A genetic sum score of effect alleles associated with serum lipid concentrations interacts with educational attainment

Carina Emmel1,2*, Mirjam Frank1, Nico Dragano3, Markus M. Nöthen3,4, Raimund Erbel1, Karl-Heinz Jöckel1 & Börge Schmidt1

High-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and total cholesterol (TC) levels are influenced by both genes and the environment. The aim was to investigate whether education and income as indicators of socioeconomic position (SEP) interact with lipid-increasing genetic effect allele scores (GES) in a population-based cohort. Using baseline data of 4516 study participants, age- and sex-adjusted linear regression models were fitted to investigate associations between GES and lipids stratified by SEP as well as including GES×SEP interaction terms. In the highest education group compared to the lowest stronger effects per GES standard deviation were observed for HDL-C (2.96 mg/dl [95%‑CI: 2.19, 3.83] vs. 2.45 mg/dl [95%‑CI: 1.12, 3.72]), LDL-C (6.57 mg/dl [95%‑CI: 4.73, 8.37] vs. 2.66 mg/dl [95%‑CI: −0.50, 5.76]) and TC (8.06 mg/dl [95%‑CI: 6.14, 9.98] vs. 4.37 mg/dl [95%‑CI: 0.94, 7.80]). Using the highest education group as reference, interaction terms showed indication of GES by low education interaction for LDL-C (ßGES×Education: −3.87; 95%‑CI: −7.47, −0.32), which was slightly attenuated after controlling for GESLDL-C×Diabetes interaction (ßGES×Education: −3.42; 95%‑CI: −6.98, 0.18). The present study showed stronger genetic effects on LDL-C in higher SEP groups and gave indication for a GESLDL-C×Education interaction, demonstrating the relevance of SEP for the expression of genetic health risks.

Elevated serum concentration of low-density lipoprotein cholesterol (LDL-C) is an important causal risk factor for cardiovascular disease (CVD)1–3 and has now largely replaced total cholesterol (TC) as the primary treatment target for dyslipidemia1. Reduced serum concentration of high-density lipoprotein cholesterol (HDL-C) is independently associated with CVD1,4,5. In genetic studies however, HDL-C has not been causally associated with CVD5. Serum lipid concentrations (in the following named as lipids) are complex human traits that are influenced by both genetic and lifestyle factors6. With regard to health inequalities, research has shown that indicators of socioeconomic position (SEP) such as educational attainment, household income and employment status are also associated with lipids7–11. However, the association has been of different strength with regard to different SEP indicators, heterogeneous between men and women and appears to be of different direction in developed and developing countries. In developed countries a more unfavorable serum lipid profile (higher LDL-C and TC levels and lower HDL-C levels) is observed in lower SEP groups8,10–13.

The genome-wide association study (GWAS) meta-analysis from the Global Lipids Genetics Consortium (GLGC) in 2013 identified 157 loci associated with lipids accounting for up to 12% of the variance in each lipid trait14. Subsequent large-scale GWAS demonstrated enhanced gene discovery from expanded sample sizes, independently reporting additional novel loci in different ancestry groups, accounting for up to 20% of the variance in lipids15–17. In contrast family-based association studies have indicated that 30–70% of the variance in lipids is genetically based18,19.

It is assumed that gene-environment (G×E) interactions, where the effect of some gene variants depend on specific environmental exposures20, may account for parts of the unexplained variance of lipids21. G×E interactions for lipids have been observed for life style risk factors such as obesity, smoking, diet, physical activity,
diabetes and sleep duration. While some studies showed stronger effects of lipid-increasing alleles in groups reporting a more favorable lifestyle such as lean or physically active individuals, Bentley et al. also reported SNPs with stronger effects in smokers as an example for an unfavorable lifestyle. So far interactions of lipid-related effect alleles with SEP have not been investigated in adults, but G×SEP interactions on overall health and other traits such as body mass index (BMI) have been indicated by twin and population based studies. Li et al. reported an interaction between one SNP related with a metabolically obese, normal weight phenotype including high TC levels and a composite score including parental education and household income on metabolically obese, normal weight children.

SEP can be considered as a context-defining variable that describes certain risk constellations such as unequally distributed environmental, psychosocial and behavioral health risk factors and may be better suited for describing health-related environments as a whole rather than single risk factors. Thus, investigating interactions between lipid-associated loci and SEP indicators may be crucial for identifying subgroups for which genetic effects show stronger signals than for the average population and who may benefit from genotype-based targeted intervention.

The aim of this study was to investigate whether the SEP indicators education and income interact with genetic sum scores of lipid-increasing effect alleles (GES) for HDL-C (GES\textsubscript{HDL-C}), LDL-C (GES\textsubscript{LDL-C}) and TC (GES\textsubscript{TC}) in a population-based cohort study. In order to explore whether any detected SEP interactions can be explained by SEP-associated lifestyle risk factors, information on smoking, BMI, physical activity, alcohol consumption and diabetes mellitus was included in the interaction analysis. Main results of GES based on the GLGC GWAS meta-analysis were compared to the results of GES also including loci of more recently published GWAS to check for differences in lipid prediction and GES×SEP interaction.

Methods

Study population. All analysis are based on baseline data of the Heinz Nixdorf Recall Study (Risk factors evaluation of coronary calcium and lifestyle cohort), a population-based prospective cohort study. Details on the rationale of the study, study design, sampling methods, response rate, and data collection have been published in detail previously. In brief, 4814 participants aged 45–74 years were randomly selected between 2000 and 2003 from mandatory registries of residence of the cities Bochum, Essen and Mülheim/Ruhr within the densely populated Ruhr metropolitan area in Germany. All participants gave written informed consent. The study has been approved by the institutional ethics committee of the University Duisburg-Essen and was conducted according to the guidelines and recommendations for ensuring Good Epidemiological Practice. An extended quality management procedure and certification according to DIN ISO 9001:2000 was established. All study participants were of European ancestry and thus genetically very homogeneous.

Data collection. At study baseline, standard enzymatic methods (homogeneous direct determination with OPERA measuring system) were used to measure HDL-C, LDL-C and TC within 12 h after blood serum collection at the central laboratory of the University Hospital of Essen, Germany. Participants were asked to fast for at least 4 h before examination resulting in 60% of subjects with fasting status > 8 h, 2% 6–8 h, 5% 4–6 h, 26% 2–4 h and 7% with < 2 h of fasting. The fasting duration was on average 9.7 h (SD: ± 4.9 h), with no difference between men and women. Information on educational attainment, household income, smoking status, physical activity and diagnosis of diabetes mellitus or use of anti-diabetic medication was collected at study baseline in standardized computer-assisted face-to-face interviews. Information on alcohol consumption was collected with a self-administered questionnaire. Education was defined by combining school and vocational training as total years of formal education according to the International Standard Classification of Education (ISCED 97). Years of education were categorized into three groups with ≤ 10 years (equivalent to a minimum compulsory school attendance with no additional vocational degree), 11–13 years (equivalent to upper secondary educational degrees or a combination of lower secondary education and vocational training), and ≥ 14 years of education (equivalent to a vocational training including additional qualification or a university degree). Income was measured as the monthly household equivalent income calculated by dividing the participants’ household net income by a weighting factor for each household member. Income was categorized into three groups using sex-specific tertiles. In order to take account for their different mechanisms in causing health inequalities, both SEP indicators were analyzed separately. The body-mass-index (BMI) was calculated based on standardized measurements of body weight (in underwear) and height (kg/m\textsuperscript{2}). Physical activity was defined as exercising one and more times per week versus no weekly engagement in physical exercise. Smoking status was dichotomized for analyses as current smoker (smoking cigarettes during the past 12 months) versus former and never smoker. Alcohol intake was estimated as gram of pure alcohol per week using information on the number of alcoholic drinks usually consumed in a week by type of drink (i.e., beer, wine, sparkling wine, spirits). Participants were classified as diabetics if they reported a diagnosis of diabetes mellitus, or if a fasting blood glucose level ≥ 126 mg/dl, a postprandial blood glucose level ≥ 200 mg/dl was found, or if the use of anti-diabetic medication was documented. All variables have been checked in an ongoing data quality control during the baseline examination period.

Genetic data. Lymphocyte DNA was isolated from EDTA anti-coagulated venous blood using the Chemagic Magnetic Separation Module I (Chemagen, Baesweiler, Germany). Genotyping was performed by matrix-assisted laser desorption ionization-time of flight mass spectrometry-based iPLEX Gold assay at the Department of Genomics, Life and Brain Center, Bonn Germany using two different Illumina microarrays (Metaochip, Global Screening Array (GSA); Illumina, San Diego, USA) according to the manufacturer’s protocols. Genotype imputation was carried out using IMPUTE v.2.3.0. Quality control was applied prior to imputation and per-
formed on subject level including sex-, ethnicity- and relatedness-checks, excluding subjects with missing genotype data > 10%. Further, single nucleotide polymorphisms (SNPs) with a missing genotype frequency > 10% were excluded.

Using the GLGC GWAS meta-analyses and all afterwards published large-scale GWAS, 152 HDL-C-, 108 LDL-C- and 131 TC-associated SNPs at genome-wide significance level of p < 5 × 10⁻⁸ in study populations of European ancestry have been selected for analysis (Supplementary Table S1). Of these, 102, 85, and 103 genotyped SNPs or a proxy were found for HDL-C, LDL-C and TC, respectively, on the Metabochip. Additionally, 3/2/3 HDL-C/LDL-C/TC SNPs were found in imputed data of the Metabochip and 18/8/11 HDL-C/LDL-C/TC SNPs or a proxy were found on the GSA (Supplementary Table S1–S4). A proxy was defined as a SNP within a linkage disequilibrium (LD) ≥ 0.8. For all SNPs included in the analysis, no deviation from Hardy–Weinberg equilibrium was found (p ≤ 1 × 10⁻⁸). LD-based SNP pruning was performed with PLINK to exclude selected SNPs correlated with an LD ≥ 0.8 before calculating the GES Lipid. However, all SNPs included represented independent loci and none of the SNPs had to be pruned out. Two different GES Lipid were calculated for each lipid trait: GES Lipid based on and an extended GES Lipid-EXT based on. In detail, the genetic sum scores of effect alleles for HDL-C (GES Lipid), LDL-C (GES Lipid), and TC (GES Lipid) were calculated by aggregating the total number of lipid-increasing effect alleles (0, 1 or 2) for each individual from the Heinz Nixdorf Recall Study population across the selected SNPs based on the GLGC GWAS meta-analyses. For comparison, extended GES Lipid-EXT have been calculated for each trait by additionally adding the number of lipid-increasing effect alleles of selected SNPs based on all afterwards published large-scale GWAS (i.e., GES Lipid, GES Lipid-EXT, GES Lipid-EXT, GES Lipid-EXT). Imputation of missing genotype information was based on the study sample’s effect allele frequencies according to the PLINK scoring routine.

### Statistical analyses.

Out of the study population (n = 4814) all participants without genetic information (n = 296) and missing values for all three lipids (HDL-C, LDL-C and TC) (n = 22) were excluded from the analysis, leading to an analysis population of 4516 study participants (50.0% women) (Supplementary Table S1). Participants with missing information on education (n = 13) and income (n = 283), were excluded from the respective analysis. Compared to the analysis population, participants with missing genetic information as well as missing values in lipids and SEP indicators did not differ substantially regarding the main variables included in the analysis.

First, sex- and age-adjusted linear regression models were fitted to calculate effect size estimates and their corresponding 95% confidence intervals (95% CIs) for the association of education, income and the respective GES Lipid (and GES Lipid-EXT) with each lipid trait. The explained variance of GES Lipid (and GES Lipid-EXT) on lipids was calculated with a non-adjusted linear regression model. Second, the GES Lipid (and GES Lipid-EXT) and SEP main effects as well as GES x SEP interaction terms were included into sex- and age-adjusted linear regression models to investigate GES x SEP interactions. Third, the genetic effect of GES Lipid (and GES Lipid-EXT) on each lipid trait was calculated stratified by education groups, income tertiles and diabetes status. Fourth, all possible combinations of GES Lipid (and GES Lipid-EXT) tertiles and SEP groups were entered into regression models as dummy variables to calculate single reference joint effects of the GES and the SEP indicators, using the group with the highest SEP and the lowest GES Lipid as tertile and accordingly the highest SEP and the highest GES Lipid tertile as reference. Additionally, absolute measures of lipids in each of the different combinations of SEP groups and GES Lipid tertiles were calculated. Fifth, to analyze whether GES x SEP interactions may be affected by underlying interactions between GES Lipid and SEP-related life style risk factors, smoking (S), BMI, physical activity (PA), alcohol consumption (A) or diabetes mellitus (D) main effects and the respective GES x S/BMI/PA/A/D interaction term were included in the interaction model separately for each life style risk factor. Single SNP main effect and single SNP interaction analysis between SNPs and education groups were performed for all SNPs used in the GES Lipid. Education is entered as a dummy variable and only the results of the lowest education group compared to the highest education group as reference were presented. Additionally, participants with lipid-lowering medication (n = 557) were considered in sensitivity analyses by (1) adjusting the main results of GES Lipid education and GES Education for lipid-lowering medication and by (2) excluding individuals with lipid-lowering medication. The LD-based pruning and the calculation of GES and single SNP analyses were performed with using PLINK v1.07 software package and RStudio v3.6.0. For all other analyses SAS software v9.4 was used. All GES Lipid and extended GES Lipid-EXT-related beta coefficients and 95% CIs (except for the single reference joint effect analysis) were standardized by multiplying the coefficients by the standard deviation of the respective GES Lipid or the respective extended GES Lipid-EXT to facilitate comparability of each GES.

### Results

In the analysis population, the mean age (± standard deviation) was 59.6 ± 7.8 years and the mean serum lipid concentration (± standard deviation) were 229.26 ± 39.18 mg/dl for TC (Table 1). 11.4% had less than or equal 10 years of education and 33.0% had more than or equal 14 years of education. The median income was 1448.7 Euro/month. The mean number of effect alleles were 78.3 ± 5.1 for the GES Lipid (GES Lipid-EXT: 136.4 ± 6.6), 55.1 ± 4.5 for the GES Lipid (GES Lipid-EXT: 126.2 ± 6.6), 615 (13.6%) participants had diabetes mellitus. Correlation matrix of lipid phenotypes, GES Lipid, education and income is shown in Supplementary Table S5.

SEP inequalities in HDL-C and LDL-C were found in the study population with worse HDL-C and LDL-C profiles observed in lower income and education groups (Table 2). SEP inequalities in TC were not seen. Participants in the lowest education group (≤ 10 years) had a 4.14 (95%-CI: 5.82, –2.47) mg/dl lower HDL-C, a 4.23 (95%-CI: –0.81, 7.49) mg/dl higher LDL-C and 3.34 (95%-CI: –0.81, 7.49) mg/dl higher TC level compared to the participants in the highest education group (≥ 14 years) with similar patterns for income (Table 2). On average,
a 2.91 (95%-CI: (2.45, 3.37)) mg/dl higher HDL-C, a 6.17 (95%-CI: 5.13, 7.20) mg/dl higher LDL-C and a 7.33 (95%-CI: 6.24, 8.48) mg/dl higher TC were seen per standard deviation of the respective GESLipid (Table 2). The explained variance (R^2) of GESHDL-C/ GESLDL-C/ GESTC on their respective lipid was 2.9/ 2.9/ 3.6%.

In the linear regression analysis including interaction terms, effect size estimates of interaction terms showed stronger indication of GESLipid by low education interaction for LDL-C (ßGES×Education: −3.87; 95%-CI: −7.47, −0.32) compared to TC (ßGES×Education: −3.64; 95%-CI: −7.44; 0.16) and HDL-C (ßGES×Education: −0.56; 95%-CI: −2.09; 1.02) using the highest education group as reference (Table 3). The negative interaction coefficient showed that in the lower education group genetic effects of GESLipid were less strong. The effect size estimates for the GESLipid by income interaction were directionally consistent, except for TC, but substantially smaller in magnitude.

In the stratified analysis, the two higher education groups compared to the lowest showed stronger genetic effect size estimates per GESLipid standard deviation for LDL-C and TC, supporting the results of the interaction analysis (Fig. 1). The results for HDL-C followed the same pattern, but the difference in effect size between the highest and the lowest education group was considerably less strong, while the 95% confidence interval of the effect in the lowest education group was completely overlapping with the 95% confidence intervals of both higher education groups. The results of the stratified analysis for income did not follow a clear pattern (Fig. 1). The partial R^2 (explained proportion of variance) of the GESLDL-C on LDL-C and GESTC on TC in the respective education groups was higher in the two higher education groups (LDL-C: high education group R^2 = 0.033, middle education group R^2 = 0.035; TC: high education group R^2 = 0.044, middle education group R^2 = 0.038) compared to the lower education group (LDL-C: R^2 = 0.005; TC: R^2 = 0.013). For Income and HDL-C this trend could not be observed.

The analysis of single reference joint effects for lipids describe the relationship between SEP and GESLipid on lipids in detail by comparing effects of different combinations of SEP groups and GESLipid tertiles. Each beta estimate represent the increase in lipids of the specific group compared to the reference group. Reference group was selected as the combination of GESLDL-C and SEP group representing the lowest CVD risk (equally applied for GESTC and GESHDL-C). For HDL-C beta estimates showed a downward trend between and within education groups with decreasing years of education and decreasing number of effect alleles. Compared to the reference group with the highest education and highest GESHDL-C participants with the lowest education and the lowest

**Table 1.** Characteristics of analysis population. nmiss = number of participants with missing values, HDL-C = high-density lipoprotein cholesterol, LDL-C = low-density lipoprotein cholesterol, TC = total cholesterol, GES = genetic effect allele sum score, diabetes mellitus is defined as self-reported diabetes mellitus, or fasting blood glucose level ≥ 126 mg/dl or postprandial blood glucose level ≥ 200 mg/dl or if the use of anti-diabetic medication was documented. *Mean ± standard deviation (SD) ^Proportion (%) †Median (first quartile-third quartile) §GESLipid based on14 + GESLipid-EXT based on14–17

| Characteristic                           | All (n = 4516)                  |
|----------------------------------------|---------------------------------|
| Age (years)* [nmiss = 0]               | 59.6 ± 7.8                     |
| Female Sex†                            | 2256 (50.0%)                   |
| Serum lipid concentration (mg/dl)*     |                                 |
| HDL-C [nmiss = 1]                      | 58.16 ± 17.29                  |
| LDL-C [nmiss = 14]                     | 145.41 ± 36.21                 |
| TC [nmiss = 0]                         | 229.26 ± 39.18                 |
| Lipid-lowering medication [nmiss = 289]*| 357 (12.2%)                    |
| Number of lipid effect alleles (GESLipid)^* |                          |
| GESHDL-C [nmiss = 0]                   | 78.3 ± 5.1                     |
| GESLDL-C [nmiss = 0]                   | 55.1 ± 4.5                     |
| GESTC [nmiss = 0]                      | 75.7 ± 5.2                     |
| Number of lipid effect alleles (GESLipid-EXT)^* |                |
| GESHDL-C-EXT [nmiss = 0]               | 136.4 ± 6.6                    |
| GESLDL-C-EXT [nmiss = 0]               | 93.6 ± 5.9                     |
| GESTC-EXT [nmiss = 0]                  | 126.2 ± 6.6                    |
| Education (years of training)^† [nmiss = 13] |                               |
| ≤ 10                                   | 515 (11.4%)                    |
| 11–13                                  | 2502 (55.6%)                   |
| ≥ 14                                   | 1486 (33.0%)                   |
| Income (€/month) † [nmiss = 283]       | 1448.7 (1107.8–1874.7)         |
| Body mass index (kg/m^2)* [nmiss = 25] | 27.92 ± 4.64                   |
| No physical activity*                  | 2302 (51.0%)                   |
| Current smoking* [nmiss = 7]           | 1069 (23.7%)                   |
| Alcohol consumption† (g/week) [nmiss = 112] | 13.9 (0–64.5)  |
| Diabetes mellitus †                    | 615 (13.6%)                    |
GESHDL-C showed a 9.85 mg/dl lower HDL-C level. Slightly smaller joint effects were observed for income and GESHDL-C on HDL-C still following the same pattern (Fig. 2). The joint effects for LDL-C and TC showed an upward trend with decreasing years of education and increasing number of effect alleles. Participants with highest CVD-risk (highest GESLipid and lowest education) had a 13.36 mg/dl higher LDL-C and a 15.40 mg/dl higher TC level than those with the lowest CVD-risk (highest education and lowest GESLipid). Slightly stronger joint effects were observed for income and GESLipid on LDL-C and TC (Figs. 3, 4). The absolute measures of lipids in each of the different combinations of SEP groups and GESLipid tertiles show the same pattern as in the single reference joint effect analysis (Supplementary Figures S3–S5). Participants in the highest GESLipid tertile and the lowest education had on average 8.8/14.0/21.1 mg/dl higher HDL-C/LDL-C/TC level than participants in the

| Lipid – Education + age + sex |        |        |        |        |        |
|-------------------------------|--------|--------|--------|--------|--------|
| n                             | 4502   | 4489   | 4503   |        |        |
| Intercept                     | 34.28  | 129.13 | 196.18 | 1.2*10^{-34} |
| Age                           | 0.05 (-0.01; 0.11) | 0.26 (0.12; 0.39) | 0.34 (0.19; 0.49) | 6.3*10^{-48} |
| Sex                           | 14.85 (13.87; 15.82) | 0.72 (-2.96; 1.52) | 0.530 | 3.1*10^{-16} |
| Education (low)               | -4.14 (-5.82; -2.47) | 1.4*10^{-44} |
| Education (high)              | -1.77 (-2.83; -0.72) | 9.6*10^{-44} |
| Education Ref                 | -       | -       | -       | -       |

| Lipid – Income + age + sex    |        |        |        |        |        |
|-------------------------------|--------|--------|--------|--------|--------|
| n                             | 4232   | 4220   | 4233   |        |        |
| Intercept                     | 36.17  | 127.00 | 194.28 | 7.6*10^{-29} |
| Age                           | 0.03 (-0.03; 0.09) | 0.27 (0.13; 0.41) | 0.36 (0.21; 0.51) | 2.4*10^{-28} |
| Sex                           | 14.11 (13.16; 15.07) | 0.01 (-2.17; 2.19) | 0.995 | 2.5*10^{-11} |
| Income (low)                  | -1.09 (-2.24; -0.05) | 3.91 (1.29; 6.53) | 2.62 (-0.17; 5.42) | 0.066 |
| Income (high)                 | -1.55 (-2.75; -0.34) | 2.71 (-0.05; 5.46) | 0.054 | 0.452 |
| Income Ref                    | -       | -       | -       | -       |

| Lipid – GESLipid + age + sex  |        |        |        |        |        |
|-------------------------------|--------|--------|--------|--------|--------|
| n                             | 4515   | 4502   | 4516   |        |        |
| Intercept                     | -8.42 (-16.33; -0.51) | 53.76 (36.51; 69.01) | 89.74 (71.21; 108.26) | 3.4*10^{-21} |
| Age                           | 0.02 (-0.04; 0.08) | 0.27 (0.14; 0.41) | 0.35 (0.21; 0.49) | 1.8*10^{-06} |
| Sex                           | 14.05 (13.15; 15.00) | 0.06 (-2.02; 2.14) | 0.956 | 2.2*10^{-11} |
| GESLipid 2.91 (2.45; 3.37) | 6.17 (5.13; 7.20) | 7.1*10^{-15} |
| GESLipid Ref                  | -       | -       | -       | -       |

| Lipid – GESLipid + Education + age + sex |        |        |        |        |        |
|------------------------------------------|--------|--------|--------|--------|--------|
| n                                        | 4502   | 4489   | 4503   |        |        |
| Intercept                                | -9.92 (-17.85; -1.98) | 54.40 (39.06; 69.75) | 90.80 (72.16; 109.44) | 2.1*10^{-12} |
| Age                                      | 0.05 (-0.01; 0.11) | 0.25 (0.11; 0.38) | 0.33 (0.18; 0.48) | 9.4*10^{-06} |
| Sex                                      | 14.79 (13.83; 15.75) | -1.03 (-3.23; 1.18) | 0.363 | 1.3*10^{-06} |
| GESLipid 2.91 (2.45; 3.36) | 6.17 (5.13; 7.20) | 6.9*10^{-16} |
| Education (low)                          | -4.12 (-5.77; -2.47) | 4.73 (0.92; 8.53) | 0.015 | 0.090 |
| Education (high)                         | -1.81 (-2.84; -0.77) | 6.2*10^{-04} |
| Education (high) Ref                     | -       | -       | -       | -       |

| Lipid – GESLipid + Income + age + sex   |        |        |        |        |        |
|-----------------------------------------|--------|--------|--------|--------|--------|
| n                                        | 4232   | 4220   | 4233   |        |        |
| Intercept                                | -7.34 (-15.49; 0.82) | 56.39 (40.66; 72.12) | 92.54 (73.60; 111.48) | 1.6*10^{-12} |
| Age                                      | 0.02 (-0.04; 0.08) | 0.27 (0.13; 0.40) | 0.35 (0.20; 0.50) | 3.0*10^{-06} |
| Sex                                      | 14.08 (13.14; 15.01) | -0.14 (-2.30; 2.01) | 0.895 | 6.1*10^{-11} |
| GESLipid 2.86 (2.40; 3.32) | 5.81 (4.73; 6.89) | 3.5*10^{-36} |
| Income (low)                             | -1.08 (-2.20; 0.05) | 3.57 (0.96; 6.15) | 2.19 (-0.56; 4.94) | 0.119 |
| Income (high)                            | -1.49 (-2.68; -0.31) | 2.29 (-0.43; 5.01) | 0.098 | 0.622 |
| Income (high) Ref                        | Ref    | Ref    | Ref    | Ref    |

Table 2. Sex- and age-adjusted effects per GESLipid standard deviation and corresponding 95% confidence intervals (95% CI) on high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) and total cholesterol (TC) in linear regression models including main effects of education groups (≤ 10 years/11–13 years/≥ 14 years), income tertiles and genetic effect allele sum scores (GESLipid) based on 19.
Table 3. Sex- and age-adjusted effects per GESLipid standard deviation and corresponding 95% confidence intervals (95% CI) on high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) and total cholesterol (TC) in linear regression models including main effects of a lipid-associated genetic effect allele sum score (GESLipid based on 18), indicators of socioeconomic position (SEP; education groups and income tertiles) and interaction terms of GESLipid and indicators of SEP. # Education, $Income.

|          | HDL-C | LDL-C | TC   |
|----------|-------|-------|------|
|          | β (95%-CI) | p     | β (95%-CI) | p     | β (95%-CI) | p   |
| Lipid – age + sex + GESLipid + Education + GESLipid*Education |       |       |       |
| n        | 4502  | 4489  | 4503  |
| Intercept| −11.50 (−23.95; 1.00) | 0.070 | 49.89 (25.51; 74.26) | 6.1*10^{-14} | 80.32 (50.25; 110.40) | 1.7*10^{-14} |
| Age      | 0.05 (−0.11; 0.21) | 0.106 | 0.25 (0.11; 0.38) | 4.1*10^{-14} | 0.33 (0.19; 0.478) | 8.1*10^{-14} |
| Sex      | 14.79 (13.83; 15.75) | 3.91*10^{-18} | −1.07 (−3.28; 1.14) | 0.341 | 7.35 (5.99; 9.72) | 1.2*10^{-14} |
| GESLipid | 3.01 (2.24; 3.77) | 5.2*10^{-14} | 6.53 (4.68; 8.37) | 5.2*10^{-12} | 8.01 (6.08; 9.98) | 1.2*10^{-15} |
| Education (low) | 0.68 (−16.00; 14.64) | 0.931 | 1.84 (−26.52; 30.20) | 0.899 | 8.88 (−27.09; 44.86) | 0.628 |
| Education (high) | Ref | Ref | Ref | Ref | Ref | Ref |
| GESLipid × Edu# (low) | −0.56 (−2.09; 1.02) | 0.487 | −3.87 (−7.47; −0.32) | 0.033 | −3.64 (−7.44; 0.16) | 0.062 |
| GESLipid × Edu# (high) | −0.05 (−1.07; 0.92) | 0.885 | 0.14 (−2.16; 2.43) | 0.911 | −0.52 (−2.96; 1.98) | 0.686 |
| Lipid – age + sex + GESLipid + Income + GESLipid*Income |       |       |       |
| n        | 4232  | 4220  | 4233  |
| Intercept| −6.60 (−21.34; 4.14) | 0.186 | 55.82 (31.47; 80.16) | 7.2*10^{-14} | 88.25 (53.43; 118.07) | 7.0*10^{-14} |
| Age      | 0.02 (−0.04; 0.08) | 0.848 | 0.27 (0.13; 0.40) | 1.8*10^{-14} | 0.35 (0.20; 0.50) | 3.2*10^{-14} |
| Sex      | 14.07 (13.14; 15.01) | 5.51*10^{-17} | −0.14 (−2.41; 1.90) | 0.897 | 7.66 (5.37; 9.95) | 6.5*10^{-14} |
| GESLipid | 2.96 (2.14; 3.72) | 5.4*10^{-13} | 5.85 (4.01; 7.70) | 6.8*10^{-10} | 7.38 (5.46; 9.31) | 1.2*10^{-13} |
| Income (low) | 0.80 (−16.72; 18.33) | 0.928 | 6.54 (−24.83; 37.91) | 0.683 | −4.95 (−44.85; 34.94) | 0.808 |
| Income (high) | 0.56 (−17.31; 18.42) | 0.951 | 0.40 (−23.81; 23.61) | 0.981 | 23.38 (−18.03; 64.79) | 0.269 |
| GESLipid × Inc# (low) | −0.10 (−1.28; 1.02) | 0.833 | −0.23 (−2.79; 2.30) | 0.852 | 0.47 (−2.11; 2.12) | 0.726 |
| GESLipid × Inc# (high) | −0.15 (−1.28; 1.02) | 0.822 | 0.14 (−2.57; 2.84) | 0.911 | −1.56 (−4.42; 1.30) | 0.283 |

After including interaction terms of smoking, BMI, physical activity, alcohol consumption and diabetes separately into the GESLipid × Education interaction model, we observed a GESLipid × Diabetes interaction effect (ßGES × Diabetes: −4.46; 95%-CI: −7.38, −1.53) indicating less strong genetic effects on LDL-C in diabetics compared to non-diabetics (Table 4), which was also observed in the stratified analysis (Supplementary Figure S6). Including the GESLipid × Diabetes interaction effect also partly explained the GESLipid × Education interaction effect of the lowest compared to the highest education group, as the effect estimate was attenuated (ßGES × Education: −3.42; 95%-CI: −6.98, 0.18) (Table 4). The GESLipid × Education interaction was not affected by other lifestyle risk factors, as the respective GESLipid × low education interaction effect size estimates did not change in magnitude after including smoking, BMI, physical activity and alcohol consumption in the regression models (Table 4). In addition, results did not indicate GESLipid by lifestyle risk factor interactions.

Results of the extended GESLipid-EXT, which included additional SNPs selected from recent large-scale GWAS, showed overall smaller effect size estimates per GESLipid-EXT standard deviation for all three lipid traits compared to the GESLipid (Supplementary Table S6). The explained variance (R²) of the extended GESLipid-EXT was slightly higher (3.2%) compared to the GESLipid-EXT (2.6%) and GESTC (3.4%) slightly lower compared to the GESLipid-EXT and GESTC. In the extended GESLipid-EXT, by SEP indicator interaction analysis, using the highest SEP groups as reference, effect size estimates of interaction terms were overall slightly smaller in magnitude for all three lipid levels compared to the GESLipid. In the extended GESLipid-EXT stratified by education groups showed similar patterns for HDL-C, LDL-C, and TC compared to the GESLipid. However, differences in the genetic effects between education groups were less strong in magnitude, while for income no difference in the genetic effects were observed (Supplementary Figure S2). The downward trend between and within education groups with decreasing years of education and decreasing number of effect alleles in the analysis of single reference joint effects for HDL-C and the upward trend for LDL-C and TC for both SEP indicators also showed the same pattern using the extended GESLipid-EXT compared to the GESLipid (Supplementary Tables S8–S9).

Results of the single SNP main effect of all SNPs used in the GESLipid are presented in Supplementary Tables S2–S4. 61 out of 71 HDL-C-, 52 out of 58 LDL-C-, 65 out of 74 TC-associated SNPs were directionally consistent. Single SNP interaction analysis for education and lipid-associated SNPs showed that some SNPs showed overall smaller effect size estimates per GESLipid-EXT standard deviation for all three lipid traits compared to the GESLipid (Supplementary Table S6). The explained variance (R²) of the extended GESLipid-EXT was slightly higher (3.2%) compared to the GESLipid-EXT (2.6%) and for the extended GESLipid-EXT (2.6%) and GESTC (3.4%) slightly lower compared to the GESLipid-EXT and GESTC. In the extended GESLipid-EXT, by SEP indicator interaction analysis, using the highest SEP groups as reference, effect size estimates of interaction terms were overall slightly smaller in magnitude for all three lipid levels compared to the GESLipid (Supplementary Table S7). Effects of the extended GESLipid-EXT Stratified by education groups showed similar patterns for HDL-C, LDL-C, and TC compared to the GESLipid. However, differences in the genetic effects between education groups were less strong in magnitude, while for income no difference in the genetic effects were observed (Supplementary Figure S2). The downward trend between and within education groups with decreasing years of education and decreasing number of effect alleles in the analysis of single reference joint effects for HDL-C and the upward trend for LDL-C and TC for both SEP indicators also showed the same pattern using the extended GESLipid-EXT compared to the GESLipid (Supplementary Tables S8–S9).
contributed with stronger effects to the observed GESLDL-C by low education interaction on LDL-C in relation to the remaining SNPs (Supplementary Table S10). Similar differences of single SNP interaction effect size estimates were present for HDL-C- and TC-associated SNPs (Supplementary Tables S11–S12). Two of the LDL-C-associated loci with the strongest indication for interaction with education (i.e., PCSK9, MAFB) were also upon the strongest TC-associated SNPs. The two HDL-C-associated loci with the strongest indication for interaction with education were OR4C46 and LPL.

In the sensitivity analysis, main results for all lipid traits did not differ in direction and only slightly in magnitude after adjustment for lipid-lowering medication as well as after the exclusion of participants on lipid-lowering medication (Supplementary Tables S13–S14). The GESLDL-C by low education interaction effect size estimates did not change substantially compared to the main analysis population.

**Figure 1.** Sex-, age-adjusted effects per GES$_{lipid}$ standard deviation and corresponding 95% confidence interval (95% CI) of the genetic effect on (a) high-density lipoprotein cholesterol (HDL-C), (b) low-density lipoprotein cholesterol (LDL-C) and (c) total cholesterol (TC), stratified by education groups (years) and income tertiles in linear regression models using the GES$_{lipid}$ based on 14.
Discussion

The aim of the study was to investigate whether the SEP indicators education and income interact with genetic sum scores of lipid-increasing effect alleles in a population-based cohort study. To the best of our knowledge, this was the first study investigating G×E interaction with SEP as environmental factor on lipids in adults. Results gave some indication for an interaction between the GESLipid and the SEP indicator education, which was strongest for LDL-C. This was supported by stratified analysis in which the strongest genetic effects on LDL-C were observed in the high education group as well as by single reference joint effect analysis. After including information on smoking, BMI, physical activity, alcohol consumption and diabetes mellitus into analysis, there was an indication that a GESLDL-C by diabetes mellitus interaction partly explained the observed GESLDL-C by low education interaction. Using the extended GESLipid-EXT in comparison to the GESLipid, effect size measures were smaller but directionally consistent. Li et al. reported an interaction between rs2206734 SNP (CDKAL1), a favorable childhood environment and birthweight on metabolically obese, normal weight phenotype (defined as the presence of hypertension, hypertriglyceridemia, low serum HDL-C or impaired fasting plasma concentrations of glucose) in Chinese children. Their findings suggest that a favorable childhood environment represented by a composite score consisting of parental education, annual household income, high physical activity and fruit consumption can further amplify a protective effect of the CDKAL1 locus in children with a pediatric metabolic syndrome and high birthweight33. However, this study investigated a composite environmental score with parental SEP on a composite children’s phenotype in Chinese and can therefore only indirectly be compared with present study results.

Figure 2. Sex- and age-adjusted effects and corresponding 95% confidence intervals (95% CI) on high-density lipoprotein cholesterol (HDL-C) in linear regression models for single reference joint effects of tertiles of a HDL-C-associated genetic effect allele sum score (GESHDL-C based on14) and socioeconomic position indicators, calculated separately for education groups and income tertiles, with the group of having a high GESHDL-C and the highest socioeconomic position as reference.
Recent studies have investigated statistical interactions between genetic risk scores and SEP-related lifestyle factors or health behaviors such as physical activity, dietary patterns and BMI and their effects on lipids\textsuperscript{27,28}. Cole et al. (2014) have demonstrated in a population of European ancestry that the effect of a genetic risk score consisting of HDL-C-increasing alleles has been stronger for lean than for obese (BMI ≥ 35 kg/m\textsuperscript{2}) study participants. These interactions have been largely driven by the SNPs rs3764261 (\textit{CETP}), rs4846914 (\textit{GALNT2}), rs7241918 (\textit{LIPG}) and rs6065906 (\textit{PLTP})\textsuperscript{28}. As high education is strongly associated with low BMI\textsuperscript{48}, these results may at least partly reflect the results of the present study. However, in the present study the SNPs representing the loci \textit{CETP}, \textit{GALNT2}, \textit{LIPG} and \textit{PLTP} did not show indication for SNP by education interaction on HDL-C.

Justesen et al. (2015) have reported an interaction between a genetic risk score of HDL-C-decreasing effect alleles and physical activity in a Danish population (n = 5961), suggesting that the genetic risk score has exerted a smaller effect in physically active compared to inactive individuals. However, this interaction was statistically not significant in a replication cohort of smaller sample size\textsuperscript{27}. As higher education is strongly associated with low BMI\textsuperscript{48}, the results of the present study may represent the same interaction signal, i.e. a stronger genetic effect of HDL-C-increasing effect alleles on HDL-C in population groups of higher education groups.

Recent SNP×E interaction analyses have identified several lipid-associated loci interacting with lifestyle factors such as smoking and diet\textsuperscript{50–52}. While Junyent et al. (2009) have reported an interaction of rs6720173 (\textit{ABCG5}) and rs11887534, rs6709904, rs4148217 (\textit{ABCG8}) with smoking\textsuperscript{50}, Lu et al. (2010) have shown an interaction of rs174546 (\textit{FADS1}) with intake of n-3 and n-6 polyunsaturated fatty acid\textsuperscript{51} and Kim et al. (2013) have demonstrated an interaction of rs2072183 (\textit{NPC1L1}) with cholesterol intake in male only\textsuperscript{52}. Although SEP is strongly associated with smoking and dietary factors, none of these loci showed indication for interaction with education in the present study. The $\beta$ $\text{GES}_{\text{LDL-C}}$ $\times$ Diabetes mellitus interaction on LDL-C observed in the present study is not statistically significant.

### Figure 3.

Sex- and age-adjusted effects and corresponding 95% confidence intervals (95% CI) on low-density lipoprotein cholesterol (LDL-C) in linear regression models for single reference joint effects of tertiles of a LDL-C-associated genetic effect allele sum score ($\text{GES}_{\text{LDL-C}}$ based on\textsuperscript{14}) and socioeconomic position indicators, calculated separately for education groups and income tertiles, with the group of having a low $\text{GES}_{\text{LDL-C}}$ and the highest socioeconomic position as reference.

| Education | $\beta$ (95% CI) | p     | n   |
|-----------|-----------------|-------|-----|
| ≤10 years |                 |       |     |
| High GES  | 13.36 (6.98; 19.74) | 4.1*10\textsuperscript{-5} | 172 |
| Middle GES | 11.79 (5.42; 18.16) | 2.9*10\textsuperscript{-5} | 174 |
| Low GES   | 7.06 (0.64; 13.49)  | 0.031 | 169 |
| 11-13 years |                |       |     |
| High GES  | 14.86 (10.76; 18.96) | 1.4*10\textsuperscript{-12} | 836 |
| Middle GES | 12.68 (8.57; 16.79) | 1.6*10\textsuperscript{-9} | 820 |
| Low GES   | 0.70 (-3.38; 4.77)  | 0.740 | 846 |
| ≥14 years |                 |       |     |
| High GES  | 12.77 (8.32; 17.23) | 2.0*10\textsuperscript{-8} | 519 |
| Middle GES | 5.50 (1.00; 10.00)  | 0.017 | 495 |
| Low GES   | ref.             |       | 472 |

| Income | $\beta$ (95% CI) | p     | n   |
|--------|-----------------|-------|-----|
| Lowest Tertile |                 |       |     |
| High GES  | 15.33 (10.94; 19.72) | 8.9*10\textsuperscript{-12} | 534 |
| Middle GES | 12.84 (8.38; 17.30) | 1.8*10\textsuperscript{-9} | 504 |
| Low GES   | 4.32 (-0.12; 8.75)  | 0.056 | 510 |
| Middle Tertile |              |       |     |
| High GES  | 14.26 (9.70; 18.82) | 9.6*10\textsuperscript{-10} | 454 |
| Middle GES | 11.50 (6.73; 16.27) | 2.3*10\textsuperscript{-9} | 388 |
| Low GES   | 2.67 (-2.06; 7.40)  | 0.268 | 397 |
| Highest Tertile |              |       |     |
| High GES  | 13.01 (8.45; 17.57) | 2.4*10\textsuperscript{-9} | 451 |
| Middle GES | 8.36 (3.92; 12.80)  | 2.3*10\textsuperscript{-9} | 502 |
| Low GES   | ref.             |       | 493 |

$\beta$ $\text{GES}_{\text{LDL-C}}$
study partly explained the GESLDL-C×Education interaction. Deng et al. found a Gene×Diabetes mellitus interaction. In their study the SNP rs16996148 (CILP2) decreased the risk of hyperlipidemia, whereas rs16996148 GT/TT and diabetes mellitus as well as rs16996148 TT and diabetes mellitus increased the risk of hyperlipidemia29.

In the present analysis, smaller effects on lipids were observed for using the extended GESLipid-EXT compared to the GESLipid. This may be caused by the overall smaller effect size of newly discovered loci as a result of larger analysis populations in recent GWAS, making it possible to detect risk alleles with very small effects. It may also be due to the recently published GWAS meta-analyses that were based on single large cohorts potentially producing less generalizable study results16,17. The overall smaller main effects of the additional SNPs included in the extended GESLipid-EXT have led to smaller interaction effect size estimates, i.e. less strong impact of SEP on the expression of LDL-C-related genetic risks. One possible mechanism that has been hypothesized in this regard is epigenetic modification. In contrast to an individual's

---

### Table: Sex- and age-adjusted effects and corresponding 95% confidence intervals (95% CI) on total cholesterol (TC) in linear regression models for single reference joint effects of tertiles of a TC-associated genetic effect allele sum score (GESTC based on14) and socioeconomic position indicators, calculated separately for education groups and income tertiles, with the group of having a low GESTC and the highest socioeconomic position as reference.

| Education | β (95% CI) | p       | n   |
|-----------|------------|---------|-----|
| ≤10 years |
| High GES  | 15.40 (8.72; 22.07) | 6.3*10^-6 | 181 |
| Middle GES| 11.99 (5.21; 18.76) | 5.3*10^-6 | 172 |
| Low GES   | 7.89 (0.96; 14.82)  | 0.026    | 162 |
| 11-13 years |
| High GES  | 17.97 (13.66; 22.29) | 4.1*10^-16 | 820 |
| Middle GES| 10.02 (5.74; 14.30) | 4.5*10^-6 | 850 |
| Low GES   | 1.16 (-3.13; 5.45)  | 0.597    | 832 |
| ≥14 years |
| High GES  | 16.69 (11.96; 21.41) | 5.1*10^-12 | 505 |
| Middle GES| 8.56 (3.75; 13.37)  | 4.9*10^-4 | 473 |
| Low GES   | ref.        |         | 508 |

### Table: Sex- and age-adjusted effects and corresponding 95% confidence intervals (95% CI) on total cholesterol (TC) in linear regression models for single reference joint effects of tertiles of a TC-associated genetic effect allele sum score (GESTC based on14) and socioeconomic position indicators, calculated separately for education groups and income tertiles, with the group of having a low GESTC and the highest socioeconomic position as reference.

| Income | β (95% CI) | p       | n   |
|--------|------------|---------|-----|
| Lowest Tertile |
| High GES | 18.34 (13.69; 23.00) | 1.5*10^-14 | 524 |
| Middle GES | 12.25 (7.64; 16.86)  | 2.0*10^-7 | 542 |
| Low GES   | 2.42 (-2.32; 7.16)   | 0.317    | 482 |
| Middle Tertile |
| High GES  | 16.96 (12.07; 21.84) | 1.1*10^-11 | 430 |
| Middle GES| 8.10 (3.08; 13.12)   | 1.6*10^-3 | 394 |
| Low GES   | 3.87 (-1.06; 8.80)   | 0.124    | 415 |
| Highest Tertile |
| High GES  | 16.50 (11.71; 21.28) | 1.6*10^-11 | 463 |
| Middle GES| 10.25 (5.48; 15.01)  | 2.5*10^-9 | 469 |
| Low GES   | ref.         |         | 514 |

---

**Figure 4.** Sex- and age-adjusted effects and corresponding 95% confidence intervals (95% CI) on total cholesterol (TC) in linear regression models for single reference joint effects of tertiles of a TC-associated genetic effect allele sum score (GESTC based on14) and socioeconomic position indicators, calculated separately for education groups and income tertiles, with the group of having a low GESTC and the highest socioeconomic position as reference.
The epigenome is subject to environmentally induced changes during the life course, but is crucial for the regulation of gene expression. Several studies have indeed reported SEP-related differences in epigenetic markers60,61. Interestingly, the lifestyle factors BMI, physical activity, smoking and alcohol consumption did not account for the observed LDL-C by education interaction, while diabetes mellitus accounted for it only partly. Consequently, it has to be assumed that other risk factors besides those included in the present analysis may have a mediating effect on the found LDL-C by education interaction. One explanation for the stronger genetic effects on LDL-C in higher education groups may be that non-genetic health risks being of lower prevalence in high education groups leading to LDL-C profiles that are stronger affected by genetic than by non-genetic risk factors. This hypothesis is supported by the explained proportion of the variance ($R^2$) of the GESLDL-C on LDL-C, which was higher in the two higher education groups compared to the lower education group.

The effect of SEP indicators on health is outcome specific and each indicator operates via different pathways linking social factors to health outcomes43,62. Even though educational attainment and income are moderately

Table 4. Sex- and age- adjusted effects per GESLDL-C standard deviation and corresponding 95% confidence intervals (95% CI) on LDL-C in linear regression models including main effects and interaction terms of a LDL-C-associated genetic effect allele score (GESLDL-C based on14), education groups and SEP-related life style risk factors (i.e., current smoking [S], BMI, physical activity [PA], alcohol consumption [A; per 100 g/week]) and diabetes mellitus [D]. # Education

| Education | LDL-C ~ age + sex + Edu + GESLDL-C | LDL-C ~ age + sex + Edu + GESLDL-C + BMI | LDL-C ~ age + sex + Edu + GESLDL-C + BMI + PA | LDL-C ~ age + sex + Edu + GESLDL-C + A | LDL-C ~ age + sex + Edu + GESLDL-C + D |
|-----------|-----------------------------------|-------------------------------------------|------------------------------------------|-----------------------------------|----------------------------------|
| Edu        | β (95%-CI)                         | β (95%-CI)                                | β (95%-CI)                                | β (95%-CI)                         | β (95%-CI)                       |
| Intercept  | 41.87 (16.61; 67.14)               | 46.66                                     | 46.69                                     | 43.82                             | 44.69                            |
| Age        | 0.01 (0.27; 0.45)                  | 0.01 (0.27; 0.45)                         | 0.01 (0.27; 0.45)                         | 0.01 (0.27; 0.45)                  | 0.01 (0.27; 0.45)                |
| Sex        | 0.50 (2.59; 3.86)                  | 0.750                                     | 0.750                                     | 0.750                             | 0.750                            |
| PA         | -0.04 (26.25; 26.16)               | 1.000                                     | 1.000                                     | 1.000                             | 1.000                            |
| A          | -0.01 (0.00; 0.00)                 | 0.960                                     | 0.960                                     | 0.960                             | 0.960                            |
| D          | -4.66 (7.38; 15.63)                | -4.66                                     | -4.66                                     | -4.66                             | -4.66                            |

The effect of SEP indicators on health is outcome specific and each indicator operates via different pathways linking social factors to health outcomes43,62. Even though educational attainment and income are moderately
correlated ($r = 0.45$) in the present study, the different strength of genetic effect in education groups on LDL-C could not be seen in income tertiles. The net effect of education is reflected among others in the ability to turn health related information into behavior and facilitates understanding of therapeutic measures\cite{10}. Which could support the hypothesis that in highly educated, due to the ability to create environments with less health risks, genetic influence on LDL-C might be stronger. Furthermore, education as a marker of childhood social environment could, due to the duration of exposure until adulthood, be more likely support epigenetic changes. Material resources do not seem to modify genetic risk on LDL-C.

Strengths of the present study were its population-based study sample and the use of two different individual SEP indicators in the analysis. Even though education and income are correlated SEP indicators, each of them represents certain aspects of SEP related to different health behaviors and risks. Moreover, two different $G_{ \text{ESLipid}}$ for each lipid trait were compared, allowing to check for differences in the genetic effects and $G \times E$ interactions between scores derived by different GWAS study populations. The sample size and the limited statistical power for single SNP analysis had to be mentioned as limitation of the present study. However, indication for interaction was based on the cumulative genetic risk of the study participants. Another limitation was the cross-sectional design of the study that does not allow for strong conclusions on causality of effects. However, educational attainment is usually acquired in adolescence or early adulthood and lipids were assessed at an older age in the present study. Due to this exposure-outcome temporality reverse causation is very unlikely. Even if the effect of education on lipids was not causal, a modification of the $G_{ \text{ESLipid}}$ effect on LDL-C by education would still be of interest, because the knowledge of the heterogeneous genetic effects in different education groups could be interesting for CVD risk prediction and genotype-based targeted interventions\cite{11}. Furthermore, this knowledge supports CVD lifestyle-based interventions in lower education groups due to lower genetic effect on LDL-C in these groups. Finally, it cannot be excluded that the indication for a $G_{ \text{ESLipid}} \times E$ Education interaction could have been observed randomly due to the number of independent tests performed. However, the number of independent tests performed are justified through the three lipid end points and the two SEP indicators and we have calculated 95% confidence intervals to report the precision of the obtained effect size estimates. Furthermore, the interaction analyses results are supported by the results of the stratified and single reference joint effect analysis, which showed a constant pattern across phenotypes and indicated difference in genetic effect sizes between the education groups.

The results of the present study gave some indication for interaction between genetic variants associated with LDL-C and education in a population-based cohort study. Stronger genetic effects were observed in groups of higher education, which seemed to be partly mediated by diabetes mellitus but not by other life style risk factors such as BMI, smoking, physical activity and alcohol consumption. This gave supporting evidence that SEP has an impact on the expression of genetic susceptibility related to LDL-C. Further research is needed to replicate our findings in independent study samples, investigate possible biological mechanisms behind the interaction and to assess the potential of the found gene by SEP interactions for improving CVD prediction. Additionally, our study included only individuals of European origin and therefore the results may not be applied to populations of other ethnicities.

Data availability
Due to data security reasons (i.e., data contain potentially participant identifying information), the Heinz Nixdorf Recall Study does not allow sharing data as a public use file. However, other authors are allowed to access data upon request, which is the same way authors of the present paper obtained the data. Data requests can be addressed to: recall@uk-essen.de.

Received: 5 October 2020; Accepted: 2 August 2021
Published online: 16 August 2021

References
1. Wadhera, R. K. et al. A review of low-density lipoprotein cholesterol, treatment strategies, and its impact on cardiovascular disease morbidity and mortality. J Clin Lipidol 10, 472–489 (2016).
2. Piepoli, M. F. et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: the Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J 37, 2315–2381 (2016).
3. Holmes, M. V. et al. Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J 36, 539–550 (2015).
4. Chapman, M. J. et al. Triglyceride-rich lipoproteins and high-density lipoprotein cholesterol in patients at high risk of cardiovascular disease: evidence and guidance for management. Eur Heart J 32, 1345–1361 (2011).
5. Rohatgi, A. et al. HDL in the 21st Century: a multifunctional roadmap for future HDL research. Circulation 143, 2293–2309 (2021).
6. Sing, C. E., Stengård, J. H. & Kardia, S. L. R. Genes, environment, and cardiovascular disease. Arterioscler Thromb Vasc Biol 23, 1190–1196 (2003).
7. Luepker, R. V. et al. Socioeconomic status and coronary heart disease risk factor trends. The Minnesota Heart Survey. Circulation 88, 2172–2179 (1993).
8. Perova, N. V. et al. Multi-country comparison of plasma lipid relationship to years of schooling in men and women. Int J Epidemiol 30, 371–379 (2001).
9. Fukuda, Y. & Hiyoshi, A. Associations of household expenditure and marital status with cardiovascular risk factors in Japanese adults: analysis of nationally representative surveys. J Epidemiol 23, 21–27 (2013).
10. Shohaimi, S. et al. Distribution of lipid parameters according to different socio-economic indicators- the EPIC-Norfolk prospective population study. BMC Public Health 14, 782 (2014).
11. Fujishoshi, N. et al. Associations between socioeconomic status and the prevalence and treatment of hypercholesterolemia in a general Japanese population: NIPPON DATA2010. J Atheroscler Thromb 25, 606–620 (2018).
12. Benetou, V. et al. Total cholesterol and HDL-cholesterol in relation to socioeconomic status in a sample of 11,645 Greek adults: the EPIC study in Greece. European Prospective Investigation into Nutrition and Cancer. Scand J Public Health 28, 260–265 (2000).
58. Cao, M. & Cui, B. Association of educational attainment with adiposity, type 2 diabetes, and coronary artery diseases: a mendelian randomization study. *Front Public Health* **8**, 112 (2020).

59. Krieger, N., Williams, D. R. & Moss, N. E. Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu Rev Public Health* **18**, 341–378 (1997).

60. Borghol, N. et al. Associations with early-life socio-economic position in adult DNA methylation. *Int J Epidemiol* **41**, 62–74 (2012).

61. McGuinness, D. et al. Socio-economic status is associated with epigenetic differences in the pSoBid cohort. *Int J Epidemiol* **41**, 131–160 (2012).

62. Fliesser M, De Witt Huberts J, Wippert P-M. The choice that matters: the relative influence of socioeconomic status indicators on chronic back pain: a longitudinal study. *BMC Health Serv Res*; 17. Epub ahead of print 2 December 2017. doi: https://doi.org/10.1186/s12913-017-2735-9.

Acknowledgements

We are indebted to all study participants and to both the dedicated personnel of the study center of the Heinz Nixdorf Recall study and to the investigative group, in particular to U. Slomiany, E. M. Beck, A. Öffner, S. Münkel, R. Peter, and H. Hirche. Advisory Board: Meinertz T., Hamburg, Germany (Chair); Bode C., Freiburg, Germany; deFeyter P. J., Rotterdam, Netherlands; Güntert B, Hall, Austria; Gutzwiller F., Bern, Switzerland; Heinen H., Bonn, Germany; Hess O., Bern, Switzerland; Klein B., Essen, Germany; Löwel H., Neuherberg, Germany; Reiser M., Munich, Germany; Schmidt G., Essen, Germany; Schwaiger M., Munich, Germany; Steinmüller C., Bonn, Germany; Theorell T., Stockholm, Sweden; Willich S. N., Berlin, Germany.

Author contributions

BS formed the concept of the work. ND, RE and KHJ contributed to data acquisition. CE performed the statistical analysis and drafted the manuscript. MN performed sequencing of the genetic data. All authors contributed to the interpretation of results, revised the paper draft and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Heinz Nixdorf Foundation; the German Ministry of Education and Science [Grants Nationales Genomforschungsnetz, 01GS0820], the German Research Council [Projects SI 236/8-1, SI 236/9-1].

Competing interests

The authors declare no competing interests.

Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-95970-z.

**Correspondence** and requests for materials should be addressed to C.E.

**Reprints and permissions information** is available at www.nature.com/reprints.

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

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

© The Author(s) 2021