



Universiteit  
Leiden  
The Netherlands

## Genetic and environmental determinants of cardiometabolic health

Bos, M.M.

### Citation

Bos, M. M. (2020, October 1). *Genetic and environmental determinants of cardiometabolic health*. Retrieved from <https://hdl.handle.net/1887/136917>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/136917>

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden

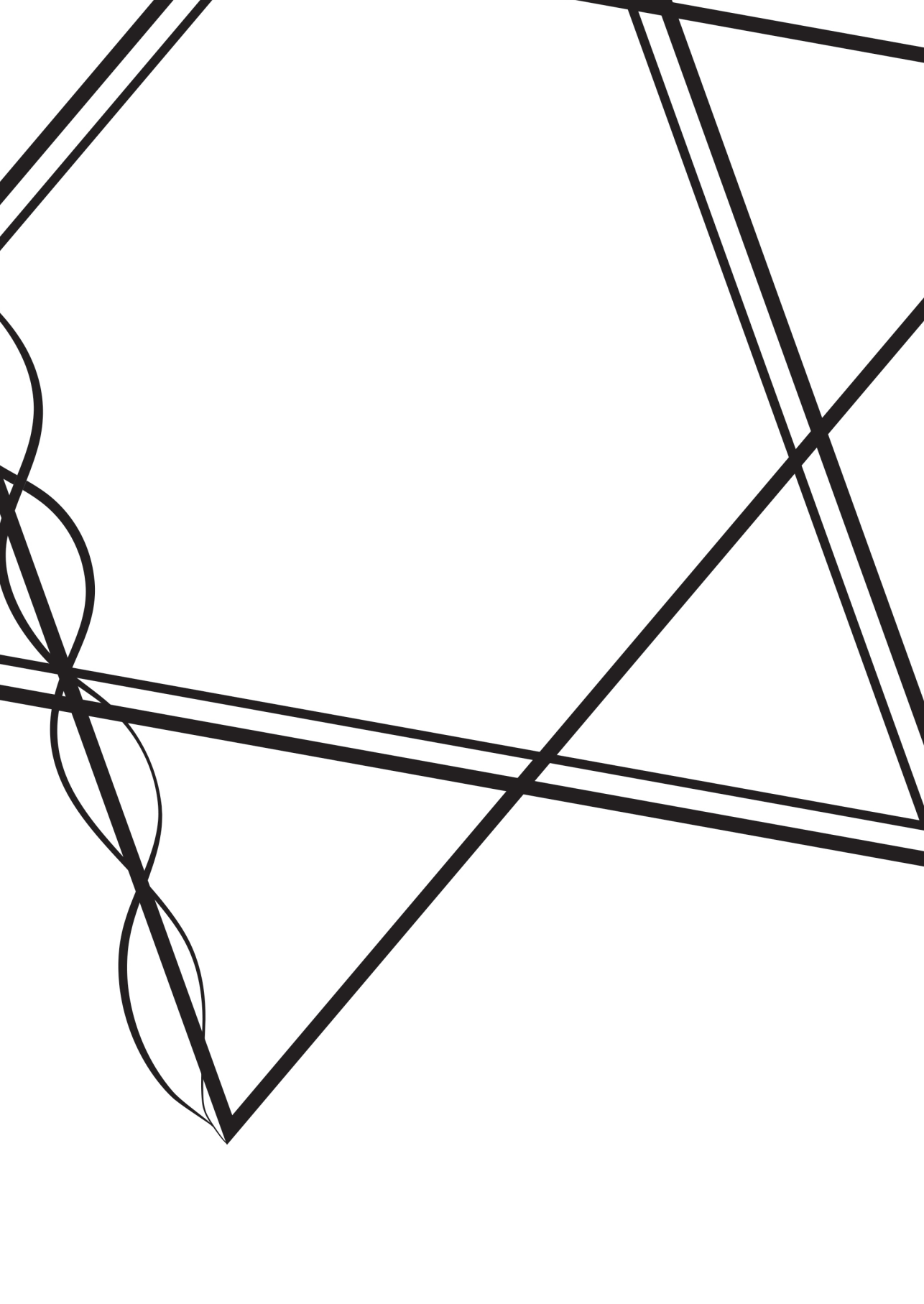


The handle <http://hdl.handle.net/1887/136917> holds various files of this Leiden University dissertation.

**Author:** Bos, M.M.

**Title:** Genetic and environmental determinants of cardiometabolic health

**Issue date:** 2020-10-01





# CHAPTER 3.3

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

Raymond Noordam, Maxime M Bos, Heming Wang, *et al.*

*Nat Commun.* 2019;10(1):5121.

## ABSTRACT

Both short and long sleep are associated with an adverse lipid profile, likely through different biological pathways. To provide new insights in the biology of sleep-associated adverse lipid profile, we conducted 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 identified 49 novel lipid loci, and 10 additional novel lipid loci in a restricted sample of European-ancestry cohorts. In addition, we identified new gene-sleep interactions for known lipid loci such as *LPL* and *PCSK9*. The novel gene-sleep interactions had a modest explained variance in lipid levels: most notable, gene-short-sleep interactions explained 4.25% of the variance in triglyceride concentration. Collectively, these findings contribute to our understanding of the biological mechanisms involved in sleep-associated adverse lipid profiles.

## INTRODUCTION

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<sup>1-3</sup>, and therefore are a primary focus for preventive and therapeutic interventions. Over 300 genetic loci have been identified to be associated with blood lipid concentrations<sup>4-10</sup>. 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<sup>10</sup>. 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<sup>11</sup>, an additional 18 novel lipid loci were identified when considering interactions with high alcohol consumption<sup>12</sup>, and 13 novel lipid loci were identified when considering interaction with smoking status<sup>13</sup>, suggesting that behavioral factors may interact with genetic loci to influence lipid levels.

Sleep is increasingly recognized as a fundamental behavior that influences a wide range of physiological processes<sup>14</sup>. A large volume of epidemiological research implicates disturbed sleep in the pathogenesis of atherosclerosis<sup>15</sup>, and specifically, both a long and short sleep duration are associated with an adverse blood lipid profile<sup>16-26</sup>. However, it is unknown whether sleep duration modifies genetic risk factors for adverse blood lipid profiles. We hypothesize 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 investigated 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<sup>27, 28</sup>. To permit the detection of both such sleep-duration-SNP interactions and lipid-SNP associations accounting for total sleep duration, a 2 degree of freedom (2df) test that jointly tests the SNP-main and SNP-interaction effect was applied<sup>29</sup>. Given that there are differences among ancestry groups in sleep behaviours and lipid levels, analysis of data from cohorts of varying ancestries may facilitate the discovery of robust interactions between genetic loci and sleep traits. We focused 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 to the remaining individuals in the study population, given that each extreme sleep trait has been associated with multiple metabolic and health outcomes<sup>15-26, 30-34</sup>. Within the present study, we report multi-ancestry sleep-by-SNP interaction analyses for blood lipid levels that successfully identified several novel loci for blood lipid traits.

## METHODS

Details regarding motivation and methodology of this and other projects of the CHARGE Gene-Lifestyle Interactions Working Group are available elsewhere<sup>35</sup>.

### *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 **Online Supplementary Materials**, and study-specific characteristics (sizes, trait distribution and data preparation) are presented in **Online Supplementary Tables 1-6**. Every effort was made to include as many studies as possible. 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<sup>36</sup> (the latter being restricted to those with TG  $\leq$  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 standardized sleep questionnaire (e.g., the Pittsburgh Sleep Quality Index questionnaire<sup>37</sup>). For the present project, we defined both STST and LTST. For the



present project, we defined both STST and LTST. To harmonize 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 programs and with local cleaning thresholds for call rates (usually > 98%) and Hardy-Weinberg equilibrium (usually p-value <1e<sup>-5</sup>). 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 (**Online 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 (**Online Supplementary Tables 1-3**). All cohorts ran statistical models according to a standardized 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:

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

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<sup>4,11,12</sup>. In addition, we examined the SNP-main effect (without incorporating LTST/STST) and the SNP-main effect stratified by the exposure:

$$E(Y) = \beta_o + \beta_G \cdot SNP + \beta_C \cdot C$$

All models were performed for each lipid trait and separately for the different ancestry groups. Consequently, per ancestry group, we requested a total of 7 GWA analyses per lipid trait. All models were adjusted for age, sex, field center (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`<sup>38, 39</sup> or `ProbABEL`<sup>40</sup>. Studies including related individuals used either generalized estimating equations (R package `geepack`<sup>41</sup>) or linear mixed models (`GenABEL`<sup>42</sup>, `MMAP`, or R package `sandwich`<sup>38, 39</sup>). Sample code provided to studies to generate these data has been previously published (see **Online Supplementary Materials** <sup>35</sup>).

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`<sup>43</sup> ([www.genepi-regensburg.de/easyqc](http://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, harmonization of alleles, comparison of allele frequencies with ancestry-appropriate 1000 Genomes reference data, and harmonization of all SNPs to a standardized 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)<sup>44</sup>.

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 3 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<sup>45</sup> (<http://genome.sph.umich.edu/wiki/METAL>). We evaluated both a 1 degree (1df) of freedom test of interaction effect and a 2 degree of freedom (2df) joint test of main and interaction effects, following previously published methods<sup>29</sup>. 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 x LTST/STST interaction<sup>46</sup>. 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  $p \leq 5 \times 10^{-7}$  for replication in the Stage 2 analysis. This cut-off was selected to minimize 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 (**Online Supplementary Tables 4-6**). As we had a limited number of individuals from non-European ancestry in the replication analyses, we did not consider 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 novel) were considered to be replicated when Stage 1 p-values  $< 5 \times 10^{-7}$ , Stage 2 p-value  $< 0.05$  with a similar direction of effect as in the discovery meta-analysis, and Stage 1+2 p-value  $< 5 \times 10^{-8}$ . Replicated SNPs were subsequently used in different bioinformatics tools for further processing. In addition, 1 df p-values (SNP-sleep interaction effect only) of the lead SNPs of both the replicated known and novel genetic 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<sup>47</sup>.

### *Bioinformatics*

Replicated SNPs were first processed using the online tool FUMA<sup>48</sup> to identify independent lead SNPs and to perform gene mapping. From the SNP that has a  $P_{\text{joint}} < 5 \times 10^{-8}$ , we determined lead SNPs that were independent from each other at  $R^2 < 0.1$  using the 1000G Phase 3 EUR as a reference panel population. Independent lead SNPs with a physical distance  $>1$  mB from a known locus were considered as novel. Regional plots of the novel loci from the European- and multi-ancestry meta-analyses were made using the online LocusZoom tool<sup>49</sup>. The explained variance of the newly identified genetic independent variants 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<sup>50</sup>. 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 newly identified genetic loci was performed using GTEx (<https://gtexportal.org/home/>)<sup>51, 52</sup>.

### *Look-ups of novel loci in publicly available databases*

Newly identified genetic loci for the three lipid traits were further explored in the GWAS catalogue (<https://www.ebi.ac.uk/gwas/>) to investigate the role of the newly identified mapped genes in other traits. Furthermore, we extracted the lead SNPs from the novel genetic 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 newly identified lead SNPs from the GWAS analyses done on accelerometer-based sleep variables, which was done in European-ancestry individuals from the UK Biobank (N = 85,670)<sup>53</sup>. In addition, we extracted the newly 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<sup>54</sup>.

## RESULTS

### *Study population*

Discovery analyses were performed in up to 62,457 individuals (40,041 European-ancestry, 14,908 African-ancestry, 4,460 Hispanic-ancestry, 2,379 Asian-ancestry, and 669 Brazilian/mixed-ancestry individuals) from 21 studies spanning 5 different ancestry groups (**Online Supplementary Tables 1-3**). 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, 3,133 Asian-ancestry, and 1,146 African-ancestry individuals) from 19 studies spanning 4 different ancestry groups (**Online Supplementary Tables 4-6**). Of the total replication analysis, 12,952 (20.1%) individuals were classified as short sleepers and 12,834 (19.9%) individuals as long sleepers.

### *Identification of novel loci for lipid traits when considering potential interaction with long or short total sleep time*

An overview of the multi-ancestry analyses process for both STST and LTST is presented in **Figure 1**. QQ plots of the combined multi-ancestry and European meta-analysis of the discovery and replication analysis are presented in **Online Supplementary Figures 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 ( $P_{\text{joint}}$  in replication  $< 0.05$  with similar directions of effect as in the discovery analyses and  $P_{\text{joint}}$  in combined discovery and replication analysis  $< 5 \times 10^{-8}$ ). Notably, we replicated 2,395 and 2,576 SNPs for HDL-c, 2,012 and 2,074 SNPs for LDL-c, and 2,643 and 2,734 SNPs for TG in the joint model with LTST and STST respectively.

Most of the replicated SNPs were mapped to known loci (**Online Supplementary Tables 7 and 8**). We looked at the 427 known lipid SNPs (**Online Supplementary Table 9**), but these did not reveal significant 1df interactions with either LTST or STST. In addition, we identified lead SNPs mapping to novel 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 novel loci for HDL-c, 12 novel loci for LDL-c, and 23 novel loci for TG ( $R^2 < 0.1$ ; **Figure 2**). Of these, 7 loci for HDL-c, 4 loci for LDL-c and 7 loci for TG were identified after considering an interaction with LTST (**Online Supplementary Table 10**). Furthermore, 7 loci for HDL-c, 8 loci for LDL-c and 16 loci for TG were identified when considering an interaction with STST (**Online Supplementary Table 11**). Importantly, none of the novel loci for the three lipid traits identified through

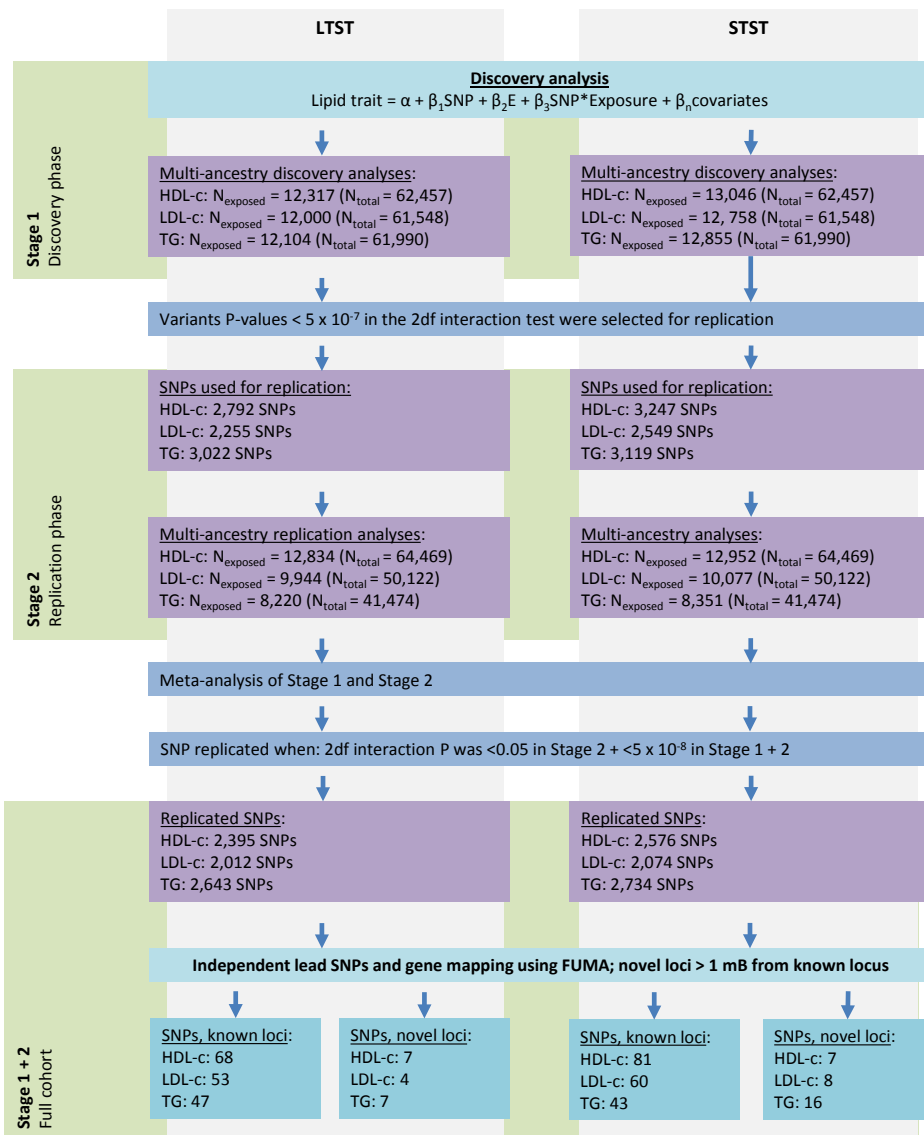
LTST were identified in the analyses with STST, and *vice versa*. Furthermore, the novel lipid loci were specific to a single lipid trait. Regional plots of the newly identified loci from the multi-ancestry analyses are presented in **Online Supplementary Figures 3-8**. Some of the novel SNPs identified through modelling a short or long sleep duration interaction (1 df) also showed suggestive evidence of association with lipid levels in the joint model (2 df 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 novel SNPs (**Online Supplementary Table 10-11**).

Using the R-based VarExp package<sup>50</sup>, we calculated the explained variance based on the summary statistics of the combined discovery and replication analysis. Collectively, novel lead 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, novel lead 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.

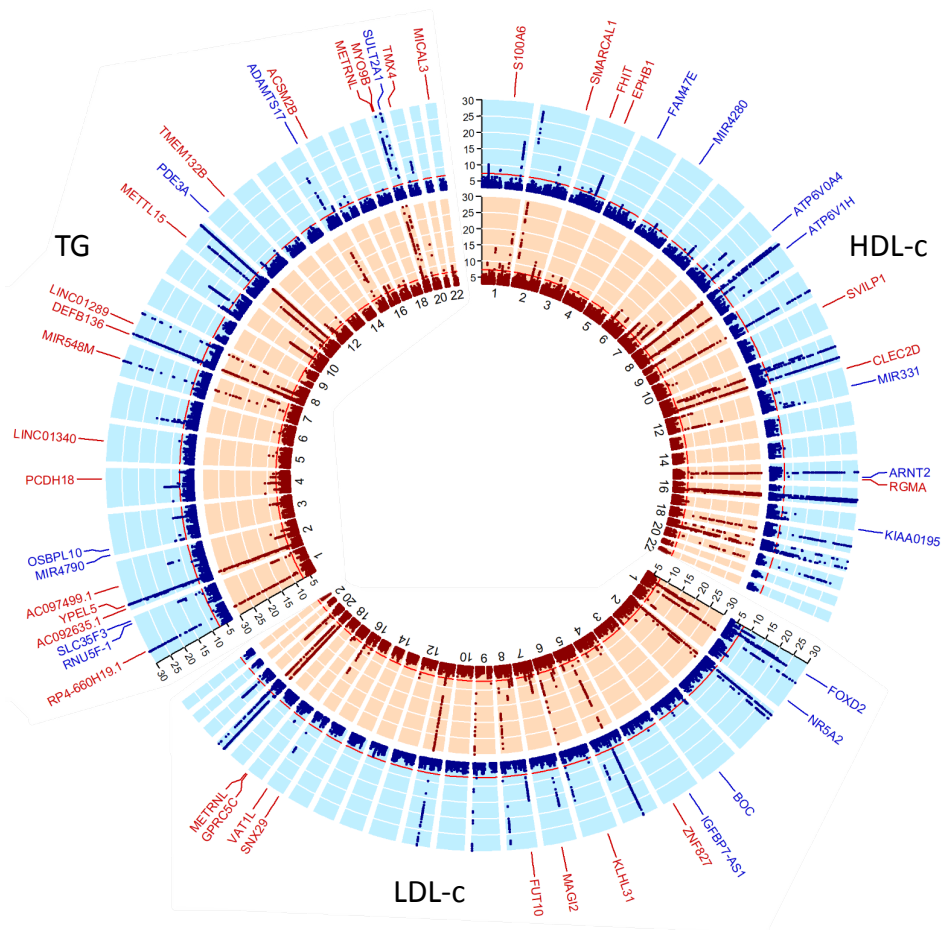
In the analyses restricted to European-ancestry individuals (overview **Online Supplementary Figure 9**), we identified 10 additional novel loci (7 novel loci with LTST and 3 novel loci with STST; **Online Supplementary Figure 10**), which were not identified in the multi-ancestry analyses. Of these, we identified 4 loci for HDL-c, 2 loci for LDL-c, and 1 locus for TG with LTST (**Online Supplementary Table 12**). In addition, we identified 1 locus for HDL-c and 2 for TG with STST (**Online Supplementary Table 13**). Again, we observed no overlapping findings between the two sleep exposures and the three lipid traits. Regional plots of the identified novel loci were presented in **Online Supplementary Figures 11-15**.

#### *Sleep × SNP interactions in identified novel and known lipid loci in the combined sample of discovery and replication studies*

Based on a total of 402 lead SNPs in known and novel 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<sup>29</sup>. We corrected the 1df interaction p-value for multiple testing using the false-discovery rate<sup>47</sup> 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 novel lipid loci show clearly stronger interaction with either LTST or STST than the loci defined as known (**Figure 3**). The majority of the newly 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.,  $MAF < 0.05$ ).



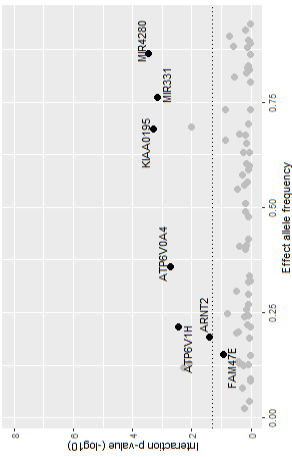
**Figure 1.** Project overview and SNP selection in the multi-ancestry analyses. Project overview of the multi-ancestry analyses of how the new lipid loci were identified in the present project. Replicated variants had to have a  $P_{\text{joint}} < 5 \times 10^{-7}$  in stage 1,  $P_{\text{joint}} < 0.05$  with similar direction of effect in stage 2, and  $P_{\text{joint}} < 5 \times 10^{-8}$  in the combined stage 1 + 2.



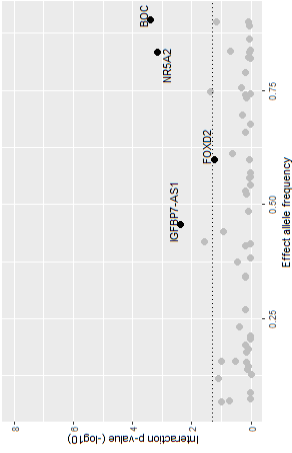
**Figure 2.** Circular  $-\log(p\text{-value})$  plots of the multi-ancestry sleep-SNP interactions analyses for the three lipid traits. Plot visualizes the  $-\log(p_{\text{joints}})$  for HDL-c, LDL-c and TG per chromosome. In red (inner circle) are the  $-\log(p\text{-value})$  plots for the analyses taking into account potential interaction with short total sleep time. In blue (outer circle) are the  $-\log(p\text{-value})$  plots for the analyses taking into account potential interaction with long total sleep time. Loci defined as novel and replicated are labeled. Replicated variants had to have a  $P_{\text{joint}} < 5 \times 10^{-7}$  in stage 1,  $P_{\text{joint}} < 0.05$  in stage 2, and  $P_{\text{joint}} < 5 \times 10^{-8}$  in the combined stage 1 + 2. Labeled gene names in red were identified in the STST analysis; Labeled gene names in blue were identified in the LTST analysis. All  $-\log(p_{\text{joints}}) > 30$  were truncated to 30 for visualization purposes only. The unlabeled regions with  $P_{\text{joint}} < 5 \times 10^{-8}$  were in known loci. Figure prepared using the R package circize<sup>55</sup>.



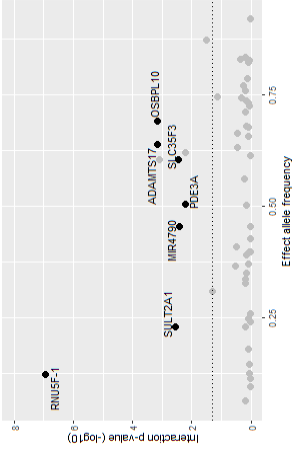
a) HDL-c, long total sleep time



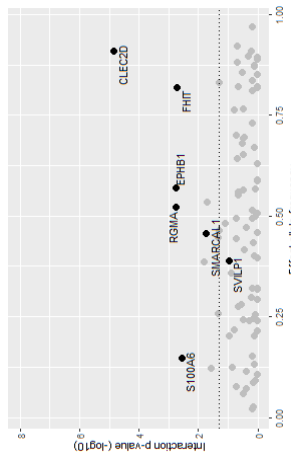
c) LDL-c, long total sleep time



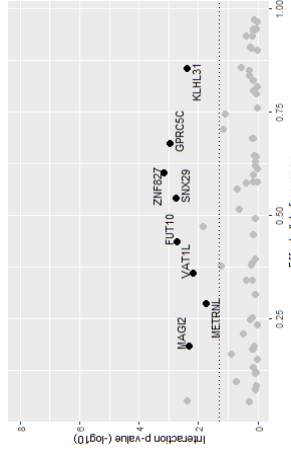
e) TG, long total sleep time



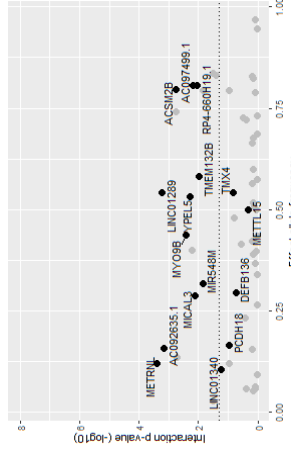
b) HDL-c, short total sleep time



d) LDL-c, short total sleep time



f) TG, short total sleep time

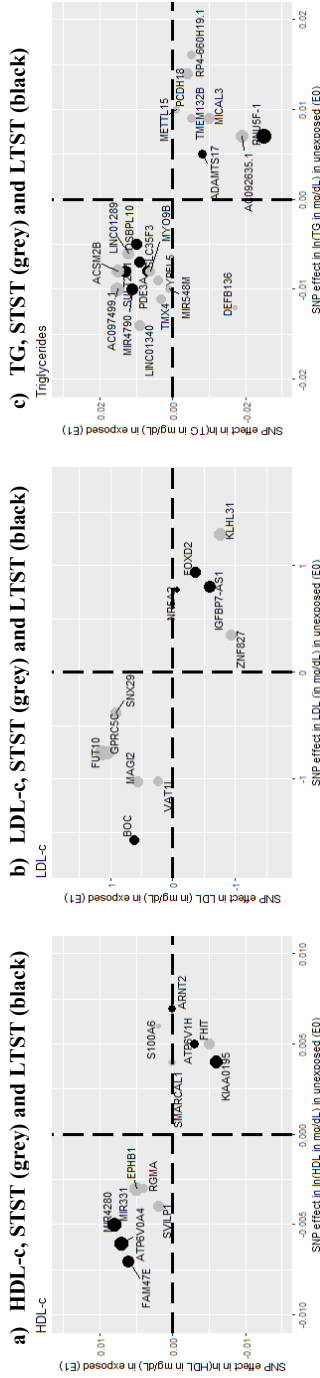


**Figure 3.** *Sleep-interactions on lipid traits in novel and known loci in the multi-ancestry meta-analyses.* Plot displaying the  $-\log(p\text{-value})$  of the 1df interaction between the SNP and either LTST or STST on the lipid trait after correction for multiple testing using false-discovery rate 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. Visualization of the plots was performed using the R package ggplot2 <sup>56</sup>.

Out of the 7 novel HDL-c loci identified in the joint model with LTST, 6 had a 1df interaction  $p\text{-value}_{\text{FDR}} < 0.05$ , 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  $\times$  SNP interaction also showed effect estimates that modestly differed between LTST exposure groups (**Online Supplementary Table 14**). Interestingly, two lead SNPs near known HDL-c loci showed a 1df interaction  $p\text{-value}_{\text{FDR}} < 0.05$ , including SNPs near *CETP* and *LIPC* (**Online Supplementary Table 7**). Out of the 7 novel HDL-c loci identified in the joint model with STST, we found 6 loci with a 1df interaction  $p\text{-value}_{\text{FDR}} < 0.05$ , notably lead SNPs mapped to *S100A6*, *SMARCAL1*, *RGMA*, *EPHB1*, *FHIT* and *CLEC2D*. Again, their effect estimates differed between the exposure groups in the discovery multi-ancestry meta-analysis (**Online Supplementary Table 15; Figure 4**). Some lead SNPs near known HDL-c loci showed evidence of a 1df interaction with STST (e.g., *MADD* and *LPL*;  $p\text{-value}_{\text{FDR}} < 0.05$ ).

For all four novel lead SNPs associated with LDL-c when considering LTST, we observed a 1df interaction  $p\text{-value}_{\text{FDR}} < 0.05$ ; 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 (**Online Supplementary Table 7**). Similarly, all 8 independent novel lead SNPs associated with LDL-c when considering STST, had a 1df interaction  $p\text{-value}_{\text{FDR}} < 0.05$ ; 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  $p\text{-value}_{\text{FDR}} < 0.05$  (**Online Supplementary Table 8**). For both analyses, we observed that effect estimates differed between the LTST and STST exposure groups in the multi-ancestry discovery analysis (**Online Supplementary Table 14 and 15; Figure 4**).

All 7 independent novel lead SNPs associated with TG when considering LTST, had a 1df interaction  $p\text{-value}_{\text{FDR}} < 0.05$ ; notably, lead SNPs mapped to *RNU5F-1*, *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* (**Online Supplementary Table 7**). Of the 16 novel lead SNPs associated with TG when considering STST, we observed 12 lead SNPs with a 1df interaction  $p\text{-value} < 5 \times 10^{-4}$  ( $p\text{-value}_{\text{FDR}} < 0.05$ ), 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* (**Online Supplementary Table 8**). For both LTST and STST analyses, we again observed differing effects dependent on the exposure group in the discovery meta-analyses (**Online Supplementary Table 14 and 15; Figure 4**).



**Figure 4.** Comparison of SNP-main effects stratified by exposure in the multi-ancestry discovery meta-analyses. 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. Visualization of the plots was performed using the R package ggplot2 <sup>56</sup>.

### *Look-ups and bioinformatics analyses*

Based on the lead SNPs mapped to novel 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 (**Online Supplementary Table 15**). 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 (7 sleep phenotypes  $\times$  49 genes; 10 SNPs were unavailable; threshold for significance =  $1.46e^{-4}$ ). Furthermore, we did not observe that any of the newly identified SNPs was associated with accelerometer-based sleep traits (**Online Supplementary Table 17**). In general, we did not find substantial evidence that the identified novel lead SNPs were associated with coronary artery disease in the CARGIoGRAMplusC4D consortium (**Online Supplementary Table 18**).

Newly identified lipid loci were further explored in the GWAS catalogue (**Online Supplementary Table 19**). Several of the mapped genes of our novel lead SNPs have previously been identified with multiple other traits, such as body mass index (*FHIT*, *KLH31*, *ADAMTS17*, *MAGI2*), mental health (*FHIT* [autism/schizophrenia, depression], *SNX13* [cognition]), gamma-glutamyltransferase (*ZNF827*, *MICAL3*), and inflammatory processes (*ZNF827*, *NR5A2*).

We additionally investigated differential expression of the novel lead SNPs using data from multiple tissues from the GTEx consortium<sup>51,52</sup> (**Online Supplementary Table 20**). 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 (**Online Supplementary Figure 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<sup>57-62</sup>, at least in part through effects on gene expression, we hypothesized 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 5 different ancestry groups, we identified 49 novel lipid loci when considering either long or short total sleep time in the analyses. An additional 10 novel lipid loci were identified in analyses in Europeans only. Of the newly identified loci, most loci at least in part were driven by differing effects in short/long sleepers compared to the rest of the study population. Multiple of the novel genes identified by our efforts 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<sup>4-10</sup>, we identified 49 additional loci associated with either HDL-c, LDL-c or TG in our multi-ancestry analysis. While for some of the novel SNPs had no neighboring 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<sup>63</sup>. Considering the novel 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. While the additionally explained variance for LDL-c (0.38% and 0.13%) and HDL-c (1.00% and 0.97%) was low/modest, the novel lead SNPs identified include genes 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<sup>64</sup>. 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<sup>65, 66</sup>. 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<sup>67</sup>, as well as with higher C-reactive

protein and interleukin-6 concentrations<sup>68</sup>. 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<sup>64, 69</sup>. 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<sup>70</sup>. Altered sleep-wake and circadian rhythms influence glucocorticoid signalling and autonomic nervous system excitation patterns across the day<sup>59</sup>, 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<sup>71</sup>. 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<sup>72</sup>. Collectively, these data suggest different biological mechanisms involved in short and long sleep-associated adverse lipid profiles.

Consistent with different hypothesized physiological effects of short and long sleep, we observed no overlap in the novel loci that were identified by modelling interactions with short or long sleep duration. Novel lipid loci that were identified after considering STST include *FHIT*, *MAGI2* and *KLH3*, which have been previously associated with body mass index (BMI)<sup>73-79</sup>. Interestingly, although not genome-wide significant, variation in *MAGI2* has been associated with sleep duration<sup>80</sup>, 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-glutamyl transferase measurement and/or aspartate aminotransferase levels<sup>81, 82</sup>, which have been implicated in cardiometabolic disturbances<sup>83-86</sup> and associated with prolonged work hours (which often results in short or irregular sleep)<sup>87</sup>. 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 novel genes identified have been previously related to inflammation-driven diseases of the intestine, blood pressure and blood count measurements, including traits influenced by circadian rhythms<sup>88, 89</sup>. However, no novel identified lipid loci with LTST directly interacted with genes involved in the central circadian clock (e.g., *PER2*, *CRY2* and *CLOCK*) in the KEGG pathways database<sup>90</sup>. The novel loci *NR5A2* and *SLC35F3* have been associated with inflammation-driven diseases of the intestine<sup>91, 92</sup>. Ulcerative colitis, an inflammatory bowel disease, has been associated with

both longer sleep duration<sup>93</sup> and circadian disruption<sup>88</sup>. *ARNT2*, also identified via a LTST interaction, heterodimerizes with transcriptional factors implicated in homeostasis and environmental stress responses<sup>94,95</sup>. A linkage association study has reported nominal association of this gene with lipids in a Caribbean Hispanic population<sup>96</sup>.

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<sup>97</sup>, but was not identified in our larger multi-ancestry analysis, possibly due to cultural differences in sleep behaviours<sup>98</sup>.

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<sup>99</sup>. LPL expression has been shown to follow a diurnal rhythm in several metabolic organs<sup>61,100</sup>, and disturbing sleeping pattern by altered light exposure can lower LPL activity, at least in brown adipose tissue<sup>61</sup>. 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<sup>101</sup>. 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 novel lipid loci have been previously associated with traits related to sleep. For example, *MAGI2* and *MYO9B*<sup>80</sup> have been suggestively associated with sleep duration and quality, respectively. Genetic variation in *TMEM132B* has been associated with excessive daytime sleepiness<sup>102</sup>, and *EPHB1* has been associated with self-reported chronotype<sup>103</sup>. 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 new 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.

The present 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, suggesting 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<sup>102</sup>. Identified variants for sleep traits have been recently successfully validated using accelerometer data<sup>104</sup>, although the overall genetic correlation with accelerometer-based sleep duration was shown to be low<sup>53</sup>. Moreover, observational studies showed only a modest correlation between the phenotypes<sup>105</sup>, which suggest that each approach characterizes somewhat different phenotypes. At this time, we did not have sufficient data to evaluate other measures of sleep duration such as polysomnography or accelerometry; however, a more comprehensive characterization and circadian traits likely 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 novel genetic loci associated with either HDL-c, LDL-c or triglycerides levels. Multiple of the novel genetic 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.

### Acknowledgments

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 acknowledgments can be found in the **Supplementary Materials**. Data on coronary artery disease have been contributed by the Myocardial Infarction Genetics and CARDIoGRAM investigators and have been downloaded from [www.CARDIOGRAMPLUSC4D.ORG](http://www.CARDIOGRAMPLUSC4D.ORG).



**Conflict of interest statement**

DOMK is a part-time research consultant for Metabolon, Inc. HJG has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm and Janssen Cilag. HJG 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)". SA reports employment and stock options with 23andMe, Inc.

## REFERENCES

1. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J*. 2015;36(9):539-50.
2. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada 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*. 2012;60(25):2631-9.
3. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*. 2012;380(9841):572-80.
4. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nature genetics*. 2013;45(11):1274-83.
5. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nature genetics*. 2013;45(11):1345-52.
6. Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitzel NO, 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*. 2014;94(2):223-32.
7. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, et al. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Human Molecular Genetics*. 2017;26(9):1770-84.
8. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466(7307):707-13.
9. Kathiresan S, Melander O, Anevski D, Guiducci C, Burt NP, Roos C, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med*. 2008;358(12):1240-9.
10. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nature genetics*. 2018;50(11):1514-23.
11. Kilpelainen TO, Bentley AR, Noordam R, Sung YJ, Schwander K, Winkler TW, et al. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat Commun*. 2019;10(1):376.
12. de Vries PS, Brown MR, Bentley AR, Sung YJ, Winkler TW, Ntalla I, et al. Multi-Ancestry Genome-Wide Association Study of Lipid Levels Incorporating Gene-Alcohol Interactions. *Am J Epidemiol*. 2019.
13. Bentley AR, Sung YJ, Brown MJ, Winkler TW, Kraja AT, Ntalla I, et al. Multi-ancestry genome-wide smoking interaction study of 387,272 individuals identifies novel lipid loci. *Nature Genetics* (accepted for publication). 2019.
14. Tobaldini E, Costantino G, Solbiati M, Cogliati C, Kara T, Nobili L, et al. Sleep, sleep deprivation, autonomic nervous system and cardiovascular diseases. *Neurosci Biobehav Rev*. 2017;74(Pt B):321-9.
15. Tobaldini E, Pecis M, Montano N. Effects of acute and chronic sleep deprivation on cardiovascular regulation. *Arch Ital Biol*. 2014;152(2-3):103-10.

16. Ford ES. Habitual sleep duration and predicted 10-year cardiovascular risk using the pooled cohort risk equations among US adults. *J Am Heart Assoc.* 2014;3(6):e001454.
17. Aggarwal S, Loomba RS, Arora RR, Molnar J. Associations between sleep duration and prevalence of cardiovascular events. *Clin Cardiol.* 2013;36(11):671-6.
18. Wu Y, Zhai L, Zhang D. Sleep duration and obesity among adults: a meta-analysis of prospective studies. *Sleep Med.* 2014;15(12):1456-62.
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.* 2014;18(4):293-7.
20. Cappuccio FP, Taggart FM, Kandala NB, Currie A, Peile E, Stranges S, et al. Meta-analysis of short sleep duration and obesity in children and adults. *Sleep.* 2008;31(5):619-26.
21. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *Eur Heart J.* 2011;32(12):1484-92.
22. Lee JA, Park HS. Relation between sleep duration, overweight, and metabolic syndrome in Korean adolescents. *Nutr Metab Cardiovasc.* 2014;24(1):65-71.
23. van den Berg JF, Miedema HM, Tulen JH, Neven AK, Hofman A, Witteman JC, et al. Long sleep duration is associated with serum cholesterol in the elderly: the Rotterdam Study. *Psychosom Med.* 2008;70(9):1005-11.
24. Petrov ME, Kim Y, Lauderdale D, Lewis CE, Reis JP, Carnethon MR, et al. Longitudinal associations between objective sleep and lipids: the CARDIA study. *Sleep.* 2013;36(11):1587-95.
25. Bos MM, Noordam R, van den Berg R, de Mutsert R, Rosendaal FR, Blauw GJ, et al. Associations of sleep duration and quality with serum and hepatic lipids: The Netherlands Epidemiology of Obesity Study. *J Sleep Res.* 2018:e12776.
26. Kaneita Y, Uchiyama M, Yoshiike N, Ohida T. Associations of usual sleep duration with serum lipid and lipoprotein levels. *Sleep.* 2008;31(5):645-52.
27. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, 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. *Circulation Cardiovascular genetics.* 2009;2(1):73-80.
28. Rao DC, Sung YJ, Winkler TW, Schwander K, Borecki I, Cupples LA, et al. Multiancestry Study of Gene-Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts: Design and Rationale. *Circ Cardiovasc Genet.* 2017;10(3).
29. Manning AK, LaValley M, Liu C-T, Rice K, An P, Liu Y, et al. Meta-analysis of Gene-Environment interaction: joint estimation of SNP and SNP×Environment regression coefficients. *Genetic Epidemiology.* 2011;35(1):11-8.
30. Lopez-Garcia E, Faubel R, Leon-Munoz L, Zuluaga MC, Banegas JR, Rodriguez-Artalejo F. Sleep duration, general and abdominal obesity, and weight change among the older adult population of Spain. *The American journal of clinical nutrition.* 2008;87(2):310-6.
31. van den Berg JF, Knvistingh Neven A, Tulen JH, Hofman A, Witteman JC, Miedema HM, et al. Actigraphic sleep duration and fragmentation are related to obesity in the elderly: the Rotterdam Study. *International journal of obesity (2005).* 2008;32(7):1083-90.
32. Wong PM, Manuck SB, DiNardo MM, Korytkowski M, Muldoon MF. Shorter sleep duration is associated with decreased insulin sensitivity in healthy white men. *Sleep.* 2015;38(2):223-31.
33. Reutrakul S, Van Cauter E. Interactions between sleep, circadian function, and glucose metabolism: implications for risk and severity of diabetes. *Annals of the New York Academy of Sciences.* 2014;1311:151-73.

34. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care*. 2010;33(2):414-20.
35. Rao DC, Sung YJ, Winkler TW, Schwander K, Borecki I, Cupples LA, et al. Multiancestry Study of Gene–Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts. Design and Rationale. 2017;10(3).
36. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972;18(6):499-502.
37. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry research*. 1989;28(2):193-213.
38. Zeileis A. Object-oriented Computation of Sandwich Estimators. 2006. 2006;16(9):16.
39. Zeileis A. Econometric COmputing with HC and HAC Covariance Matrix Estimators. *J of Stat Software*. 2004;11(10).
40. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*. 2010;11:134.
41. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. *Journal of Statistical Software*. 2006;15(2):1-11.
42. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007;23(10):1294-6.
43. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Magi R, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9(5):1192-212.
44. Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics*. 2015;31(2):259-61.
45. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-1.
46. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting Gene–Environment Interaction to Detect Genetic Associations. *Human Heredity*. 2007;63(2):111-9.
47. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*. 1995;57(1):289-300.
48. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826.
49. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-7.
50. Laville V, Bentley AR, Prive F, Zhu X, Gauderman J, Winkler TW, et al. VarExp: estimating variance explained by genome-wide GxE summary statistics. *Bioinformatics*. 2018;34(19):3412-4.
51. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-60.
52. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nature genetics*. 2013;45(6):580-5.
53. Jones SE, van Hees VT, Mazzotti DR, Marques-Vidal P, Sabia S, van der Spek A, et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. *Nat Commun*. 2019;10(1):1585.

54. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015;47(10):1121-30.
55. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinformatics.* 2014;30(19):2811-2.
56. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. 2016.
57. Aho V, Ollila HM, Kronholm E, Bondia-Pons I, Soininen P, Kangas AJ, et al. Prolonged sleep restriction induces changes in pathways involved in cholesterol metabolism and inflammatory responses. *Sci Rep.* 2016;6:24828.
58. Chua EC, Shui G, Cazenave-Gassiot A, Wenk MR, Gooley JJ. Changes in Plasma Lipids during Exposure to Total Sleep Deprivation. *Sleep.* 2015;38(11):1683-91.
59. Gooley JJ. Circadian regulation of lipid metabolism. *Proc Nutr Soc.* 2016;75(4):440-50.
60. Huang T, Zeleznik OA, Poole EM, Clish CB, Deik AA, Scott JM, et al. Habitual sleep quality, plasma metabolites and risk of coronary heart disease in post-menopausal women. *Int J Epidemiol.* 2018.
61. van den Berg R, Kooijman S, Noordam R, Ramkisoensing A, Abreu-Vieira G, Tambyrajah LL, et al. A Diurnal Rhythm in Brown Adipose Tissue Causes Rapid Clearance and Combustion of Plasma Lipids at Wakening. *Cell Rep.* 2018;22(13):3521-33.
62. van den Berg R, Noordam R, Kooijman S, Jansen SWM, Akintola AA, Slagboom PE, et al. Familial longevity is characterized by high circadian rhythmicity of serum cholesterol in healthy elderly individuals. *Aging Cell.* 2017;16(2):237-43.
63. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012;40(Database issue):D930-4.
64. Yang L, Yang H, He M, Pan A, Li X, Min X, 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.* 2016;39(3):645-52.
65. Anafi RC, Pellegrino R, Shockley KR, Romer M, Tufik S, Pack AI. Sleep is not just for the brain: transcriptional responses to sleep in peripheral tissues. *BMC Genomics.* 2013;14:362.
66. Moller-Levet CS, Archer SN, Bucca G, Laing EE, Slak A, Kabiljo R, et al. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc Natl Acad Sci U S A.* 2013;110(12):E1132-41.
67. Carson V, Tremblay MS, Chaput JP, Chastin SF. Associations between sleep duration, sedentary time, physical activity, and health indicators among Canadian children and youth using compositional analyses. *Appl Physiol Nutr Metab.* 2016;41(6 Suppl 3):S294-302.
68. Patel SR, Zhu X, Storfer-Isser A, Mehra R, Jenny NS, Tracy R, et al. Sleep duration and biomarkers of inflammation. *Sleep.* 2009;32(2):200-4.
69. Ayas NT, White DP, Manson JE, Stampfer MJ, Speizer FE, Malhotra A, et al. A prospective study of sleep duration and coronary heart disease in women. *Arch Intern Med.* 2003;163(2):205-9.
70. Wefers J, van Moorsel D, Hansen J, Connell NJ, Havekes B, Hoeks J, et al. Circadian misalignment induces fatty acid metabolism gene profiles and compromises insulin sensitivity in human skeletal muscle. *Proc Natl Acad Sci U S A.* 2018;115(30):7789-94.
71. Adamovich Y, Aviram R, Asher G. The emerging roles of lipids in circadian control. *Biochim Biophys Acta.* 2015;1851(8):1017-25.
72. Galman C, Angelin B, Rudling M. Bile acid synthesis in humans has a rapid diurnal variation that is asynchronous with cholesterol synthesis. *Gastroenterology.* 2005;129(5):1445-53.

73. Akiyama M, Okada Y, Kanai M, Takahashi A, Momozawa Y, Ikeda M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nature genetics*. 2017;49(10):1458-67.
74. Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, 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*. 2015;11(10):e1005378.
75. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206.
76. Justice AE, Winkler TW, Feitosa MF, Graff M, Fisher VA, Young K, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun*. 2017;8:14977.
77. Hoffmann TJ, Choquet H, Yin J, Banda Y, Kvale MN, Glymour M, et al. A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci. *Genetics*. 2018;210(2):499-515.
78. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, et al. Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. *PLoS Genet*. 2017;13(4):e1006528.
79. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015;518(7538):187-96.
80. Spada J, Scholz M, Kirsten H, Hensch T, Horn K, Jawinski P, et al. Genome-wide association analysis of actigraphic sleep phenotypes in the LIFE Adult Study. *J Sleep Res*. 2016;25(6):690-701.
81. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature genetics*. 2011;43(11):1131-8.
82. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nature genetics*. 2018;50(3):390-400.
83. Kunutsor SK, Abbasi A, Adler AI. Gamma-glutamyl transferase and risk of type II diabetes: an updated systematic review and dose-response meta-analysis. *Ann Epidemiol*. 2014;24(11):809-16.
84. Kunutsor SK, Apekey TA, Cheung BM. Gamma-glutamyltransferase and risk of hypertension: a systematic review and dose-response meta-analysis of prospective evidence. *J Hypertens*. 2015;33(12):2373-81.
85. Kunutsor SK, Apekey TA, Seddoh D. Gamma glutamyltransferase and metabolic syndrome risk: a systematic review and dose-response meta-analysis. *Int J Clin Pract*. 2015;69(1):136-44.
86. 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*. 2017;12(2):e0172631.
87. Park SG, Lee YJ, Ham JO, Jang EC, Kim SW, Park H. 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*. 2014;26(1):40.
88. Swanson GR, Burgess HJ, Keshavarzian A. Sleep disturbances and inflammatory bowel disease: a potential trigger for disease flare? *Expert Rev Clin Immunol*. 2011;7(1):29-36.
89. Giles TD. Circadian rhythm of blood pressure and the relation to cardiovascular events. *J Hypertens Suppl*. 2006;24(2):S11-6.

90. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27-30.
91. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature genetics.* 2017;49(2):256-61.
92. Maguire LH, Handelman SK, Du X, Chen Y, Pers TH, Speliotes EK. Genome-wide association analyses identify 39 new susceptibility loci for diverticular disease. *Nature genetics.* 2018;50(10):1359-65.
93. Ananthakrishnan AN, Khalili H, Konijeti GG, Higuchi LM, de Silva P, Fuchs CS, et al. Sleep duration affects risk for ulcerative colitis: a prospective cohort study. *Clin Gastroenterol Hepatol.* 2014;12(11):1879-86.
94. Sullivan AE, Raimondo A, Schwab TA, Bruning JB, Froguel P, Farooqi IS, et al. Characterization of human variants in obesity-related SIM1 protein identifies a hot-spot for dimerization with the partner protein ARNT2. *Biochem J.* 2014;461(3):403-12.
95. Hao N, Bhakti VL, Peet DJ, Whitelaw ML. Reciprocal regulation of the basic helix-loop-helix/Per-Arnt-Sim partner proteins, Arnt and Arnt2, during neuronal differentiation. *Nucleic Acids Res.* 2013;41(11):5626-38.
96. Dong C, Beecham A, Wang L, Slifer S, Wright CB, Blanton SH, et al. Genetic loci for blood lipid levels identified by linkage and association analyses in Caribbean Hispanics. *J Lipid Res.* 2011;52(7):1411-9.
97. Khor SS, Miyagawa T, Toyoda H, Yamasaki M, Kawamura Y, Tani H, et al. Genome-wide association study of HLA-DQB1\*06:02 negative essential hypersomnia. *PeerJ.* 2013;1:e66.
98. Egan KJ, Knutson KL, Pereira AC, von Schantz M. The role of race and ethnicity in sleep, circadian rhythms and cardiovascular health. *Sleep Med Rev.* 2017;33:70-8.
99. 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 (Lond).* 2018;15:1.
100. Tsutsumi K, Inoue Y, Kondo Y. The relationship between lipoprotein lipase activity and respiratory quotient of rats in circadian rhythms. *Biol Pharm Bull.* 2002;25(10):1360-3.
101. Persson L, Cao G, Stahle L, Sjoberg BG, Troutt JS, Konrad RJ, 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.* 2010;30(12):2666-72.
102. Lane JM, Liang J, Vlasac I, Anderson SG, Bechtold DA, Bowden J, et al. Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nature genetics.* 2017;49(2):274-81.
103. Hu Y, Shmygelska A, Tran D, Eriksson N, Tung JY, Hinds DA. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat Commun.* 2016;7:10448.
104. Dashti H, Jones SE, Wood AR, Lane J, van Hees VT, Wang H, et al. GWAS in 446,118 European adults identifies 78 genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *bioRxiv.* 2018:274977.
105. Jackson CL, Patel SR, Jackson WB, 2nd, Lutsey PL, 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.* 2018;41(6).

