## **Supplementary Methods**

## **Study populations and data sources for observational and MR analyses**

## Study of Health in Pomerania (SHIP)

The Study of Health in Pomerania (SHIP) is a cross-sectional survey in West Pomerania, the northeastern area of Germany<sup>1</sup>. A sample from the population aged 20 to 79 years was drawn from population registries. First, the three cities of the region (with 17,076 to 65,977 inhabitants) and the 12 towns (with 1516 to 3044 inhabitants) were selected, and then 17 out of 97 smaller towns (with less than 1500 inhabitants), were drawn at random. Second, from each of the selected communities, subjects were drawn at random, proportional to the population size of each community and stratified by age and gender. Only individuals with German citizenship and main residency in the study area were included. 7008 subjects aged 20 to 79 years were sampled, with 292 persons of each gender in each of twelve five-year age strata. The net sample (without out-migrated or deceased persons) comprised 6267 eligible subjects, of whom 4308 eventually participated in the baseline SHIP study (SHIP-0) with a response rate of 68.8%. Of these subjects, individuals with missing genotype or phenotype data were excluded. Valid data were available in 3812 individuals of the SHIP-0 population. At the SHIP-0 visit (baseline), non-fasting serum IGF-I and IGFBP-3 concentrations were determined by automated two-site chemiluminescence immunoassays (Nichols Advantage®; Nichols Institute Diagnostica GmbH, Bad Vilbel, Germany). In both assays, in the first step biotin labelled monoclonal antibodies were employed as capture, and in the second step the acridinium-ester labelled monoclonal antibodies were used for detection. To determine IGF-I concentration the samples were acidified to separate IGF-I from IGF-binding proteins. The analytical sensitivity was 6 ng/mL. The IGF-I assay has been calibrated against the World Health Organization International Reference Reagent 1988, Insulin-like Growth Factor-I 87/518. The analytical sensitivity of the IGFBP-3 assay was 20 ng/mL. The assay reference standard was analytically prepared with glycosylated recombinant human IGFBP-3. Only one lot of reagents was used for all IGFBP-3 measurements2. Total cholesterol and HDL concentrations were measured photometrically (Hitachi 704, Roche, Mannheim, Germany), and serum LDL was measured by applying a precipitation procedure using dextran sulphate (Immuno, Heidelberg, Germany) on an Epos 5060 (Eppendorf, Hamburg, Germany). The serum aliquots were prepared for immediate analysis and for storage at -80 °C in the Integrated Research Biobank (Liconic, Liechtenstein).

## IGF working group of the CHARGE consortium

The most recent GWAS meta-analysis (n=30,884 for IGF-I and n= 18,995 for IGFBP-3 of individuals of European descent) conducted by the IGF working group of the CHARGE consortium identified 7 and 4 SNPs that reached genome-wide significance  $(p<5x10^{-8})$  for circulating IGF-I and IGFBP-3 levels, respectively<sup>2</sup>. The per-allele effect of each SNP on IGF-I and IGFBP-3 was reported (in SD units). An SD unit of IGF-I and IGFBP-3 is equivalent to 57.6 ng/ml and 506 ng/ml, respectively (calculated using data from the SHIP cohort). The methods applied to measure IGF-I and IGFBP-3 levels differed among the participating cohorts and include enzyme-linked immunosorbent assays (ELISA), radioimmunoassays (RIA) and chemiluminescent immunoassays, as previously described2. Participants were genotyped on genome-wide arrays and SNPs were imputed using the HapMap2 reference panel. Association analyses in individual studies were performed on IGF-I or IGFBP-3 levels measured in ng/mL using a multiple linear regression with an additive genetic model based on allele dosages adjusted for age and stratified by sex. All cohorts accounted for relatedness, population substructure using genetic principal components, study centre and laboratory batch of IGF-I or IGFBP-3 measurement where applicable. Individuals of non-European ancestry, with missing phenotype data, a diagnosis of growth hormone deficiency or known use of human growth hormones were excluded from the analyses. Betas and SEs were estimated from the p-values using a previously described method3. As an approximation, the phenotypic SD was set to 1, resulting in an effect change per unit SD IGF.

## Global lipids consortium (GLGC)

The most recent GWAS meta-analysis conducted by the Global Lipids Genetic Consortium (GLCL) and Do et al. identified 185 independent SNPs (linkage disequilibrium  $r^2 < 0.05$ ) (in 157 loci) that met genome-wide significance ( $p < 5x10^{-8}$ ) for circulating HDL-C, LDL-C and TG in a sample of 188,577 individuals of European descent<sup>4,5</sup>. These SNPs explained approximately 13.7%, 14.6% and 11.7% of the variance in circulating levels of HLD-C, LDL-C and TG, respectively. Blood lipid levels were typically measured after >8hr of fasting. Individuals taking lipid-lowering medications were excluded when possible. In most studies, total cholesterol, HDL-C and TG were measured after >8hr of fasting. LDL cholesterol levels were directly measured in 10 studies and were estimated using the Friedewald formula in the remaining studies<sup>6</sup>. The per-allele effect of each SNP on each lipid trait was reported in SD units based on inverse normal transformed residuals of lipid concentrations, after adjustment for age and sex. An SD unit of LDL-C, HDL-C and TG on the observed scale is equivalent to approximately 39.9mg/dL, 17.9mg/dL and 81.8mg/dL, respectively7. Participants were genotyped using the Illumina iSelect Metabochip genotyping array. Individual SNP association tests were performed using linear regression with the inverse normal transformed trait values as the dependent variable and the expected allele count for each individual as the independent variable. Meta-analysis was performed using the Stouffer method with weights proportional to the square root of the sample size for each sample.

#### Breast Cancer Association Consortium (BCAC)

The BCAC consortium conducted a GWAS meta-analysis to identify SNPs associated with breast cancer risk involving 122,977 breast cancer cases and 105,974 controls. Genotyping of BCAC samples were performed using an Illumina Custom Infinium genotyping array (OncoArray) consisting of ~570 000 single nucleotide polymorphisms (SNPs) and imputed using the 1000 Genomes Project as a reference panel as previously described<sup>8</sup>. All SNPs with poor imputation quality ( $r^2$ <0.30), a minor allele frequency of <1%, a call rate of <95% or evidence of Hardy-Weinberg equilibrium were removed. All participating studies were approved by their appropriate ethics review board and all individuals provided informed consent.

## UK Biobank

UK Biobank is a population-based health research resource consisting of approximately 500 000 people, aged between 38 years and 73 years, who were recruited between the years 2006 and 2010 from across the UK<sup>9</sup>. Particularly focused on identifying determinants of human diseases in middleaged and older individuals, participants provided a range of information (such as demographics, health status, lifestyle measures, cognitive testing, personality self-report, and physical and mental health measures) via questionnaires and interviews; anthropometric measures, BP readings and samples of blood, urine and saliva were also taken (data available at www.ukbiobank.ac.uk). A full description of the study design, participants and quality control (QC) methods have been described in detail previously<sup>10</sup>. UK Biobank received ethical approval from the Research Ethics Committee (REC reference for UK Biobank is 11/NW/0382). The full data release contains the cohort of successfully genotyped samples (n=488,377). 49,979 individuals were genotyped using the UK BiLEVE array and 438 398 using the UK Biobank axiom array. Pre-imputation QC, phasing and imputation are described elsewhere11. In brief, prior to phasing, multiallelic SNPs or those with MAF ≤1% were removed. Phasing of genotype data was performed using a modified version of the SHAPEIT2 algorithm<sup>12</sup>. Genotype imputation to a reference set combining the UK10K haplotype and HRC reference panels<sup>13</sup> was performed using IMPUTE2 algorithms<sup>14</sup>. The analyses presented here were restricted to autosomal variants within the HRC site list using a graded filtering with varying imputation quality for different allele frequency ranges<sup>15</sup>. Therefore, rarer genetic variants are required to have a higher imputation INFO score (Info>0.3 for MAF >3%; Info>0.6 for MAF 1-3%; Info>0.8 for MAF 0.5-1%; Info>0.9 for MAF 0.1-0.5%) with MAF and Info scores having been recalculated on an in-house derived 'European' subset<sup>15</sup>. Individuals with sex-mismatch (derived by comparing genetic sex and reported sex) or individuals with sex-chromosome aneuploidy were excluded from the analysis (n=814). We restricted the sample to individuals of 'European' ancestry as defined by an in-house kmeans cluster analysis performed using the first 4 principal components provided by UK Biobank in the statistical software environment R. The current analysis includes the largest cluster from this analysis (n=462,933)<sup>15</sup>. The UK Biobank database contained records of all cancer including their subtypes using the International Classification of Diseases (ICD 10 and ICD9) and their self-reported cancer data.

We utilized the ICD10 and ICD9 fields, and records of self-reported status of breast cancer to identify individuals with breast cancer within UKBB, as described previously<sup>16</sup>. We did not stratify by whether the case was a prevalent or incident case. Female participants were defined as controls if they had no record of any cancer, in-situ carcinoma or an undefined neoplasm. In total, 15,349 BCa cases and 215,605 controls were included in the breast cancer GWAS. We also carried out GWAS for breast cancer risk stratified by menopausal status (pre- and post-menopausal). The criteria for premenopausal were females aged ≤55 years old who reported that they still had periods, did not report a history of hysterectomy or bilateral oophorectomy and menarche age ≥7 years old. Postmenopausal females were defined as those who reported no longer having periods and did not report a history of hysterectomy or bilateral oophorectomy and their menopause age ≥40 years old. These criteria were used to minimise inclusion of both pre-mature and medically induced pre and postmenopausal women. In total, 59,124 participants (1781 cases and 57,343 controls) were in the premenopause group and 112,245 participants (8750 cases and 103,495 controls) were in the postmenopause group.

Genome-wide association analysis (GWAS) was conducted using linear mixed model (LMM) association method as implemented in BOLT-LMM  $(v2.3)^{17}$ . To model population structure and relatedness in the sample we used 143,006 directly genotyped SNPs, obtained after filtering on MAF > 0.01; genotyping rate > 0.015; Hardy-Weinberg equilibrium p-value < 0.0001 and LD pruning to an

 $r^2$  threshold of 0.1 using PLINKv2.00. Genotype array and sex were adjusted for in the model. BOLT-LMM association statistics are on the linear scale. As such, test statistics (betas and their corresponding standard errors) were transformed to log odds ratios and their corresponding 95% confidence intervals on the liability scale using a Taylor transformation expansion series<sup>18</sup>.

# Statistical analysis

Multivariable MR allows genetic variants associations with other variables included in the model (i.e. a genetic variant may associate with more than one lipid trait, so long as any association with the outcome is via a given lipid trait)<sup>22,23</sup>. A weighted regression-based method was performed by regressing the SNP-outcome association estimate on each of the SNP-exposure association estimates in a multivariable weighted regression model, using the standard error as weights (inversevariance weighted). This weighted linear regression-based method is equivalent to a two-stage least squares regression analysis<sup>22</sup>.

MR-Egger regression is used to test overall directional pleiotropy to provide valid causal estimates, taking into account the presence of pleiotropy<sup>24</sup>. The weighted median method provides a consistent estimate of causal effect if at least 50% of the information in the analysis comes from variants that are valid instrumental variables $^{25,26}$ .

## **References**

- 1. Volzke H, Alte D, Schmidt CO, et al. Cohort Profile: The Study of Health in Pomerania. *Int J Epidemiol*. 2011 Apr 1;**40**(2):294–307.
- 2. Teumer A, Qi Q, Nethander M, et al. Genomewide meta-analysis identifies loci associated with IGF-I and IGFBP-3 levels with impact on age-related traits. *Aging Cell*. 2016 Oct;**15**(5):811– 824.
- 3. Rietveld CA, Medland SE, Derringer J, et al. GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science (80- )*. 2013 Jun 21;**340**(6139):1467–1471.
- 4. Do R, Willer CJ, Schmidt EM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet*. 2013;**45**(11):1345–1352.
- 5. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013 Nov 6;**45**(11):1274–1283.
- 6. 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*. 1972/06/01. 1972;**18**(6):499–502.
- 7. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;**45**(11):1274–1283.
- 8. Amos CI, Dennis J, Wang Z, et al. The OncoArray Consortium: a Network for Understanding

the Genetic Architecture of Common Cancers. *Cancer Epidemiol Prev Biomarkers*. 2016;**26**(1):126–135.

- 9. Allen NE, Sudlow C, Peakman T, Collins R, UK Biobank. UK Biobank Data: Come and Get It. *Sci Transl Med*. 2014 Feb 19;**6**(224):224ed4-224ed4.
- 10. Collins R. What makes UK Biobank special? *Lancet*. 2012 Mar 31;**379**(9822):1173–1174.
- 11. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*. 2017 Jul 20;
- 12. O'Connell J, Sharp K, Shrine N, et al. Haplotype estimation for biobank-scale data sets. *Nat Genet*. Nature Publishing Group; 2016 Jul 6;**48**(7):817–820.
- 13. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun*. Nature Publishing Group; 2015 Nov 14;**6**(1):8111.
- 14. Howie B, Marchini J, Stephens M. Genotype Imputation with Thousands of Genomes. *G3: Genes|Genomes|Genetics*. 2011 Nov 4;**1**(6):457–470.
- 15. Mitchell, R, Hemani, G, Dudding, T, Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, Version 1. 2017;
- 16. Al-Ajmi K, Lophatananon A, Ollier W, Muir KR. Risk of breast cancer in the UK biobank female cohort and its relationship to anthropometric and reproductive factors. Meyre D, editor. *PLoS One*. Public Library of Science; 2018 Jul 26;**13**(7):e0201097.
- 17. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015 Mar 2;**47**(3):284–290.
- 18. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobankscale datasets. *Nat Genet*. Nature Publishing Group; 2018 Jul 11;**50**(7):906–908.
- 19. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013 Feb 1;**42**(1):97–110.
- 20. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013 Feb;**42**(1):111–127.
- 21. UK10K Consortium, Walter K, Min JL, et al. The UK10K project identifies rare variants in health and disease. *Nature*. 2015 Oct 14;**526**(7571):82–90.
- 22. Burgess S, Thompson SG. Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *Am J Epidemiol*. Oxford University Press; 2015 Feb 15;**181**(4):251–260.
- 23. Burgess S, Dudbridge F, Thompson SG. Re: "Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects". *Am J Epidemiol*. 2015 Feb 15;**181**(4):290–291.
- 24. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;**40**(4):304–314.
- 25. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;**40**(4):304–314.
- 26. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, Consortium E-I. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol*. 2015;**30**(7):543–552.

<b>SNP</b>	<b>Effect allele</b>	Other allele	<b>Beta</b> $(LDL-C)$	se $(LDL-C)$	pval $(LDL-C)$	<b>Beta</b> $(HDL-C)$	se $(HDL-C)$	pval $(HDL-C)$	beta (TG)	se (TG)	pval (TG)
rs10019888	A	G	$-1.80E-02$	4.98E-03	3.00E-04	2.70E-02	4.64E-03	6.00E-09	$-2.30E-02$	4.58E-03	5.00E-07
rs10029254	Τ	$\overline{c}$	5.80E-03	4.33E-03	1.80E-01	$-8.50E-03$	4.08E-03	3.70E-02	2.70E-02	3.97E-03	1.00E-11
rs1010167	$\mathsf{C}$	G	$-2.50E-02$	3.97E-03	3.00E-10	4.40E-03	3.67E-03	2.30E-01	$-1.60E-03$	3.64E-03	6.60E-01
rs10102164	A	G	3.20E-02	4.59E-03	3.00E-12	$-5.00E-04$	3.98E-03	9.00E-01	1.10E-02	3.95E-03	5.40E-03
rs10282707	$\overline{\mathsf{T}}$	$\overline{c}$	$-8.40E-03$	3.75E-03	2.50E-02	$-2.50E-02$	3.49E-03	8.00E-13	9.20E-03	3.42E-03	7.20E-03
rs103294	$\overline{\mathsf{T}}$	$\overline{c}$	7.30E-03	4.70E-03	1.20E-01	5.20E-02	4.34E-03	4.00E-33	$-2.10E-03$	4.12E-03	6.10E-01
rs1035744	$\mathsf T$	C	6.90E-03	4.13E-03	9.50E-02	$-5.50E-03$	3.82E-03	1.50E-01	2.10E-02	3.82E-03	4.00E-08
rs10401969	T	C	1.20E-01	7.32E-03	2.00E-60	$-1.30E - 02$	6.83E-03	5.70E-02	1.20E-01	6.49E-03	3.00E-76
rs10493326	Α	G	2.10E-02	4.21E-03	6.00E-07	$-1.30E-03$	3.92E-03	7.40E-01	3.10E-02	3.87E-03	1.00E-15
rs10513688	A	G	2.20E-02	6.09E-03	3.00E-04	$-4.90E-03$	5.58E-03	3.80E-01	3.10E-02	5.65E-03	4.00E-08
rs10790162	A	G	7.60E-02	7.17E-03	3.00E-26	$-9.50E-02$	6.65E-03	3.00E-46	2.30E-01	6.47E-03	1.00E-276
rs10832962	T	C	3.20E-02	4.03E-03	2.00E-15	4.30E-03	3.74E-03	2.50E-01	1.10E-02	3.68E-03	2.80E-03
rs10903129	Α	$\overline{G}$	$-3.30E - 02$	3.69E-03	4.00E-19	$-9.00E - 04$	3.38E-03	7.90E-01	$-8.00E - 03$	3.35E-03	1.70E-02
rs11220462	A	G	5.90E-02	5.94E-03	3.00E-23	$-1.60E-02$	5.56E-03	4.00E-03	1.90E-02	5.37E-03	4.00E-04
rs11246602	T	$\overline{c}$	$-1.90E-03$	5.51E-03	7.30E-01	$-3.40E - 02$	5.20E-03	6.00E-11	9.10E-03	5.18E-03	7.90E-02
rs11563251	T	$\overline{C}$	3.50E-02	6.24E-03	2.00E-08	5.80E-03	5.71E-03	3.10E-01	8.30E-03	5.62E-03	1.40E-01
rs11660468	T.	$\overline{c}$	1.10E-02	3.68E-03	2.80E-03	3.90E-02	3.44E-03	9.00E-30	$-8.00E - 04$	3.52E-03	8.20E-01
rs12133576	A	G	1.00E-02	3.64E-03	6.00E-03	2.40E-02	3.49E-03	6.00E-12	$-9.00E - 03$	3.44E-03	8.80E-03
rs12145743	$\overline{\mathsf{T}}$	G	4.20E-03	3.97E-03	2.90E-01	$-2.00E-02$	3.56E-03	2.00E-08	1.20E-02	3.50E-03	6.00E-04
rs12226802	A	G	$-2.00E - 04$	5.32E-03	9.70E-01	$-3.30E - 02$	4.92E-03	2.00E-11	6.70E-03	5.00E-03	1.80E-01
rs1250229	$\overline{\mathsf{T}}$	$\overline{c}$	$-2.40E - 02$	4.16E-03	8.00E-09	3.40E-03	3.96E-03	3.90E-01	$-8.90E-03$	3.79E-03	1.90E-02
rs12525163	$\overline{\mathsf{T}}$	$\mathsf{C}$	4.30E-03	4.06E-03	2.90E-01	$-2.20E-02$	3.83E-03	9.00E-09	8.60E-03	3.64E-03	1.80E-02
rs1260326	T	$\overline{C}$	2.10E-02	3.79E-03	3.00E-08	$-1.10E-02$	3.34E-03	1.00E-03	1.10E-01	3.23E-03	2.00E-254
rs12670798	$\overline{\mathsf{T}}$	C	$-3.40E-02$	4.21E-03	7.00E-16	1.40E-03	4.06E-03	7.30E-01	$-1.00E-02$	3.82E-03	8.90E-03

**Supplementary Table 1. Genetic variants used to construct genetic risk scores for lipid traits used in multivariable MR analyses**













Beta, regression coefficient; se, standard error; pval, p-value; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; TG, triglycerides; SNP, single nucleotide polymorphism.



**Supplementary Table 2. Genetic variants used to construct genetic risk scores for IGF-I and IGFBP-3**

Beta, regression coefficient; se, standard error; pval, p-value; HDL-C, IGF-I, insulin-like growth factor 1; IGFBP-3, insulin-like growth factor binding protein 3; eaf, effect allele frequency; SNP, single nucleotide polymorphism





BMI, body mass index; yrs, years; kg, kg/m2, kilograms per metres squared; LDL-C, low density lipoprotein; HDL-C, high density lipoprotein; TG, triglycerides;SD, standard deviation



**Supplementary Table 4.** Association between IGF or lipid traits and potential confounders measured in the SHIP study.

Beta, effect size; CI, confidence interval; p, p-value; IGF-I, insulin-like growth factor 1; IGFBP-3, insulin-like growth factor binding protein 3; LDL-C, lowdensity lipoprotein; HDL-C, high-density lipoprotein; BMI, body mass index.

<sup>a</sup>Beta refers to unit change in IGF-I (ng/ml), IGFBP-3 (ng/ml), LDL-C (mmol/l), HDL-C (mmol/l) and triglycerides (mmol/l) levels per unit increase in confounders. Calculated using unadjusted linear regression.