Ulcerative colitis, an inflammatory bowel disease, has been associated with both longer sleep duration[75] and circadian disruption[70]. *ARNT2*, also identified via a LTST interaction, heterodimerizes with transcriptional factors implicated in homoeostasis and environmental stress responses[76,77]. A linkage association study has reported nominal association of this gene with lipids in a Caribbean Hispanic population[78].

We identified a number of additional genetic lead SNPs in the meta-analyses performed in European-Americans only. For example, we identified rs3938236 mapped to *SPRED1* to be associated with HDL-c after accounting for potential interaction with LTST. Interestingly, this gene has been previously associated with hypersomnia in Caucasian and Japanese populations[79], but was not identified in our larger multi-ancestry analysis, possibly due to cultural differences in sleep behaviours[80].

We additionally found evidence, amongst others, in the known lipid loci *APOB*, *PCSK9* and *LPL* for interaction with either short or long sleep. Associations have been observed previously between short sleep and ApoB concentrations, have been observed previously[81]. LPL expression has been shown to follows a diurnal rhythm in several metabolic organs[43,82], and disturbing sleeping pattern by altered light exposure can lower LPL activity, at least in brown adipose tissue[43]. Similar effects of sleep on hepatic secretion of ApoB and PCSK9 may be expected. Indeed, in humans PCSK9 has a diurnal rhythm synchronous with hepatic cholesterol synthesis[83]. Although the interaction effects we observed were rather weak, the supporting evidence from the literature suggests that sleep potentially modifies the effect of some of the well-known lipid regulators that are also targets for therapeutic interventions.

Some of the previously unreported lipid loci have been previously associated with traits related to sleep. For example, *MAGI2* and *MYO9B*[62] have been suggestively associated with sleep duration and quality, respectively. Genetic variation in *TMEM132B* has been associated with excessive daytime sleepiness[84], and *EPHB1* has been associated with self-reported chronotype[85]. These findings suggest some shared genetic component of lipid regulation and sleep biology. However, with the exception of the *METTL15*-mapped rs7924896 variant in relation to snoring, none of the lead SNPs mapped to the previously unreported lipid loci were associated with any of the investigated sleep phenotypes in the UK Biobank population, suggesting no or minimal shared component in sleep and lipid biology but rather that sleep duration specifically modifies the effect of the variant on the lipid traits.

This study was predominantly comprised of individuals of European ancestry, despite our efforts to include as many studies of diverse ancestries as possible. For this reason, additional efforts are required to specifically study gene–sleep interactions in those of African, Asian and Hispanic ancestry once more data becomes available. In line, we identified several loci that were identified only in the European-ancestry analysis, and not in the multi-ancestry analysis, suggestion that there might be ancestry-specific effects. The multi-ancestry analysis highlighted the genetic regions that are more likely to play a role in sleep-associated adverse lipid profiles across ancestries. In addition, our study used questionnaire-based data on sleep duration. Although the use of questionnaires likely increased measurement error and decreased statistical power, questionnaire-based assessments of sleep duration have provided important epidemiological data, including the identification of genetic variants for sleep traits in genome-wide association studies[84]. Identified variants for sleep traits have been recently successfully validated using accelerometer data[86], although the overall genetic correlation with accelerometer-based sleep duration was shown to be low[87]. Moreover, observational studies showed only a modest correlation between the phenotypes[88], which suggest that each approach characterises somewhat different phenotypes. At this time, we did not have sufficient data to evaluate other measures of sleep duration such as polysomnography or accelerometery. A more comprehensive characterisation, additional circadian traits as well as larger study samples (e.g., embedded in the large biobanks that become increasingly available for research) will refine our understanding of the interaction of these fundamental phenotypes and lipid biology.

In summary, the gene–sleep interaction efforts described in the present multi-ancestry study identified many lipid loci previously unreported to be associated with either HDL-c, LDL-c or triglycerides levels. Multiple of the these loci were driven by interactions with either short or long sleep duration, and were mapped to genes also associated with adiposity, inflammatory or neuropsychiatric traits. Collectively, the results highlight the interactions between extreme sleep–wake exposures and lipid biology.

## Methods

**Participants**. Analyses were performed locally by the different participating studies. Discovery and replication analyses comprised men and women between the age of 18 and 80 years, and were conducted separately for the different contributing (self-defined) ancestry groups, including: European, African, Asian, Hispanic and Brazilian (discovery analysis only). Descriptions of the different participating studies are described in detail in the Supplementary Notes 1 and 3, and study-specific characteristics (sizes, trait distribution and data preparation) are presented in Supplementary Tables 1–6. Every effort was made to include as many studies as possible.

**Ethical regulations**. The present work was approved by the Institutional Review Board of Washington University in St. Louis and complies with all relevant ethical regulations. Each participating study obtained written informed consent from all participants and received approval from the appropriate local institutional review boards.

**Lipid traits**. We conducted all analyses on the following lipid traits: HDL-c, LDL-c and TG. TG and LDL-c concentrations were measured in samples from individuals who had fasted for at least 8 hours. LDL-c could be either directly assayed or derived using the Friedewald equation[89] (the latter being restricted to those with TG ≤ 400 mg/dL). We furthermore corrected LDL-c for the use of lipid-lowering drugs, defined as any use of a statin drug or any unspecified lipid-lowering drug

after the year 1994 (when statin use became common in general practice). If LDL-c was directly assayed, the concentration of LDL-c was corrected by dividing the LDL-c concentration by 0.7. If LDL-c was derived using the Friedewald equation, we first divided the concentration of total cholesterol by 0.8 before LDL-c was calculated by the Friedewald equation. Due to the skewed distribution of HDL-c and TG, we ln-transformed the concentration prior to the analyses; no transformation for LDL-c was required. When an individual cohort measured the lipid traits during multiple visits, the visit with the largest available sample and concurrent availability of the sleep questions was selected.

**Nocturnal total sleep time**. Contributing cohorts collected information on the habitual sleep duration using either a single question such as 'on an average night, how long do you sleep?' or as part of a standardised sleep questionnaire (e.g., the Pittsburgh Sleep Quality Index questionnaire[90]). For the present project, we defined both STST and LTST. To harmonise the sleep duration data across cohorts from different countries, cultures and participants with different physical characteristics, in whom sleep duration was assessed using various questions, we defined STST and LTST using cohort-specific residuals, adjusting for age and sex. An exception was for AGES and HANDLS cohorts, we used a cohort-specific definition due to limited response categories in relationship to the available question on sleep duration. Instead, we defined STST or LTST based on expert input. Exposure to STST was defined as the lowest 20% of the sex- and age-adjusted sleep-time residuals (coded as '1'). Exposure to LTST was defined as the highest 20% of the sex- and age-adjusted sleep-time residuals (coded as '1'). For both sleep-time definitions, we considered the remaining 80% of the population as being unexposed to either STST or LTST (coded as '0').

**Genotype data**. Genotyping was performed by each participating study locally using genotyping arrays from either Illumina (San Diego, CA, USA) or Affymetrix (Santa Clara, CA, USA). Each study conducted imputation using various software programmes and with local cleaning thresholds for call rates (usually > 98%) and Hardy–Weinberg equilibrium (usually $P$-value < $1e^{-5}$). The cosmopolitan reference panel from the 1000 Genomes Project Phase I Integrated Release Version 3 Haplotypes (2010–11 data freeze, 2012-03-14 haplotypes) was specified for imputation. Only SNPs on the autosomal chromosomes with a minor allele frequency of at least 0.01 were considered in the analyses. Specific details of each participating study's genotyping platform and imputation software are described (Supplementary Tables 3 and 6).

**Stage 1 analysis (discovery phase)**. The discovery phase of the present project included 21 cohorts contributing data from 28 study/ancestry groups, and included up to 62,457 participants of EUR, AFR, ASN, HISP and BR ancestry (Supplementary Tables 1–3). All cohorts ran statistical models according to a standardised analysis protocol. The main model for this project examined the SNP-main effect and the multiplicative interaction term between the SNP and either LTST or STST:

$$\mathrm{E}(Y) = \beta_0 + \beta_E E + \beta_G SNP + \beta_{GE} E * SNP + \beta_C C \quad (1)$$

in which E is the sleep exposure variable (LTST/STST) and C are the (study-specific) covariates, which was similar to what we have done in previous studies[4,11,12]. In addition, we examined the SNP-main effect (without incorporating LTST/STST) and the SNP-main effect stratified by the exposure:

$$\mathrm{E}(Y) = \beta_{0^*} + \beta_{G^*} SNP + \beta_{C^*} C \quad (1)$$

All models were performed for each lipid trait and separately for the different ancestry groups. Consequently, per ancestry group, we requested a total of seven GWA analyses per lipid trait. All models were adjusted for age, sex, field centre (if required), and the first principal components to correct for population stratification. The number of principal components included in the model was chosen according to cohort-specific preferences (ranging from 0 to 10). All studies were asked to provide the effect estimates (SNP-main and -interaction effect) with accompanying robust estimates of the standard error for all requested models. A robust estimate of the covariance between the main and interaction effects was also provided. To obtain robust estimates of covariance matrices and standard errors, studies with unrelated participants used R packages such as either sandwich[91,92] or ProbABEL[93]. Studies including related individuals used either generalised estimating equations (R package geepack[94]) or linear mixed models (GenABEL[95], MMAP or R package sandwich[91,92]). Sample code provided to studies to generate these data has been previously published[96].

Upon completion of the analyses by local institution, all summary data were stored centrally for further processing and meta-analyses. We performed estimative quality control (QC) using the R-based package EasyQC[97] (www.genepi-regensburg.de/easyqc) at the study level (examining the results of each study individually), and subsequently at the ancestry level (after combining all ancestry-specific cohorts using meta-analyses). Study-level QC consisted of excluding all SNPs with MAF < 0.01, harmonisation of alleles, comparison of allele frequencies with ancestry-appropriate 1000 Genomes reference data, and harmonisation of all SNPids to a standardised nomenclature according to chromosome and position. Ancestry-level QC included the compilation of summary statistics on all effect estimates, standard errors and p-values across studies to identify potential outliers, and production of SE-N and QQ plots to identify analytical problems (such as improper trait transformations)[98].

Prior to the ancestry-specific meta-analyses, we excluded the following SNPs from the cohort-level data files: all SNPs with an imputation quality < 0.5, and all SNPs with a minor allele count in the exposed group (LTST or STST equals '1') x imputation quality of less than 20. SNPs in the European-ancestry and multi-ancestry analyses had to be present in at least three cohorts and 5000 participants. Due to the limited sample size of the non-European ancestries (either discovery or replication), we did not take into account this filter in those ancestry-level meta-analyses.

Meta-analyses were conducted for all models using the inverse variance-weighted fixed effects method as implemented in METAL[99] (http://genome.sph.umich.edu/wiki/METAL). We evaluated both a 1df of freedom test of interaction effect and a 2df joint test of main and interaction effects, following previously published methods[29]. A 1df Wald test was used to evaluate the 1df interaction, as well as the main effect in models without an interaction term. A 2df Chi-squared test was used to jointly test the effects of both the variant and the variant × LTST/STST interaction[100]. Meta-analyses were conducted within each ancestry separately. Multi-ancestry meta-analyses were conducted on all ancestry-specific meta-analyses. Genomic control correction was applied on all cohorts incorporated in the ancestry-level meta-analyses as well as on the final meta-analyses for the publication. From this effort, we selected all SNPs associated with any of the lipid traits with $P \le 5 \times 10^{-7}$ in the 2df interaction test for replication in the Stage 2 analysis. This cut-off was selected to minimise false-negative results.

**Stage 2 analysis (replication phase)**. All SNPs selected in Stage 1 for replication were evaluated in the interaction model in up to 18 cohorts contributing data from 20 study groups totalling up to 64,469 individuals (Supplementary Tables 4–6). As we had a limited number of individuals from non-European ancestry in the replication analyses, we did not consider an the non-European ancestries separately and only focussed on a European-ancestry and multi-ancestry analysis.

Study- and ancestry-level QC was carried out as in stage 1. In contrast to stage 1, no additional filters were included for the number of studies or individuals contributing data to stage 2 meta-analyses, as these filters were implemented to reduce the probability of false positives, and were less relevant in stage 2. Stage 2 SNPs were evaluated in all ancestry groups and for all traits, no matter what specific meta-analysis met the $P$-value threshold in the stage 1 analysis. We did not apply genomic control to any of the Stage 2 analyses given the expectation of association.

An additional meta-analysis was performed combining the Stage 1 and 2 meta-analyses. SNPs (irrespective of being known or previously unreported) were considered to be replicated when the 2df interaction test $P$-values of Stage $1 < 5 \times 10^{-7}$, Stage $2 < 0.05$ with a similar direction of effect as in the discovery meta-analysis, and Stage $1 + 2 < 5 \times 10^{-8}$. Replicated SNPs were subsequently used in different bioinformatics tools for further processing. In addition, 1df $P$-values (SNP-sleep interaction effect only) of the lead SNPs of both the replicated known and previously unreported loci were calculated to explore whether genetic variant were specifically driven by SNP-main or SNP-interaction effects. Based on the total number of lead SNPs across all analyses, we performed correction using the false discovery rate to quantify statistical significance[36].

**Bioinformatics**. Replicated SNPs were first processed using the online tool FUMA[101] to identify independent lead SNPs and to perform gene mapping. From the SNP that has a $P$-value in the 2df interaction test $< 5 \times 10^{-8}$, we determined lead SNPs that were independent from each other at $R^2 < 0.1$ using the 1000 G Phase 3 EUR as a reference panel population. Independent lead SNPs with a physical distance > 1 mB from a known locus were considered as previously unreported. Regional plots of these loci were made using the online LocusZoom tool[102]. The explained variance of the identified genetic lead SNPs mapped to previously unreported lipid regions was calculated based on the summary statistics of the combined analysis of Stage 1 and 2 using the R-based VarExp package, which has been previously validated to provide similar results to individual participant data[35]. This package calculates the variance explained on the basis of the combined (joint) SNP-main and SNP-interaction effect. Differential expression analyses of the lead SNPs in the identified genetic loci was performed using GTEx [https://gtexportal.org/home/][37,38].

**Look-ups of previously unreported loci in other databases**. The genetic loci for the three lipid traits previously unreported were further explored in the GWAS catalogue [https://www.ebi.ac.uk/gwas/] to investigate the role of these mapped genes in other traits. Furthermore, we extracted the lead SNPs from the previously unreported lipid loci from publically available GWAS data from the UK Biobank [http://www.nealelab.is/uk-biobank/] for different questionnaire-based sleep phenotypes, notably 'daytime snoozing/sleeping (narcolepsy)', 'getting up in the morning', 'morning/evening person (chronotype)', 'nap during the day', 'sleep duration', 'sleeplessness/insomnia' and 'snoring'. Analyses on these phenotypes were generally done using continuous outcomes; the variable 'sleep duration' was expressed in hours of total sleep per day. GWAS in the UK Biobank were done in European-ancestry individuals only ($N$ up to 337,074). We furthermore extracted

the identified lead SNPs from the previously unreported regions for lipid traits from the GWAS analyses done on accelerometer-based sleep variables, which was done in European-ancestry individuals from the UK Biobank ($N = 85{,}670$; [http://sleepdisordergenetics.org/])[87]. In addition, we extracted the these identified lead SNPs from publically available summary-statistics data on coronary artery disease of the CARDIoGRAMplusC4D consortium, which included 60,801 cases of coronary artery disease and 123,504 controls [http://www.cardiogramplusc4d.org][103].

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

## Data availability

Due to restrictions in the written informed consent and local regulations, no individual genotype-level data could be shared that were part of this project. Summary results files from both the trans-ancestry and European meta-analyses are available to the public via the CHARGE (Cohorts for Heart and Ageing Research in Genomics Epidemiology) dbGaP summary site (phs000930 [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000930.v1.p1]). We acknowledge the use of publically available data sources for summary-based statistics, which includes the gTex portal [https://gtexportal.org/home/], Nealelab [http://www.nealelab.is/uk-biobank/], Sleep Disorder Genetics [http://sleepdisordergenetics.org/] and the CARDIoGRAMplusC4D consortium [http://www.cardiogramplusc4d.org].

Received: 8 March 2019; Accepted: 4 October 2019;
Published online: 12 November 2019

## References

1. Holmes, M. V. et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur. Heart J.* **36**, 539–550 (2015).
2. Ference, B. A. et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. *J. Am. Coll. Cardiol.* **60**, 2631–2639 (2012).
3. Voight, B. F. et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* **380**, 572–580 (2012).
4. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
5. Do, R. et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.* **45**, 1345–1352 (2013).
6. Peloso, G. M. et al. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am. J. Hum. Genet.* **94**, 223–232 (2014).
7. Spracklen, C. N. et al. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum. Mol. Genet.* **26**, 1770–1784 (2017).
8. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
9. Kathiresan, S. et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N. Engl. J. Med.* **358**, 1240–1249 (2008).
10. Klarin, D. et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat. Genet.* **50**, 1514–1523 (2018).
11. Kilpelainen, T. O. et al. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat. Commun.* **10**, 376 (2019).
12. de Vries, P. S. et al. Multi-ancestry genome-wide association study of lipid levels incorporating gene-alcohol interactions. *Am. J. Epidemiol.* **188**, 1033–1054 (2019).
13. Bentley, A. R. et al. Multi-ancestry genome-wide smoking interaction study of 387,272 individuals identifies novel lipid loci. *Nat. Genet.* **51**, 636–648 (2019).
14. Tobaldini, E. et al. Sleep, sleep deprivation, autonomic nervous system and cardiovascular diseases. *Neurosci. Biobehav. Rev.* **74**, 321–329 (2017).
15. Tobaldini, E., Pecis, M. & Montano, N. Effects of acute and chronic sleep deprivation on cardiovascular regulation. *Arch. Ital. Biol.* **152**, 103–110 (2014).
16. Ford, E. S. Habitual sleep duration and predicted 10-year cardiovascular risk using the pooled cohort risk equations among US adults. *J. Am. Heart Assoc.* **3**, e001454 (2014).
17. Aggarwal, S., Loomba, R. S., Arora, R. R. & Molnar, J. Associations between sleep duration and prevalence of cardiovascular events. *Clin. Cardiol.* **36**, 671–676 (2013).
18. Wu, Y., Zhai, L. & Zhang, D. Sleep duration and obesity among adults: a meta-analysis of prospective studies. *Sleep. Med.* **15**, 1456–1462 (2014).
19. Xi, B., He, D., Zhang, M., Xue, J. & Zhou, D. Short sleep duration predicts risk of metabolic syndrome: a systematic review and meta-analysis. *Sleep. Med. Rev.* **18**, 293–297 (2014).
20. Cappuccio, F. P. et al. Meta-analysis of short sleep duration and obesity in children and adults. *Sleep* **31**, 619–626 (2008).
21. Cappuccio, F. P., Cooper, D., D'Elia, L., Strazzullo, P. & Miller, M. A. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *Eur. Heart J.* **32**, 1484–1492 (2011).
22. Lee, J. A. & Park, H. S. Relation between sleep duration, overweight, and metabolic syndrome in Korean adolescents. *Nutr. Metab. Cardiovasc. Dis.* **24**, 65–71 (2014).
23. van den Berg, J. F. et al. Long sleep duration is associated with serum cholesterol in the elderly: the Rotterdam Study. *Psychosom. Med.* **70**, 1005–1011 (2008).
24. Petrov, M. E. et al. Longitudinal associations between objective sleep and lipids: the CARDIA study. *Sleep* **36**, 1587–1595 (2013).
25. Bos, M. M. et al. Associations of sleep duration and quality with serum and hepatic lipids: The Netherlands Epidemiology of Obesity Study. *J. Sleep Res.* **28**, e12776 (2018).
26. Kaneita, Y., Uchiyama, M., Yoshiike, N. & Ohida, T. Associations of usual sleep duration with serum lipid and lipoprotein levels. *Sleep* **31**, 645–652 (2008).
27. Psaty, B. M. et al. Cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium: design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circ. Cardiovasc. Genet.* **2**, 73–80 (2009).
28. Rao, D.C. et al. Multiancestry study of gene-lifestyle interactions for cardiovascular traits in 610 475 individuals from 124 cohorts: design and rationale. *Circ. Cardiovasc. Genet.* **10**, e001649 (2017).
29. Manning, A. K. et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP× environment regression coefficients. *Genet. Epidemiol.* **35**, 11–18 (2011).
30. Lopez-Garcia, E. et al. Sleep duration, general and abdominal obesity, and weight change among the older adult population of Spain. *Am. J. Clin. Nutr.* **87**, 310–316 (2008).
31. van den Berg, J. F. et al. Actigraphic sleep duration and fragmentation are related to obesity in the elderly: the Rotterdam Study. *Int. J. Obes.* **32**, 1083–1090 (2008).
32. Wong, P. M., Manuck, S. B., DiNardo, M. M., Korytkowski, M. & Muldoon, M. F. Shorter sleep duration is associated with decreased insulin sensitivity in healthy white men. *Sleep* **38**, 223–231 (2015).
33. Reutrakul, S. & Van Cauter, E. Interactions between sleep, circadian function, and glucose metabolism: implications for risk and severity of diabetes. *Ann. N. Y. Acad. Sci.* **1311**, 151–173 (2014).
34. Cappuccio, F. P., D'Elia, L., Strazzullo, P. & Miller, M. A. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care* **33**, 414–420 (2010).
35. Laville, V. et al. VarExp: estimating variance explained by genome-wide GxE summary statistics. *Bioinformatics* **34**, 3412–3414 (2018).
36. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
37. GTEx Consortium. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
38. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
39. Aho, V. et al. Prolonged sleep restriction induces changes in pathways involved in cholesterol metabolism and inflammatory responses. *Sci. Rep.* **6**, 24828 (2016).
40. Chua, E. C., Shui, G., Cazenave-Gassiot, A., Wenk, M. R. & Gooley, J. J. Changes in plasma lipids during exposure to total sleep deprivation. *Sleep* **38**, 1683–1691 (2015).
41. Gooley, J. J. Circadian regulation of lipid metabolism. *Proc. Nutr. Soc.* **75**, 440–450 (2016).
42. Huang, T. et al. Habitual sleep quality, plasma metabolites and risk of coronary heart disease in post-menopausal women. *Int. J. Epidemiol.* **48**, 1262–1274 (2018).
43. van den Berg, R. et al. A diurnal rhythm in brown adipose tissue causes rapid clearance and combustion of plasma lipids at wakening. *Cell Rep.* **22**, 3521–3533 (2018).
44. van den Berg, R. et al. Familial longevity is characterized by high circadian rhythmicity of serum cholesterol in healthy elderly individuals. *Aging Cell.* **16**, 237–243 (2017).
45. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).

46. Yang, L. et al. Longer sleep duration and midday napping are associated with a higher risk of CHD incidence in middle-aged and older Chinese: the Dongfeng-Tongji Cohort Study. *Sleep* **39**, 645–652 (2016).
47. Anafi, R. C. et al. Sleep is not just for the brain: transcriptional responses to sleep in peripheral tissues. *BMC Genomics* **14**, 362 (2013).
48. Moller-Levet, C. S. et al. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc. Natl Acad. Sci. USA* **110**, E1132–E1141 (2013).
49. Carson, V., Tremblay, M. S., Chaput, J. P. & Chastin, S. F. Associations between sleep duration, sedentary time, physical activity, and health indicators among Canadian children and youth using compositional analyses. *Appl. Physiol. Nutr. Metab.* **41**, S294–S302 (2016).
50. Patel, S. R. et al. Sleep duration and biomarkers of inflammation. *Sleep* **32**, 200–204 (2009).
51. Ayas, N. T. et al. A prospective study of sleep duration and coronary heart disease in women. *Arch. Intern. Med.* **163**, 205–209 (2003).
52. Wefers, J. et al. Circadian misalignment induces fatty acid metabolism gene profiles and compromises insulin sensitivity in human skeletal muscle. *Proc. Natl Acad. Sci. USA* **115**, 7789–7794 (2018).
53. Adamovich, Y., Aviram, R. & Asher, G. The emerging roles of lipids in circadian control. *Biochim. Biophys. Acta* **1851**, 1017–1025 (2015).
54. Galman, C., Angelin, B. & Rudling, M. Bile acid synthesis in humans has a rapid diurnal variation that is asynchronous with cholesterol synthesis. *Gastroenterology* **129**, 1445–1453 (2005).
55. Akiyama, M. et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49**, 1458–1467 (2017).
56. Winkler, T. W. et al. The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. *PLoS Genet.* **11**, e1005378 (2015).
57. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
58. Justice, A. E. et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat. Commun.* **8**, 14977 (2017).
59. Hoffmann, T. J. et al. A large multiethnic genome-wide association study of adult body mass index identifies novel loci. *Genetics* **210**, 499–515 (2018).
60. Graff, M. et al. Genome-wide physical activity interactions in adiposity—a meta-analysis of 200,452 adults. *PLoS Genet.* **13**, e1006528 (2017).
61. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
62. Spada, J. et al. Genome-wide association analysis of actigraphic sleep phenotypes in the LIFE adult study. *J. Sleep. Res.* **25**, 690–701 (2016).
63. Chambers, J. C. et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat. Genet.* **43**, 1131–1138 (2011).
64. Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
65. Kunutsor, S. K., Abbasi, A. & Adler, A. I. Gamma-glutamyl transferase and risk of type II diabetes: an updated systematic review and dose-response meta-analysis. *Ann. Epidemiol.* **24**, 809–816 (2014).
66. Kunutsor, S. K., Apekey, T. A. & Cheung, B. M. Gamma-glutamyltransferase and risk of hypertension: a systematic review and dose-response meta-analysis of prospective evidence. *J. Hypertens.* **33**, 2373–2381 (2015).
67. Kunutsor, S. K., Apekey, T. A. & Seddoh, D. Gamma glutamyltransferase and metabolic syndrome risk: a systematic review and dose-response meta-analysis. *Int. J. Clin. Pract.* **69**, 136–144 (2015).
68. Wang, J., Zhang, D., Huang, R., Li, X. & Huang, W. Gamma-glutamyltransferase and risk of cardiovascular mortality: A dose-response meta-analysis of prospective cohort studies. *PLoS. One.* **12**, e0172631 (2017).
69. Park, S. G. et al. Association between long working hours and serum gamma-glutamyltransferase levels in female workers: data from the fifth Korean National Health and Nutrition Examination Survey (2010–2011). *Ann. Occup. Environ. Med.* **26**, 40 (2014).
70. Swanson, G. R., Burgess, H. J. & Keshavarzian, A. Sleep disturbances and inflammatory bowel disease: a potential trigger for disease flare? *Expert Rev. Clin. Immunol.* **7**, 29–36 (2011).
71. Giles, T. D. Circadian rhythm of blood pressure and the relation to cardiovascular events. *J. Hypertens. Suppl.* **24**, S11–S16 (2006).
72. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
73. de Lange, K. M. et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
74. Maguire, L. H. et al. Genome-wide association analyses identify 39 new susceptibility loci for diverticular disease. *Nat. Genet.* **50**, 1359–1365 (2018).
75. Ananthakrishnan, A. N. et al. Sleep duration affects risk for ulcerative colitis: a prospective cohort study. *Clin. Gastroenterol. Hepatol.* **12**, 1879–1886 (2014).
76. Sullivan, A. E. et al. Characterization of human variants in obesity-related SIM1 protein identifies a hot-spot for dimerization with the partner protein ARNT2. *Biochem. J.* **461**, 403–412 (2014).
77. Hao, N., Bhakti, V. L., Peet, D. J. & Whitelaw, M. L. Reciprocal regulation of the basic helix-loop-helix/Per-Arnt-Sim partner proteins, Arnt and Arnt2, during neuronal differentiation. *Nucleic Acids Res.* **41**, 5626–5638 (2013).
78. Dong, C. et al. Genetic loci for blood lipid levels identified by linkage and association analyses in Caribbean Hispanics. *J. Lipid Res.* **52**, 1411–1419 (2011).
79. Khor, S. S. et al. Genome-wide association study of HLA-DQB1*06:02 negative essential hypersomnia. *PeerJ* **1**, e66 (2013).
80. Egan, K. J., Knutson, K. L., Pereira, A. C. & von Schantz, M. The role of race and ethnicity in sleep, circadian rhythms and cardiovascular health. *Sleep. Med. Rev.* **33**, 70–78 (2017).
81. Ren, H., Liu, Z., Zhou, X. & Yuan, G. Association of sleep duration with apolipoproteins and the apolipoprotein B/A1 ratio: the China health and nutrition survey. *Nutr. Metab.* **15**, 1 (2018).
82. Tsutsumi, K., Inoue, Y. & Kondo, Y. The relationship between lipoprotein lipase activity and respiratory quotient of rats in circadian rhythms. *Biol. Pharm. Bull.* **25**, 1360–1363 (2002).
83. Persson, L. et al. Circulating proprotein convertase subtilisin kexin type 9 has a diurnal rhythm synchronous with cholesterol synthesis and is reduced by fasting in humans. *Arterioscler. Thromb. Vasc. Biol.* **30**, 2666–2672 (2010).
84. Lane, J. M. et al. Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat. Genet.* **49**, 274–281 (2017).
85. Hu, Y. et al. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Commun.* **7**, 10448 (2016).
86. Dashti, H. S. et al. Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat. Commun.* **10**, 1100 (2019).
87. Jones, S. E. et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. *Nat. Commun.* **10**, 1585 (2019).
88. Jackson, C.L., Patel, S.R., Jackson, W.B., 2nd, Lutsey, P.L. & Redline, S. Agreement between self-reported and objectively measured sleep duration among white, black, Hispanic, and Chinese adults in the United States: multi-ethnic study of atherosclerosis. *Sleep* **41**, zsy057 (2018).
89. Friedewald, W. T., Levy, R. I. & Fredrickson, D. S. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* **18**, 499–502 (1972).
90. Buysse, D. J., Reynolds, C. F. 3rd, Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193–213 (1989).
91. Zeileis, A. Object-oriented computation of sandwich estimators. *J. Stat. Softw.* **16**, 16 (2006).
92. Zeileis, A. Econometric computing with HC and HAC covariance matrix estimators. *J. Stat. Softw.* **11** https://www.jstatsoft.org/article/view/v011i10 (2004).
93. Aulchenko, Y. S., Struchalin, M. V. & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinforma.* **11**, 134 (2010).
94. Halekoh, U., Højsgaard, S. & Yan, J. The R package geepack for generalized estimating equations. *J. Stat. Softw.* **15**, 1–11 (2006).
95. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
96. Rao, D.C. et al. Multiancestry study of gene–lifestyle interactions for cardiovascular traits in 610 475 individuals from 124 cohorts: design and rationale. *Circ. Cardiovasc. Interv.* **10**, e001649 (2017).
97. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
98. Winkler, T. W. et al. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259–261 (2015).
99. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
100. Kraft, P., Yen, Y. C., Stram, D. O., Morrison, J. & Gauderman, W. J. Exploiting gene-environment interaction to detect genetic associations. *Hum. Hered.* **63**, 111–119 (2007).
101. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
102. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).

103. Nikpay, M. et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
104. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
105. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York, 2016).

## Acknowledgements

This project was supported by a grant from the US National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (R01HL118305). This research was supported in part by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health. Tuomas O. Kilpeläinen was supported by the Danish Council for Independent Research (DFF–6110-00183) and the Novo Nordisk Foundation (NNF18CC0034900, NNF17OC0026848 and NNF15CC0018486). Diana van Heemst was supported by the European Commission funded project HUMAN (Health-2013-INNOVATION-1-602757). Susan Redline was supported in part by NIH R35HL135818 and HL11338. Study-specific acknowledgements can be found in the Supplementary Notes 2 and 4. The data on coronary artery disease have been contributed by the Myocardial Infarction Genetics and CARDIoGRAM investigators, and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG.

## Author contributions

R.N. and M.M.B. conducted the centralised data analysis, which included quality control checks, meta-analyses and bioinformatics. R.N., M.M.B. and S.R. drafted the initial version of the paper. R.N., M.M.B., H.W., T.W.W., A.R.B., T.O.K., P.B.M., C.T.L., A.C. M., D.C.R., D.v.H. and S.R. were part of the writing group and were mainly responsible for the study design, interpretation of the data and critical commenting on the initial draft versions of the paper. All other co-authors were responsible for cohort-level data collection, cohort-level data analysis and critical reviews of the draft paper. All authors approved the final version of the paper that was submitted to the journal.

## Competing interests

D.O.M.K. is a part-time research consultant for Metabolon, Inc. H.J.G. has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm and Janssen Cilag. H.J.G. has received research funding from the German Research Foundation (DFG), the German Ministry of Education and Research (BMBF), the DAMP Foundation, Fresenius Medical Care, the EU "Joint Programme Neurodegenerative Disorders (JPND) and the European Social Fund (ESF)". S.A. reports employment and stock options with 23andMe, Inc.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41467-019-12958-0.

**Correspondence** and requests for materials should be addressed to R.N. or S.R.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer 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.



Raymond Noordam[1,116]*, Maxime M. Bos[1,116], Heming Wang[2,116], Thomas W. Winkler[3,116], Amy R. Bentley[4,116], Tuomas O. Kilpeläinen[5,6,116], Paul S. de Vries[7], Yun Ju Sung[8], Karen Schwander[8], Brian E. Cade[2], Alisa Manning[9,10], Hugues Aschard[11,12], Michael R. Brown[7], Han Chen[7,13], Nora Franceschini[14], Solomon K. Musani[15], Melissa Richard[16], Dina Vojinovic[17], Stella Aslibekyan[18], Traci M. Bartz[19], Lisa de las Fuentes[8,20], Mary Feitosa[21], Andrea R. Horimoto[22], Marjan Ilkov[23], Minjung Kho[24], Aldi Kraja[21], Changwei Li[25], Elise Lim[26], Yongmei Liu[27], Dennis O. Mook-Kanamori[28,29], Tuomo Rankinen[30], Salman M. Tajuddin[31], Ashley van der Spek[17], Zhe Wang[7], Jonathan Marten[32], Vincent Laville[12], Maris Alver[33,34], Evangelos Evangelou[35,36], Maria E. Graff[14], Meian He[37], Brigitte Kühnel[38,39], Leo-Pekka Lyytikäinen[40,41], Pedro Marques-Vidal[42], Ilja M. Nolte[43], Nicholette D. Palmer[44], Rainer Rauramaa[45], Xiao-Ou Shu[46], Harold Snieder[43], Stefan Weiss[47], Wanqing Wen[46], Lisa R. Yanek[48], Correa Adolfo[15], Christie Ballantyne[49,50], Larry Bielak[24], Nienke R. Biermasz[51,52], Eric Boerwinkle[7,53], Niki Dimou[36], Gudny Eiriksdottir[23], Chuan Gao[54], Sina A. Gharib[55], Daniel J. Gottlieb[2,10,56], José Haba-Rubio[57], Tamara B. Harris[58], Sami Heikkinen[59,60], Raphaël Heinzer[57], James E. Hixson[7], Georg Homuth[47], M. Arfan Ikram[17,61], Pirjo Komulainen[45], Jose E. Krieger[22], Jiwon Lee[2], Jingmin Liu[62], Kurt K. Lohman[63], Annemarie I. Luik[17], Reedik Mägi[33], Lisa W. Martin[64], Thomas Meitinger[65], Andres Metspalu[33,34], Yuri Milaneschi[64], Mike A. Nalls[66,67], Jeff O'Connell[68,69], Annette Peters[39,70], Patricia Peyser[24], Olli T. Raitakari[71,72], Alex P. Reiner[62], Patrick C.N. Rensen[51,52], Treva K. Rice[8], Stephen S. Rich[73], Till Roenneberg[74], Jerome I. Rotter[75], Pamela J. Schreiner[76], James Shikany[77], Stephen S. Sidney[78], Mario Sims[15], Colleen M. Sitlani[79], Tamar Sofer[2,80], Konstantin Strauch[81,82], Morris A. Swertz[83], Kent D. Taylor[75],

André G. Uitterlinden[17,84], Cornelia M. van Duijn[17,85], Henry Völzke[86], Melanie Waldenberger[38,39,70], Robert B. Wallance[87], Ko Willems van Dijk[51,52,88], Caizheng Yu[37], Alan B. Zonderman[89], Diane M. Becker[48], Paul Elliott[34,90,91,92], Tõnu Esko[33,93], Christian Gieger[38,94], Hans J. Grabe[95], Timo A. Lakka[45,60,96], Terho Lehtimäki[40,41], Kari E. North[14], Brenda W.J.H. Penninx[97], Peter Vollenweider[42], Lynne E. Wagenknecht[98], Tangchun Wu[37], Yong-Bing Xiang[99], Wei Zheng[46], Donna K. Arnett[100], Claude Bouchard[30], Michele K. Evans[31], Vilmundur Gudnason[23,101], Sharon Kardia[24], Tanika N. Kelly[102], Stephen B. Kritchevsky[103], Ruth J.F. Loos[104,105], Alexandre C. Pereira[22], Mike Province[21], Bruce M. Psaty[106,107], Charles Rotimi[4], Xiaofeng Zhu[108], Najaf Amin[17], L. Adrienne Cupples[26,109], Myriam Fornage[16], Ervin F. Fox[110], Xiuqing Guo[75], W. James Gauderman[111], Kenneth Rice[112], Charles Kooperberg[62], Patricia B. Munroe[113,114], Ching-Ti Liu[26], Alanna C. Morrison[7], Dabeeru C. Rao[8], Diana van Heemst[1] & Susan Redline[2,115]*

[1]Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands. [2]Division of Sleep and Circadian Disorders, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA. [3]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. [4]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [5]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2200, Denmark. [6]Department of Environmental Medicine and Public Health, The Icahn School of Medicine at Mount Sinai, New York, New York, USA. [7]Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA. [8]Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA. [9]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA, USA. [10]Department of Medicine, Harvard Medical School, Boston, MA, USA. [11]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA. [12]Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), Institut Pasteur, Paris, France. [13]Center for Precision Health, School of Public Health & School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA. [14]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA. [15]Jackson Heart Study, Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [16]Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA. [17]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands. [18]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA. [19]Cardiovascular Health Research Unit, Biostatistics and Medicine, University of Washington, Seattle, WA, USA. [20]Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA. [21]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. [22]Laboratory of Genetics and Molecular Cardiology, Heart Institute (InCor), University of São Paulo Medical School, São Paulo, SP, Brazil. [23]Icelandic Heart Association, Kopavogur, Iceland. [24]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [25]Epidemiology and Biostatistics, University of Georgia at Athens College of Public Health, Athens, GA, USA. [26]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [27]Public Health Sciences, Epidemiology and Prevention, Wake Forest University Health Sciences, Winston-Salem, NC, USA. [28]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands. [29]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands. [30]Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA, USA. [31]Health Disparities Research Section, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA. [32]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [33]Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia. [34]Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [35]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [36]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece. [37]Department of Occupational and Environmental Health and State Key Laboratory of Environmental Health for Incubating, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [38]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [39]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [40]Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. [41]Department of Clinical Chemistry, Finnish Cardiovascular Research Center—Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [42]Medicine, Internal Medicine, Lausanne University Hospital, Lausanne, Switzerland. [43]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands. [44]Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. [45]Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. [46]Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA. [47]Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [48]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [49]Section of Cardiovascular Research, Baylor College of Medicine, Houston, TX, USA. [50]Houston Methodist Debakey Heart and Vascular Center, Houston, TX, USA. [51]Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands. [52]Einthoven Laboratory for Experimental Vascular Medicine, Leiden, The Netherlands. [53]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. [54]Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC, USA. [55]Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Medicine, University of Washington, Seattle, WA, USA. [56]VA Boston Healthcare System, Boston, MA, USA. [57]Medicine, Sleep Laboratory, Lausanne University Hospital, Lausanne, Switzerland. [58]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [59]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland. [60]Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio Campus, Finland. [61]Department of Radiology and Nuclear Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands. [62]Fred Hutchinson Cancer Research Center, University of Washington School of Public Health, Seattle, WA, USA. [63]Public Health Sciences, Biostatistical Sciences, Wake Forest University Health Sciences, Winston-Salem, NC, USA. [64]Cardiology, School of Medicine and Health Sciences, George Washington University, Washington, D.C., USA. [65]Institute of Human Genetics, Helmholtz Zentrum München, German Research

Center for Environmental Health, Neuherberg, Germany. [66]Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA. [67]Data Tecnica International, Glen Echo, MD, USA. [68]Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA. [69]Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [70]DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Neuherberg, Germany. [71]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [72]University of Turku, Turku, Finland. [73]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [74]Institute of Medical Psychology, Ludwig-Maximilians-Universitat Munchen, Munich, Germany. [75]Genomic Outcomes, Department of Pediatrics, Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical Center, Torrance, CC, USA. [76]Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA. [77]Division of Preventive Medicine, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. [78]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [79]Cardiovascular Health Research Unit, Medicine, University of Washington, Seattle, WA, USA. [80]Institute of Human Genetics, Technische Universität München, Munich, Germany. [81]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [82]Institute for Medical Informatics Biometry and Epidemiology, Ludwig-Maximilians-Universitat Munchen, Munich, Germany. [83]University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands. [84]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands. [85]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [86]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [87]Department of Epidemiology, University of Iowa College of Public Health, Iowa City, IA, USA. [88]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. [89]Behavioral Epidemiology Section, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA. [90]MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [91]National Institute of Health Research Imperial College London Biomedical Research Centre, London, UK. [92]UK-DRI Dementia Research Institute at Imperial College London, London, UK. [93]Broad Institute of the Massachusetts Institute of Technology and Harvard University, Boston, MA, USA. [94]German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany. [95]Department Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [96]Department of Clinical Phsiology and Nuclear Medicine, Kuopia University Hospital, Kuopio, Finland. [97]Department of Psychiatry, Amsterdam Neuroscience and Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands. [98]Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. [99]SKLORG & Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, P. R. China. [100]Dean's Office, University of Kentucky College of Public Health, Lexington, KS, USA. [101]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [102]Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA. [103]Sticht Center for Healthy Aging and Rehabilitation, Gerontology and Geriatric Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. [104]The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY, USA. [105]The Mindich Child Health Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY, USA. [106]Cardiovascular Health Research Unit, Epidemiology, Medicine and Health Services, University of Washington, Seattle, WA, USA. [107]Kaiser Permanente Washington, Health Research Institute, Seattle, WA, USA. [108]Department of Population Quantitative and Health Sciences, Case Western Reserve University, Cleveland, OH, USA. [109]NHLBI Framingham Heart Study, Framingham, MA, USA. [110]Cardiology, Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [111]Biostatistics, Preventive Medicine, University of Southern California, Los Angeles, CA, USA. [112]Department of Biostatistics, University of Washington, Seattle, WA, USA. [113]Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [114]NIHR Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, London, UK. [115]Division of Pulmonary Medicine, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. [116]These authors contributed equally: Raymond Noordam, Maxime M. Bos, Heming Wang, Thomas W. Winkler, Amy R. Bentley, Tuomas O. Kilpeläinen. [117]These authors jointly directed this work: Patricia B. Munroe, Ching-Ti Liu, Alanna C. Morrison, Dabeeru C. Rao, Diana van Heemst, Susan Redline. *email: r.noordam@lumc.nl; sredline@bwh.harvard.edu

**Supplementary Information, R.Noordam et al**

**Supplementary Note 1 STAGE 1 STUDY DESCRIPTIONS:**
Brief descriptions are provided below for each of the discovery studies, some of which are based outside the United States:

**AGES (Age Gene/Environment Susceptibility Reykjavik Study):** The AGES Reykjavik study originally comprised a random sample of 30,795 men and women born in 1907-1935 and living in Reykjavik, Iceland in 1967. A total of 19,381 people attended, resulting in a 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within month. One group was designated for longitudinal follow up and was examined in all stages; another was designated as a control group and was not included in examinations until 1991. Other groups were invited to participate in specific stages of the study. Between 2002 and 2006, the AGES Reykjavik study re-examined 5,764 survivors of the original cohort who had participated before in the Reykjavik Study. Blood samples were drawn from all participants after overnight fasting. Serum total cholesterol and HDL were analyzed on a Hitachi 912 using reagents from Roche Diagnostics (Mannheim, Germany) and following the manufacturer's instructions (CV% was 2.5% and 3.5% respectively). Serum LDL cholesterol was calculated using the Friedewald equation when triglycerides < 4.5 mmol/L. Information on sleep duration was collected using the single question for both summer and winter: "During a typical 24-hour period in the summer time/winter time, how many hours do you spend sleeping at night?" Genotyping was performed using the Illumina 370CNV BeadChip array on 3,664 participants. Sample exclusion criteria included sample failure, genotype mismatch with reference panel, and sex mismatch, resulting in clean genotype data on 3,219 individuals. Standard protocols for working with Illumina data were followed with clustering score greater than 0.4. From a total of 353,202 SNPS, 325,094 were used for imputation after exclusion of SNPs with call rate < 97%, HWE deviation < $1 \times 10^{-6}$, mishap (PLINK haplotype-based test for non-random missing genotype data[2]) $p < 1 \times 10^{-9}$, and mismatched positions between Illumina, dbSNP and/or HapMap. Imputation was done using MACH1.0.16 against all the HapMap CEPH haplotypes (release 22/NCBI build 36) resulting in 2,533,153 total SNPs for analysis. All participants signed an informed consent form. The AGES-Reykjavik Study GWAS was approved by the National Bioethics Committee (VSN: 00-063) and the Data Protection Authority. A detailed description of the AGES Reykjavik Study has been described previously (Harris TB *et al.* Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics.Am J Epidemiol. 2007 May 1;165(9):1076-87. Epub 2007 Mar 10. PMID:17351290)

**ARIC (Atherosclerosis Risk in Communities):** The ARIC study is a population-based prospective cohort study of cardiovascular disease sponsored by the National Heart, Lung, and Blood Institute (NHLBI). ARIC included 15,792 individuals, predominantly European American and African American, aged 45-64 years at baseline (1987-89), chosen by probability sampling from four US communities. Cohort members completed three additional triennial follow-up examinations, a fifth exam in 2011-2013, and a sixth exam in 2016-2017. The ARIC study has been described in detail previously (The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) study: Design and objectives. Am J Epidemiol. 1989;129:687-702).

**Baependi Heart Study (Brazil):** The Baependi Heart Study, is an ongoing family-based cohort conducted in a rural town of the state of Minas Gerais. The study has enrolled approximate 2,200 individuals (over 10% of the town's adult population) and 10-year follow up period of longitudinal data. Briefly, probands were selected at random across 11 out of the 12 census districts in Baependi. After enrolment, the proband's first-degree (parents, siblings, and offspring), second-degree (half-siblings, grandparents/grandchildren, uncles/aunts, nephews/nieces, and double cousins), and third-degree (first cousins, great uncles/aunts, and great nephews/nieces) relatives, and his/her respective spouse's relatives resident both within Baependi (municipal and rural area) and surrounding towns were invited to participate. Only individuals age 18 and older were eligible to participate in the study. The study is conducted from a clinic/office in an easily accessible sector of the town, where the questionnaires were completed. A broad range of phenotypes ranging from cardiovascular, neurocognitive, psychiatric, imaging, physiologic and several layers of endophenotypes like metabolomics and lipidomics have been collected throughout the years Details about follow-up visits and available data can be found in the cohort profile paper (PMID: 18430212). DNA samples were genotyped using the Affymetrix 6.0 genechip. After quality control, the data were prephased using SHAPEIT and imputed using IMPUTE2 based on 1000 Genomes haplotypes.

**CARDIA (Coronary Artery Risk Development in Young Adults):** CARDIA is a prospective multicenter study with 5,115 adults Caucasian and African American participants of the age group 18-30 years, recruited from four centers at the baseline examination in 1985-1986. The recruitment was done from the total community in Birmingham, AL, from selected census tracts in Chicago, IL and Minneapolis, MN; and from the Kaiser Permanente health plan membership in Oakland, CA. The details of the study design for the CARDIA study have been previously published[1]. Nine examinations have been completed since initiation of the study, respectively in the years 0, 2, 5, 7, 10, 15, 20, 25 and 30. Written informed consent was obtained from participants at each examination and all study protocols were approved by the institutional review boards of the participating institutions. Age and race were self-reported using standardized questionnaires, as were use of cholesterol-lowering medication. participants were asked to fast for 12 hours before each clinic visit.

1. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hully SB, Jacobs DR Jr., Liu K, Savage PJ. CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol. 1988;41:1105–1116.

**CHS (Cardiovascular Health Study):** CHS is a population-based cohort study of risk factors for cardiovascular disease in adults 65 years of age or older conducted across four field centers (1). The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists and an additional predominately African-American cohort of 687 persons was enrolled in 1992-93 for a total sample of 5,888. Blood samples were drawn from all participants at their baseline examination and DNA was subsequently extracted from available samples. European ancestry participants were excluded from the GWAS study sample due to prevalent coronary heart disease, congestive heart failure,

peripheral vascular disease, valvular heart disease, stroke, or transient ischemic attack at baseline. After QC, genotyping was successful for 3271 European ancestry and 823 African-American participants. CHS was approved by institutional review committees at each site and individuals in the present analysis gave informed consent including consent to use of genetic information for the study of cardiovascular disease. Information on habitual sleep duration was collected 4 years after the cholesterol measurement.

1. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol 1991; 1:263-76.

**ERF (Erasmus Rucphen Family study):** Erasmus Rucphen Family is a family based study that includes inhabitants of a genetically isolated community in the South-West of the Netherlands, studied as part of the Genetic Research in Isolated Population (GRIP) program (Padro LM, et al; Ann Hum Genet. 2015l 69(Pt 3):288-95). The goal of the study is to identify the risk factors in the development of complex disorders. Study population includes approximately 3,000 individuals who are living descendants of 22 couples who lived in the isolate between 1850 and 1900 and had at least six children baptized in the community church (Aulchenko Y, et al; Eur J Hum Genet. 2015; 12 527-534). All data were collected between 2002 and 2005. All participants gave informed consent, and the Medical Ethics Committee of the Erasmus University Medical Centre approved the study.

**FHS (Framingham Heart Study):** FHS began in 1948 with the recruitment of an original cohort of 5,209 men and women (mean age 44 years; 55 percent women). In 1971 a second generation of study participants was enrolled; this cohort (mean age 37 years; 52% women) consisted of 5,124 children and spouses of children of the original cohort. A third generation cohort of 4,095 children of offspring cohort participants (mean age 40 years; 53 percent women) was enrolled in 2002-2005 and are seen every 4 to 8 years. Details of study designs for the three cohorts are summarized elsewhere. At each clinic visit, a medical history was obtained with a focus on cardiovascular content, and participants underwent a physical examination including measurement of height and weight from which BMI was calculated.

**GenSalt (Genetic Epidemiology Network of Salt Sensitivity):** GenSalt is a multi-center, family based study designed to identify, through dietary sodium and potassium intervention, salt-sensitivity susceptibility genes which may underlie essential hypertension in rural Han Chinese families. Approximately 629 families with at least one 'proband' with high blood pressure were recruited and tested for a wide variety of physiological, metabolic and biochemical measures at baseline and at multiple times during the 3-week intervention. The intervention consisted of one week on a low sodium diet, followed by one week on a high sodium diet, and finally one week on a high sodium diet with a potassium supplement.

**HANDLS (Healthy Aging in Neighborhoods of Diversity across the Life Span):** HANDLS is a community-based, longitudinal epidemiologic study examining the influences of race and socioeconomic status (SES) on the development of age-related health disparities among a sample of socioeconomically diverse African Americans and whites. This unique study will assess over a

20-year period physical parameters and also evaluate genetic, biologic, demographic, and psychosocial, parameters of African American and white participants in higher and lower SES to understand the driving factors behind persistent black-white health disparities in overall longevity, cardiovascular disease, and cognitive decline. The study recruited 3,722 participants from Baltimore, MD with a mean age of 47.7 years, 2,200 African Americans and 1,522 whites, with 41% reporting household incomes below the 125% poverty delimiter. Genotyping was done on a subset of self-reporting African American participants by the Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health (NIH). A larger genotyping effort included a small subset of self-reporting European ancestry samples. This research was supported by the Intramural Research Program of the NIH, NIA and the National Center on Minority Health and Health Disparities.

**Health ABC (Health, Aging, and Body Composition):** Cohort description: The Health ABC study is a prospective cohort study investigating the associations between body composition, weight-related health conditions, and incident functional limitation in older adults. Health ABC enrolled well-functioning, community-dwelling black (n=1281) and white (n=1794) men and women aged 70-79 years between April 1997 and June 1998. Participants were recruited from a random sample of white and all black Medicare eligible residents in the Pittsburgh, PA, and Memphis, TN, metropolitan areas. Participants have undergone annual exams and semi-annual phone interviews. The current study sample consists of 1559 white participants who attended the second exam in 1998-1999 with available genotyping data.
Genotyping: Genotyping was performed by the Center for Inherited Disease Research (CIDR) using the Illumina Human1M-Duo BeadChip system. Samples were excluded from the dataset for the reasons of sample failure, genotypic sex mismatch, and first-degree relative of an included individual based on genotype data. Genotyping was successful in 1663 Caucasians. Analysis was restricted to SNPs with minor allele frequency $\geq$ 1%, call rate $\geq$97% and HWE p$\geq$10-6. Genotypes were available on 914,263 high quality SNPs for imputation based on the HapMap CEU (release 22, build 36) using the MACH software (version 1.0.16). A total of 2,543,888 imputed SNPs were analyzed for association with vitamin D levels.

**HERITAGE (Health, Risk Factors, Exercise Training and Genetics):** The HERITAGE is the only known family-based study of exercise intervention to evaluate the role of genes and sequence variants involved in the response to a physically active lifestyle. The current study is based on the data collected at baseline of the study from 99 White families (244 males, 255 females). All subjects were required to be sedentary and free of chronic diseases at baseline. There are over 18 trait domains (e.g. dietary, lipids and lipoproteins, glucose and insulin metabolism [fasting and IVGTT], steroids, body composition and body fat distribution, cardiorespiratory fitness), for a grand total of over one thousand variables. Moreover, most of the outcome traits were measured twice on two separate days both at baseline and after exercise training was completed. Marker data include a genome-wide linkage scan and GWAS, in addition to a large number of candidate genes.

**JHS (Jackson Heart Study):** The Jackson Heart Study is a longitudinal, community-based observational cohort study investigating the role of environmental and genetic factors in the development of cardiovascular disease in African Americans. Between 2000 and 2004, a total of 5,306 participants were recruited from a tri-county area (Hinds, Madison, and Rankin Counties) that encompasses Jackson, MS. Details of the design and recruitment for the Jackson Heart Study cohort has been previously published.1-3 Briefly, approximately 30% of participants were former members of the Atherosclerosis Risk in Communities (ARIC) study. The remainder were recruited by either 1) random selection from the Accudata list, 2) commercial listing, 3) a constrained volunteer sample, in which recruitment was distributed among defined demographic cells in proportions designed to mirror those in the overall population, or through the Jackson Heart Study Family Study.


1. Wyatt SB, Diekelmann N, Henderson F, Andrew ME, Billingsley G, Felder SH et al. A community-driven model of research participation: the Jackson Heart Study Participant Recruitment and Retention Study. Ethn Dis 2003; 13(4):438-455.
2. Taylor HA, Jr., Wilson JG, Jones DW, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. Ethn Dis 2005; 15:S6-17.
3. Fuqua SR, Wyatt SB, Andrew ME, et al. Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. Ethn Dis 2005; 15:S6-29.


**MESA (Multi-Ethnic Study of Atherosclerosis):** The Multi-Ethnic Study of Atherosclerosis (MESA) is a study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. MESA consisted of a diverse, population-based sample of an initial 6,814 asymptomatic men and women aged 45-84. 38 percent of the recruited participants were white, 28 percent African American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and University of California - Los Angeles. Participants are being followed for identification and characterization of cardiovascular disease events, including acute myocardial infarction and other forms of coronary heart disease (CHD), stroke, and congestive heart failure; for cardiovascular disease interventions; and for mortality. The first examination took place over two years, from July 2000 - July 2002. It was followed by four examination periods that were 17-20 months in length. Participants have been contacted every 9 to 12 months throughout the study to assess clinical morbidity and mortality.


Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR Jr, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol. 2002 Nov 1;156(9):871-81. PubMed PMID: 12397006.


**NEO (The Netherlands Epidemiology of Obesity study):** The NEO was designed for extensive phenotyping to investigate pathways that lead to obesity-related diseases. The NEO study is a

population-based, prospective cohort study that includes 6,671 individuals aged 45–65 years, with an oversampling of individuals with overweight or obesity. At baseline, information on demography, lifestyle, and medical history have been collected by questionnaires. In addition, samples of 24-h urine, fasting and postprandial blood plasma and serum, and DNA were collected. Genotyping was performed using the Illumina HumanCoreExome chip, which was subsequently imputed to the 1000 genome reference panel. Participants underwent an extensive physical examination, including anthropometry, electrocardiography, spirometry, and measurement of the carotid artery intima-media thickness by ultrasonography. The collection of data started in September 2008 and completed at the end of September 2012.

**RS (Rotterdam Study):** The Rotterdam Study is a prospective, population-based cohort study among individuals living in the well-defined Ommoord district in the city of Rotterdam in The Netherlands (Ikram MA, et al; Eur J Epidemiol. 2017; 32(9): 807-850). The aim of the study is to determine the occurrence of cardiovascular, neurological, ophthalmic, endocrine, hepatic, respiratory, and psychiatric diseases in elderly people. The cohort was initially defined in 1990 with 7,983 persons, aged 55 years and older, who underwent a home interview and extensive physical examination at the baseline and during follow-up rounds every 3-4 years (RS-I). Cohort was extended in 2000/2001 (RS-II, 3,011 individuals aged 55 years and older) and 2006/2008 (RS-III, 3,932 subjects, aged 45 and older). The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). All participants provided written informed consent to participate in the study and to have their medical information obtained from treating physicians.

**WHI (Women's Health Initiative):** WHI is a long-term national health study that focuses on strategies for preventing common diseases such as heart disease, cancer and fracture in postmenopausal women. A total of 161,838 women aged 50–79 years old were recruited from 40 clinical centers in the US between 1993 and 1998. WHI consists of an observational study, two clinical trials of postmenopausal hormone therapy (HT, estrogen alone or estrogen plus progestin), a calcium and vitamin D supplement trial, and a dietary modification trial [1]. Study recruitment and exclusion criteria have been described previously [1]. Recruitment was done through mass mailing to age-eligible women obtained from voter registration, driver's license and Health Care Financing Administration or other insurance list, with emphasis on recruitment of minorities and older women [2]. Exclusions included participation in other randomized trials, predicted survival < 3 years, alcoholism, drug dependency, mental illness and dementia. Study protocols and consent forms were approved by the IRB at all participating institutions. Socio-demographic characteristics, lifestyle, medical history and self-reported medications were collected using standardized questionnaires at the screening visit. The genome wide association study (GWAS) non-overlapping samples are composed of a case-control study (WHI Genomics and Randomized Trials Network – GARNET, which included all coronary heart disease, stroke, venous thromboembolic events and selected diabetes cases that happened during the active intervention phase in the WHI HT clinical trials and aged matched controls), women selected to be "representative" of the HT trial (mostly younger white HT subjects that were also enrolled in the WHI memory study - WHIMS) and the WHI SNP Health Association Resource (WHI SHARe), a randomly selected sample of 8,515 African American and 3,642 Hispanic women

from WHI. GWAS was performed using Affymetrix 6.0 (WHI-SHARe), HumanOmniExpressExome-8v1_B (WHIMS), Illumina HumanOmni1-Quad v1-0 B (GARNET). Extensive quality control (QC) of the GWAS data included alignment ("flipping") to the same reference panel, imputation to the 1000G data (using the recent reference panel - v3.20101123), identification of genetically related individuals, and computations of principal components (PCs) using methods developed by Price et al. (using EIGENSOFT software 53), and finally the comparison with self-reported ethnicity. After QC and exclusions from analysis protocol, the number of women included in analysis is 4,423 whites for GARNET, 5,202 white for WHIMS, 7,919 for SHARe African American and 3,377 for SHARe Hispanics.

1. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, Rossouw JE. The women's health initiative recruitment methods and results. Ann Epidemiol. 2003;13:S18-77
2. Design of the women's health initiative clinical trial and observational study. The women's health initiative study group. Control Clin Trials. 1998;19:61-109

**Supplementary Note 2 STAGE 1 STUDY ACKNOWLEDGMENTS:**

Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. Infrastructure for the Gene-Lifestyle Working Group is supported by the National Heart, Lung, and Blood Institute grant R01HL118305.

**AGES (Age Gene/Environment Susceptibility Reykjavik Study):** This study has been funded by NIH contract N01-AG012100, HSSN271201200022C, the NIA Intramural Research Program, an Intramural Research Program Award (ZIAEY000401) from the National Eye Institute, an award from the National Institute on Deafness and Other Communication Disorders (NIDCD) Division of Scientific Programs (IAA Y2-DC_1004-02), Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, VSN: 00-063. The researchers are indebted to the participants for their willingness to participate in the study.

**ARIC (Atherosclerosis Risk in Communities) study:** The ARIC study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL059367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

**Baependi Heart Study (Brazil):** The Baependi Heart Study was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) (Grant 2013/17368-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Hospital Samaritano Society (Grant 25000.180.664/2011-35), through Ministry of Health to Support Program Institutional Development of the Unified Health System (SUS-PROADI).

**CARDIA (Coronary Artery Risk Development in Young Adults):** The CARDIA Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging. Genotyping was funded as part of the NHLBI Candidate-gene Association Resource (N01-HC-65226) and the NHGRI Gene Environment Association Studies (GENEVA) (U01-HG004729, U01-HG04424, and U01-HG004446). This manuscript has been reviewed and approved by CARDIA for scientific content.

**CHS (Cardiovascular Health Study):** This Cardiovascular Health Study (CHS) research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268200960009C, HHSN268201800001C N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, U01HL130114, R01HL087652, R01HL105756, R01HL103612, R01HL085251, and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**ERF (Erasmus Rucphen Family study):** The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). The ERF study was further supported by ENGAGE consortium and CMSB. High-throughput analysis of the ERF data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). This study was further supported by the European Union's Horizon 2020 research and innovation programme as part of the Common mechanisms and pathways in Stroke and Alzheimer's disease (CoSTREAM) project (www.costream.eu, grant agreement No 667375); the Netherlands Organisation for Health Research and Development (ZonMW) as part of the Joint Programming for Neurological Disease (JPND) project PERADES (Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease using multiple powerful cohorts, focused Epigenetics and Stem cell metabolomics - grant number 733051021); the European Union's Horizon 2020 research and innovation programme Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) under the grant agreement No 645740 as part of the Personalized pREvention of Chronic DIseases (PRECeDI) project and the CardioVasculair Onderzoek Nederland (CVON 2012-03). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work, P. Snijders for his help in data collection and E.M. van Leeuwen for genetic imputation.

**FHS (Framingham Heart Study):** This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual

input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract Nos. N01-HC-25195 and HHSN268201500001I) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. This research was partially supported by grant R01-DK089256 from the National Institute of Diabetes and Digestive and Kidney Diseases (MPIs: Michael Province, L. Ching-Ti Liu, Kari North) and training grant T32GM074905-14 from NIH/National Institute of General Medical Sciences.

**GenSalt (Genetic Epidemiology Network of Salt Sensitivity):** The Genetic Epidemiology Network of Salt Sensitivity is supported by research grants (U01HL072507, R01HL087263, and R01HL090682) from the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD.

**HANDLS (Healthy Aging in Neighborhoods of Diversity across the Life Span):** The Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study was supported by the Intramural Research Program of the NIH, National Institute on Aging and the National Center on Minority Health and Health Disparities (project # Z01-AG000513 and human subjects protocol number 09-AG-N248). Data analyses for the HANDLS study utilized the high-performance computational resources of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (http://biowulf.nih.gov; http://hpc.nih.gov)).

**Health ABC (Health, Aging, and Body Composition):** Health ABC was funded by the National Institutes of Aging. This research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The GWAS was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

**HERITAGE (Health, Risk Factors, Exercise Training and Genetics):** The HERITAGE Family Study was supported by National Heart, Lung, and Blood Institute grant HL-45670.

**JHS (Jackson Heart Study):** The Jackson Heart Study is supported by contracts HSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung, and Blood Institute on Minority Health and Health Disparities. The authors acknowledge the Jackson Heart Study team institutions (University of Mississippi Medical Center, Jackson State University and Tougaloo College) and participants for their long-term commitment that continues to improve our understanding of the genetic epidemiology of cardiovascular and other chronic diseases among African Americans.

**MESA (Multi-Ethnic Study of Atherosclerosis):** This research was supported by the Multi-Ethnic Study of Atherosclerosis (MESA) contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Funding for MESA Share genotyping was provided by NHLBI Contract N02-HL-6-4278. This publication was partially developed under a STAR research assistance agreement, No. RD831697 (MESA Air), awarded by the U.S Environmental Protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and the EPA does not endorse any products or commercial services mentioned in this publication. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The authors thank the participants of the MESA study, the Coordinating Center, MESA investigators, and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

**NEO (The Netherlands Epidemiology of Obesity study):** The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Petra Noordijk, Pat van Beelen and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023).

**RS (Rotterdam Study):** The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.
The generation and management of GWAS genotype data for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics

Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera, Marjolein Peters and Carolina Medina-Gomez for their help in creating the GWAS database, and Karol Estrada, Yurii Aulchenko and Carolina Medina-Gomez for the creation and analysis of imputed data. This study was further supported by the European Union's Horizon 2020 research and innovation programme as part of the Common mechanisms and pathways in Stroke and Alzheimer's disease (CoSTREAM) project (www.costream.eu, grant agreement No 667375); the Netherlands Organisation for Health Research and Development (ZonMW) as part of the Joint Programming for Neurological Disease (JPND) project PERADES (Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease using multiple powerful cohorts, focused Epigenetics and Stem cell metabolomics - grant number 733051021); the European Union's Horizon 2020 research and innovation programme Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) under the grant agreement No 645740 as part of the Personalized pREvention of Chronic DIseases (PRECeDI) project and the CardioVasculair Onderzoek Nederland (CVON 2012-03).

**WHI (Women's Health Initiative):** The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at:
http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

**Supplementary Note 3 STAGE 2 STUDY DESCRIPTIONS:**
Brief descriptions are provided below for each of the replication studies/cohorts:

**Airwave (The Airwave Health Monitoring Study):** The Airwave Health Monitoring Study (22) was established to evaluate possible health risks associated with use of TETRA, a digital communication system used by police forces and other emergency services in Great Britain since 2001. The study has been broadened to investigate more generally the health of the work force. From 2004, participants from each force who agreed to participate were enrolled either with an enrolment questionnaire or a comprehensive health screening performed locally. This includes questionnaire, 7-day food diaries, anthropometry, measurements of cardiovascular and cognitive function, blood chemistry, coagulation and hematology. By March 2015, the study had recruited 53,606 participants, of whom 45,433 had attended the health screening, and 14,002 have genotype data (1000G imputed).
Ref: Elliott, P. et al. The Airwave Health Monitoring Study of police officers and staff in Great Britain: rationale, design and methods. Environ Res 134, 280-5 (2014).

**CFS (Cleveland Family Study):** The Cleveland Family Study (CFS) is a family-based, longitudinal study designed to characterize the genetic and non-genetic risk factors for sleep apnea. In total, 2534 individuals (46% African American) from 352 families were studied on up to 4 occasions over a period of 16 years (1990-2006). The initial aim of the study was to quantify the familial aggregation of sleep apnea. 632 African Americans were genotyped on the Affymetrix array 6.0 platform through the CARe Consortium with suitable genotying quality control. A further 122 African-Americans had genotyping based on the Illumina OmniExpress + Exome platform. Genomes were imputed separately for each chip based on a 1000 Genomes Project Phase 3 Version 5 cosmopolitan template using SHAPEIT and IMPUTE2.

**Colaus (Cohorte Lausannoise):** The cohort is a random sample of the population aged 35-75 years of the city of Lausanne, Switzerland. Recruitment began in June 2003 and ended in May 2006, and the first follow-up was conducted between April 2009 and September 2012. The CoLaus study was approved by the Institutional Ethics Committee of the University of Lausanne and informed consent was appropriately obtained by all participants. Both at baseline and follow-up, all participants attended the outpatient clinic of the University Hospital of Lausanne in the morning after an overnight fast. Data were collected by trained field interviewers in a single visit lasting about 60 min.

**DFTJ (Dongfeng-Tongji Cohort Study):** The DFTJ-cohort study includes 27,009 retired employees from a state-owned automobile enterprise in China. This study was launched in 2008 and will be followed up every 5 years. In 2013 we conducted the first follow-up. By using semi-structural questionnaire and health examination, those having cancer or severe diseases were excluded. Fasting blood samples and detailed epidemiology data were collected. The main goal of the cohort was to identify the environmental and genetic risk factors and the gene-environment interactions on chronic diseases, and to find novel biomarkers for chronic disease and mortality

prediction. Finally, 1,461 included in the present study with GWAS data. All of the participants wrote informed consent and the ethical committees in the Tongji Medical College approved this research project. Detailed information has been described in elsewhere(1).

QC criteria and imputation methods:

We did the GWAS scan on the DFTJ-cohort with Affymetrix Genome-Wide Human SNP Array 6.0 chips. In total, we genotyped 906,703 SNPs among 1,461 subjects. After stringent QC filtering, SNPs with MAF < 0.01, Hardy-Weinberg Equilibrium (HWE) < 0.0001, and SNP call rate < 95% were excluded. Individuals with call rates < 95% were also not included for further analysis. In total, we retained 1,452 subjects with 658,288 autosomal SNPs for statistical analyses, with an overall call rate of 99.68%. We used MACH 1.0 software to impute untyped SNPs using the LD information from the HapMap phase II database (CHB+JPT as a reference set (2007-08_rel22, released 2007-03-02). Imputed SNPs with high genotype information content (Rsq > 0.3 for MACH) were kept for the further association analysis.

Reference

1) Wang, F., Zhu, J., Yao, P., Li, X., He, M., Liu, Y., Yuan, J., Chen, W., Zhou, L., Min, X. et al. (2012) Cohort profile: The Dongfeng-Tongji cohort study of retired workers. International journal of epidemiology.

**DR's EXTRA (Dose-Responses to Exercise Training):** The Dose-Responses to Exercise Training (DR's EXTRA) Study is a 4-year RCT on the effects of regular physical exercise and healthy diet on endothelial function, atherosclerosis and cognition in a randomly selected population sample (n=3000) of Eastern Finnish men and women, identified from the national population register, aged 55-74 years. Of the eligible sample, 1410 individuals were randomized into one of the 6 groups: aerobic exercise, resistance exercise, diet, combined aerobic exercise and diet, combined resistance exercise and diet, or reference group following baseline assessments. During the four year intervention the drop-out rate was 15%.

**EGCUT (Estonian Genome Center - University of Tartu (Estonian Biobank)):** The Estonian Biobank is a population-based biobank of the Estonian Genome Center at the University of Tartu (www.biobank.ee; EGCUT). The entire project is conducted according to the Estonian Gene Research Act and all of the participants have signed a broad informed consent. The cohort comprises 51535 individuals, aged 18+ and closely reflects the age, sex and geographical distribution of the Estonian population. All of the subjects are recruited randomly by general practitioners and physicians at hospitals. A Computer Assisted Personal interview was filled within 1-2 hours at a doctor's office, which included personal, genealogical, educational, occupational history and lifestyle data. Anthropometric measurements, blood pressure and resting heart rate were measured and venous blood taken during the visit. Medical history and current health status are recorded according to ICD-10 codes.

**GeneSTAR (Genetic Studies of Atherosclerosis Risk):** GeneSTAR is a family-based prospective study of more than 4000 participants begun in 1983 to determine phenotypic and genetic causes of premature cardiovascular disease. Families were identified from 1983-2006 from probands with a premature coronary disease event prior to 60 years of age who were

identified at the time of hospitalization in any of 10 hospitals in the Baltimore, Maryland area. Their apparently healthy 30-59 year old siblings without known coronary disease were recruited and screened between 1983 and 2006. From 2003-2006, adult offspring over 21 years of age of all participating siblings and probands, as well as the coparents of the offspring were recruited and screened. Genotyping was performed in 3,232 participants on the Illumina 1Mv1_c platform.

**HCHS/SOL (Hispanic Community Health Study/ Study of Latinos):** The HCHS/SOL is a community-based cohort study of 16,415 self-identified Hispanic/Latino persons aged 18–74 years and selected from households in predefined census-block groups across four US field centers (in Chicago, Miami, the Bronx, and San Diego). The census-block groups were chosen to provide diversity among cohort participants with regard to socioeconomic status and national origin or background. The HCHS/SOL cohort includes participants who self-identified as having a Hispanic/Latino background; the largest groups are Central American (n = 1,730), Cuban (n = 2,348), Dominican (n = 1,460), Mexican (n = 6,471), Puerto Rican (n = 2,728), and South American (n = 1,068). The HCHS/SOL baseline clinical examination occurred between 2008 and 2011 and included comprehensive biological, behavioral, and sociodemographic assessments. Consenting HCHS/SOL subjects were genotyped at Illumina on the HCHS/SOL custom 15041502 B3 array. The custom array comprised the Illumina Omni 2.5M array (HumanOmni2.5-8v.1-1) ancestry-informative markers, known GWAS hits and drug absorption, distribution, metabolism, and excretion (ADME) markers, and additional custom content including ~150,000 SNPs selected from the CLM (Colombian in Medellin, Colombia), MXL (Mexican Ancestry in Los Angeles, California), and PUR (Puerto Rican in Puerto Rico) samples in the 1000Genomes phase 1 data to capture a greater amount of Amerindian genetic variation. QA/QC procedures yielded a total of 12,803 unique study participants for imputation and downstream association analyses.

**IRAS Family Study (Insulin Resistance Atherosclerosis Study):** The IRASFS was a family study designed to examine the genetic and epidemiologic basis of glucose homeostasis traits and abdominal adiposity. Briefly, self-reported Mexican American pedigrees were recruited in San Antonio, TX and San Luis Valley, CO. Probands with large families were recruited from the initial non-family-based IRAS, which was modestly enriched for impaired glucose tolerance and T2D. Inclusion of IRASFS data is limited to 1040 normoglycemic individuals in 88 pedigrees with genotype data from the Illumina OmniExpress and Omni 1S arrays and imputation to the 1000 Genome Integrated Reference Panel (phase I).

**KORA (Cooperative Health Research in the Augsburg Region):** The KORA study is a series of independent population-based epidemiological surveys of participants living in the region of Augsburg, Southern Germany. All survey participants are residents of German nationality identified through the registration office and were examined in 1994/95 (KORA S3) and 1999/2001 (KORA S4). In the KORA S3 and S4 studies 4,856 and 4,261 subjects have been examined implying response rates of 75% and 67%, respectively. 3,006 subjects participated in a 10-year follow-up examination of S3 in 2004/05 (KORA F3), and 3080 of S4 in 2006/2008 (KORA F4). The age range of the participants was 25 to 74 years at recruitment. Informed

consent has been given by all participants. The study has been approved by the local ethics committee. Individuals for genotyping in KORA F3 and KORA F4 were randomly selected and these genotypes are taken for the analysis of the phenotypes in KORA S3 and KORA S4.

**Lifelines (Netherlands Biobank):** Lifelines (https://lifelines.nl/) is a multi-disciplinary prospective population-based cohort study using a unique three-generation design to examine the health and health-related behaviors of 165,000 persons living in the North East region of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multimorbidity. In addition, the Lifelines project comprises a number of cross-sectional sub-studies which investigate specific age-related conditions. These include investigations into metabolic and hormonal diseases, including obesity, cardiovascular and renal diseases, pulmonary diseases and allergy, cognitive function and depression, and musculoskeletal conditions. All survey participants are between 18 and 90 years old at the time of enrollment. Recruitment has been going on since the end of 2006, and over 130,000 participants had been included by April 2013. At the baseline examination, the participants in the study were asked to fill in a questionnaire (on paper or online) before the first visit. During the first and second visit, the first or second part of the questionnaire, respectively, are checked for completeness, a number of investigations are conducted, and blood and urine samples are taken. Lifelines is a facility that is open for all researchers. Information on application and data access procedure is summarized on www.lifelines.nl.
Scholtens S, Smidt N, Swertz MA, Bakker SJ, Dotinga A, Vonk JM, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol. 2014 Dec 14.

LifeLines Cohort Study: group author genetics acknowledgement
Behrooz Z Alizadeh (1), H Marike Boezen (1), Lude Franke (2), Pim van der Harst (3), Gerjan Navis (4), Marianne G. Rots (5), Morris Swertz (2), Bruce HR Wolffenbuttel (6), Cisca Wijmenga (2)

(1) Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands
(2) Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands
(3) Department of Cardiology, University of Groningen, University Medical Center Groningen, The Netherlands
(4) Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, The Netherlands
(5) Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, The Netherlands
(6) Department of Endocrinology, University of Groningen, University Medical Center Groningen, The Netherlands

**NESDA (Netherlands Study of Depression and Anxiety):** NESDA is a multi-center study designed to examine the long-term course and consequences of depressive and anxiety disorders (http://www.nesda.nl)[1]. NESDA included both individuals with depressive and/or anxiety disorders and controls without psychiatric conditions. Inclusion criteria were age 18-65 years and

self-reported western European ancestry while exclusion criteria were not being fluent in Dutch and having a primary diagnosis of another psychiatric condition (psychotic disorder, obsessive compulsive disorder, bipolar disorder, or severe substance use disorder).

Reference: Penninx, B.W. et al. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int J Methods Psychiatr Res 17, 121-40 (2008).

**SHIP (Study of Health in Pomerania):** The Study of Health In Pomerania (SHIP) is a prospective longitudinal population-based cohort study in Mecklenburg-West Pomerania assessing the prevalence and incidence of common diseases and their risk factors (PMID: 20167617). SHIP encompasses the two independent cohorts SHIP and SHIP-TREND. Participants aged 20 to 79 with German citizenship and principal residency in the study area were recruited from a random sample of residents living in the three local cities, 12 towns as well as 17 randomly selected smaller towns. Individuals were randomly selected stratified by age and sex in proportion to population size of the city, town or small towns, respectively. A total of 4,308 participants were recruited between 1997 and 2001 in the SHIP cohort. Between 2008 and 2012 a total of 4,420 participants were recruited in the SHIP-TREND cohort. Individuals were invited to the SHIP study centre for a computer-assisted personal interviews and extensive physical examinations. The study protocol was approved by the medical ethics committee of the University of Greifswald. Oral and written informed consents was obtained from each of the study participants

Genome-wide SNP-typing was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 or the Illumina Human Omni 2.5 array (SHIP-TREND samples). Array processing was carried out in accordance with the manufacturer's standard recommendations. Genotypes were determined using GenomeStudio Genotyping Module v1.0 (GenCall) for SHIP-TREND and the Birdseed2 clustering algorithm for SHIP. Imputation of genotypes in SHIP and SHIP-TREND was performed with the software IMPUTE v2.2.2 based on 1000 Genomes release March 2012.

**SWHS/SMHS (Shanghai Women's Health Study/ Shanghai Men's Health Study):** The Shanghai Women's Health Study (SWHS) is an ongoing population-based cohort study of approximately 75,000 women who were aged 40-70 years at study enrollment and resided in in urban Shanghai, China; 56,832 (75.8%) provided a blood samples. Recruitment for the SWHS was initiated in 1997 and completed in 2000. The self-administered questionnaire includes information on demographic characteristics, disease and surgery histories, personal habits (such as cigarette smoking, alcohol consumption, tea drinking, and ginseng use), menstrual history, residential history, occupational history, and family history of cancer.

The Shanghai Men's Health Study (SMHS) is an ongoing population-based cohort study of 61,480 Chinese men who were aged between 40 and 74 years, were free of cancer at enrollment, and lived in urban Shanghai, China; 45,766 (74.4%) provided a blood samples. Recruitment for the SMHS was initiated in 2002 and completed in 2006. The self-administered questionnaire includes information on demographic characteristics, disease and surgery histories, personal habits (such as cigarette smoking, alcohol consumption, tea drinking, and ginseng use), residential history, occupational history, and family history of cancer.

Genotyping and imputation: Genomic DNA was extracted from buffy coats by using a Qiagen DNA purification kit (Valencia, CA) or Puregene DNA purification kit (Minneapolis, MN)

according to the manufacturers' instructions and then used for genotyping assays. The GWAS genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affy6.0) platform or Illumina 660, following manufacturers' protocols. After sample quality control, we exclude SNPs with 1) MAF <0.01; 2) call rate <95%; 2) bad genotyping cluster; and 3) concordance rate <95% among duplicated QC samples. Genotypes were imputed using the program MACH (http://www.sph.umich.edu/csg/abecasis/MACH/download/), which determines the probable distribution of missing genotypes conditional on a set of known haplotypes, while simultaneously estimating the fine-scale recombination map. Phased autosome SNP data from HapMap Phase II Asians (release 22) were used as the reference. To test for associations between the imputed SNP data with BMI, linear regression (additive model) was used, in which SNPs were represented by the expected allele count, an approach that takes into account the degree of uncertainty of genotype imputation (http://www.sph.umich.edu/csg/abecasis/MACH/download/).

**YFS (The Cardiovascular Risk in Young Finns Study):** The YFS is a population-based follow up-study started in 1980. The main aim of the YFS is to determine the contribution made by childhood lifestyle, biological and psychological measures to the risk of cardiovascular diseases in adulthood. In 1980, over 3,500 children and adolescents all around Finland participated in the baseline study. The follow-up studies have been conducted mainly with 3-year intervals. The latest 30-year follow-up study was conducted in 2010-11 (ages 33-49 years) with 2,063 participants. The study was approved by the local ethics committees (University Hospitals of Helsinki, Turku, Tampere, Kuopio and Oulu) and was conducted following the guidelines of the Declaration of Helsinki. All participants gave their written informed consent.

**Supplementary Note 4 STAGE 2 STUDY ACKNOWLEDGMENTS:**

Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. Infrastructure for the Gene-Lifestyle Working Group is supported by the National Heart, Lung, and Blood Institute grant R01HL118305.

**Airwave (The Airwave Health Monitoring Study):** We thank all participants in the Airwave Health Monitoring Study. The study is funded by the Home Office (Grant number 780-TETRA) with additional support from the National Institute for Health Research (NIHR), Imperial College Healthcare NHS Trust (ICHNT) and Imperial College Biomedical Research Centre (BRC). The study has ethical approval from the National Health Service Multi-site Research Ethics Committee (MREC/13/NW/0588). This work used computing resources provided by the MRC-funded UK MEDical Bioinformatics partnership programme (UK MED-BIO) (MR/L01632X/1). P.E. would like to acknowledge support from the Medical Research Council and Public Health England for the MRC-PHE Centre for Environment and Health (MR/L01341X/1) and from the NIHR NIHR Health Protection Research Unit in Health Impact of Environmental Hazards (HPRU-2012-10141). P.E. is supported by the UK-DRI Dementia Research Institute at Imperial College London which receives funding from UK DRI Ltd funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK. P.E. is associate director of Health Data Research UK-London which receives funding from a consortium led by the UK Medical Research Council.

**CFS (Cleveland Family Study):** The CFS was supported by the National Institutes of Health, the National Heart, Lung, Blood Institute grant HL113338, R01HL098433, HL46380.

**CoLaus (Cohorte Lausannoise):** The CoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 33CSCO-122661, 33CS30-139468 and 33CS30-148401).

**DFTJ (Dongfeng-Tongji Cohort Study):** This work was supported by grants from the Foundation of National Key Program of Research and Development of China (2016YFC0900800), the Programme of Introducing Talents of Discipline, the grants from the National Natural Science Foundation (grant NSFC-81473051, 81522040 and 81230069), and the Program for the New Century Excellent Talents in University (NCET-11-0169).

**DR's EXTRA (Dose-Responses to Exercise Training):** The study was supported by grants from Ministry of Education and Culture of Finland (722 and 627; 2004-2010); Academy of Finland (102318, 104943, 123885, 211119); European Commission FP6 Integrated Project (EXGENESIS), LSHM-CT-2004-005272; City of Kuopio; Juho Vainio Foundation; Finnish Diabetes Association; Finnish Foundation for Cardiovascular Research; Kuopio University Hospital; Päivikki and Sakari Sohlberg Foundation; Social Insurance Institution of Finland 4/26/2010.

**EGCUT (Estonian Genome Center - University of Tartu (Estonian Biobank)):** This study was supported by EU H2020 grants 692145, 676550, 654248, Estonian Research Council Grant IUT20-60 and PUT1660, NIASC, EIT – Health and NIH-BMI Grant No: 2R01DK075787-06A1 and EU through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012 GENTRANSMED.

**GeneSTAR (Genetic Studies of Atherosclerosis Risk):** GeneSTAR was supported by National Institutes of Health grants from the National Heart, Lung, and Blood Institute (HL49762, HL59684, HL58625, HL071025, U01 HL72518, HL087698, HL092165, HL099747, and K23HL105897), National Institute of Nursing Research (NR0224103), National Institute of Neurological Disorders and Stroke (NS062059), and by grants from the National Center for Research Resources to the Johns Hopkins General Clinical Research Center (M01-RR000052) and the Johns Hopkins Institute for Clinical & Translational Research (UL1 RR 025005).

**HCHS/SOL (Hispanic Community Health Study/ Study of Latinos):** The baseline examination of HCHS/SOL was supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research (NIDCR), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Neurological Disorders and Stroke, and NIH Office of Dietary Supplements additionally contributed funding to HCHS/SOL. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). Additional analysis support was provided by 1R01DK101855-01 and 13GRNT16490017. Genotyping was also supported by National Center for Advancing Translational Sciences UL1TR000124 and NIDDK DK063491 to the Southern California Diabetes Endocrinology Research Center. This research was also supported in part by the Intramural Research Program of the NIDDK, contract no. HHSB268201200054C, and Illumina.

**IRAS Family Study (Insulin Resistance Atherosclerosis Study):** The IRASFS is supported by the National Heart Lung and Blood Institute (HL060944, HL061019, and HL060919). Genotyping for this study was supported by the GUARDIAN Consortium with grant support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; DK085175) and in part by UL1TR000124 (CTSI) and DK063491 (DRC). The authors thank study investigators, staff, and participants for their valuable contributions.

**KORA (Cooperative Health Research in the Augsburg Region):** The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as

part of LMUinnovativ. This work was further supported by a grant (WA 4081/1-1) from the German Research Foundation.

**LifeLines (Netherlands Biobank):** The Lifelines Cohort Study, and generation and management of GWAS genotype data for the Lifelines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation.
The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants.

**NESDA (Netherlands Study of Depression and Anxiety):** Funding was obtained from the Netherlands Organization for Scientific Research (Geestkracht program grant 10-000-1002); the Center for Medical Systems Biology (CSMB, NWO Genomics), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL), VU University's Institutes for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam, University Medical Center Groningen, Leiden University Medical Center, National Institutes of Health (NIH, R01D0042157-01A, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health.Computing was supported by BiG Grid, the Dutch e-Science Grid, which is financially supported by NWO.

**SHIP (Study of Health in Pomerania):** SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data were supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG.

**SWHS/SMHS (Shanghai Women's Health Study/ Shanghai Men's Health Study):** We thank all the individuals who took part in these studies and all the researchers who have enabled this work to be carried out. The Shanghai Women's Health Study and the Shanghai Men's Health Study are supported by research grants UM1CA182910 and UM1CA173640 from the U.S. National Cancer Institute, respectively.

**YFS (The Cardiovascular Risk in Young Finns Study):** The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (grant X51001); Juho

Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association.
The expert technical assistance in the statistical analyses by Leo-Pekka Lyytikäinen and Irina Lisinen is gratefully acknowledged.

**Supplementary Table 1:** Sample Sizes of Studies Participating in Stage 1

| | | HDL | | | LDL | | | TG | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Ancestry** | | **N STST** | **N LTST** | **N Total** | **N STST** | **N LTST** | **N Total** | **N STST** | **N LTST** | **N Total** |
| **African** | ARIC | 173 | 176 | 855 | 160 | 168 | 793 | 160 | 169 | 803 |
| | CARDIA | 170 | 167 | 816 | 168 | 164 | 811 | 170 | 167 | 816 |
| | CHS | 87 | 94 | 434 | 86 | 93 | 427 | 87 | 93 | 429 |
| | HANDLS | 98 | NA | 485 | 92 | NA | 461 | 94 | NA | 464 |
| | HABC | 216 | 214 | 1066 | 213 | 211 | 1057 | 216 | 215 | 1067 |
| | JHS | 389 | 370 | 1942 | 383 | 366 | 1921 | 389 | 370 | 1942 |
| | MESA | 233 | 237 | 1183 | 231 | 236 | 1178 | 233 | 237 | 1184 |
| | WHI-SHARe | 1641 | 1675 | 8127 | 1641 | 1675 | 8127 | 1641 | 1675 | 8127 |
| | **Total** | **3007** | **2933** | **14908** | **2974** | **2913** | **14775** | **2990** | **2926** | **14832** |
| **Asian** | GenSalt | 369 | 364 | 1813 | 365 | 358 | 1790 | 369 | 364 | 1813 |
| | MESA | 114 | 112 | 566 | 110 | 110 | 554 | 114 | 112 | 566 |
| | Total | **483** | **476** | **2379** | **475** | **468** | **2344** | **483** | **476** | **2379** |
| **European** | AGES | 879 | 263 | 3202 | 876 | 261 | 3190 | 879 | 263 | 3202 |
| | ARIC | 629 | 655 | 3131 | 610 | 622 | 3003 | 612 | 632 | 3032 |
| | CARDIA | 310 | 319 | 1528 | 307 | 312 | 1507 | 310 | 319 | 1528 |
| | CHS | 310 | 346 | 1544 | 305 | 339 | 1510 | 311 | 340 | 1535 |
| | ERF | 180 | 176 | 892 | 178 | 174 | 880 | 178 | 176 | 886 |
| | FHS | 1523 | 1494 | 7561 | 1367 | 1317 | 6668 | 1385 | 1334 | 6726 |
| | HABC | 334 | 328 | 1611 | 325 | 321 | 1577 | 334 | 328 | 1612 |
| | HERITAGE | 97 | 98 | 483 | 97 | 98 | 483 | 97 | 98 | 483 |
| | MESA | 398 | 409 | 1993 | 394 | 403 | 1973 | 398 | 409 | 1993 |
| | NEO | 1004 | 992 | 4953 | 990 | 973 | 4884 | 1004 | 992 | 4953 |
| | RS1 | 166 | 166 | 820 | 158 | 161 | 792 | 160 | 161 | 795 |
| | RS2 | 225 | 223 | 1123 | 222 | 215 | 1091 | 222 | 216 | 1096 |
| | RS3 | 573 | 583 | 2263 | 555 | 569 | 2832 | 564 | 578 | 2872 |
| | WHI-GARNET | 760 | 761 | 3767 | 760 | 761 | 3767 | 760 | 761 | 3767 |
| | WHI-WHIMS | 1086 | 1057 | 5170 | 1086 | 1057 | 5170 | 1086 | 1057 | 5170 |
| | **Total** | **8474** | **7870** | **40041** | **8230** | **7583** | **39327** | **8300** | **7664** | **39650** |
| **Hispanic** | MESA | 216 | 216 | 1076 | 214 | 214 | 1054 | 216 | 216 | 1076 |
| | WHI-SHARe | 733 | 696 | 3384 | 733 | 696 | 3384 | 733 | 696 | 3384 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Total** | **949** | **912** | **4460** | **947** | **910** | **4438** | **949** | **912** | **4460** |
| **Brazilian** | Baependi | 133 | 126 | 669 | 132 | 126 | 664 | 133 | 126 | 669 |
| | **Total** | **133** | **126** | **669** | **132** | **126** | **664** | **133** | **126** | **669** |
| **Transancestry** | **Total** | **13046** | **12317** | **62457** | **12758** | **12000** | **61548** | **12855** | **12104** | **61990** |

**Supplementary Table 2**: Genotyping and Imputation in Stage 1 Studies

| Ancestry | | Genotyping Platform | Imputation Software |
|---|---|---|---|
| **African** | ARIC | Affymetrix 6.0 | IMPUTE2 |
| | CARDIA | Affymetrix 6.0 | MaCH/minimac |
| | CHS | Illumina HumanOmni-Quad_v1 BeadChip | IMPUTE version 2.2.2 |
| | HANDLS | Illumina 1M and 1Mduo arrays | MaCH/minimac |
| | HABC | Illumina HumanCoreExome BeadChip | MaCH (version 1.0.16) |
| | JHS | Affymetrix 6.0 | MaCH (version 1.0.16) |
| | MESA | Affymetrix 6.0 | IMPUTE2 |
| | WHI-SHARe | Affymetrix 6.0 | MaCH (version 1.0.16) |
| **Asian** | GenSalt | Affymetrix 6.0 | MaCH/minimac |
| | MESA | Affymetrix 6.0 | IMPUTE2 |
| **European** | AGES | Illumina Hu370CNV | MaCH (version 1.0.16) |
| | ARIC | Affymetrix 6.0 | IMPUTE2 |
| | CARDIA | Affymetrix 6.0 | BEAGLE version 3.3.2 |
| | CHS | Illumina 370CNV BeadChip (merged with ITMAT-Broad-CARe (IBC) Illumina iSELECT chip | MaCH/minimac |
| | ERF | Illumina 6k, Illumina 318K, Affymetrix 250K, Illumina 350K, Illumina 610K | MaCH 1.0.18.c |
| | FHS | Affymetrix Nsp, Sty and 50K gene centric | MaCH/minimac |
| | HABC | Illumina HumanCoreExome BeadChip | MaCH (version 1.0.16) |
| | HERITAGE | Illumina 370CNV | minimac |
| | MESA | Affymetrix 6.0 | IMPUTE2 |
| | NEO | Illumina HumanCoreExome-24v1_A Beadchip | IMPUTE2 |
| | RS1 | Illumina 550 (+duo), Illumina 610 quad | MaCH (version 1.0) |
| | RS2 | Illumina 550 duo | MaCH 1.0 |
| | RS3 | Illumina 610 quad | MaCH/minimac |
| | WHI-GARNET | Illumina HumanOmni1-Quad v1-0 B | MaCH (version 1.0.16) |
| | WHI-WHIMS | HumanOmniExpressExome-8v1_B | MaCH (version 1.0.16) |
| **Hispanic** | MESA | Affymetrix 6.0 | IMPUTE2 |
| | WHI-SHARe | Affymetrix 6.0 | MaCH (version 1.0.16) |
| **Brazilian** | Baependi | Genome-wide SNP Human Array 6.0 (Affymetrix 6.0) | SHAPEIT and IMPUTE2 |

**Supplementary Table 3:** Sample Sizes of Studies Participating in Stage 2

| Ancestry | | HDL | | | LDL | | | TG | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | N STST | N LTST | N Total | N STST | N LTST | N Total | N STST | N LTST | N Total |
| **African** | CFS | 52 | 48 | 301 | 49 | 46 | 291 | 52 | 48 | 302 |
| | GeneSTAR | 169 | 169 | 845 | 165 | 168 | 836 | 168 | 169 | 843 |
| | **Total** | **221** | **217** | **1146** | **214** | **214** | **1127** | **220** | **217** | **1145** |
| **Asian** | DFTJ | 279 | 266 | 1387 | 279 | 266 | 1387 | 279 | 266 | 1387 |
| | SMWHS | 350 | 350 | 1746 | 350 | 350 | 1746 | 47 | 51 | 231 |
| | **Total** | **629** | **616** | **3133** | **629** | **616** | **3133** | **326** | **317** | **1618** |
| **European** | AIRWAVE | 2810 | 2805 | 14006 | NA | NA | NA | NA | NA | NA |
| | CFS | 68 | 61 | 413 | 68 | 60 | 411 | 68 | 61 | 413 |
| | COLAUS | 929 | 925 | 4641 | 923 | 917 | 4613 | 929 | 925 | 4641 |
| | DR EXTRA | 246 | 246 | 1230 | 246 | 246 | 1230 | 246 | 246 | 1230 |
| | EGCUT - Omni Express | 166 | 119 | 803 | 166 | 119 | 803 | 131 | 90 | 654 |
| | EGCUT - Human370CNV | 103 | 99 | 512 | 103 | 99 | 512 | NA | NA | NA |
| | GeneSTAR | 247 | 249 | 1237 | 245 | 244 | 1218 | 248 | 249 | 1240 |
| | KORA - S3 | 611 | 621 | 3044 | 611 | 619 | 3038 | 73 | 47 | 246 |
| | KORA - F4 | 226 | 225 | 1128 | 226 | 225 | 1128 | 225 | 226 | 1127 |
| | Lifelines | 2470 | 2430 | 12229 | 2470 | 2430 | 12230 | 2470 | 2430 | 12230 |
| | NESDA | 274 | 274 | 1369 | 273 | 273 | 1365 | 275 | 275 | 1374 |
| | SHIP | 812 | 810 | 4023 | 807 | 803 | 4000 | NA | NA | NA |
| | SHIP-Trend | 198 | 197 | 983 | 198 | 197 | 983 | 198 | 197 | 983 |
| | YFS | 425 | 425 | 1994 | 425 | 425 | 1994 | 425 | 425 | 1994 |
| | **Total** | **9585** | **9486** | **47612** | **6761** | **6657** | **33525** | **5288** | **5171** | **26132** |
| **Hispanic** | IRASFS | 192 | 189 | 951 | 189 | 189 | 943 | 192 | 189 | 951 |
| | SOL | 2325 | 2326 | 11627 | 2284 | 2268 | 11384 | 2325 | 2326 | 11628 |
| | **Total** | **2517** | **2515** | **12578** | **2473** | **2457** | **12327** | **2517** | **2515** | **12579** |
| **Transancestry** | **Total** | **12952** | **12834** | **64469** | **10077** | **9944** | **50112** | **8351** | **8220** | **41474** |

**Supplementary Table 4:** Genotyping and Imputation in Stage 2 Studies

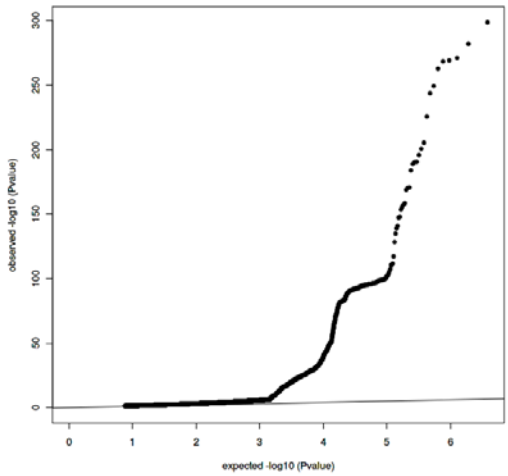
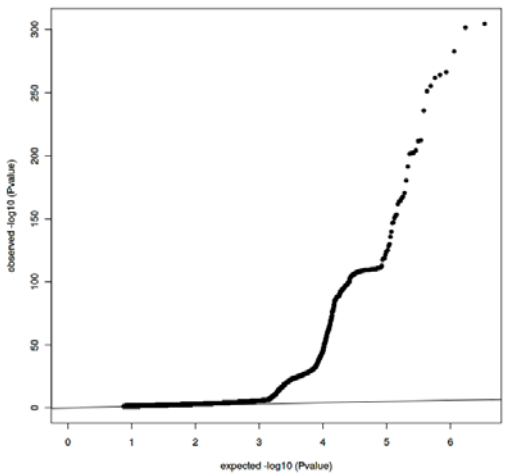
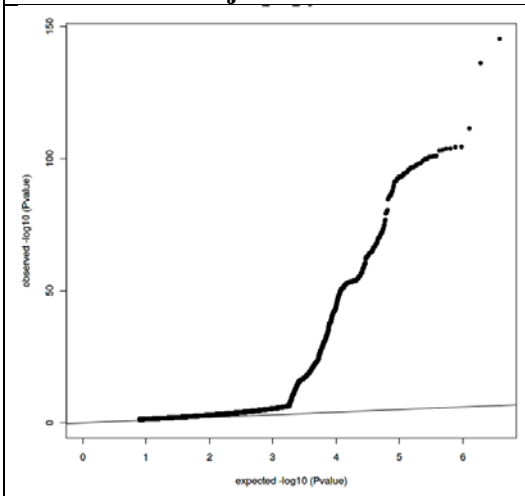
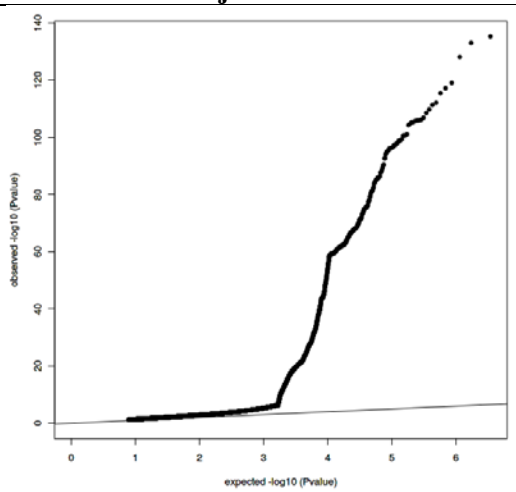
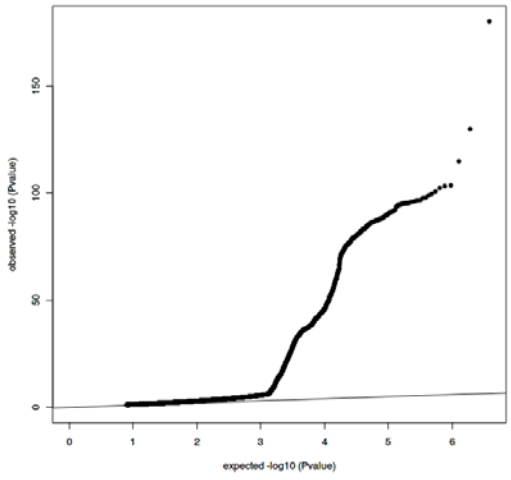
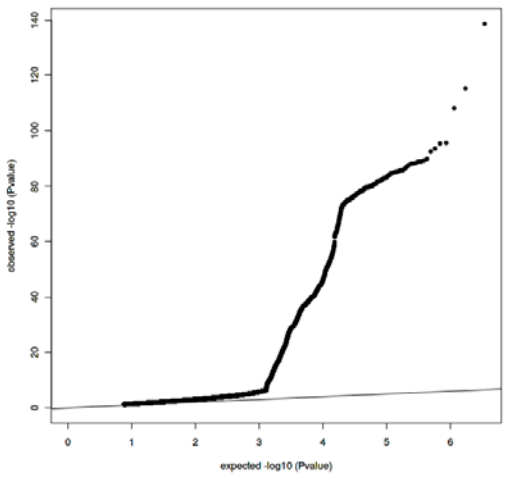
| Ancestry | | Genotyping Platform | Imputation Software |
|---|---|---|---|
| **African** | CFS | Affymetrix | MACH-ADMIX |
| | GeneSTAR | Illumina 1M_v1C | IMPUTE2 |
| **Asian** | DF-TJ | Affymetrix 6.0 | MaCH/minimac |
| | SMWHS | Affymetrix 6.0, Illumina Omni Express, Illumina 550, Illumina 1M | MaCH/minimac |
| **European** | AIRWAVE | Illumina HumanCoreExome- 12v1-1 | Minimac3 |
| | CFS | Illumina Omni | IMPUTE2 |
| | COLAUS | Affymetrix Human Mapping 500K | minimac |
| | DR EXTRA | Illumina Cardiometabochip | MaCH/minimac |
| | EGCUT - Omni Express | Illumina OmniExpress | IMPUTE2 |
| | EGCUT - Human370CNV | Illumina HumanCNV370 | IMPUTE2 |
| | GeneSTAR | Illumina 1M_v1C | IMPUTE2 |
| | KORA - S3 | Illumina Omni 2.5/Illumina Omni Express | IMPUTE v2.3.0 |
| | KORA - F4 | Affymetrix Axiom | IMPUTE v2.3.0 |
| | Lifelines | Illumina Cyto SNP12 v2 | MaCH/minimac |
| | NESDA | Affymetrix 5.0, Affymetrix 6.0 | MACH/minimac |
| | SHIP-O | Affymetrix Genome-wide Human SNP Array 6.0 | IMPUTE v2.2.2 |
| | SHIP-Trend | Illumina HumanOmni2.5 BeadChip | IMPUTE v2.2.2 |
| | YFS | Illumina 670k custom | IMPUTE2 |
| **Hispanic** | IRASFS | Illumina OmniExpress+1S | IMPUTE2 |
| | SOL | Illumina SOL HCHS Custom 15041502 B3 array | IMPUTE2 |

**Supplementary Table 5**: Association of novel SNPs with coronary artery disease (CAD) from the CARDIoGRAMplusC4D 1000 Genomes-based GWAS meta-analysis (60,801 CAD cases and 123,504 controls).

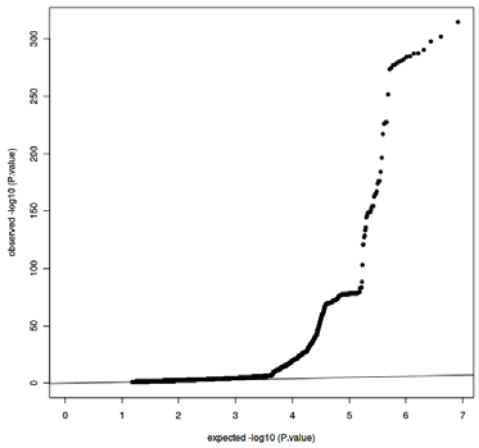
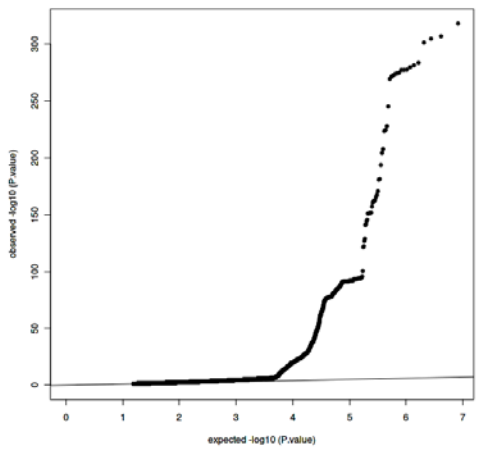
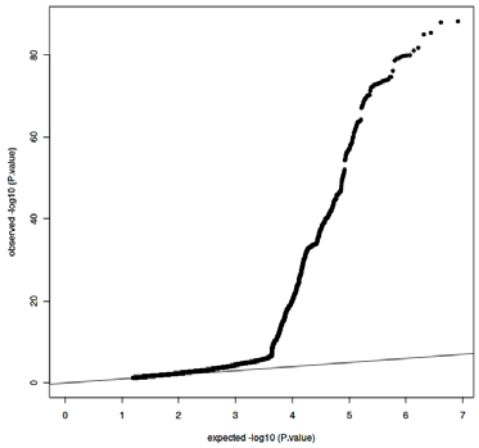
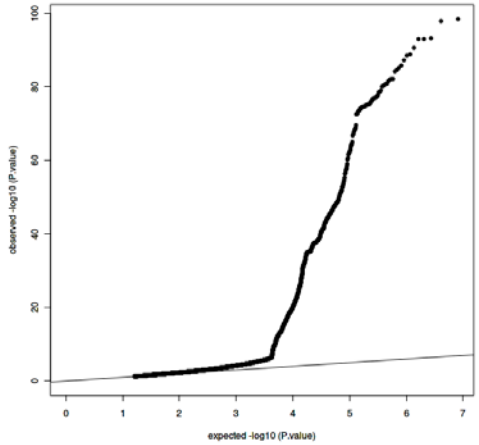
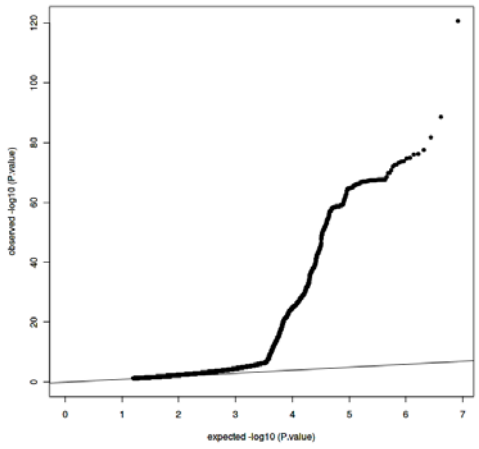
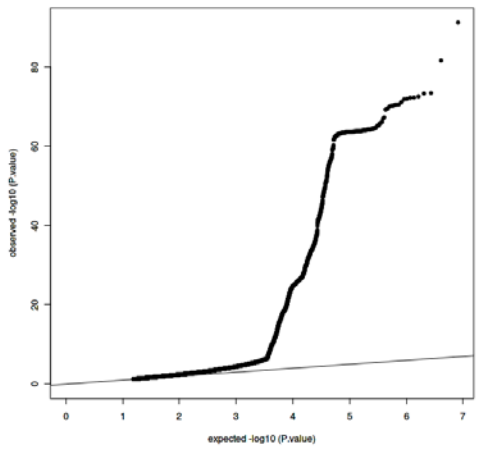
| Trait identified | SNP | mapped gene | Effect allele | Exposure | beta | se | p-value | analysis |
|---|---|---|---|---|---|---|---|---|
| ln(HDL-c) | rs12593988 | ARNT2 | a | LTST | 0.007 | 0.013 | 5.92E-01 | MULTI/EA |
| ln(HDL-c) | rs1348206 | FAM47E | a | LTST | 0.009 | 0.013 | 5.18E-01 | MULTI |
| ln(HDL-c) | rs2544681 | MIR4280 | a | LTST | 0.000 | 0.014 | 9.74E-01 | MULTI |
| ln(HDL-c) | rs4296102 | MIR331 | t | LTST | -0.020 | 0.011 | 6.70E-02 | MULTI |
| ln(HDL-c) | rs6501801 | KIAA0195 | t | LTST | 0.009 | 0.011 | 4.15E-01 | MULTI/EA |
| ln(HDL-c) | rs7462612 | ATP6V1H | a | LTST | 0.001 | 0.011 | 9.33E-01 | MULTI |
| ln(HDL-c) | rs7799249 | ATP6V0A4 | a | LTST | 0.005 | 0.011 | 6.09E-01 | MULTI/EA |
| LDL-c | rs1466848 | BOC | t | LTST | -0.013 | 0.018 | 4.71E-01 | MULTI |
| LDL-c | rs2821357 | NR5A2 | t | LTST | 0.005 | 0.014 | 7.02E-01 | MULTI |
| LDL-c | rs4075349 | IGFBP7-AS1 | t | LTST | 0.014 | 0.009 | 1.23E-01 | MULTI |
| LDL-c | rs4926854 | FOXD2 | t | LTST | 0.005 | 0.009 | 5.64E-01 | MULTI |
| ln(TG) | rs1857237 | *RNU5F-1* | t | LTST | -0.006 | 0.014 | 6.63E-01 | MULTI |
| ln(TG) | rs2801439 | *SLC35F3* | t | LTST | -0.003 | 0.010 | 7.74E-01 | MULTI |
| ln(TG) | rs296363 | *SULT2A1* | c | LTST | 0.017 | 0.011 | 1.43E-01 | MULTI |
| ln(TG) | rs6550067 | *OSBPL10* | t | LTST | 0.006 | 0.010 | 5.65E-01 | MULTI |
| ln(TG) | rs6800190 | *MIR4790* | c | LTST | 0.006 | 0.010 | 5.67E-01 | MULTI |
| ln(TG) | rs7965852 | *PDE3A* | a | LTST | 0.008 | 0.009 | 3.85E-01 | MULTI |
| ln(TG) | rs8041815 | *ADAMTS17* | a | LTST | -0.020 | 0.010 | 3.45E-02 | MULTI |
| ln(HDL-c) | rs1111341 | *SMARCAL1* | t | STST | -0.002 | 0.009 | 7.97E-01 | MULTI |
| ln(HDL-c) | rs12695617 | *EPHB1* | a | STST | 0.007 | 0.010 | 4.41E-01 | MULTI |
| ln(HDL-c) | rs2080208 | *CLEC2D* | a | STST | -0.020 | 0.018 | 2.62E-01 | MULTI |
| ln(HDL-c) | rs2594136 | *FHIT* | a | STST | -0.004 | 0.013 | 7.53E-01 | MULTI/EA |
| ln(HDL-c) | rs4778087 | *RGMA* | a | STST | 0.006 | 0.010 | 5.85E-01 | MULTI |
| ln(HDL-c) | rs6672390 | *S100A6* | t | STST | 0.006 | 0.014 | 6.78E-01 | MULTI |
| ln(HDL-c) | rs903269 | *SVILP1* | t | STST | 0.015 | 0.009 | 1.09E-01 | MULTI |
| LDL-c | rs10244093 | *MAGI2* | t | STST | 0.003 | 0.011 | 8.12E-01 | MULTI |
| LDL-c | rs12598569 | *SNX29* | t | STST | 0.002 | 0.010 | 8.24E-01 | MULTI/EA |
| LDL-c | rs429921 | *VAT1L* | a | STST | 0.003 | 0.010 | 7.70E-01 | MULTI/EA |
| LDL-c | rs4733156 | *FUT10* | a | STST | -0.004 | 0.010 | 7.07E-01 | MULTI |
| LDL-c | rs681554 | *KLHL31* | t | STST | 0.010 | 0.013 | 4.55E-01 | MULTI |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| LDL-c | rs7679068 | *ZNF827* | t | STST | -0.029 | 0.010 | 2.27E-03 | MULTI |
| LDL-c | rs8077967 | *METRNL* | t | STST | 0.005 | 0.014 | 7.39E-01 | MULTI |
| LDL-c | rs9906564 | *GPRC5C* | t | STST | 0.011 | 0.011 | 3.24E-01 | MULTI |
| ln(TG) | rs10019234 | *PCDH18* | a | STST | -0.005 | 0.012 | 6.81E-01 | MULTI |
| ln(TG) | rs1058029 | *TMX4* | a | STST | 0.014 | 0.010 | 1.59E-01 | MULTI |
| ln(TG) | rs10744213 | *TMEM132B* | a | STST | -0.006 | 0.011 | 5.99E-01 | MULTI |
| ln(TG) | rs10789347 | *RP4-660H19.1* | t | STST | 0.015 | 0.017 | 3.79E-01 | MULTI |
| ln(TG) | rs11648341 | *ACSM2B* | a | STST | -0.018 | 0.011 | 1.17E-01 | MULTI |
| ln(TG) | rs11774568 | *DEFB136* | a | STST | 0.010 | 0.011 | 3.65E-01 | MULTI |
| ln(TG) | rs1447523 | *YPEL5* | a | STST | -0.020 | 0.009 | 3.15E-02 | MULTI |
| ln(TG) | rs1606045 | *AC092635.1* | t | STST | -0.004 | 0.013 | 7.52E-01 | MULTI |
| ln(TG) | rs2714658 | *MIR548M* | c | STST | 0.008 | 0.010 | 4.07E-01 | MULTI |
| ln(TG) | rs291821 | *LINC01340* | t | STST | -0.012 | 0.015 | 4.29E-01 | MULTI |
| ln(TG) | rs3826692 | *MYO9B* | a | STST | -0.011 | 0.010 | 2.73E-01 | MULTI |
| ln(TG) | rs4849021 | *AC097499.1* | a | STST | -0.015 | 0.012 | 2.08E-01 | MULTI |
| ln(TG) | rs5746495 | *MICAL3* | t | STST | -0.004 | 0.010 | 7.28E-01 | MULTI |
| ln(TG) | rs7924896 | *METTL15* | a | STST | -0.012 | 0.009 | 2.10E-01 | MULTI |
| ln(TG) | rs970908 | *LINC01289* | a | STST | 0.000 | 0.009 | 9.68E-01 | MULTI |
| ln(TG) | rs1830079 | *ZNF273* | t | STST | -0.006 | 0.021 | 7.74E-01 | EU |
| ln(TG) | rs4394754 | *INPP5A* | t | STST | 0.001 | 0.011 | 9.20E-01 | EU |
| ln(HDL-c) | rs7689238 | *AFAP1* | t | STST | -0.003 | 0.009 | 7.61E-01 | EU |
| ln(TG) | rs7903921 | GLRX3 | a | STST | -0.006 | 0.010 | 5.43E-01 | EU |
| ln(HDL-c) | rs1436345 | *ALCAM* | a | LTST | -0.004 | 0.010 | 6.97E-01 | EU |
| LDL-c | rs181559417 | *DAPL1* | t | LTST | 0.052 | 0.052 | 3.15E-01 | EU |
| ln(HDL-c) | rs1873027 | *LRRN1* | a | LTST | 0.002 | 0.010 | 8.20E-01 | EU |
| ln(HDL-c) | rs3938236 | *SPRED1* | a | LTST | 0.006 | 0.015 | 6.70E-01 | EU |
| LDL-c | rs6461337 | *SNX13* | t | LTST | -0.003 | 0.009 | 7.40E-01 | EU |
| ln(HDL-c) | rs75543966 | *DHX8* | c | LTST | 0.071 | 0.049 | 1.48E-01 | EU |
| ln(TG) | rs970223 | *AKR1C7P* | a | LTST | -0.006 | 0.013 | 6.53E-01 | EU |

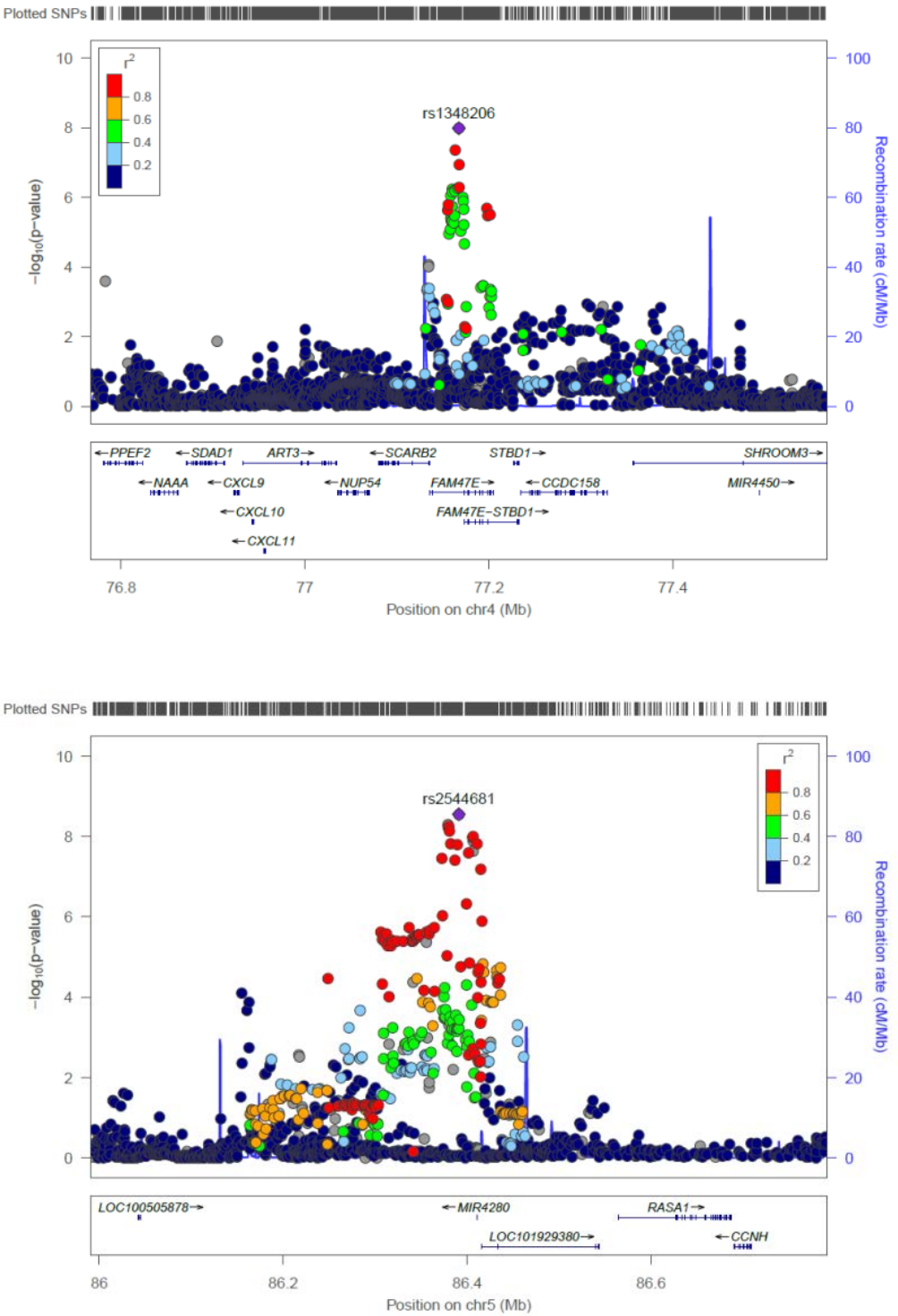
**Supplementary Figure 1:** QQ plots of the multi-ancestry meta-analyses of stage 1 and 2 studies

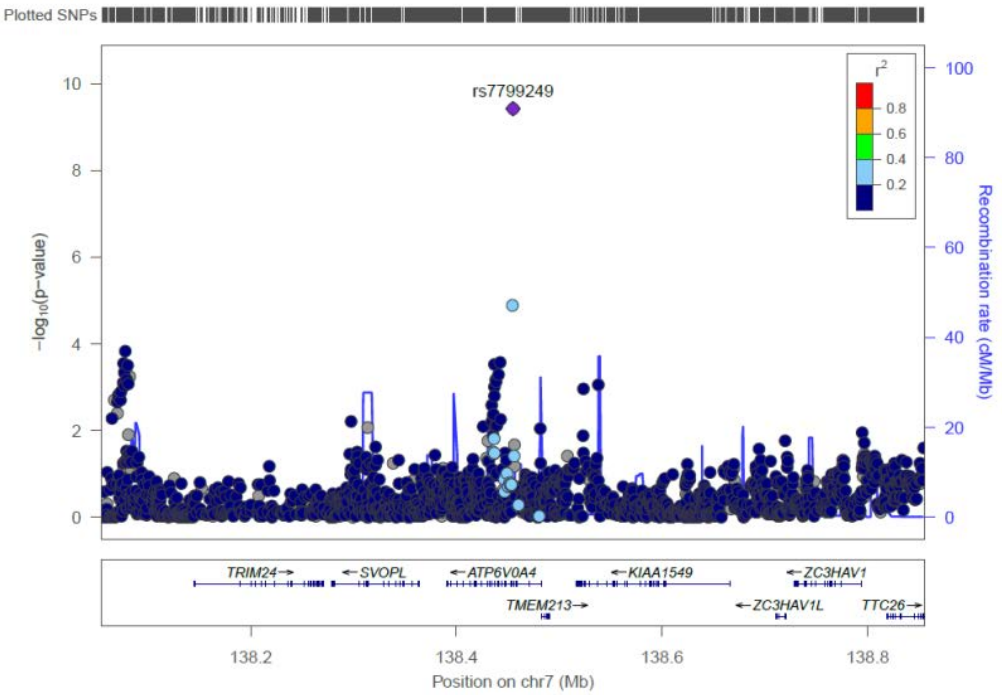
| HDL-c – STST joint effect | HDL – LTST joint effect |
|---|---|
|  |  |
| **LDL-c – STST joint effect** | **LDL-c – LTST joint effect** |
|  |  |
| **TG – STST joint effect** | **TG – LTST joint effect** |
|  |  |

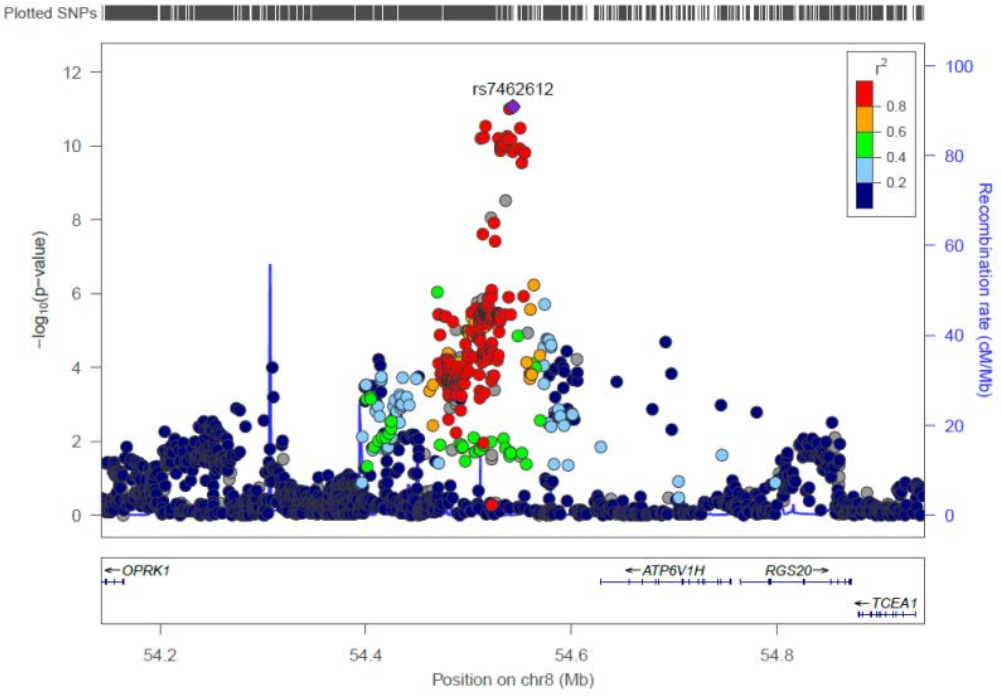
**Supplementary Figure 2:** QQ plots of the European meta-analyses of stage 1 and 2 studies

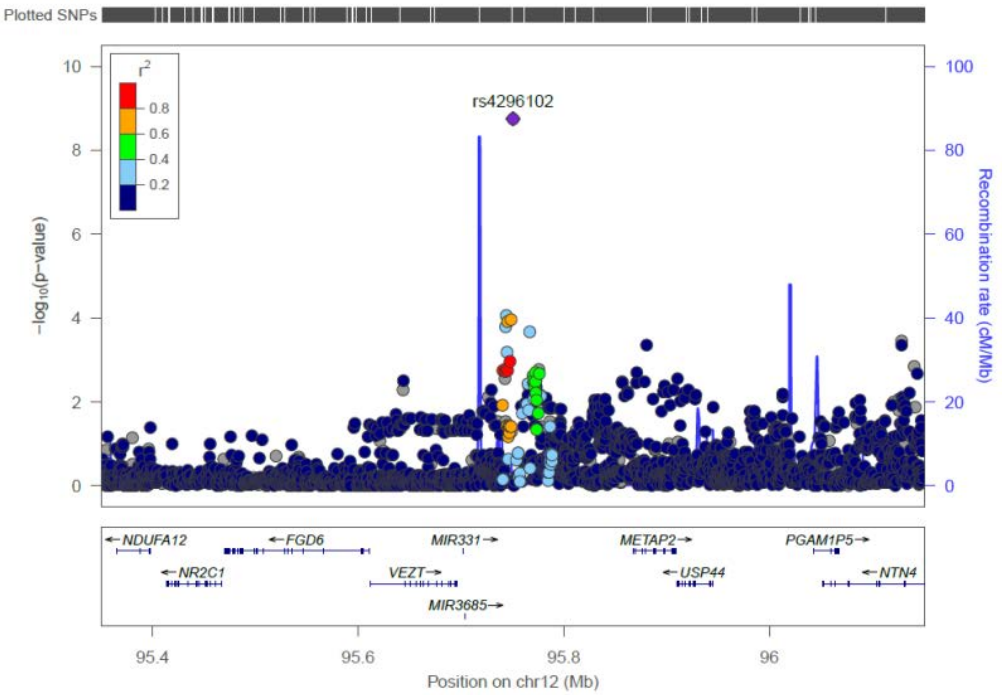
| HDL-c – STST joint effect | HDL – LTST joint effect |
|---|---|
|  |  |
| **LDL-c – STST joint effect** | **LDL-c – LTST joint effect** |
|  |  |
| **TG – STST joint effect** | **TG – LTST joint effect** |
|  |  |

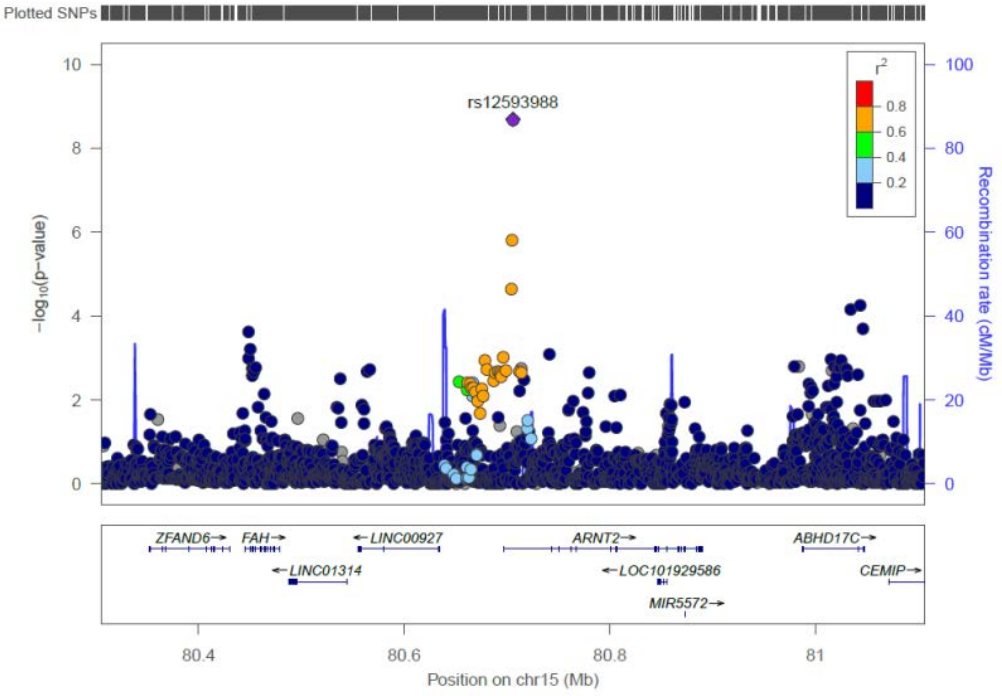
**Supplementary Figure 3:** Regional plots of the replicated novel hits of the joint effect of long total sleep time on HDL cholesterol in the multi-ancestry meta-analyses

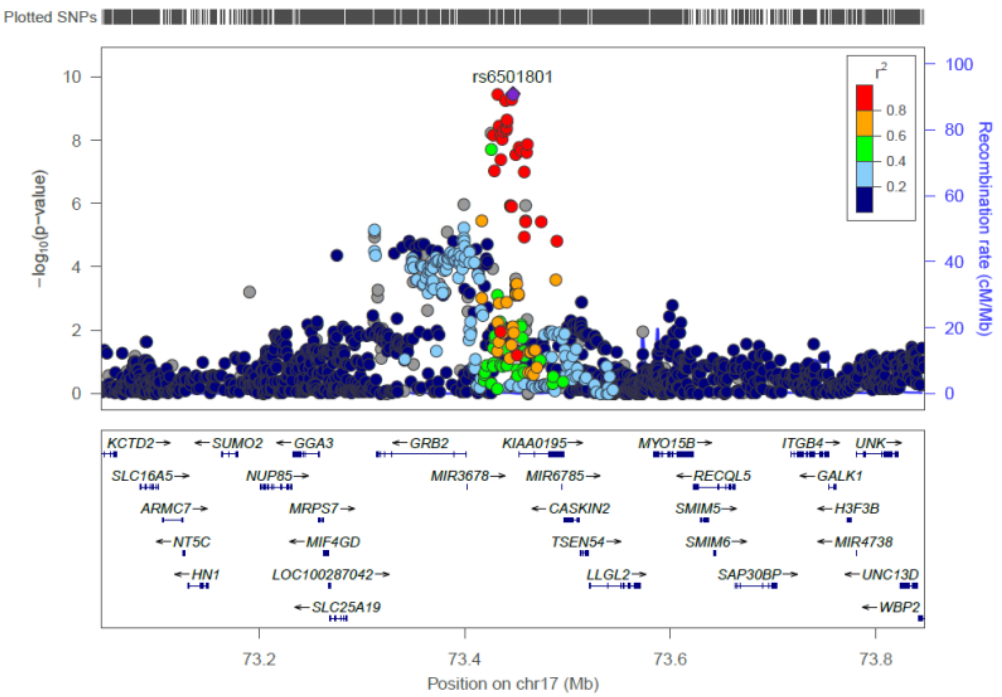



Plotted SNPs
rs7799249
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
TRIM24
SVOPL
ATP6V0A4
KIAA1549
ZC3HAV1
TMEM213
ZC3HAV1L
TTC26
Position on chr7 (Mb)

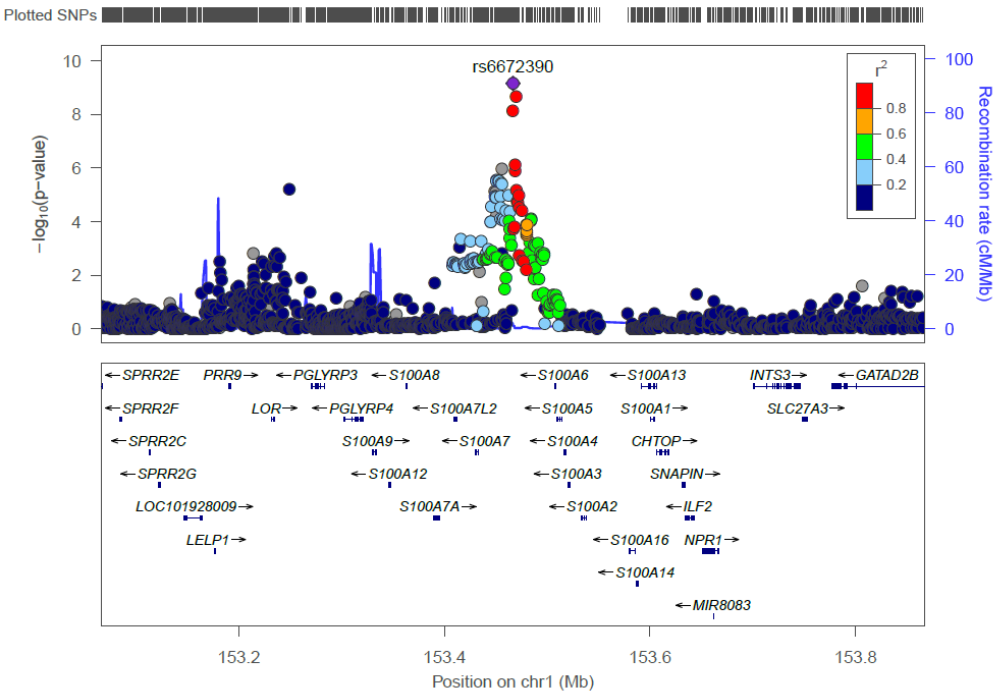

Plotted SNPs
rs7462612
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
OPRK1
ATP6V1H
RGS20
TCEA1
Position on chr8 (Mb)

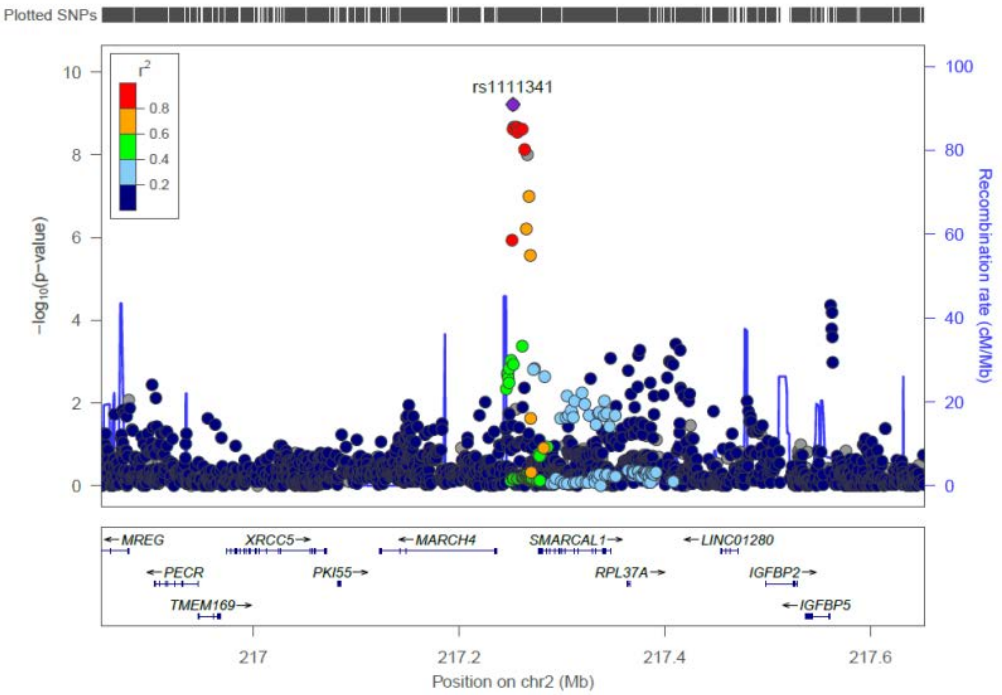

Plotted SNPs
rs4296102
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
NDUFA12
FGD6
MIR331
METAP2
PGAM1P5
NR2C1
VEZT
USP44
NTN4
MIR3685
Position on chr12 (Mb)

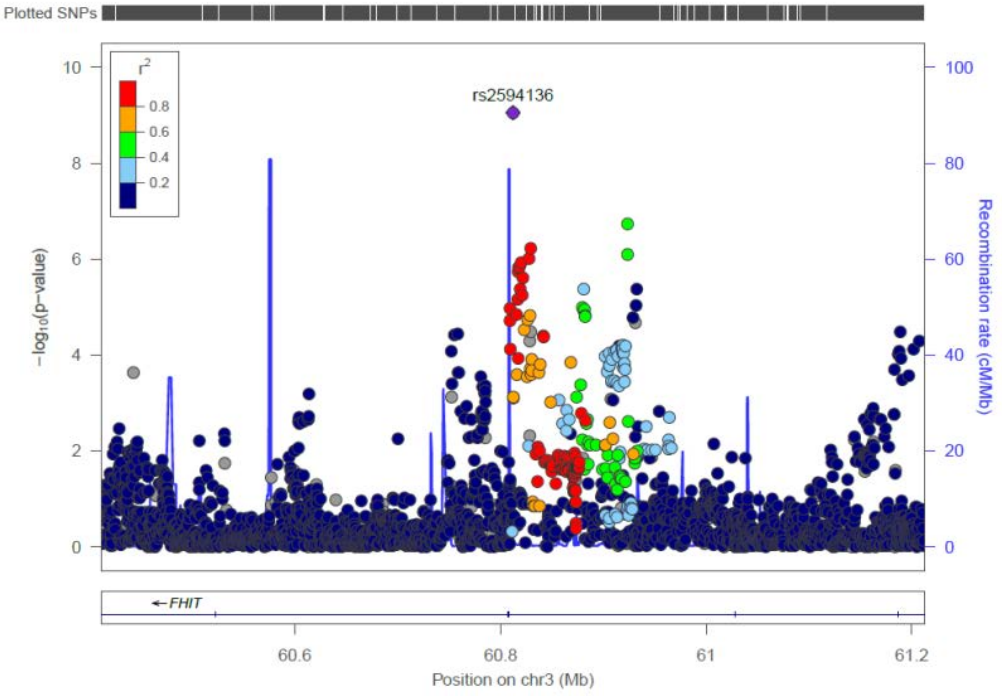

Plotted SNPs
rs12593988
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ZFAND6
FAH
LINC00927
ARNT2
ABHD17C
LINC01314
LOC101929586
CEMIP
MIR5572
Position on chr15 (Mb)

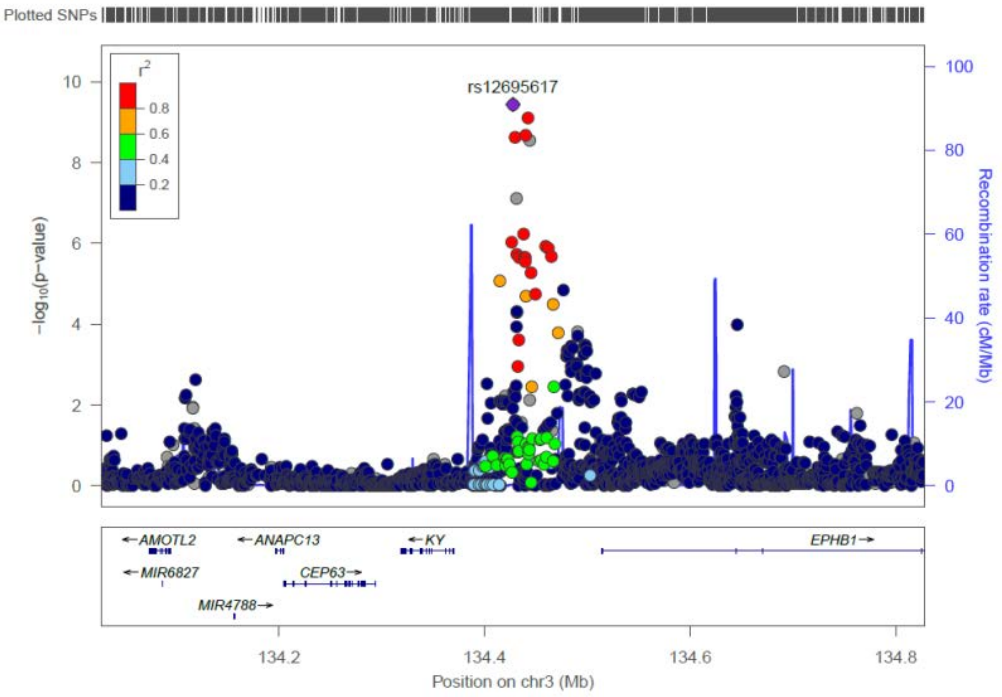

Plotted SNPs
rs6501801
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
KCTD2
SUMO2
GGA3
GRB2
KIAA0195
MYO15B
ITGB4
UNK
SLC16A5
NUP85
MIR3678
MIR6785
RECQL5
GALK1
ARMC7
MRPS7
CASKIN2
SMIM5
H3F3B
NT5C
MIF4GD
TSEN54
SMIM6
MIR4738
HN1
LOC100287042
LLGL2
SAP30BP
UNC13D
SLC25A19
WBP2
73.2
73.4
73.6
73.8
Position on chr17 (Mb)

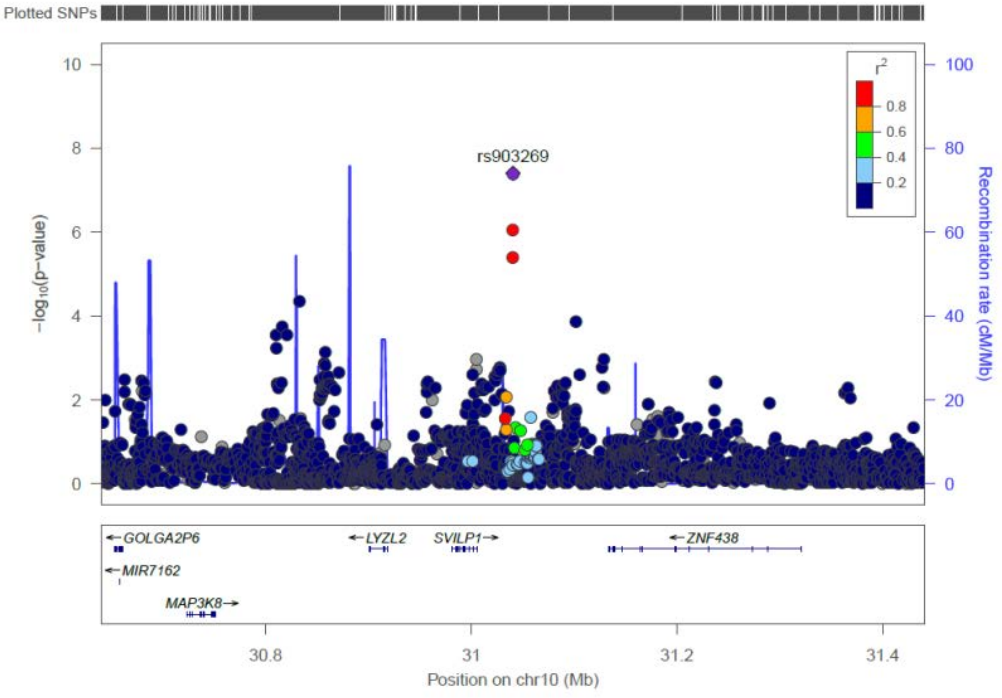
**Supplementary Figure 4:** Regional plots of the replicated novel hits of the joint effect of short total sleep time on HDL cholesterol in the multi-ancestry meta-analyses

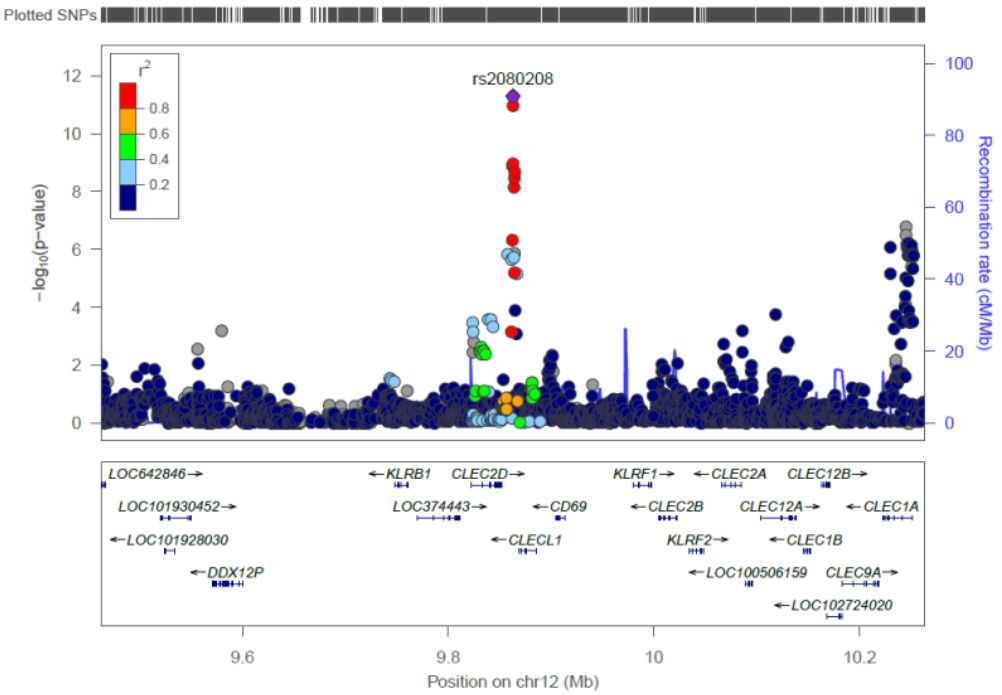


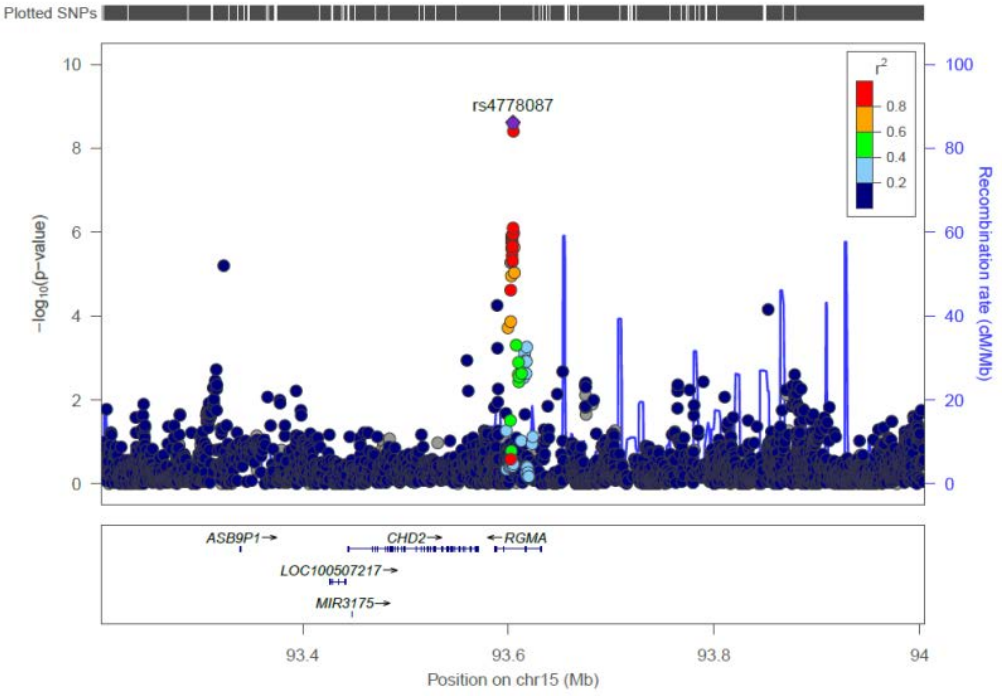

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs1111341
−log10(p−value)
Recombination rate (cM/Mb)
MREG
XRCC5
MARCH4
SMARCAL1
LINC01280
PECR
PKI55
RPL37A
IGFBP2
TMEM169
IGFBP5
Position on chr2 (Mb)


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs2594136
−log10(p−value)
Recombination rate (cM/Mb)
FHIT
Position on chr3 (Mb)

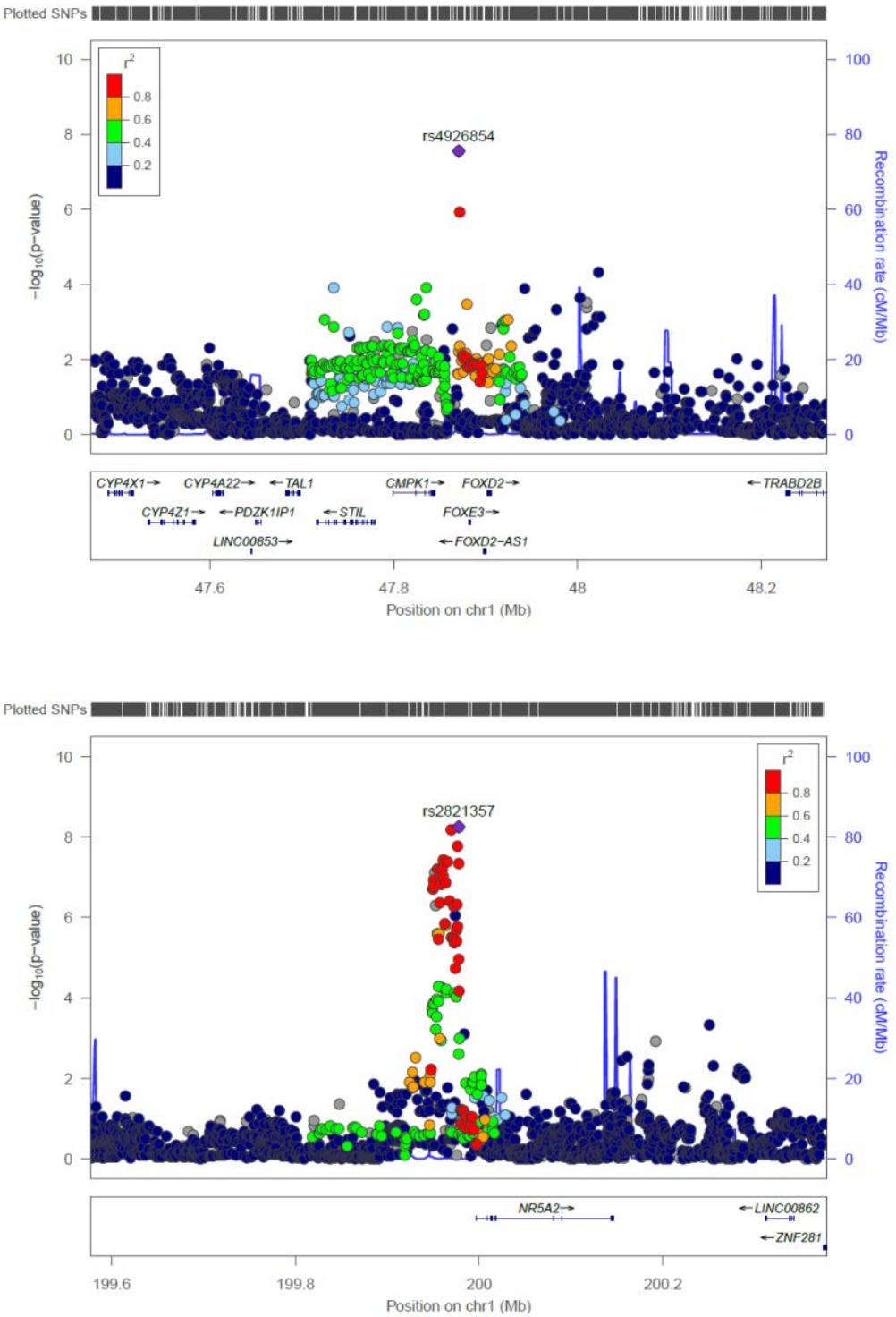

Plotted SNPs
rs12695617
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
AMOTL2
ANAPC13
KY
EPHB1
MIR6827
CEP63
MIR4788
134.2
134.4
134.6
134.8
Position on chr3 (Mb)

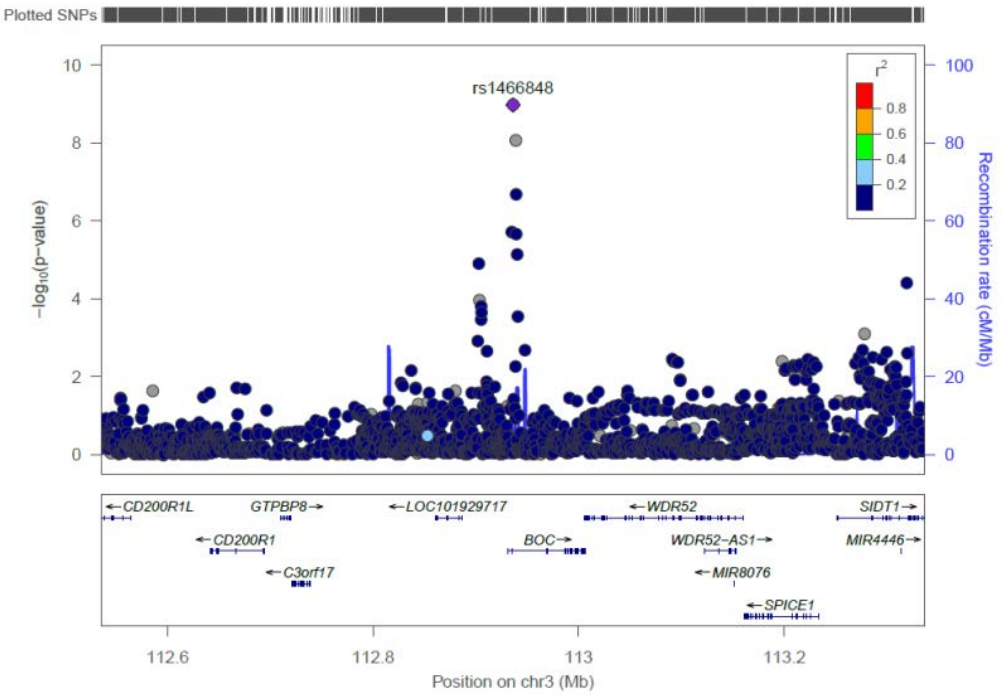

Plotted SNPs
rs903269
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
GOLGA2P6
LYZL2
SVILP1
ZNF438
MIR7162
MAP3K8
30.8
31
31.2
31.4
Position on chr10 (Mb)

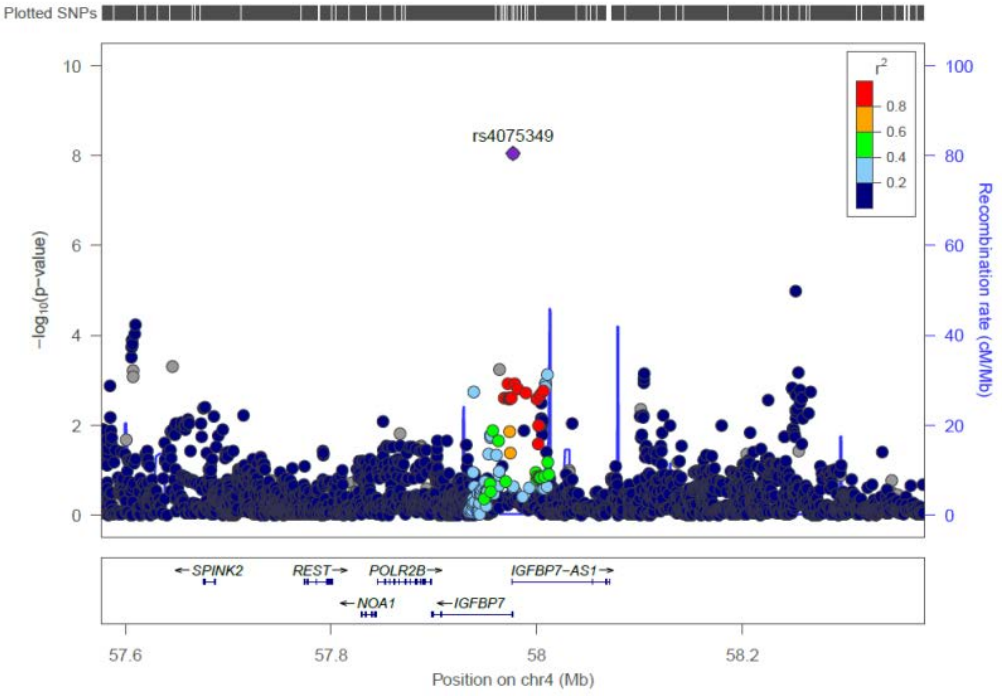

Plotted SNPs
rs2080208
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
LOC642846
KLRB1
CLEC2D
KLRF1
CLEC2A
CLEC12B
LOC101930452
LOC374443
CD69
CLEC2B
CLEC12A
CLEC1A
LOC101928030
CLECL1
KLRF2
CLEC1B
DDX12P
LOC100506159
CLEC9A
LOC102724020
Position on chr12 (Mb)


Plotted SNPs
rs4778087
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ASB9P1
CHD2
RGMA
LOC100507217
MIR3175
Position on chr15 (Mb)

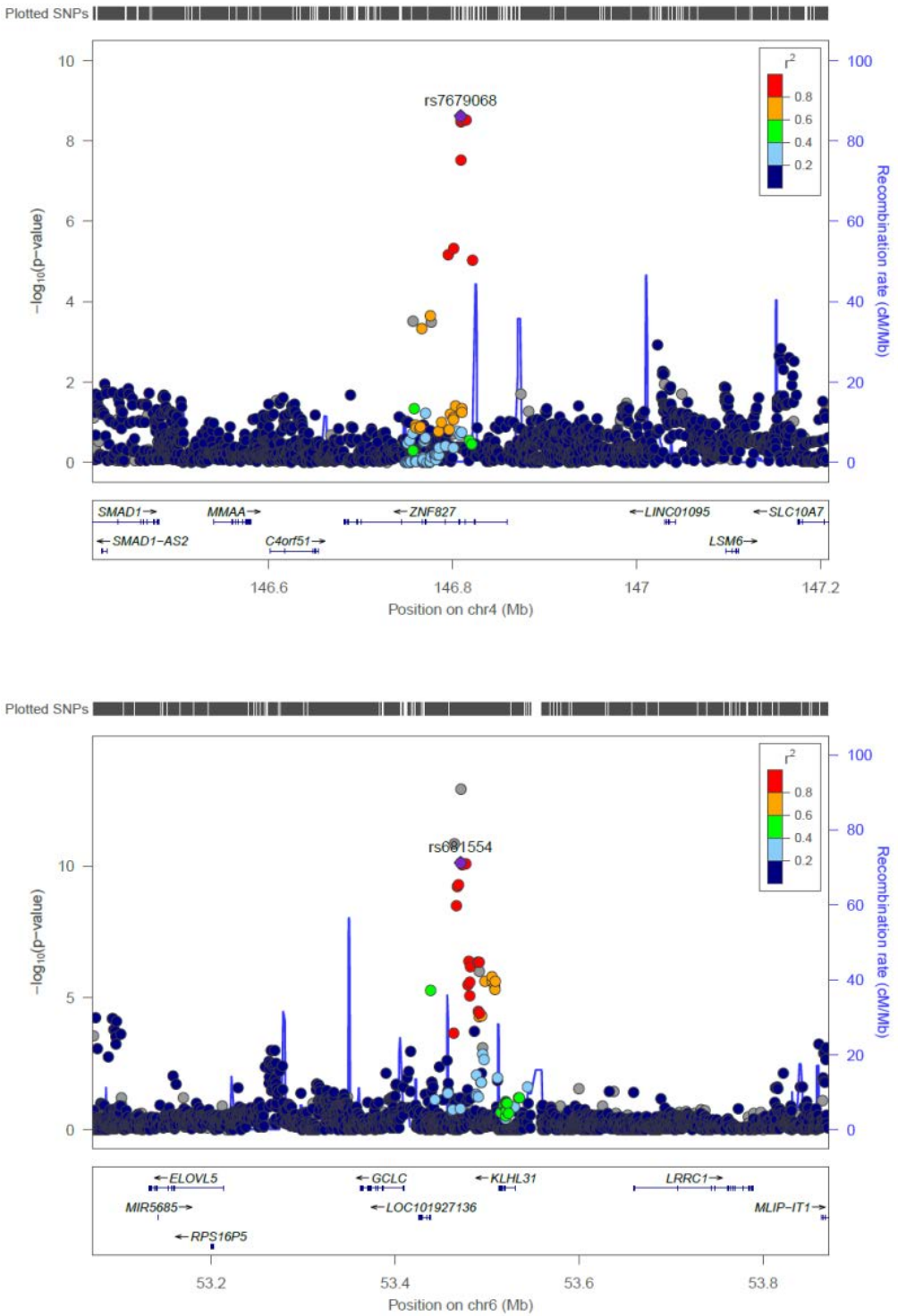
**Supplementary Figure 5:** Regional plots of the replicated novel hits of the joint effect of long total sleep time on LDL cholesterol in the multi-ancestry meta-analyses

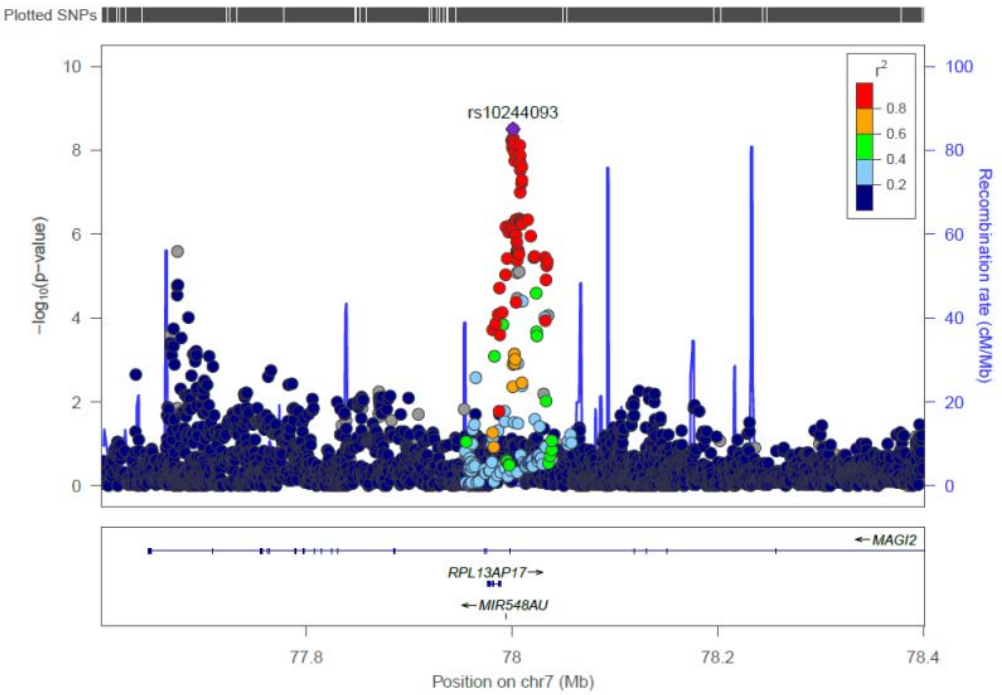


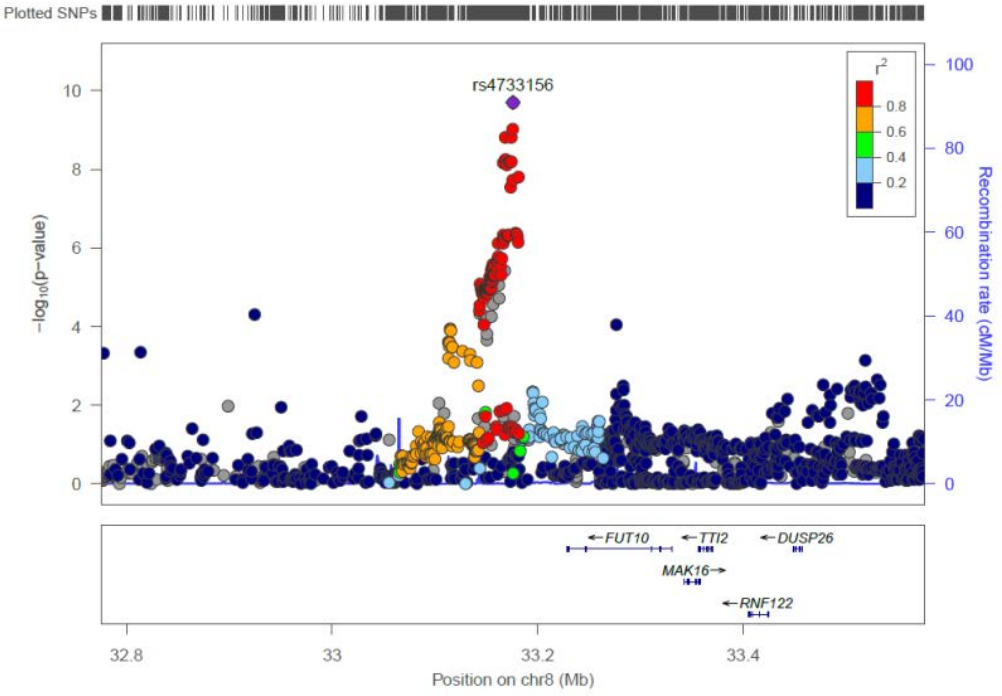

Plotted SNPs
rs1466848
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
CD200R1L
GTPBP8
LOC101929717
WDR52
SIDT1
CD200R1
BOC
WDR52−AS1
MIR4446
C3orf17
MIR8076
SPICE1
112.6
112.8
113
113.2
Position on chr3 (Mb)

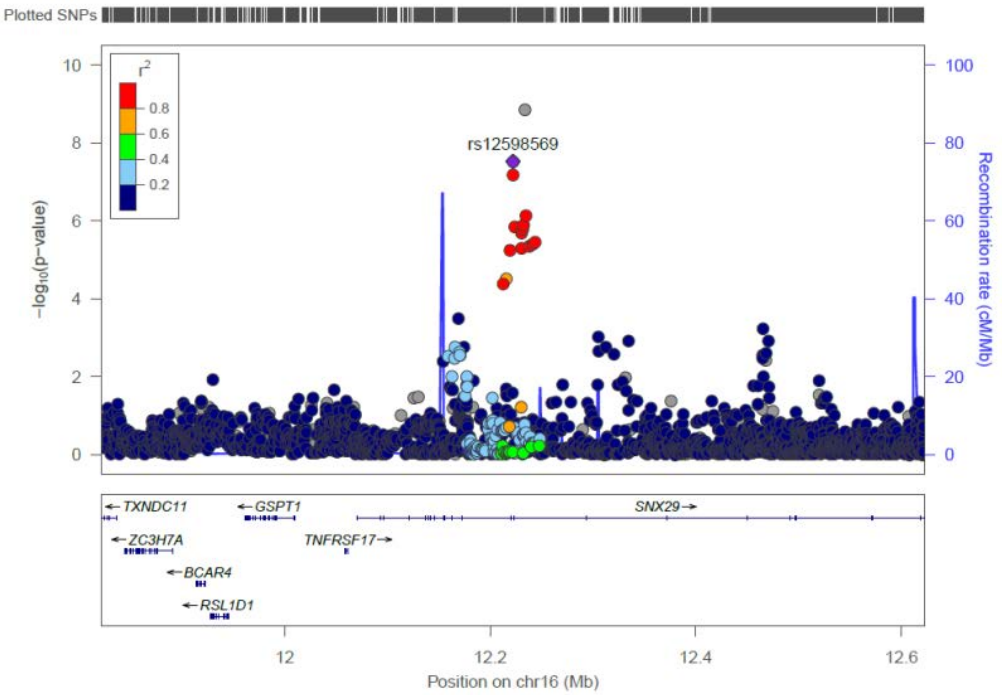

Plotted SNPs
rs4075349
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
SPINK2
REST
POLR2B
IGFBP7−AS1
NOA1
IGFBP7
57.6
57.8
58
58.2
Position on chr4 (Mb)

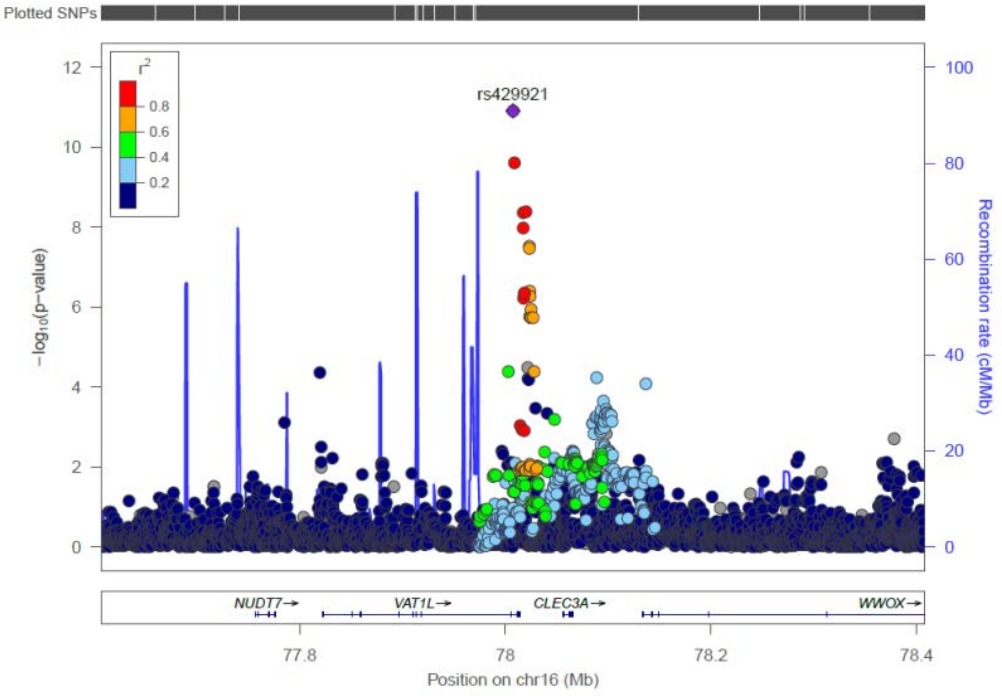
**Supplementary Figure 6:** Regional plots of the replicated novel hits of the joint effect of short total sleep time on LDL cholesterol in the multi-ancestry meta-analyses

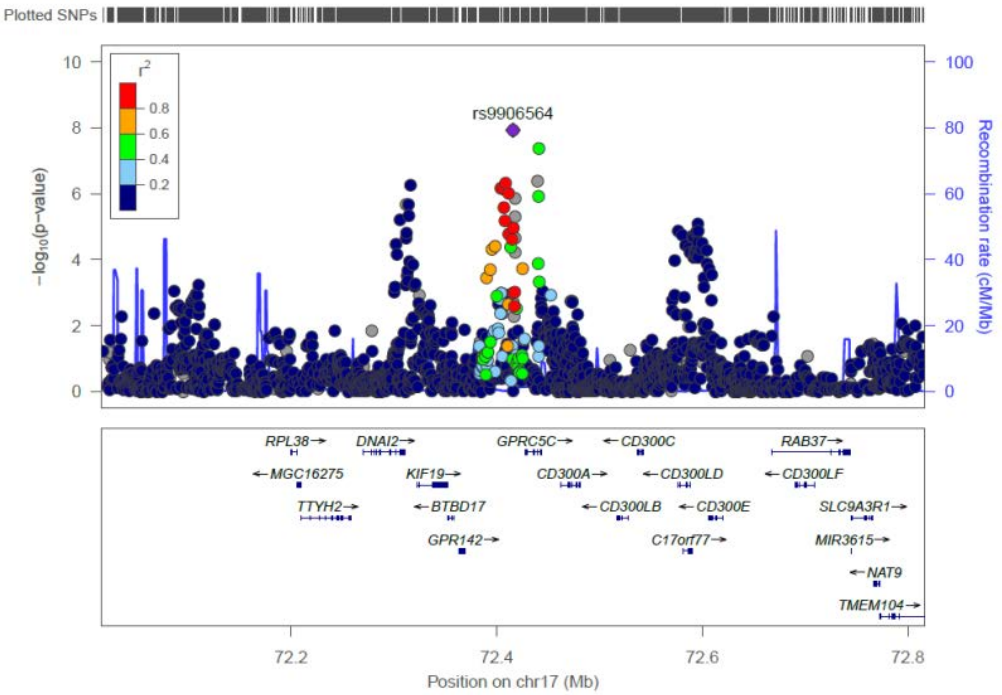


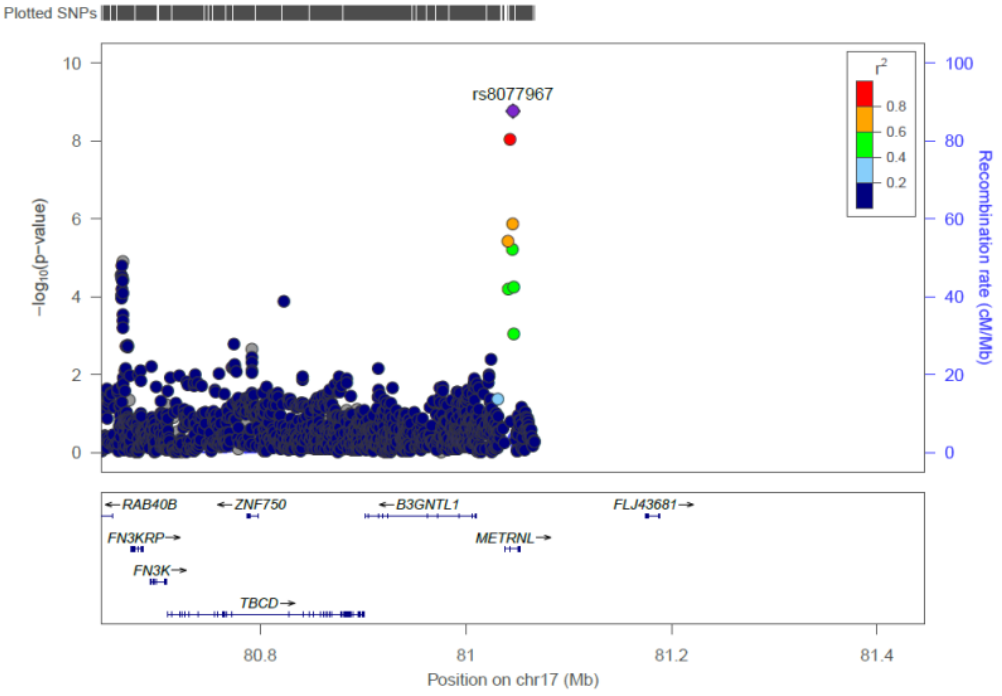

Plotted SNPs
rs10244093
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
MAGI2
RPL13AP17
MIR548AU
77.8
78
78.2
78.4
Position on chr7 (Mb)


Plotted SNPs
rs4733156
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
FUT10
TTI2
DUSP26
MAK16
RNF122
32.8
33
33.2
33.4
Position on chr8 (Mb)

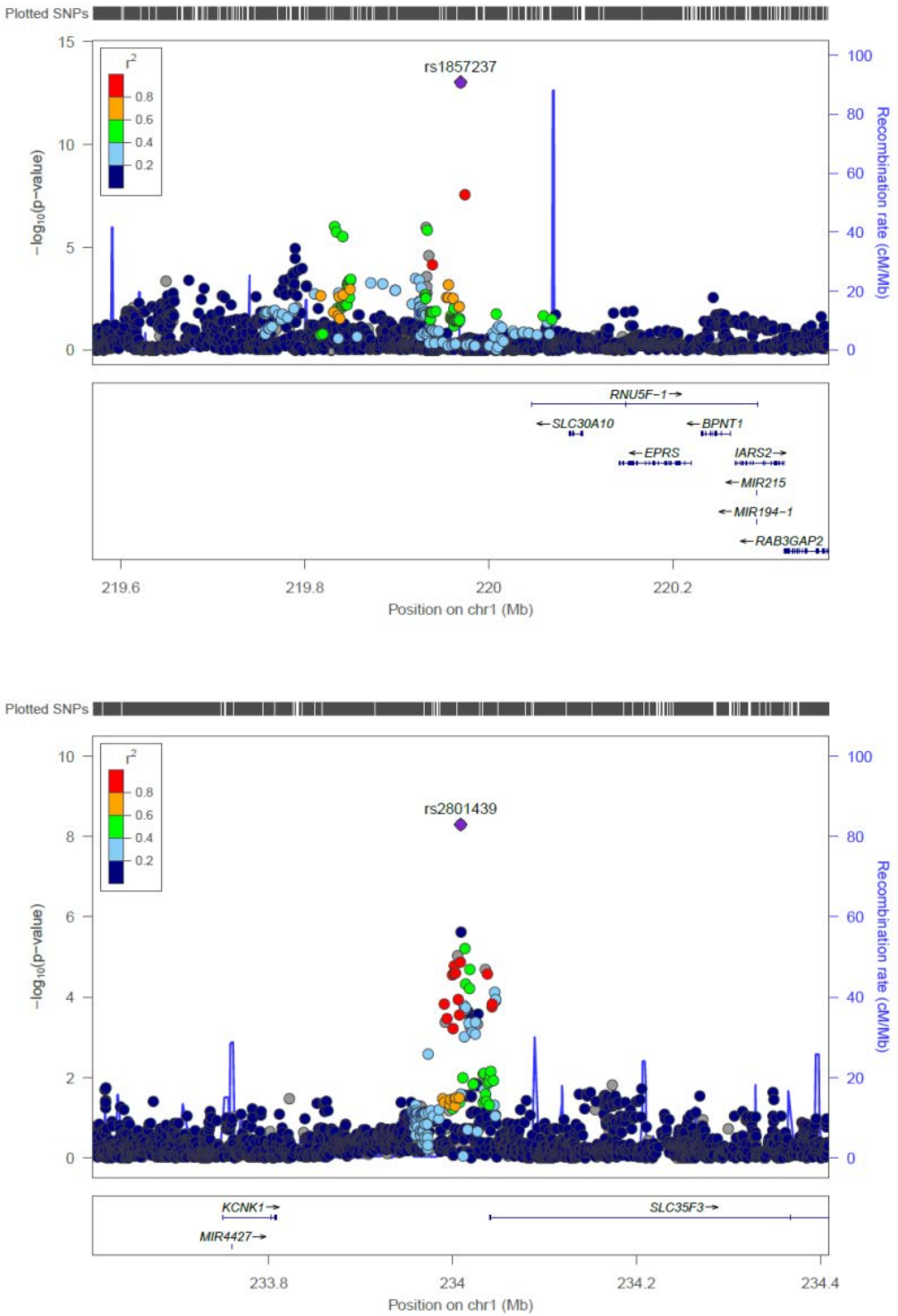

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs12598569
−log10(p−value)
Recombination rate (cM/Mb)
TXNDC11
GSPT1
SNX29
ZC3H7A
TNFRSF17
BCAR4
RSL1D1
Position on chr16 (Mb)

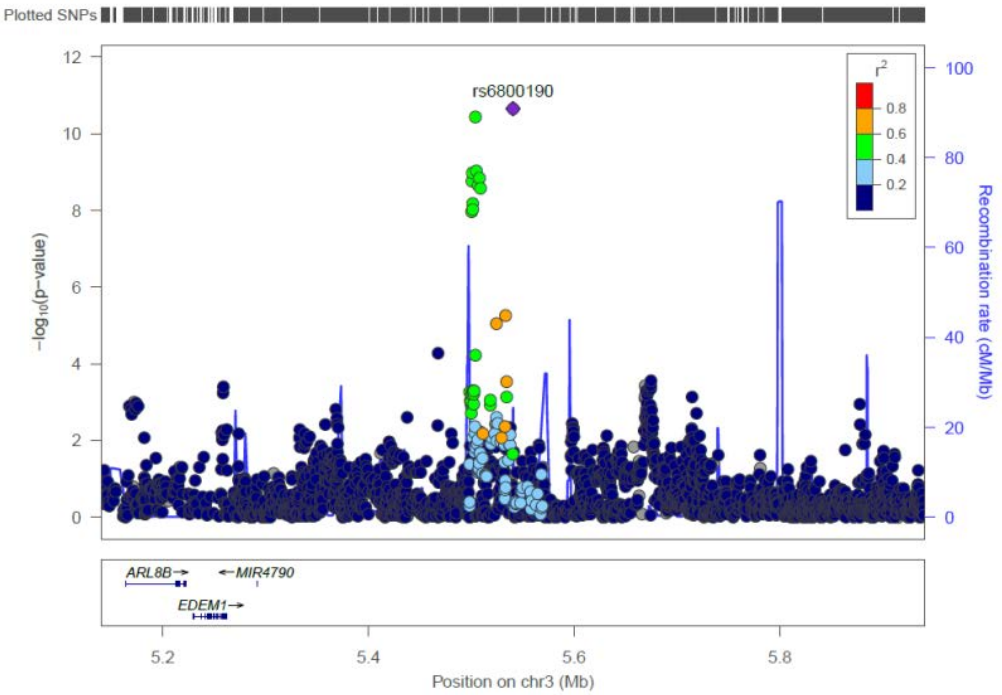

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs429921
−log10(p−value)
Recombination rate (cM/Mb)
NUDT7
VAT1L
CLEC3A
WWOX
Position on chr16 (Mb)

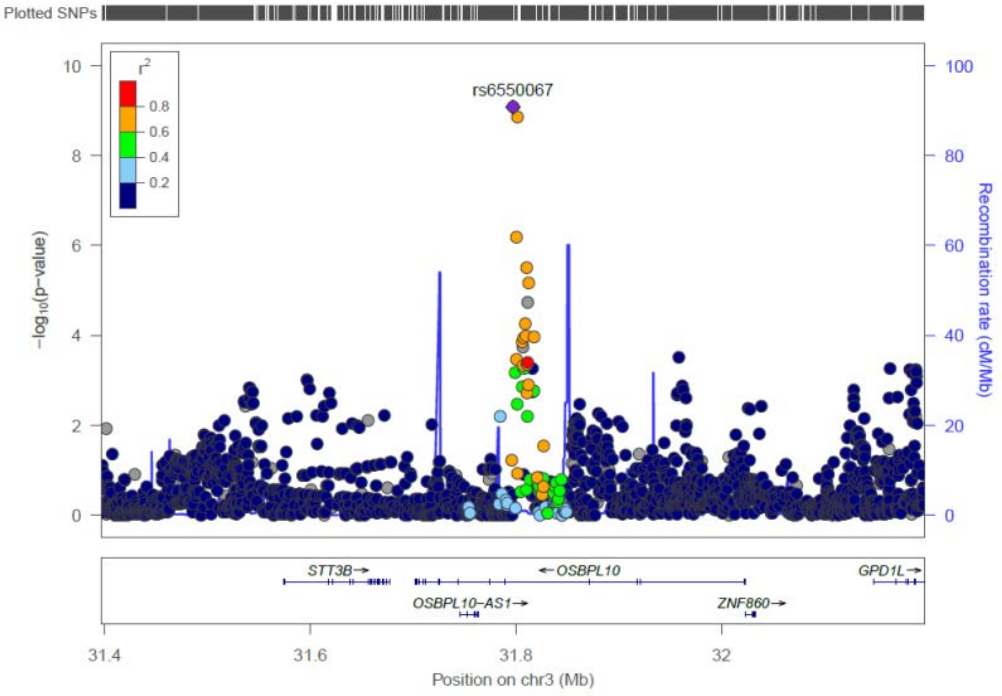

Plotted SNPs
rs9906564
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
RPL38→
DNAI2→
GPRC5C→
←CD300C
RAB37→
←MGC16275
KIF19→
CD300A→
←CD300LD
←CD300LF
TTYH2→
←BTBD17
←CD300LB
←CD300E
SLC9A3R1→
GPR142→
C17orf77→
MIR3615→
←NAT9
TMEM104→
Position on chr17 (Mb)

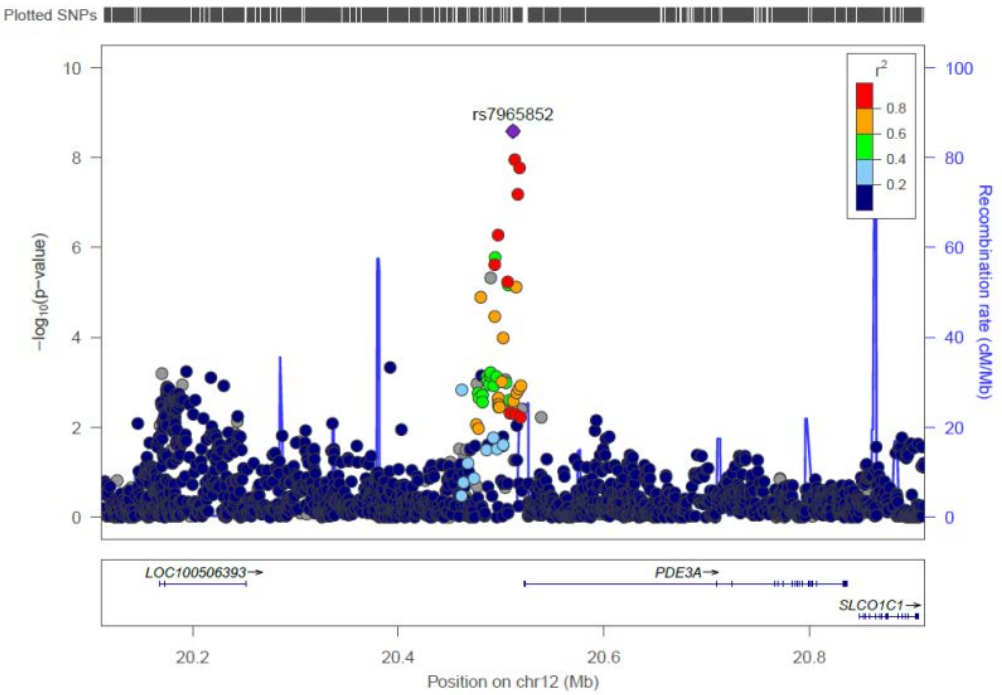

Plotted SNPs
rs8077967
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
←RAB40B
←ZNF750
←B3GNTL1
FLJ43681→
FN3KRP→
METRNL→
FN3K→
TBCD→
Position on chr17 (Mb)

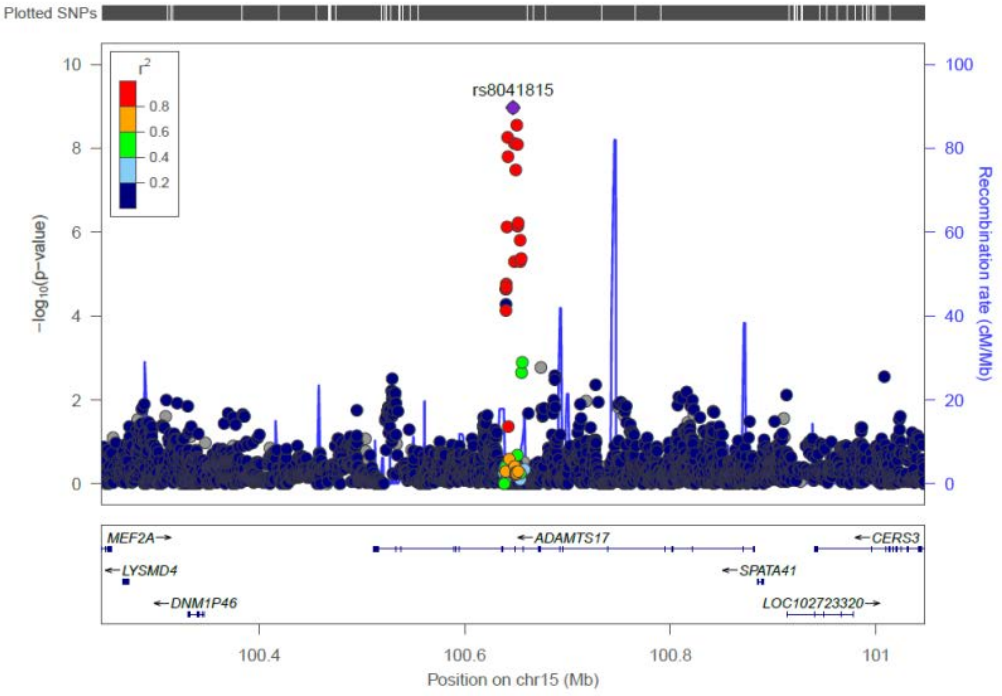
**Supplementary Figure 7:** Regional plots of the replicated novel hits of the joint effect of long total sleep time on Triglycerides in the multi-ancestry meta-analyses

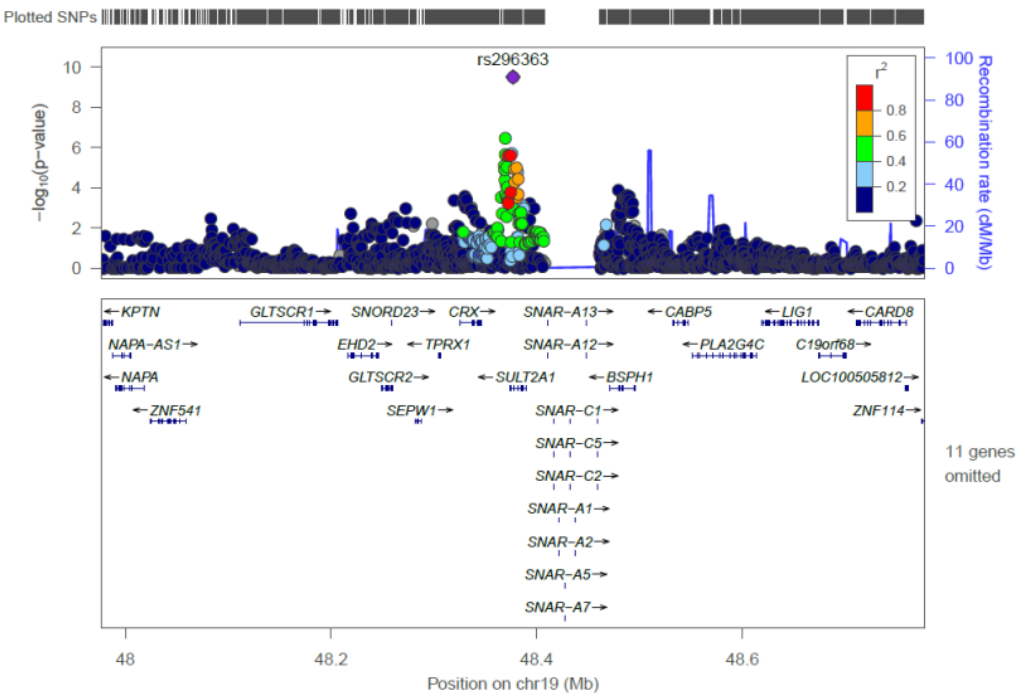



Plotted SNPs
rs6800190
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ARL8B→
←MIR4790
EDEM1→
5.2
5.4
5.6
5.8
Position on chr3 (Mb)

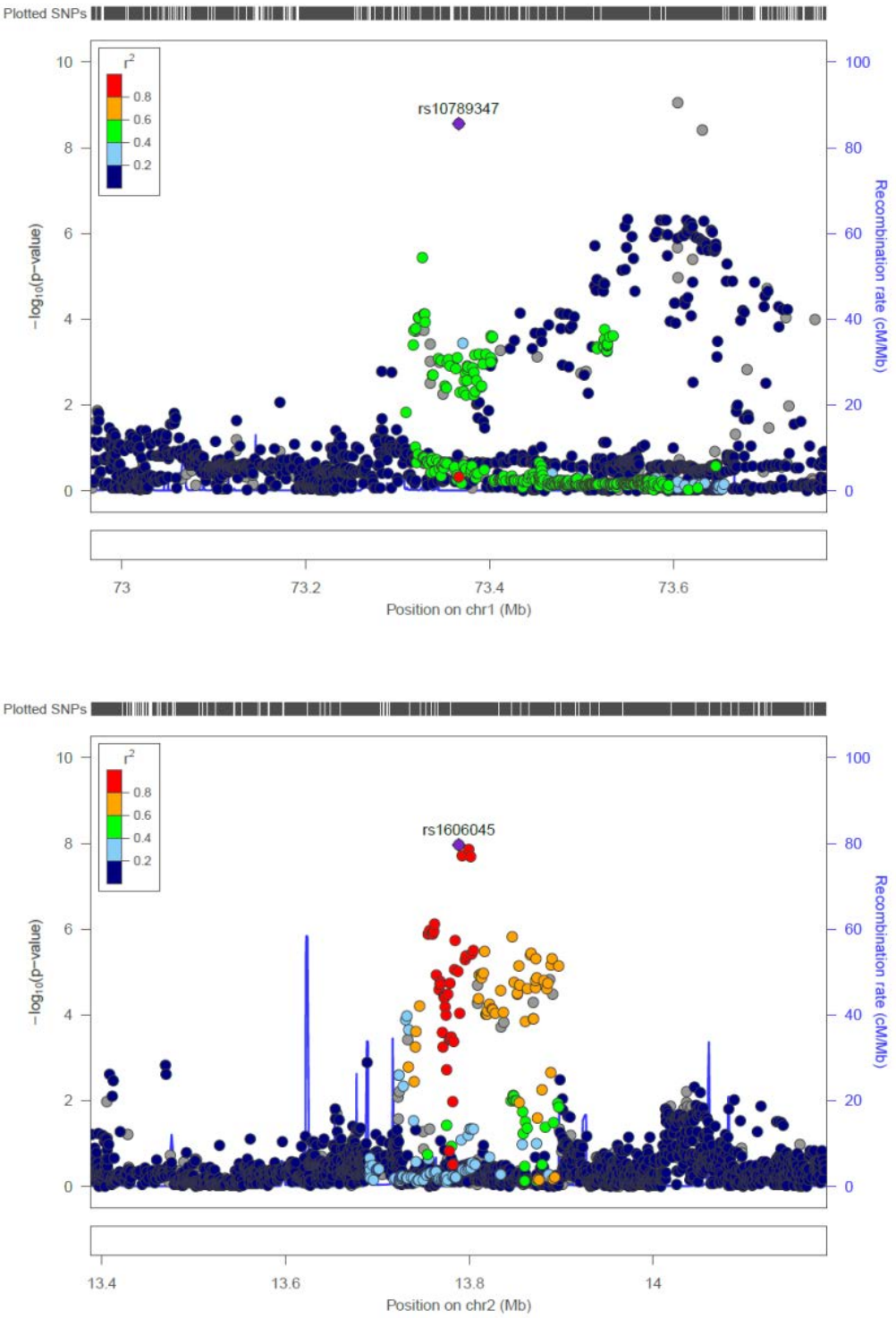

Plotted SNPs
rs6550067
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
STT3B→
←OSBPL10
GPD1L→
OSBPL10−AS1→
ZNF860→
31.4
31.6
31.8
32
Position on chr3 (Mb)

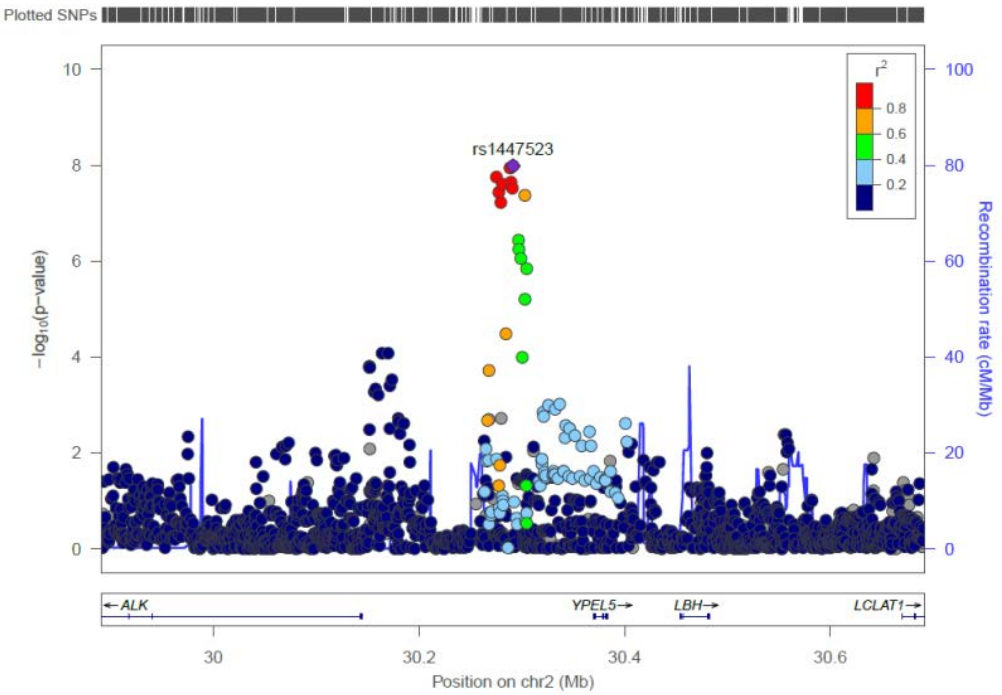

Plotted SNPs
rs7965852
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
LOC100506393
PDE3A
SLCO1C1
20.2
20.4
20.6
20.8
Position on chr12 (Mb)

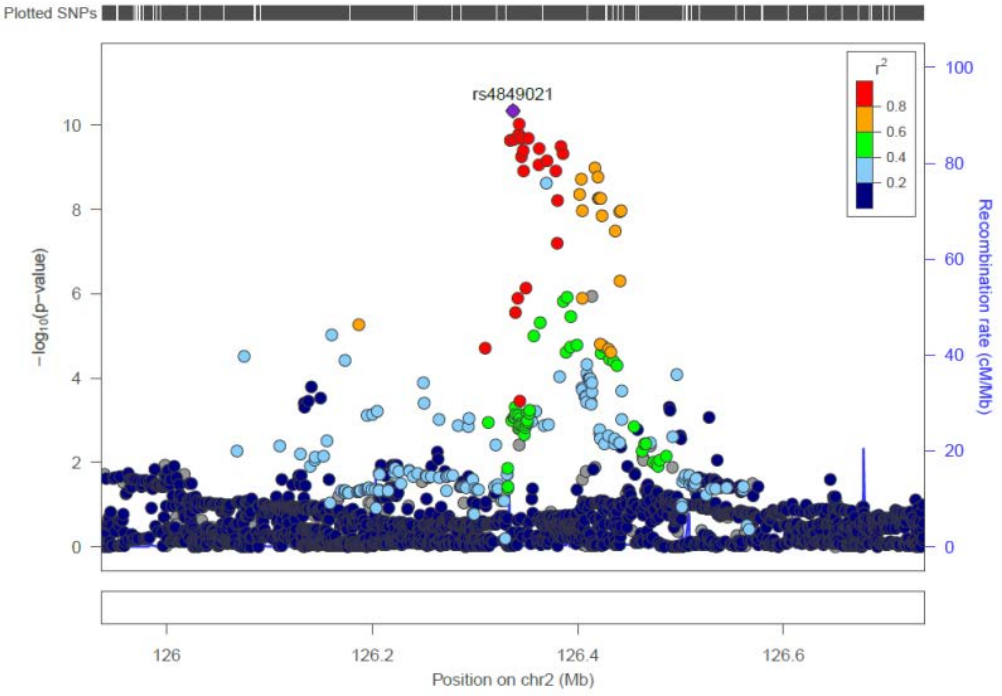

Plotted SNPs
rs8041815
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
MEF2A
ADAMTS17
CERS3
LYSMD4
SPATA41
DNM1P46
LOC102723320
100.4
100.6
100.8
101
Position on chr15 (Mb)

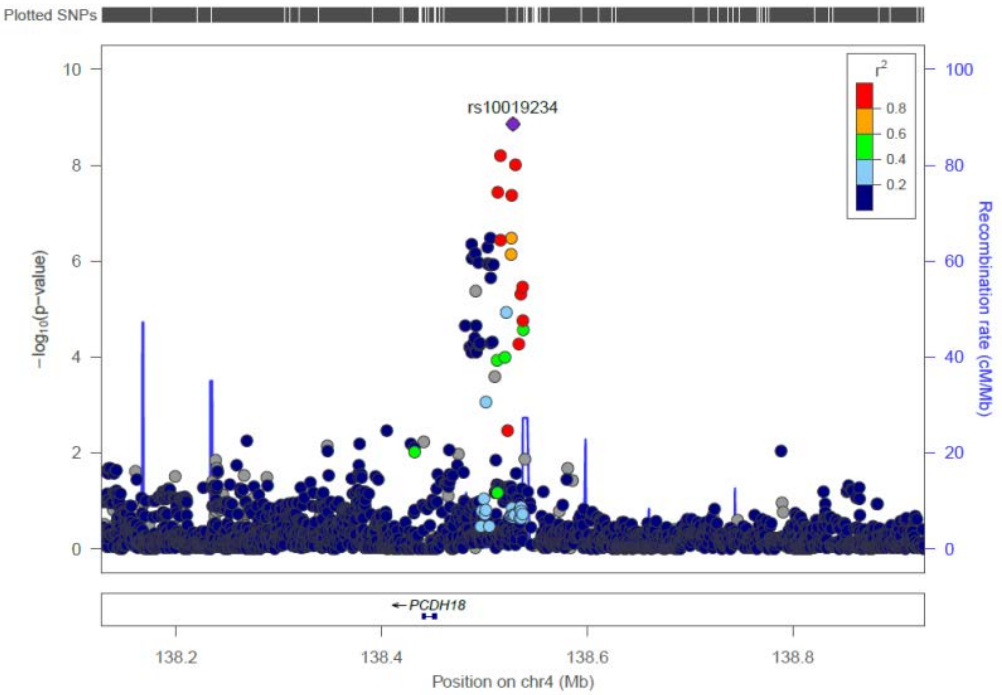

Plotted SNPs
rs296363
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
KPTN
GLTSCR1
SNORD23
CRX
SNAR−A13
CABP5
LIG1
CARD8
NAPA−AS1
EHD2
TPRX1
SNAR−A12
PLA2G4C
C19orf68
NAPA
GLTSCR2
SULT2A1
BSPH1
LOC100505812
ZNF541
SEPW1
SNAR−C1
ZNF114
SNAR−C5
SNAR−C2
SNAR−A1
SNAR−A2
SNAR−A5
SNAR−A7
11 genes omitted
48
48.2
48.4
48.6
Position on chr19 (Mb)

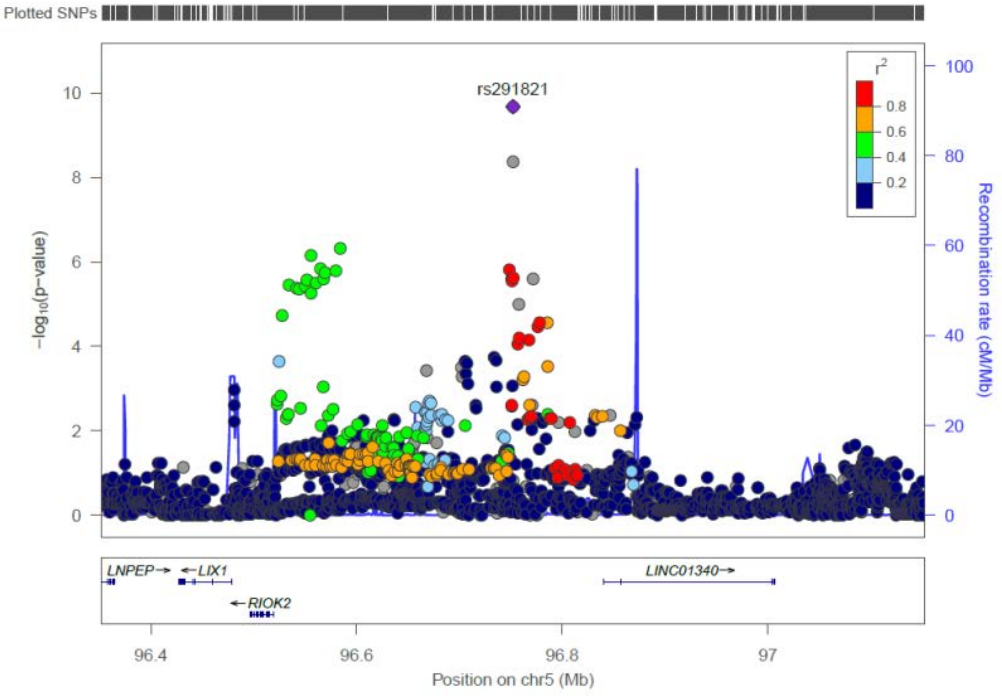
**Supplementary Figure 8:** Regional plots of the replicated novel hits of the joint effect of short total sleep time on Triglycerides in the multi-ancestry meta-analyses

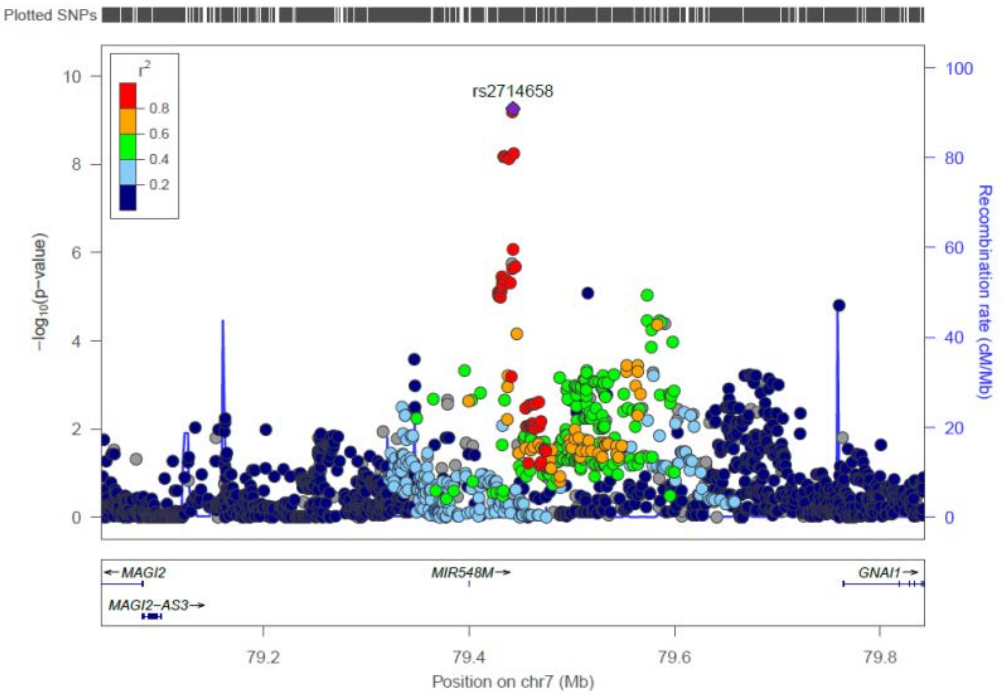


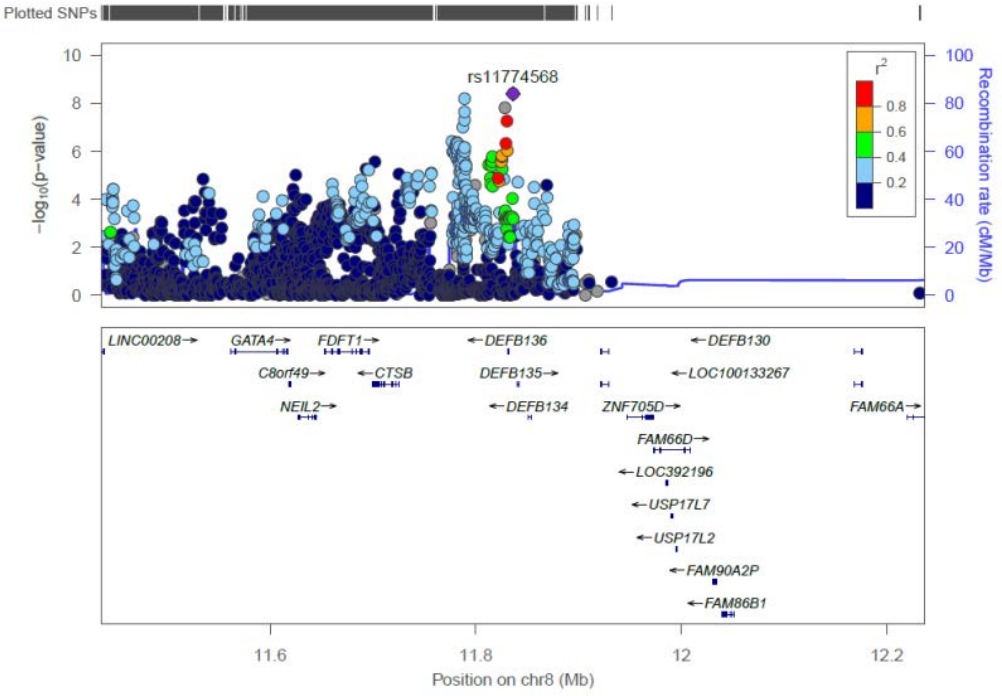

Plotted SNPs
rs1447523
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ALK
YPEL5
LBH
LCLAT1
30
30.2
30.4
30.6
Position on chr2 (Mb)

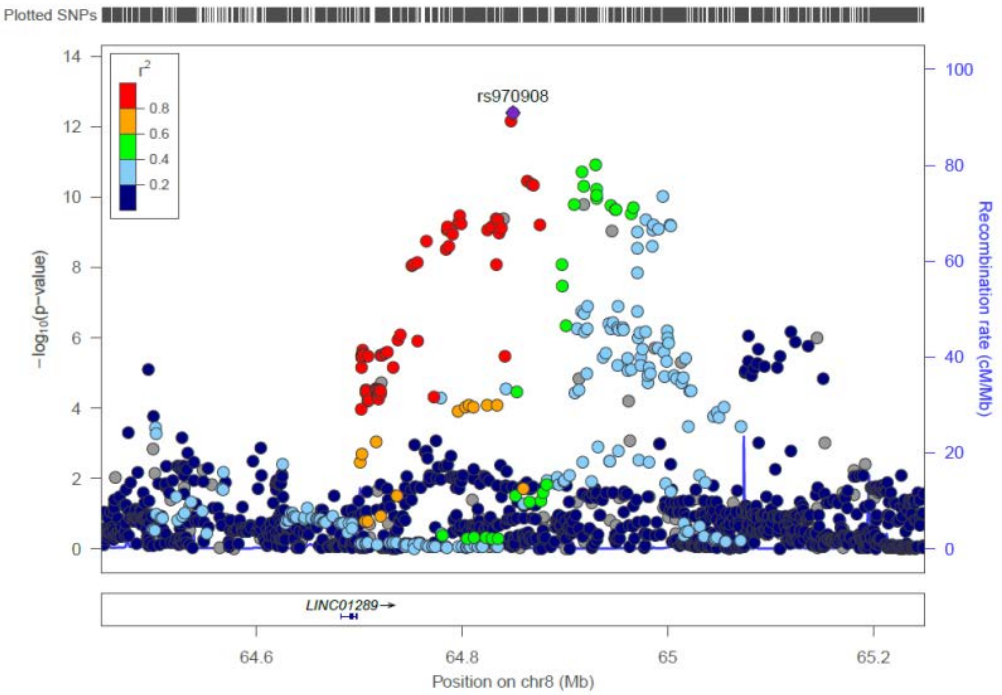

Plotted SNPs
rs4849021
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
126
126.2
126.4
126.6
Position on chr2 (Mb)

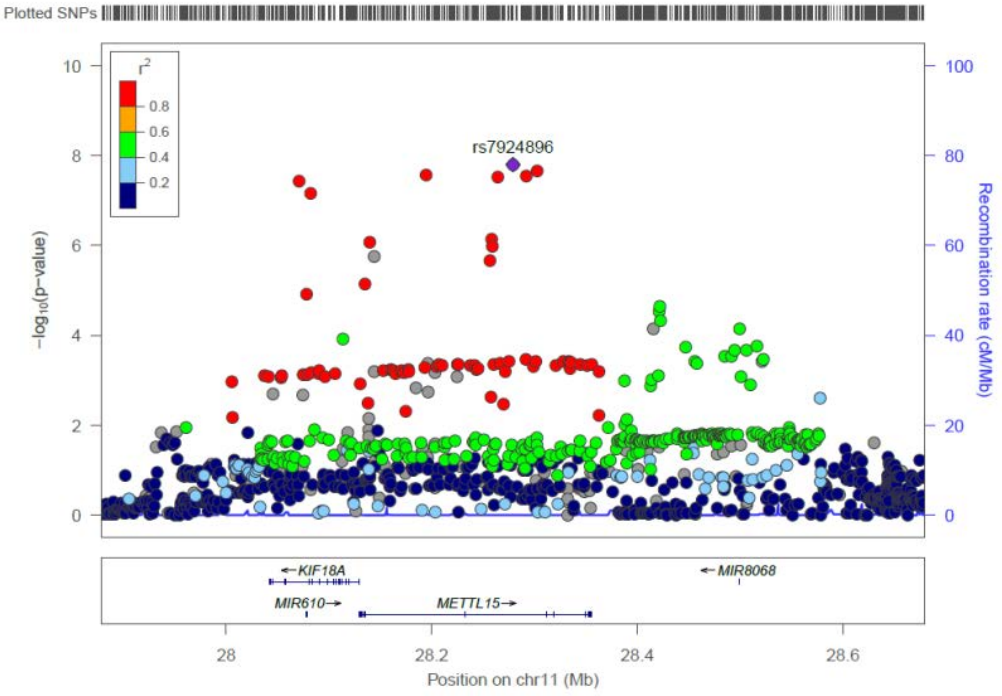

Plotted SNPs
rs10019234
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
PCDH18
138.2
138.4
138.6
138.8
Position on chr4 (Mb)

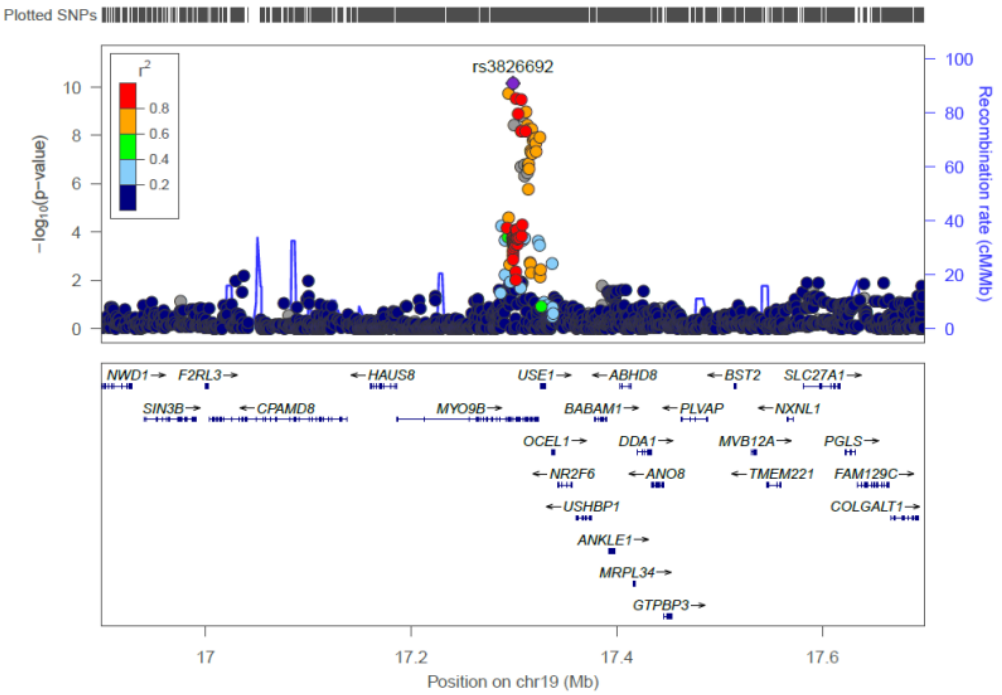

Plotted SNPs
rs291821
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
LNPEP
LIX1
RIOK2
LINC01340
96.4
96.6
96.8
97
Position on chr5 (Mb)

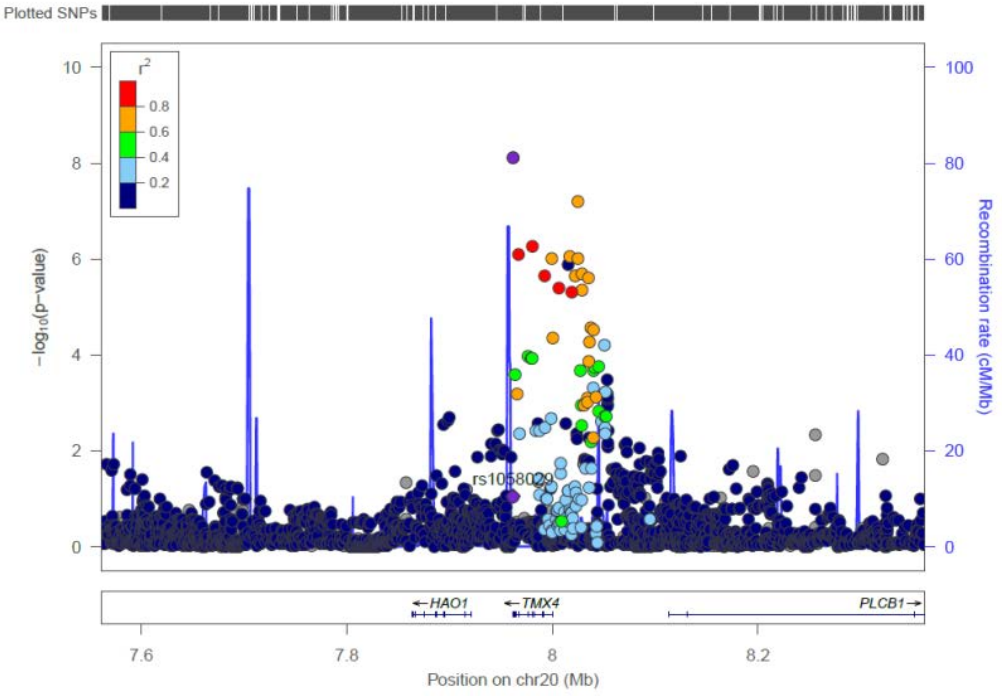

Plotted SNPs
rs2714658
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
MAGI2
MIR548M
GNAI1
MAGI2−AS3
79.2
79.4
79.6
79.8
Position on chr7 (Mb)

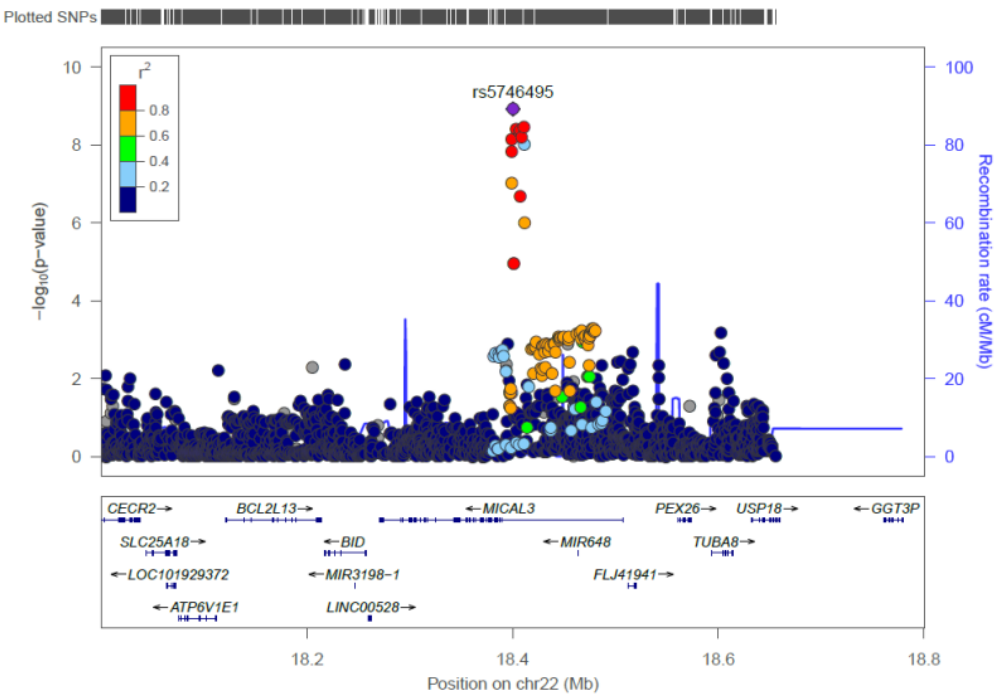

Plotted SNPs
rs11774568
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
LINC00208
GATA4
FDFT1
DEFB136
DEFB130
C8orf49
CTSB
DEFB135
LOC100133267
NEIL2
DEFB134
ZNF705D
FAM66A
FAM66D
LOC392196
USP17L7
USP17L2
FAM90A2P
FAM86B1
11.6
11.8
12
12.2
Position on chr8 (Mb)


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs970908
−log10(p−value)
Recombination rate (cM/Mb)
LINC01289
64.6
64.8
65
65.2
Position on chr8 (Mb)

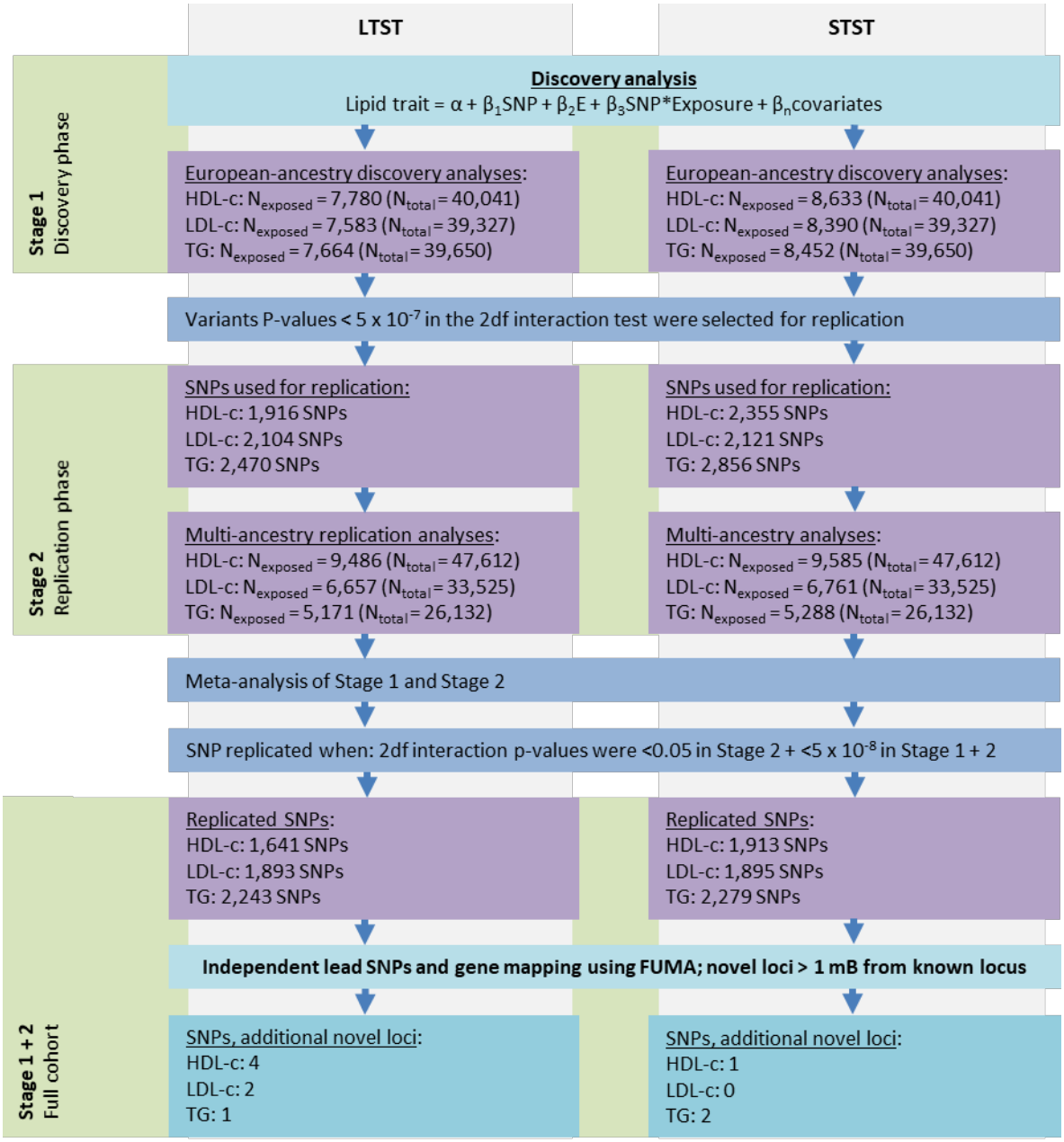

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs7924896
−log10(p−value)
Recombination rate (cM/Mb)
KIF18A
MIR8068
MIR610
METTL15
28
28.2
28.4
28.6
Position on chr11 (Mb)


Plotted SNPs
rs3826692
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
NWD1→ F2RL3→ ←HAUS8 USE1→ ←ABHD8 ←BST2 SLC27A1→
SIN3B→ ←CPAMD8 MYO9B→ BABAM1→ ←PLVAP ←NXNL1
OCEL1→ DDA1→ MVB12A→ PGLS→
←NR2F6 ←ANO8 ←TMEM221 FAM129C→
←USHBP1 COLGALT1→
ANKLE1→
MRPL34→
GTPBP3→
Position on chr19 (Mb)


Plotted SNPs
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
rs1058029
←HAO1 ←TMX4 PLCB1→
Position on chr20 (Mb)


Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs5746495
−log10(p−value)
Recombination rate (cM/Mb)
CECR2
BCL2L13
MICAL3
PEX26
USP18
GGT3P
SLC25A18
BID
MIR648
TUBA8
LOC101929372
MIR3198−1
FLJ41941
ATP6V1E1
LINC00528
Position on chr22 (Mb)

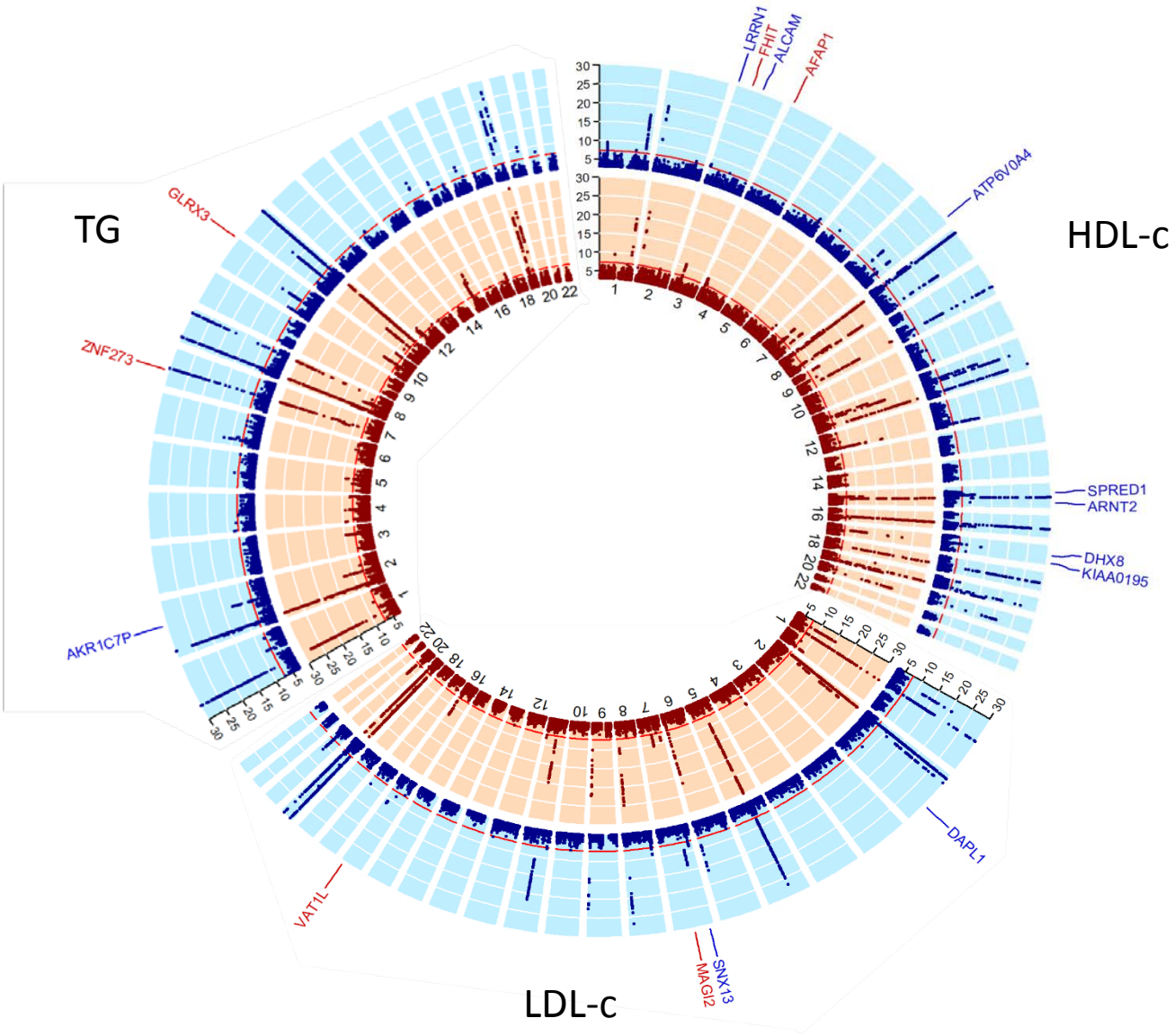
**Supplementary Figure 9:** Project overview and SNP selection in the European-ancestry analyses




Project overview of the multi-ancestry analyses of how the new lipid loci were identified in the present project. Replicated variants had to have 2df interaction test p-values of Stage 1 $<5x10^{-7}$,

Stage 2 <0.05 with a similar direction of effect as in the discovery meta-analysis, and Stage 1+2 $<5x10^{-8}$.

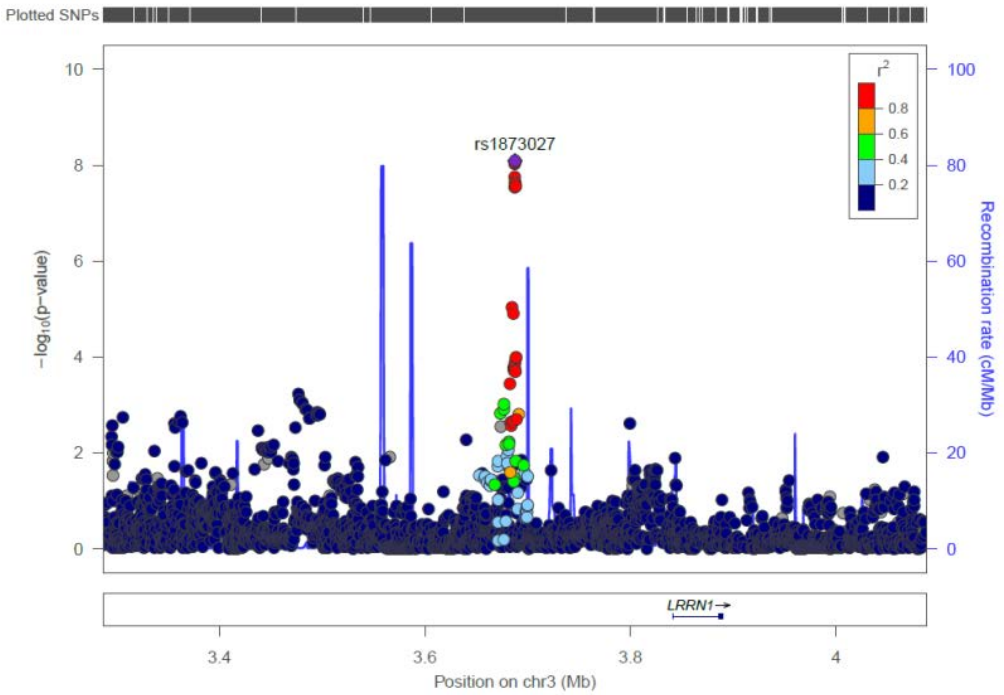
**Supplementary Figure 10:** Circular –log(p-value) plots of the European-ancestry sleep-SNP interactions analyses for the three lipid traits

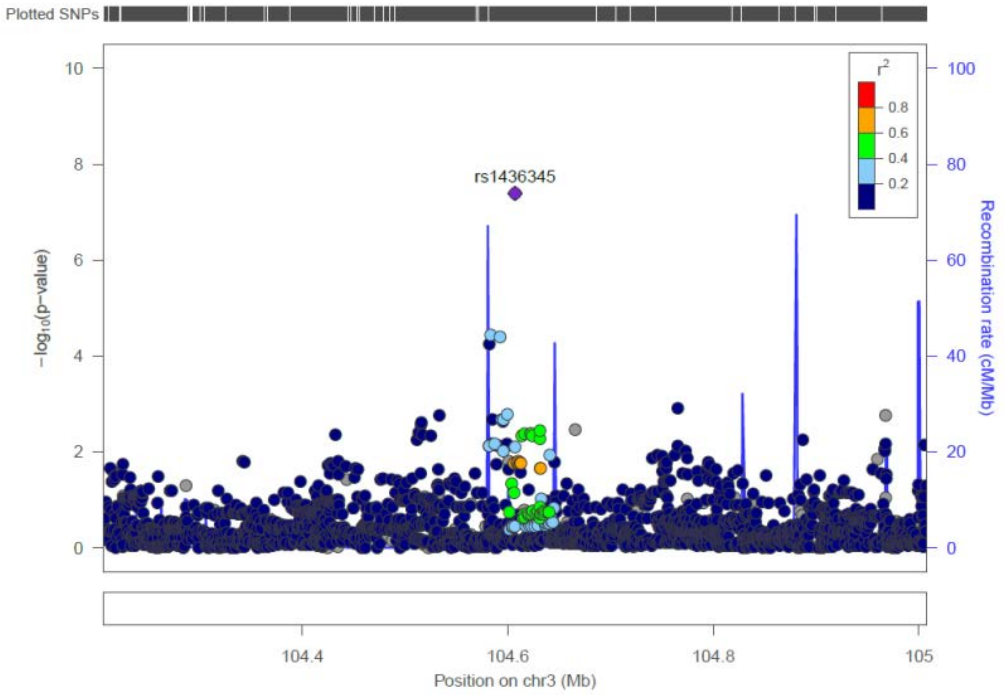


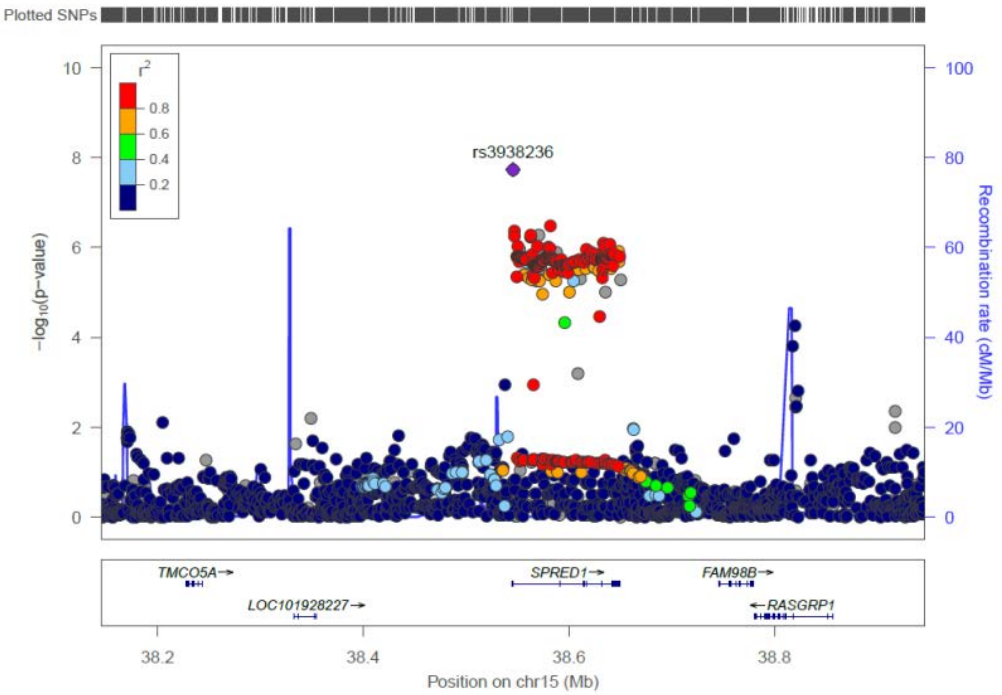

Plot visualizes the –log(P-values in the 2df interaction test) for HDL-c, LDL-c and TG per chromosome. In red (inner circle) are the –log(p-value) plots for the analyses taking into account potential interaction with short total sleep time. In blue (outer circle) are the –log(p-value plots for the analyses taking into account potential interaction with long total sleep time. Loci defined as novel and being replicated are labeled. Replicated variants had to have 2df interaction test p-

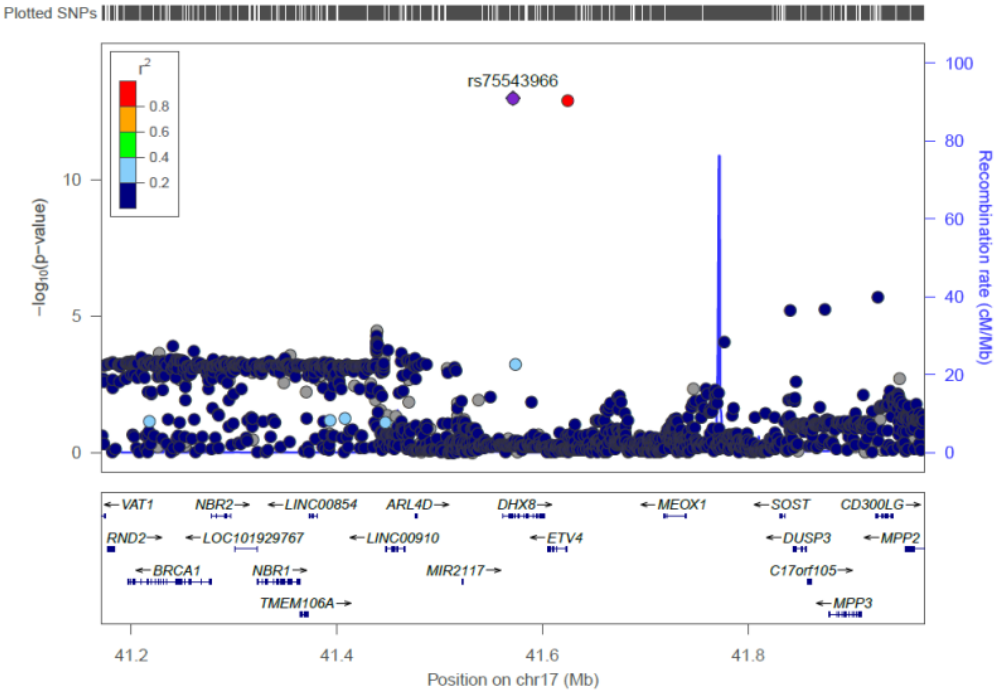
values of Stage 1 $<5x10^{-7}$, Stage 2 <0.05 with a similar direction of effect as in the discovery meta-analysis, and Stage 1+2 $<5x10^{-8}$. Labeled gene names in red were identified in the STST analysis; Labeled gene names in blue were identified in the LTST analysis. –log(p-values) > 30 were truncated for presentation purposes only. All $-\log(p_{joints}) > 30$ were truncated to 30 for visualization purposes only. Figure prepared using the R package circlize[38].

**Supplementary Figure 11:** Regional plots of the replicated additional novel hits of the joint effect of long total sleep time on HDL cholesterol in the European meta-analyses

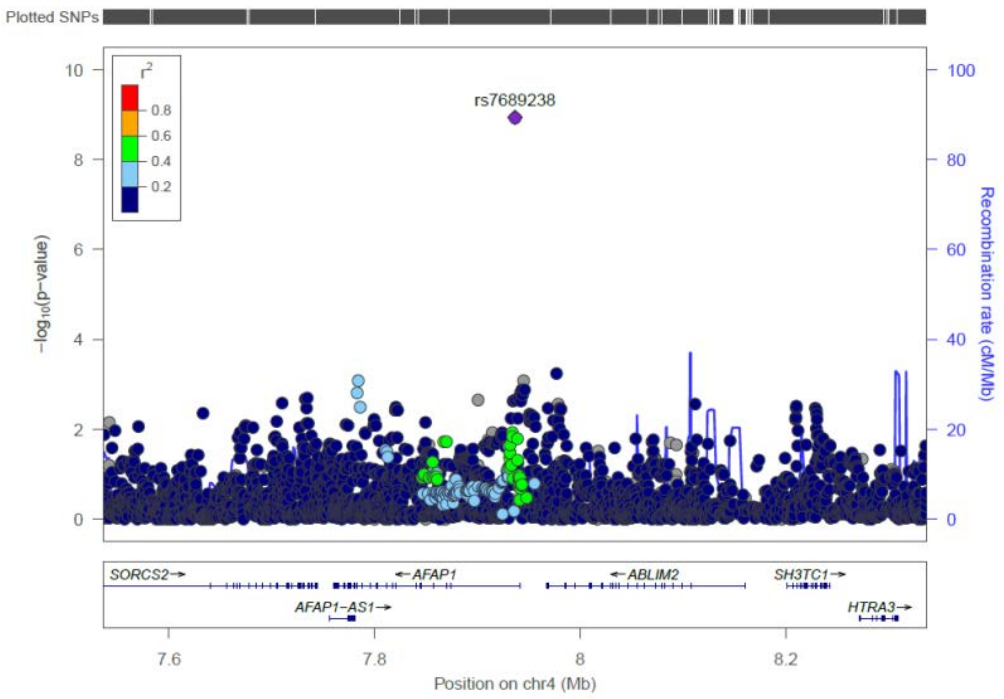


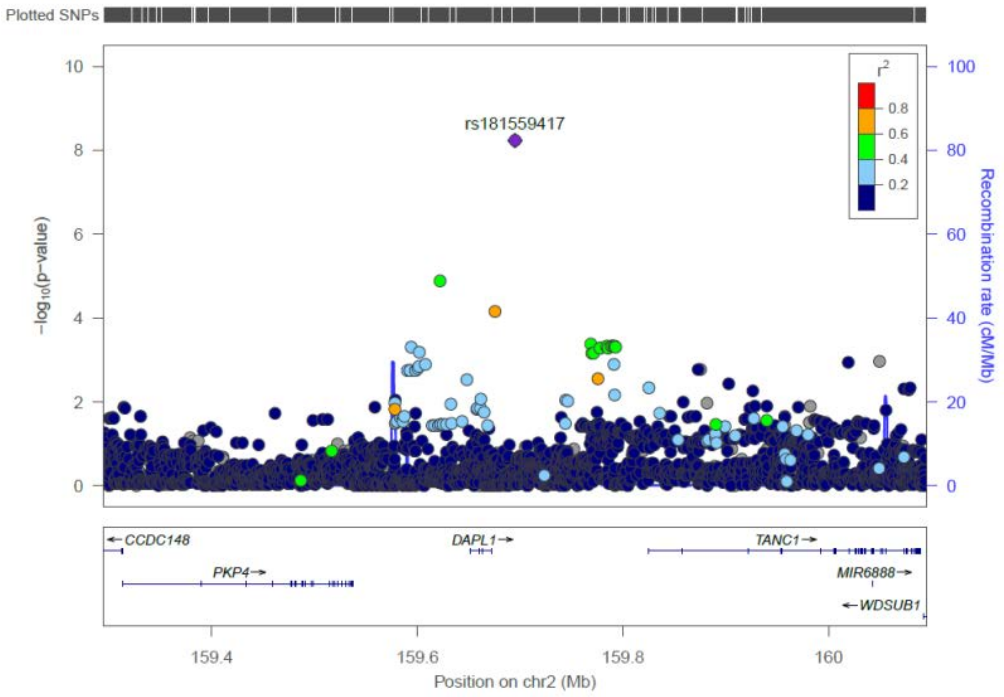

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs3938236
−log10(p−value)
Recombination rate (cM/Mb)
TMCO5A
SPRED1
FAM98B
LOC101928227
RASGRP1
38.2
38.4
38.6
38.8
Position on chr15 (Mb)

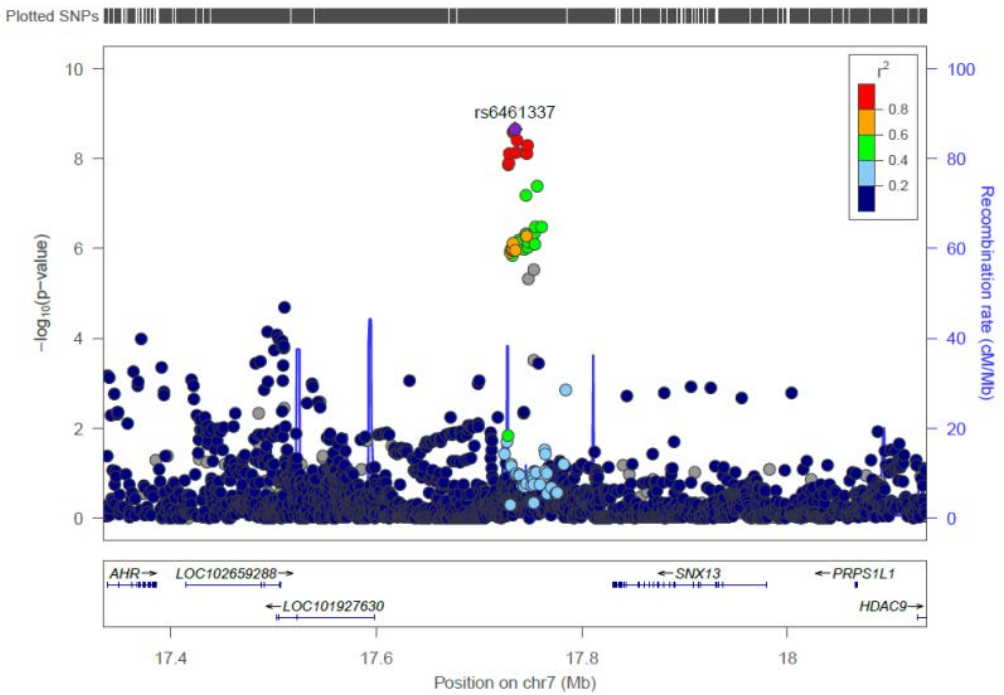

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs75543966
−log10(p−value)
Recombination rate (cM/Mb)
VAT1
NBR2
LINC00854
ARL4D
DHX8
MEOX1
SOST
CD300LG
RND2
LOC101929767
LINC00910
ETV4
DUSP3
MPP2
BRCA1
NBR1
MIR2117
C17orf105
TMEM106A
MPP3
41.2
41.4
41.6
41.8
Position on chr17 (Mb)

**Supplementary Figure 12:** Regional plots of the replicated additional novel hits of the joint effect of short total sleep time on HDL cholesterol in the European meta-analyses

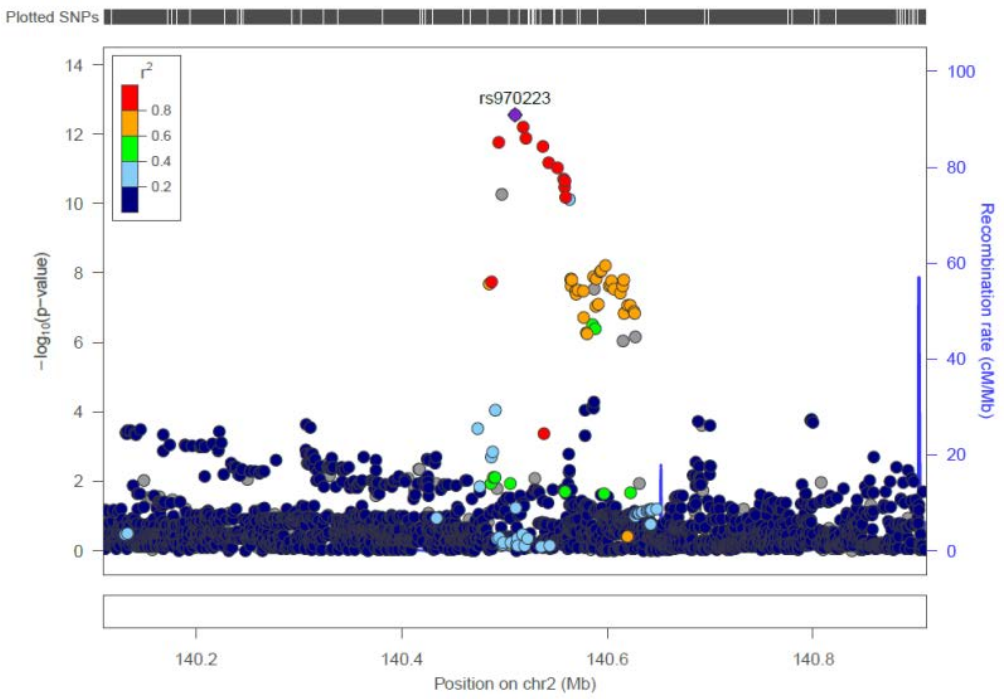


**Supplementary Figure 13:** Regional plots of the replicated additional novel hits of the joint effect of long total sleep time on LDL cholesterol in the European meta-analyses



**Supplementary Figure 14:** Regional plots of the replicated additional novel hits of the joint effect of long total sleep time on Triglycerides in the European meta-analyses



**Supplementary Figure 15:** Regional plots of the replicated additional novel hits of the joint effect of short total sleep time on Triglycerides in the European meta-analyses