Supplementary Online Contents

Supplementary Methods

Study populations

The Northern Finland Birth Cohorts (NFBC1986 and NFBC1966)

The Northern Finland Birth Cohorts 1986 and 1966 were initiated to study factors affecting preterm birth and subsequent morbidity in the two northernmost provinces in Finland. For NFBC1986, the number of deliveries in the birth cohort was 9362, which was 99% of all the deliveries taking place in the target period (July 1985-June 1986). Data collection in 2001-2002 included clinical examination and serum sampling at the age of 15-16 years for 6621 adolescent boys and girls; data from this time point were used for the secondary analyses of the present study.[1,2] Attendees in the 16-year field study (71%) were representative of the original cohort.[2] Individuals using oral contraceptives were excluded. In total, 5099 adolescents had the systemic NMR-based metabolite profile measured, of which 95% of the serum samples were drawn after an overnight fast.

The NFBC1966 included 12 058 children, comprising 96% of all births during 1966 in the region.[3,4] The collection in 1997 included clinical examination and serum sampling at the age of 31 years for 6007 individuals. Data from this time point were analysed in the present study. Attendees in the 31-year field study (52% of target population) were representative of the original cohort.[3] Pregnant women and those using oral contraceptives were excluded. In total, 4841 persons had their systemic NMR-based metabolite profile measured, of which 96% of the serum samples were drawn after an overnight fast.[5]

C-reactive protein and plasma insulin were measured by standard clinical assays. SHBG was analysed by time-resolved fluoroimmunoassay (AutoDelfia, PerkinElmer, Turku, Finland). The intra- and inter-assay coefficients of variation for the assay were 3.9 and 4.6 % at 101 nmol/l, respectively. Serum testosterone was measured by an automated chemiluminescence system (ACS-180, Ciba-Corning, Inc., Medfield, MA, USA). Informed written consent was obtained from all participants. The research protocols were approved by the Ethics Committee of Northern Ostrobotnia Hospital District, Finland.

The Cardiovascular Risk in Young Finns Study (YFS)

The Cardiovascular Risk in Young Finns Study was designed to study associations of childhood risk factors to cardiovascular disease in adulthood.[6] The baseline study conducted in 1980 included 3596 children and adolescents at the ages of 3–18 years. In this study we used data from the 2001 survey that included 2247 individuals with an overnight fasting NMR-based systemic metabolite profile (response rate 63%). These individuals were representative of the baseline cohort.[6] Pregnant women and those using oral contraceptives were excluded. In total, data for 1870 individuals who had the metabolic profiles measured were used in the study; of those 1377 individuals also had information at 6-year follow-up.

C-reactive protein and plasma insulin were measured by standard clinical assays. SHBG was measured by a Spectria SHBG IRMA kit (Orion Diagnostica, Espoo, Finland) and total testosterone

by a Spectria Testosterone kit (Orion Diagnostica, Espoo, Finland). All participants gave written informed consent, and the local ethics committees of the study sites approved the study.

The National Finnish FINRISK study

The FINRISK 1997 study was conducted to monitor the health of the Finnish population among persons at the ages of 25–74 years at the recruitment.[7] In total, 8444 individuals were recruited to represent the general population of the study areas. Pregnant women and those using oral contraceptives were excluded from the analyses of the present study. The NMR-based systemic metabolite profiling from serum samples were available for 7200 individuals. The median fasting time was 5h (interquartile range 4–6h). C-reactive protein and plasma insulin were assayed by standard clinical assays. Participants gave written informed consent and the FINRISK study was approved by the Coordinating Ethical Committee of the Helsinki University Hospital District, Helsinki, Finland.

Statistical analyses, multiple testing correction

The rationale for the modified Bonferroni correction is as follows: 1) Bonferroni adjustment is designed for independent tests, i.e., the outcome of one test is not correlated with the others. 2) By breaking a dataset into principal components (PCs), we effectively create new variables that, by definition, are not correlated.[8] The assumption here is that the independence of the PC scores is equivalent to the degree of freedom of the original dataset — with >95% variance explained, this can only cause a small inaccuracy. This logic also works for the extreme cases: for perfectly correlated variables only one needs to be tested, hence the Bonferroni correction is for one independent test. If all the variables are orthogonal, then PCA will need a large number of components to explain most of the variance, and the Bonferroni correction approaches the full number of dimensions.

Mendelian randomization analysis

Mendelian randomisation analysis using ratio estimator

The causal effect estimates for the metabolic measures were calculated via the ratio estimator, and the corresponding standard errors via the delta method using the following formulas: [9,10]

$$\beta_{IV} = \frac{\beta_{GS-metabolite}}{\beta_{Gs-SHBG}}$$

$$se_{IV} = abs(\beta_{IV}) \sqrt{(\frac{se_{GS-SHBG}}{\beta_{GS-SHBG}})^2 + (\frac{se_{GS-metabolite}}{\beta_{GS-metabolite}})^2}$$

where *abs* refers to an absolute value, $\beta_{GS\text{-}metabolite}$ is the effect size and $se_{GS\text{-}metabolite}$ is the standard

error of the association between the gene score and the metabolic measure meta-analysed using a fixed-effect model across NFBC1966, YFS and FINRISK (n=10 895). $\beta_{GS\text{-}SHBG}$ denotes the effect size and $se_{GS\text{-}SHBG}$ the corresponding standard error of the association between the gene score and the circulating SHBG level meta-analysed using a fixed-effect model across NFBC1966 and YFS (n=5860).

Mendelian randomisation analysis using summary statistics

The causal effect estimates and corresponding standard error for type 2 diabetes were calculated via the method proposed by Burgess et al.,[11] using the following formulas:

$$\beta_{IV} = \frac{\sum_k X_k Y_k \sigma_{Yk}^{-2}}{\sum_k X_k^2 \, \sigma_{Yk}^{-2}}$$

$$se_{IV} = \sqrt{\frac{1}{\sum_k X_k^2 \sigma_{Yk}^{-2}}}$$

where X_k is the effect size of the association between the genetic variant k and SHBG obtained from the SHBG-GWAS study;[12] Y_k is the log-odds ratio and σ_{Yk} the standard error between the genetic variant k and type 2 diabetes for 110 452 individuals (cases/controls=26 488/83 964) reported by DIAGRAM consortium.[13]

The causal estimates based on the two multi-SNP gene-scores in the DIAGRAM consortium were further meta-analysed with effect estimates previously published by Ding *et al*[14] (based on rs6257 and rs6259) and by Perry *et al*[15] based on rs1799941 near the *SHBG* gene. Only studies not part of the DIAGRAM consortium data were part included in the meta-analysis of the causal effects. Details and clinical characteristic of the additional cohorts included in the meta-analysis have been described in the original publications.[14,15]

References

1. Kantomaa MT, Stamatakis E, Kankaanpää A, Kaakinen M, Rodriguez A, Taanila A, et al. Physical activity and obesity mediate the association between childhood motor function and adolescents' academic achievement. Proc Natl Acad Sci USA 2013;110:1917–22.

2. Veltsista A, Laitinen J, Sovio U, Roma E, Järvelin M-R, Bakoula C. Relationship between eating behavior, breakfast consumption, and obesity among Finnish and Greek adolescents. J Nutr Educ Behav 2010;42:417–21.

3. Sabatti C, Service SK, Hartikainen A-L, Pouta A, Ripatti S, Brodsky J, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet 2009;41:35–46.

4. Järvelin M-R, Sovio U, King V, Lauren L, Xu B, McCarthy MI, et al. Early life factors and

blood pressure at age 31 years in the 1966 northern Finland birth cohort. Hypertension 2004;44:838–46.

5. Würtz P, Mäkinen V-P, Soininen P, Kangas AJ, Tukiainen T, Kettunen J, et al. Metabolic signatures of insulin resistance in 7,098 young adults. Diabetes 2012;61:1372–80.

6. Raitakari OT, Juonala M, Rönnemaa T, Keltikangas-Järvinen L, Räsänen L, Pietikäinen M, et al. Cohort profile: the cardiovascular risk in Young Finns Study. Int J Epidemiol 2008;37:1220–6.

7. Borodulin K, Vartiainen E, Peltonen M, Jousilahti P, Juolevi A, Laatikainen T, et al. Forty-year trends in cardiovascular risk factors in Finland. Eur J Public Health 2014, in press.

8. Kujala UM, Mäkinen V-P, Heinonen I, Soininen P, Kangas AJ, Leskinen TH, et al. Long-term leisure-time physical activity and serum metabolome. Circulation 2013;127:340–8.

9. Thomas DC, Lawlor DA, Thompson JR. Re: Estimation of bias in nongenetic observational studies using "Mendelian triangulation" by Bautista et al. Ann Epidemiol 2007;17:511–3.

10. Fall T, Hägg S, Mägi R, Ploner A, Fischer K, Horikoshi M, et al. The role of adiposity in cardiometabolic traits: a mendelian randomization analysis. PLoS Med 2013;10:e1001474.

11. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37:658–65.

12. Coviello AD, Haring R, Wellons M, Vaidya D, Lehtimäki T, Keildson S, et al. A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. PLoS Genet 2012;8:e1002805.

13. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Asian Genetic Epidemiology Network–Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D) Consortium, Mexican American Type 2 Diabetes (MAT2D) Consortium, Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium, Mahajan A, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 2014;46:234–44.

14. Ding EL, Song Y, Manson JE, Hunter DJ, Lee CC, Rifai N, et al. Sex hormone-binding globulin and risk of type 2 diabetes in women and men. N Engl J Med 2009;361:1152–63.

15. Perry JRB, Weedon MN, Langenberg C, Jackson AU, Lyssenko V, Sparsø T, et al. Genetic evidence that raised sex hormone binding globulin (SHBG) levels reduce the risk of type 2 diabetes. Hum Mol Genet 2010;**19:**535–44.

16. Chen BH, Brennan K, Goto A, Song Y, Aziz N, You N-CY, et al. Sex hormone-binding globulin and risk of clinical diabetes in American black, Hispanic, and Asian/Pacific Islander postmenopausal women. Clin Chem 2012;58:1457–66.

17. Lakshman KM, Bhasin S, Araujo AB. Sex hormone-binding globulin as an independent predictor of incident type 2 diabetes mellitus in men. J Gerontol A Biol Sci Med Sci 2010;65:503–9.

18. Vikan T, Schirmer H, Njølstad I, Svartberg J. Low testosterone and sex hormone-binding globulin levels and high estradiol levels are independent predictors of type 2 diabetes in men. Eur J Endocrinol 2010;162:747–54.

Table S1. Genetic variants for the SHBG gene score and their associations with type 2 diabetes.

	SNP	Gene	Chromosome	Effect[a] allele	Weight[b]	Association with type 2 diabetes[c]		
						OR (95%CI)	*P*	*N*
1	rs780093	*GCKR*	2	C	0.032	1.05 (1.02-1.08)	0.00014	104,493
2	rs17496332	*PRMT6*	1	G	0.028	0.98 (0.96-1.01)	0.23	104,087
3	rs440837	*ZBTB10*	8	G	0.028	0.99 (0.96-1.02)	0.45	104,505
4	rs7910927	*JMJD1C*	10	G	0.048	0.99 (0.97-1.02)	0.65	104,575
5	rs4149056	*SLCO1B1*	12	T	0.029	1.00 (0.96-1.04)	1.00	104,229
6	rs8023580	*NR2F2*	15	C	0.030	1.03 (1.00-1.06)	0.029	99,991
7	rs2411984	*ZNF652*	17	A	0.033	0.98 (0.95-1.01)	0.28	94,519
8	rs1573036	*TDGF3*	23 (X)	T	0.028			
9	rs12150660	*SHBG*	17	T	0.082	0.97 (0.93-1.01)	0.16	72,377
10	rs6258	*SHBG*	17	C	0.261	0.90 (0.78-1.04)	0.14	63,390
11	rs1641537	*SHBG*	17	C	0.081	0.99 (0.95-1.02)	0.48	96,111
12	rs1625895	*SHBG*	17	C	0.052	0.99 (0.94-1.03)	0.55	95,557

a: The effect allele is the SHBG-raising allele.

b: Weights of the individual genetic variants are based on a recent genome-wide analyses study.[12]

c: Associations between the individual genetic variants and the risk of type 2 diabetes are reported by the DIAGRAM consortium; results for the variant rs1573036 in the X chromosome is not available.[13]

Table S2. Prospective association magnitudes of SHBG for incident type 2 diabetes in four studies.

Publication[a]	Study design	Age at baseline[b]	Follow-up (y)	Baseline number	Incident diabetes number	Log[c]	HR (95%CI)[d]
Women							
Ding et al, 2009[14]	Nested case-control	60.3 (6.1)	10	718	359	Yes	0.40 (0.31-0.51)
Chen et al, 2012[16]	Nested case-control	50-79	5.9	1,928	642	Yes	0.48 (0.43-0.54)
					Overall (Women)		**0.45 (0.38-0.54)**
Men							
Ding et al, 2009[14]	Nested case-control	63.7 (7.6)	8	340	170	Yes	0.43 (0.31-0.59)
Lakshman et al, 2010[17]	Cohort	40-70	13	1,128	90	No	0.50 (0.35-0.70)
Vikan et al, 2010[18]	Cohort	59.4 (10.3)	9.1	1,454	76	No	0.55 (0.39-0.79)
					Overall (Men)		**0.49 (0.40-0.59)**
					Overall		**0.47 (0.41-0.53)**

a: Publications were searched from PubMed with key words "SHBG" and "type 2 diabetes" with a search criterion the publication year 2000 or later. The prospective studies were included if the results were reported as risk of type 2 diabetes per 1-SD increment SHBG. All studies listed here are based on independent cohorts.

b: Age at baseline is given as the range of age or mean (SD).

c: Log indicates whether the SHBG concentration was natural log-transformed or not.

d: When there are multiple adjusted models, the hazard ratio is from the model in which least covariates were used.

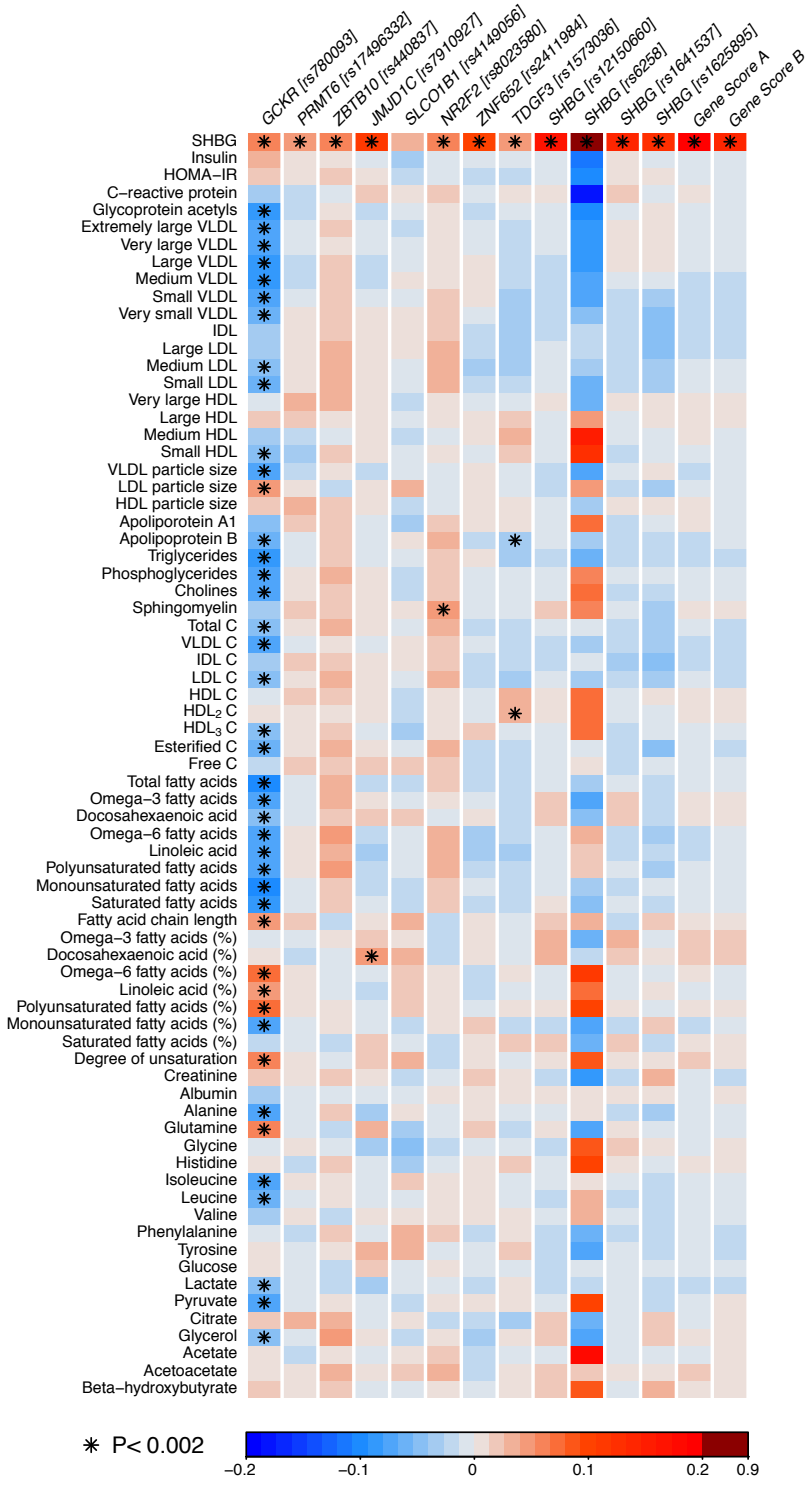

Figure S1. Pleiotropy assessment for the individual SNPs and gene scores associated with circulating SHBG. The associations between individual SNPs and the 72 metabolic measures are meta-analyses for 10 895 individuals from the NFBC1966, YFS and FINRISK. The associations were adjusted for sex, age and BMI, and calculated as SD-units per each SHBG-raising allele. Associations with P<0.002 are marked with an asterisks. The associations between the gene scores and the metabolic measures are also provided as a reference (SD-units per 1-SD gene score). Gene Score A is composed of 11 SHBG-correlated SNPs after excluding the pleiotropic *GCKR* variant. Gene Score B is composed of 4 SNPs within the *SHBG* gene, including rs12150660, rs6258, rs1641537 and rs1625895.

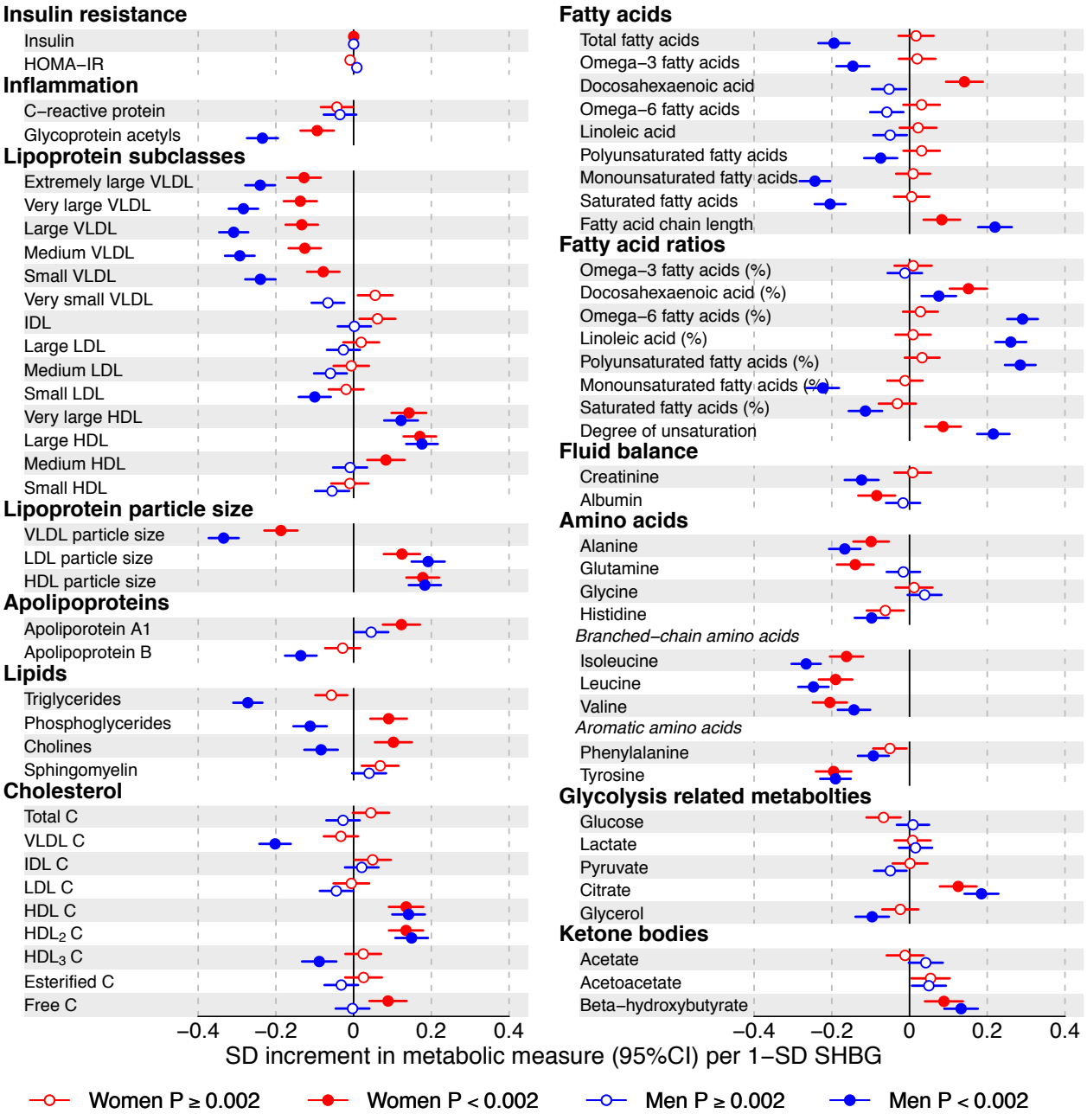

Figure S2. Cross-sectional associations of SHBG with systemic lipid and metabolite measures for men and women; additionally adjusted for insulin and testosterone. The associations were meta-analysed for 6475 individuals from the NFBC1966 and YFS. The associations were adjusted for insulin and testosterone, in addition to age and body mass index.

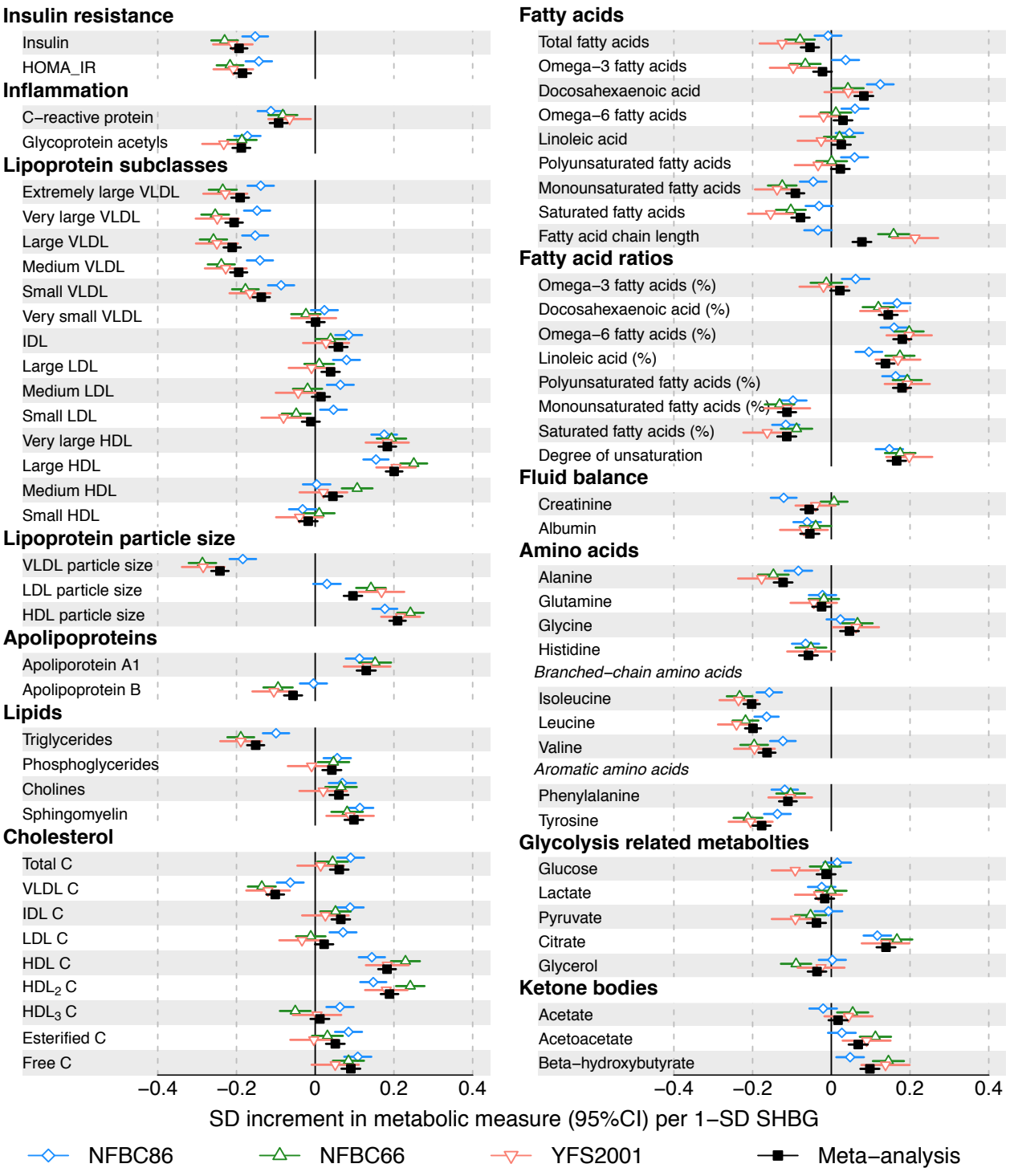

Figure S3. Cross-sectional associations of SHBG with systemic lipid and metabolite measures in individual cohorts. The associations between SHBG and the metabolic measures were further replicated in 4834 adolescents from the NFBC1986. Women and men were combined to obtain a clear visualisation. The associations were adjusted for sex, age and body mass index.

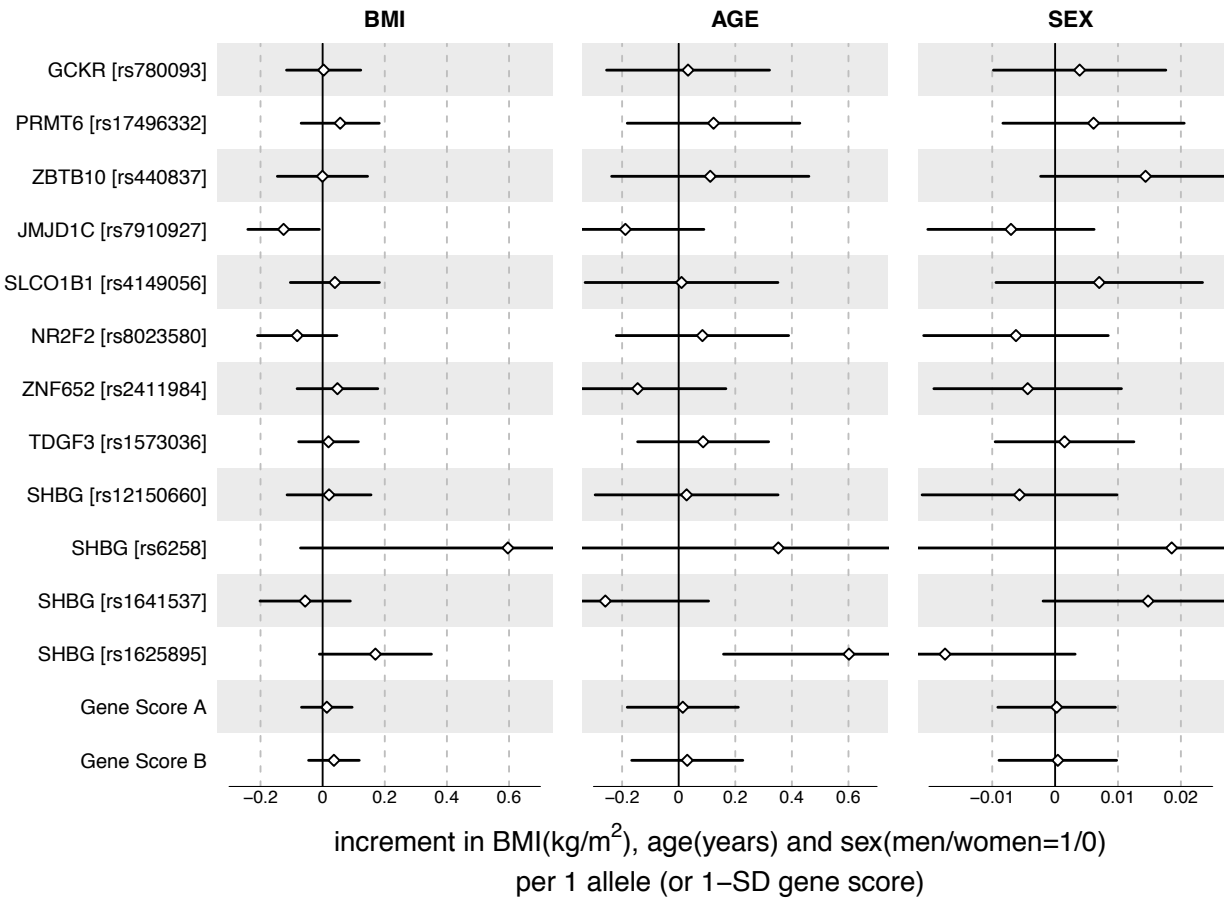

Figure S4. The associations of individual genetic variants and the SHBG gene scores with potential confounders. The associations were assessed for 10 895 individuals across three cohorts. P≥0.002 for all the associations. The large confidence interval estimated for rs6258 is due to its low minor allele frequency (<1%). The genetic variants included in Gene score A and B are as described in the legend of Figure S1.

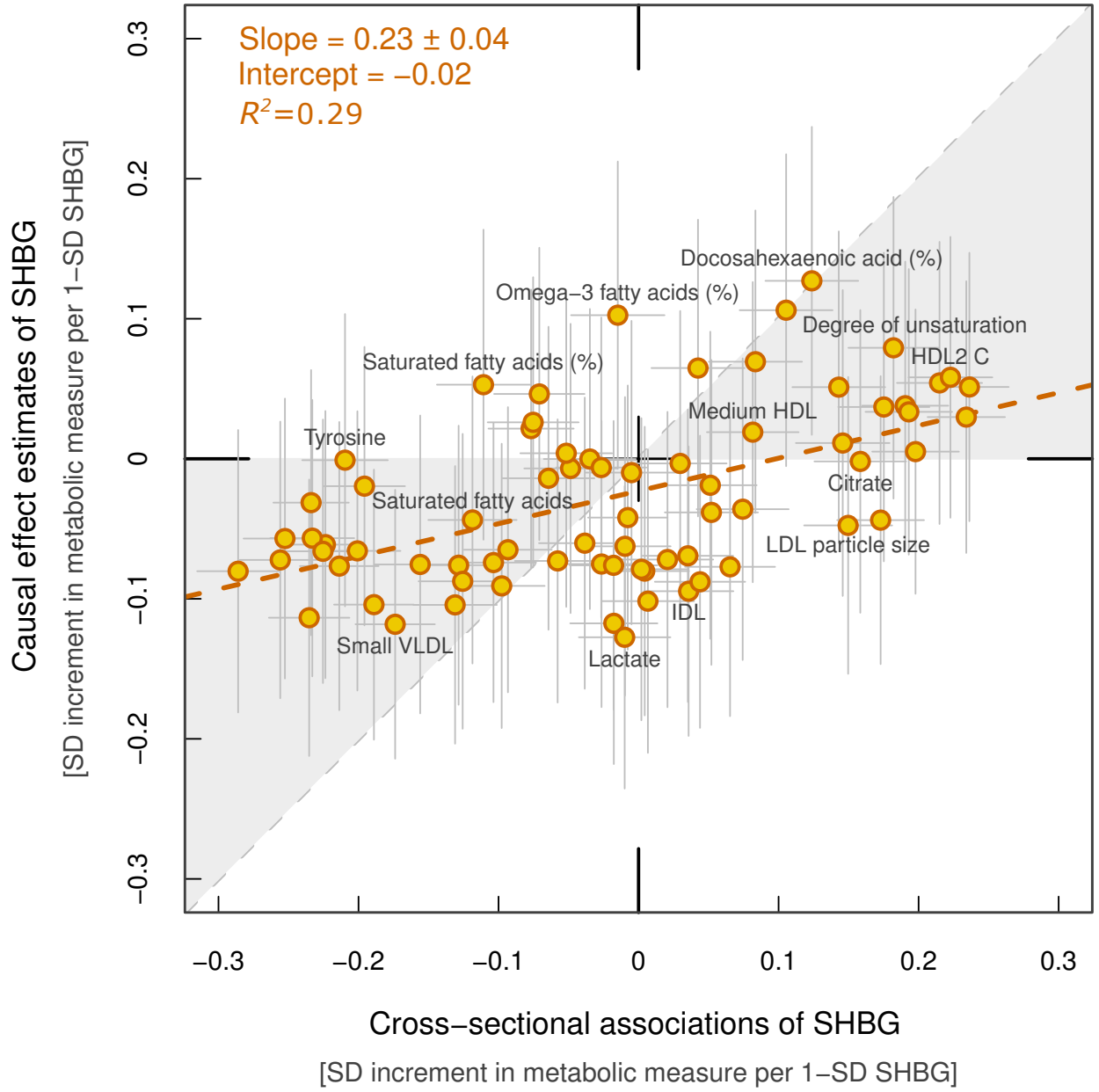

Figure S6. The correspondence between cross-sectional associations and causal effect estimates based on Gene Score A. The causal estimates were calculated via Mendelian randomization analyses using the gene score as the instrument. Gene Score A is composed of 11 established genetic variants for elevating SHBG (the *GCKR*-variant excluded due to its wide pleiotropic effects). Each point represents a metabolic measure. The horizontal grey lines behind each point denote the confidence intervals for the cross-sectional associations and the vertical grey lines indicate the confidence intervals for the causal effect estimates. A linear fit of the overall correspondence was made to summarize the match between observational (cross-sectional) associations and causal estimates, and R^2 denotes the goodness of the fit. Since the observational and genetic associations were scaled to the same units (SD increment of metabolite concentration per 1-SD SHBG), a slope of ~1 would be expected for a match between observational and causal effect estimates.

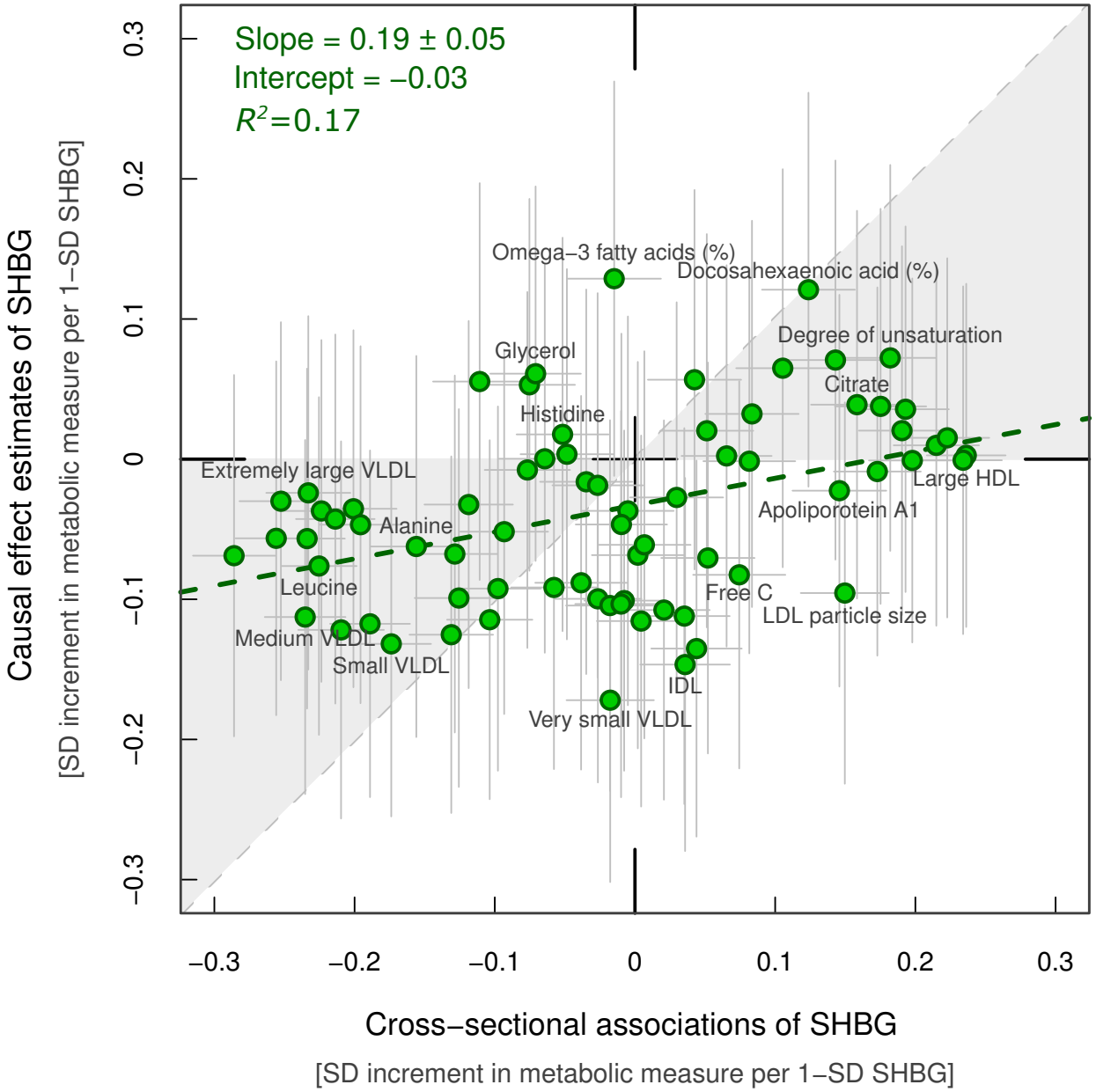

Figure S7. The correspondence between cross-sectional associations and causal effect estimates based on Gene Score B. The causal estimates were calculated via Mendelian randomization analyses using the gene score as the instrument. Gene Score B is composed of the 4 genetic variants at the *SHBG*-encoding gene (rs12150660, rs6258, rs1641537 and rs1625895). Each point represents a metabolic measure. The horizontal grey lines behind each point denote the confidence intervals for the cross-sectional associations and the vertical grey lines indicate the confidence intervals for the causal effect estimates. A linear fit of the overall correspondence was made to summarize the match between observational (cross-sectional) associations and causal estimates, and R^2 denotes the goodness of the fit. Since the observational and genetic associations were scaled to the same units (SD increment of metabolite concentration per 1-SD SHBG), a slope of ~1 would be expected for match between observational and causal effect estimates.