The American Journal of Human Genetics

Supplemental Data

Adjusting for Heritable Covariates Can Bias Effect Estimates in Genome-Wide Association Studies

Hugues Aschard, Bjarni J. Vilhjálmsson, Amit D. Joshi, Alkes L. Price, Peter Kraft

Figure S1. Illustrative examples of direct acyclic graphs.

Y is the outcome, *C* is a potential covariate, *g* is a predictor of interest, and *U* is an unmeasured variable that can results or depends on *Y*, *C* and *g*.

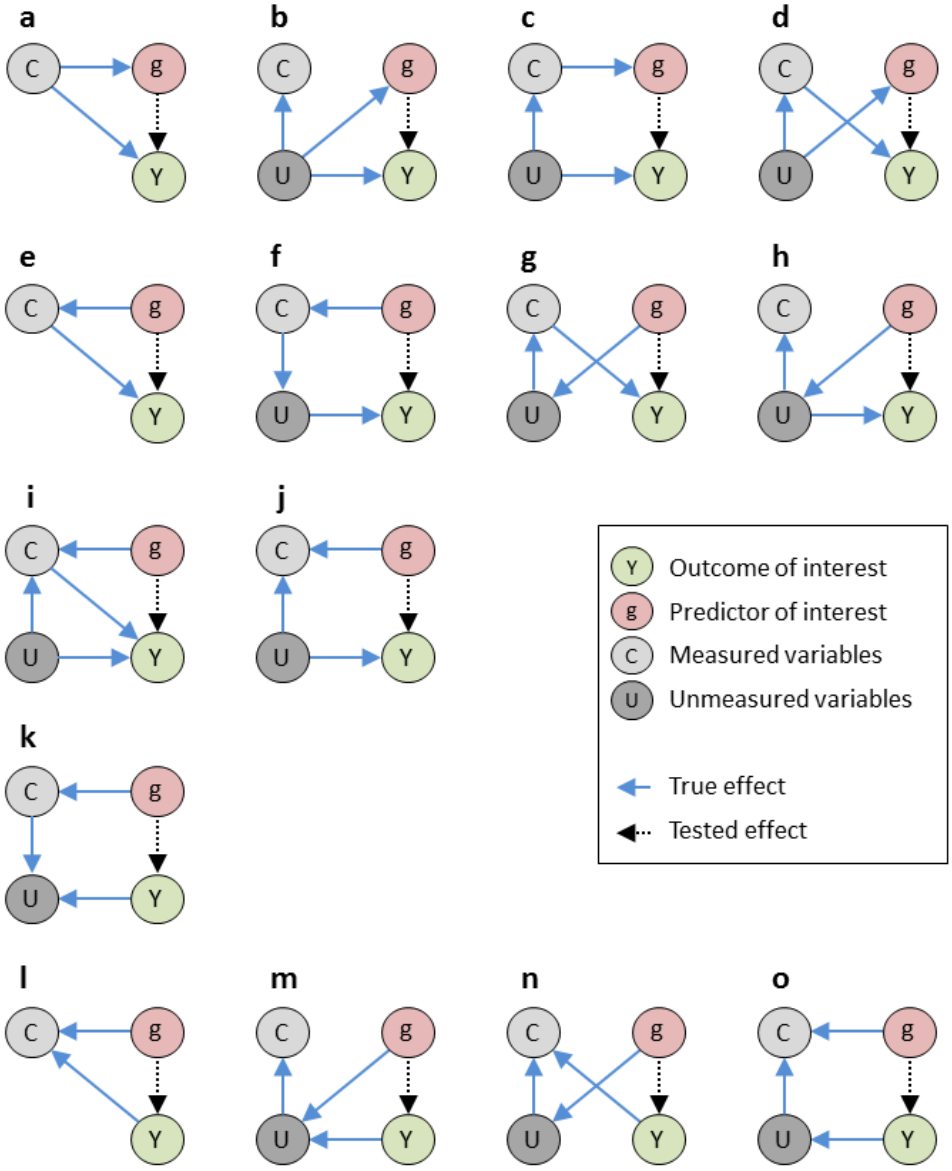

Figure S2. Statistical power in adjusted analysis

Statistical power when the genetic variant is differentially associated with two correlated phenotypes. Two traits Y and C are simulated with correlation ρ_{CY} varying from 0 to 0.9. A genetic variant g has a positive effect β_Y on Y and explains 0.015% (left panel) or 0.04% (right panel) of its variance; we vary the effect of g on C between $-\beta_Y$ and β_Y. a) We compared the power at $5x10^{-8}$ significance level to detect the g/Y association between the marginal model (unadjusted for C, red line) to the adjusted model (light grey to black plain line). b) We plotted the power to detect the marginal g/C association (light blue to dark blue plain line) while varying α, the significance level.

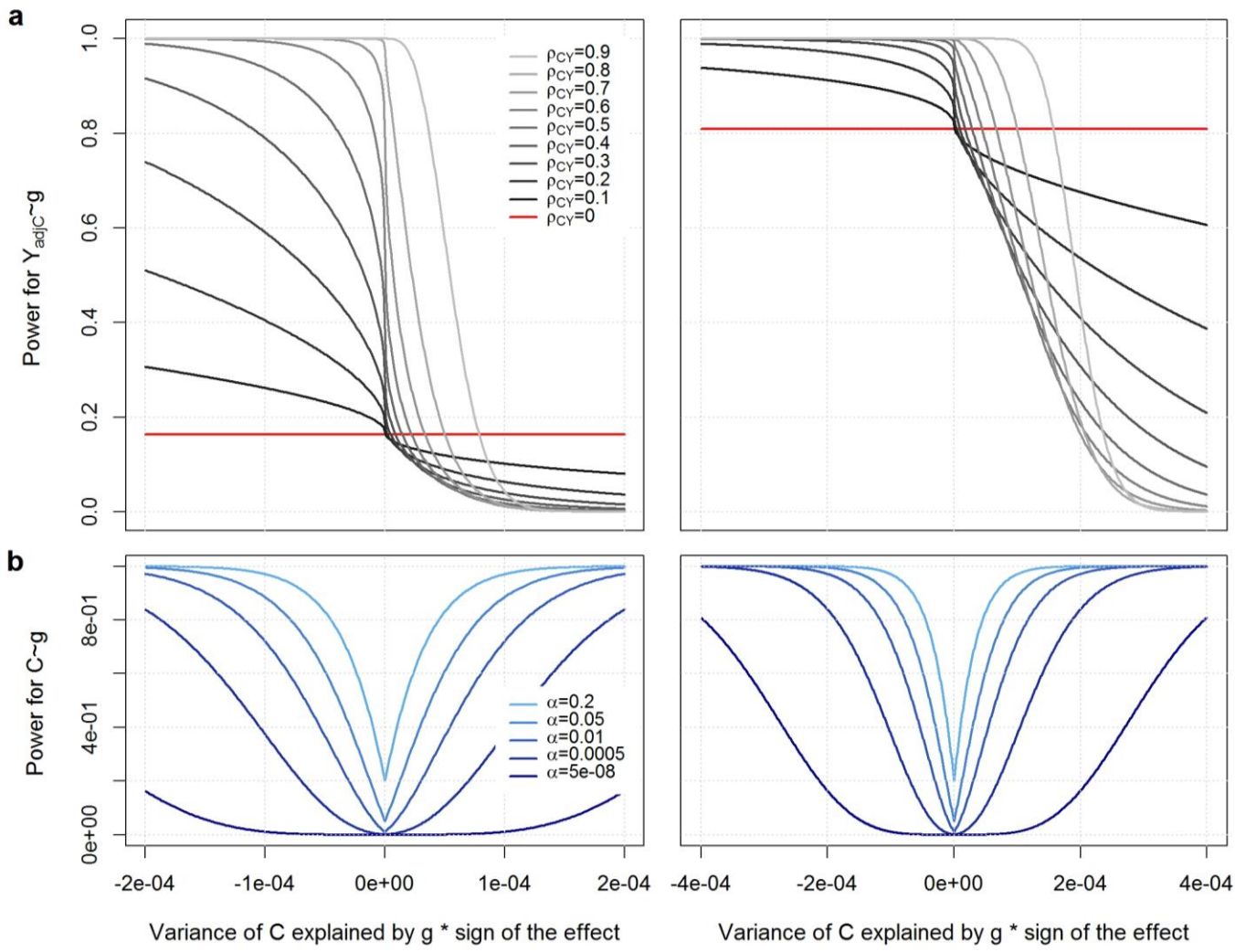

Figure S3. Effect of the adjustment on estimate

Upper panel presents $\hat{\beta}_Y$ the estimated effect of G on Y in the unadjusted model as a function of $\hat{\beta}_C$ the estimated effect of G on C across 1,000 replicates of 100,000 subjects (grey dots) and the regression line (black) when the correlation between C and Y is 0.5, $\beta_C = 0$ and $\beta_Y = 0$ (a) or $\beta_Y = 0.3$ (b) . Upper plots also display $\Delta_{\hat{\beta}_Y} = \hat{\beta}_Y - \hat{\beta}_{Yadj}$the change in estimates before and after the adjustment for C (blue dots). This change corresponds on average to the deviation of $\hat{\beta}_Y$ from β_Y. Bottom panel shows the correlation between $\hat{\beta}_C$ and the p-values of $\hat{\beta}_{Yadj}$, the estimate of G on Y after adjustment for C for the same replicates.

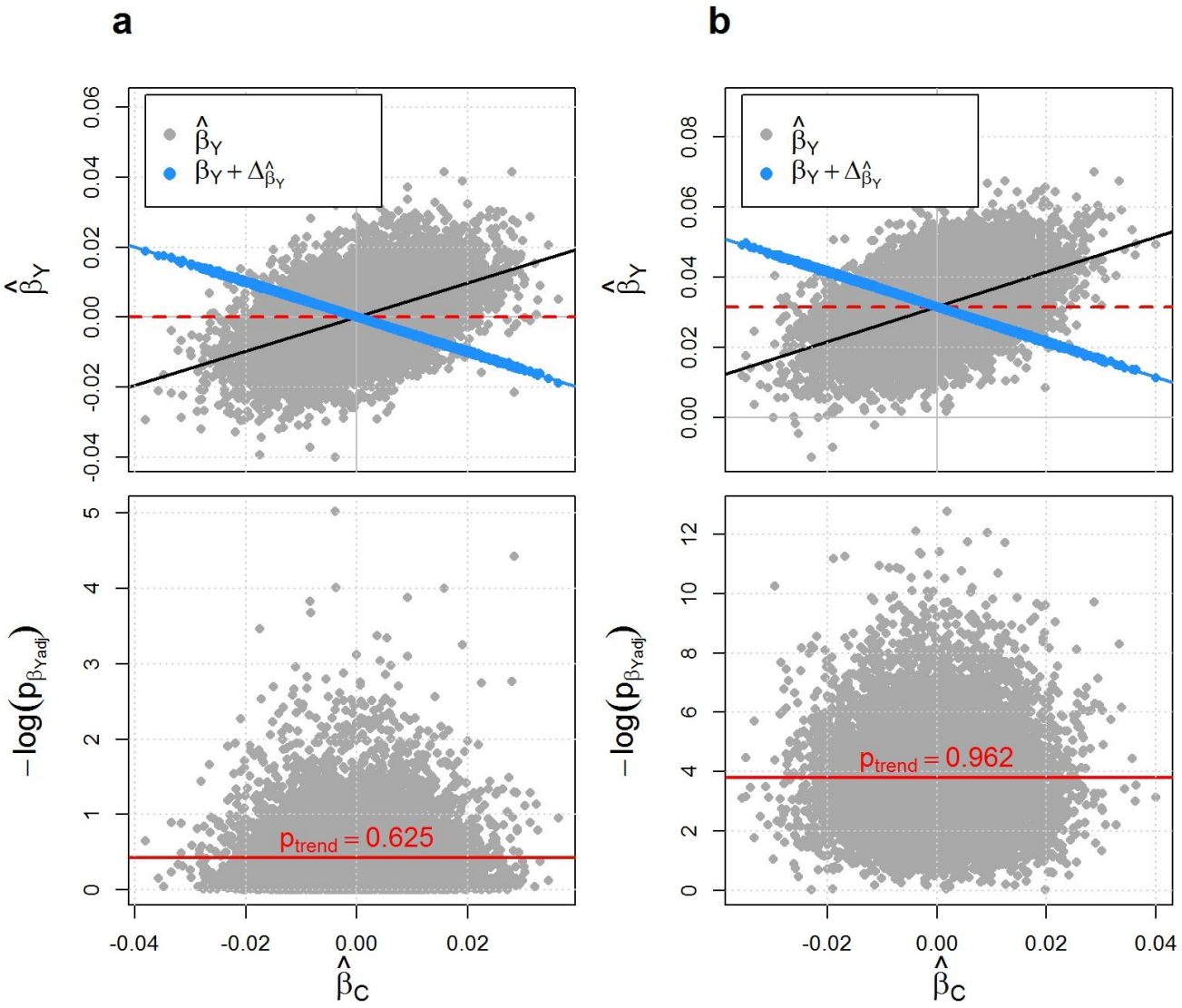

Figure S4. Opposite genetic effects between BMI and WHR/WC

Log_{10} p-value of the observed proportion of SNP with effects in opposite direction when assuming a binomial distribution with probability 0.5, when filtering out SNPs based on their *p*-value for association with BMI.

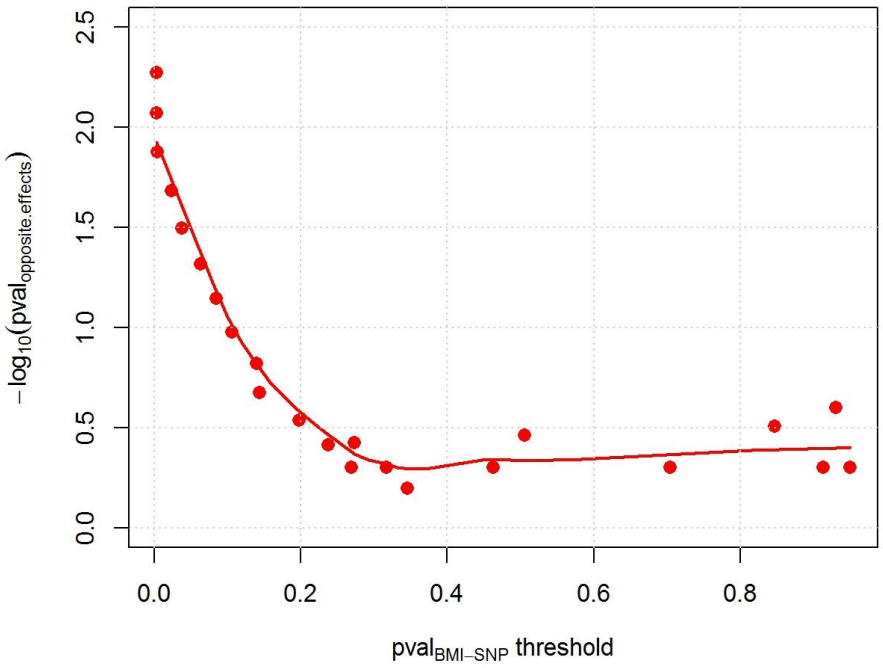

Figure S5. QQplot under the null for three corrections.

We generated 10,000 replicates of 1,000 subjects. For each replicates we simulated normally distributed variables *Y*, *C* and *g,* so that *C* and *Y* display a correlation of 0.5 and *g* explained 1% of the variance of *C* but none of *Y*. We derived the inflation factor λ and plotted the observed distribution of *p*-value for tests of association between *Y* and *g* against its expected distribution. We considered a marginal model in the whole sample (black), a model adjusted for the raw values of *C* (grey) or after removing either the true effect of *g* on *C* (blue) or its estimated effect (green), a *C*-specific strata of the marginal model ($C \geq 0$, dark red dots, and $C < 0$, light red), and a two-step procedure proposed by Vansteelandt et al[1] (orange).

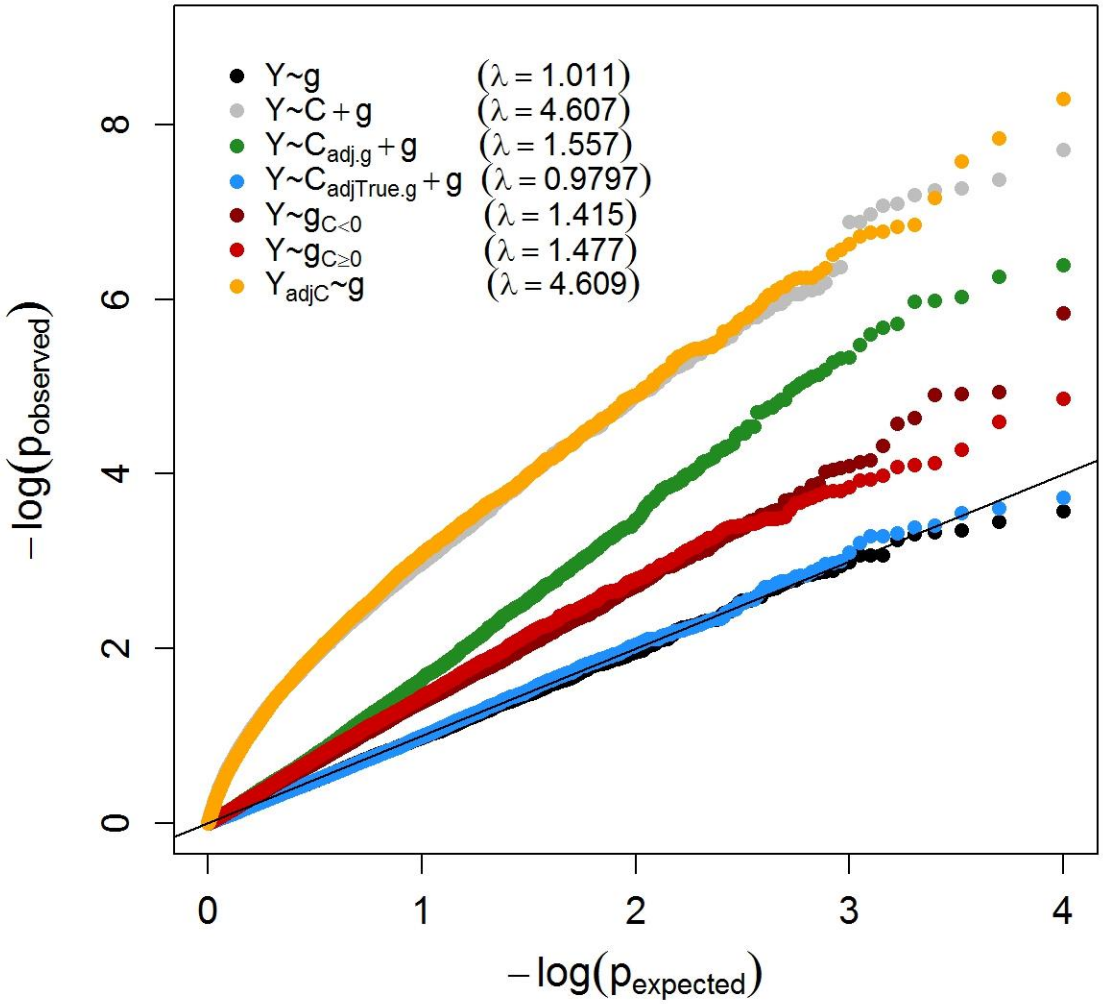

Figure S6. Bias in covariate stratified analysis.

Outcome values (*Y*) as a function of a continuous predictor (*G*) in the total sample and in strata defined by a covariates (*C*). For simplicity all variables, *G, C* and *Y* are normally distributed and generated under model from Figure 1B. The plain lines are the regression lines for the association between *G* and *Y*. While there is no correlation between *Y* and *G* in the whole sample (a), *Y* and *G* display a negative correlation in each *C* strata (b).

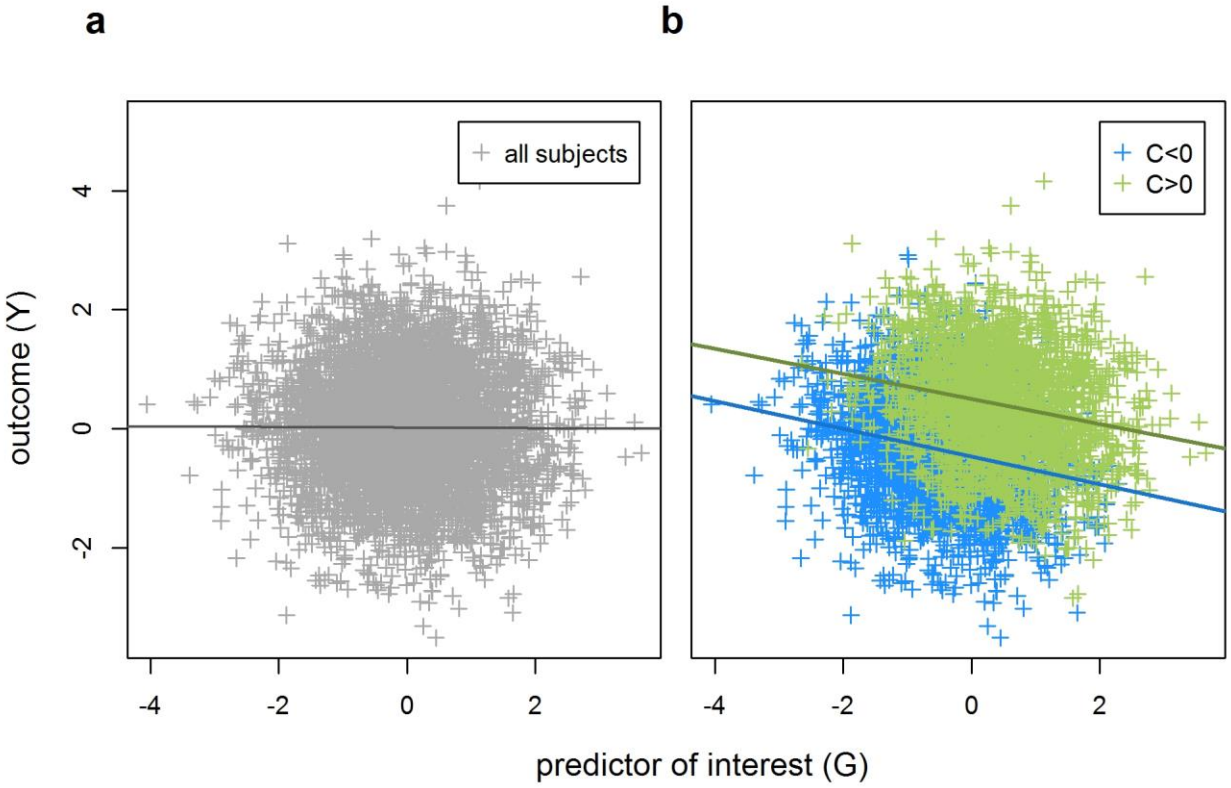

Table S1. Genetic effect estimates and p-value for association with T2D before and after adjustment for BMI

SNP[a]	Risk Allele	BMI *(pval)*	T2D *(pval)*	$T2D_{adjBMI}$ *(pval)*
rs2568958	A	3.77 *($1.2x10^{-11}$)*	0.0253 *(0.038)*	0.0170 *(0.15)*
rs10913469	**C**	**3.36** ***($6.2x10^{-8}$)***	**-0.0044** ***(0.87)***	**-0.0177** ***(0.27)***
rs7561317	G	6.12 *($4.2x10^{-17}$)*	0.0212 *(0.22)*	0.0000 *(0.9)*
rs7647305	C	4.42 *($7.2x10^{-11}$)*	0.0334 *(0.036)*	0.0128 *(0.35)*
rs2844479	**T**	**1.38** ***(0.003)***	**-0.0223** ***(0.11)***	**-0.0315** ***(0.028)***
rs6265	**G**	**4.58** ***($5.1x10^{-10}$)***	**-0.0088** ***(0.64)***	**-0.0269** ***(0.12)***
rs925946	T	3.85 *($8.5x10^{-10}$)*	0.0170 *(0.15)*	0.0043 *(0.73)*
rs7138803	A	3.28 *($1.2x10^{-7}$)*	0.0453 *(0.00023)*	0.0334 *(0.009)*
rs7498665	G	3.63 *($3.2x10^{-10}$)*	0.0212 *(0.069)*	0.0128 *(0.25)*
rs6499640	A	5.26 *($4.0x10^{-13}$)*	0.0294 *(0.027)*	0.0086 *(0.48)*
rs8050136	A	8.04 *($1.1x10^{-47}$)*	0.0531 *($1.6x10^{-5}$)*	0.0212 *(0.086)*
rs12970134	A	4.38 *($1.2x10^{-12}$)*	0.0170 *(0.18)*	0.0000 *(0.91)*
rs29941	C	4.18 *($7.3x10^{-12}$)*	0.0043 *(0.7)*	0.0000 *(0.96)*

[a]*We extracted SNPs found associated with BMI and/or weight and their estimated effect on BMI and T2D before and after adjustment for BMI from Thorleifsson et al[2].*

SNPs with decreased p-value for association with T2D after the adjustment for BMI are indicated in bold.

Table S2. Additional examples of genetically-associated covariate GWAS

		Effect on the outcome[a]			Effect on covariate[b]		
Outcome	**SNP**	**trait-increasing allele /alt allele**	**Unadjusted Pvalue[c]**	**Adjusted Pvalue**	**trait-increasing allele /alt allele**	**pValue**	**opposite effect**
Glycemic traits adjusted for BMI (positive correlation with the outcome)[3]							
	rs2943645	T / C	1.4×10^{-7}	2.3×10^{-19}	C / T	0.0069	x
	rs10195252	T / C	4.9×10^{-10}	1.3×10^{-16}	C / T	0.0093	x
	rs2126259	T / C	1.5×10^{-10}	3.3×10^{-13}	T / C	0.60	
	rs4865796	A / G	2.1×10^{-8}	2.2×10^{-12}	G / A	5.1×10^{-5}	x
	rs17036328	T / C	1.9×10^{-5}	3.6×10^{-12}	C / T	0.019	x
Fasting Insulin	rs731839	G / A	1.7×10^{-8}	5.1×10^{-12}	A / G	0.0052	x
	rs974801	G / A	2.1×10^{-10}	3.3×10^{-11}	G / A	0.89	
	rs459193	G / A	6.6×10^{-8}	1.2×10^{-10}	A / G	0.22	x
	rs6822892	A / G	3.2×10^{-5}	2.6×10^{-10}	G / A	0.37	x
	rs4846565	G / A	2.0×10^{-8}	1.8×10^{-9}	A / G	0.29	x
	rs3822072	A / G	1.2×10^{-4}	1.8×10^{-8}	G / A	0.12	x
	rs6912327	T / C	2.8×10^{-6}	2.3×10^{-8}	C / T	0.066	x
2hGlu	rs7651090	G / A	1.1 x 10+	4.5×10^{-8}	A / G	0.026	x
Fasting Glucose	rs17762454	T / C	1.9×10^{-7}	9.6×10^{-9}	T / C	0.46	
	rs7708285	G / A	4.9×10^{-6}	1.2×10^{-8}	A / G	0.034	x
	rs2657879	G / A	5.7×10^{-6}	3.9×10^{-8}	A / G	0.24	x
Body mass index adjusted for height (positive correlation with the outcome)[4]							
BMI	rs11676272	G/A	6.0×10^{-7}	6.0×10^{-9}	A/G	1.09×10^{-6}	x
Forced vital capacity (FVC) adjusted for height (positive correlation with the outcome)[5]							
	rs6923462	T/C	*na*	5.89×10^{-13}	C/T	4.86×10^{-6}	x
	rs1430193	A/T	*na*	1.86×10^{-12}	A/T	6.70×10^{-7}	
	rs2863171	C/A	*na*	8.97×10^{-10}	C/A	0.28	
	rs6501431	T/C	*na*	2.94×10^{-9}	T/C	0.23	
	rs1079572	G/A	*na*	9.95×10^{-9}	A/G	0.0098	x
	rs4237643	G/T	*na*	3.53×10^{-8}	G/T	0.91	x
Ratio of FVC over forced expiratory Volume in 1 second (FEV1) adjusted for height (negative correlation with the outcome)[6]							
A	rs1980057	T/C	*na*	3.21×10^{-20}	T/C	0.00073	x
T	rs1032295	G/T	*na*	4.37×10^{-15}	G/T	0.00013	x
G	rs2070600	T/C	*na*	3.15×10^{-14}	T/C	0.41	x
T	rs10947233	T/G	*na*	6.66×10^{-12}	T/G	0.52	x
T	rs11168048	C/T	*na*	1.08×10^{-11}	C/T	0.0065	x
T	rs7735184	T/G	*na*	6.23×10^{-11}	T/G	0.0073	x
T	rs2277027	A/C	*na*	9.93×10^{-11}	A/C	0.053	x
G	rs1422795	T/C	*na*	2.62×10^{-10}	T/C	0.046	x
G	rs3817928	G/A	*na*	1.17×10^{-9}	A/G	2.0×10^{-11}	
C	rs7776375	G/A	*na*	6.71×10^{-9}	A/G	9.4×10^{-18}	
G	rs6937121	G/T	*na*	1.25×10^{-8}	T/G	1.2×10^{-18}	

[a] *We extracted SNPs found associated with human phenotypes at genome-wide significance level in population of European ancestry after adjusting for either body mass index or height, for which effect estimates can be extracted from the GIANT consortium database[7].*

[b] *For each SNP selected we extracted from the GIANT consortium database[7] their p-value for association with BMI or height when relevant, and the trait-increasing allele.*

[c] *GWAS the unadjusted p-value and estimates were not available (noted "na").*

Table S3. Mean chi-square for adjusted and unadjusted analysis across 15 null models

Case-scenario	Generative DAG[a]	Mean chi-square for the test of association between G and Y[b]	
		unadjusted	Adjusted for C
G and Y share a common cause	Figure S1a	989.9	1.0
	Figure S1b	989.9	563.1
	Figure S1c	278.8	1.0
	Figure S1d	278.4	1.0
Effect of G on Y is mediated through C	Figure S1e	989.7	1.0
	Figure S1f	279.0	1.0
	Figure S1g	278.8	1.0
	Figure S1h	989.6	563.3
An unmeasured variable influences both C and Y	Figure S1i	727.5	676.2
	Figure S1j	1.0	818.0
C and Y both influence an unmeasured variable	Figure S1k	1.0	1.0
C is a descendant of Y	Figure S1l	1.0	2251.7
	Figure S1m	1.0	270.8
	Figure S1n	1.0	2143.4
	Figure S1o	1.0	444.0

[a] *DAG: direct acyclic graph. For each DAG from FigureS1 we generated 10,000 replicates of 10,000 individuals. For simplicity all true effects were defined so that independent variables explain 30% of the variance of the dependent variable. All effects were positives.*

[b] *G, Y and C are the genetic variant, the outcome and the covariate included in the adjusted model, respectively.*

References

1. Vansteelandt, S., Goetgeluk, S., Lutz, S., Waldman, I., Lyon, H., Schadt, E.E., Weiss, S.T., and Lange, C. (2009). On the adjustment for covariates in genetic association analysis: a novel, simple principle to infer direct causal effects. Genetic epidemiology 33, 394-405.
2. Thorleifsson, G., Walters, G.B., Gudbjartsson, D.F., Steinthorsdottir, V., Sulem, P., Helgadottir, A., Styrkarsdottir, U., Gretarsdottir, S., Thorlacius, S., Jonsdottir, I., et al. (2009). Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nature genetics 41, 18-24.
3. Scott, R.A., Lagou, V., Welch, R.P., Wheeler, E., Montasser, M.E., Luan, J., Magi, R., Strawbridge, R.J., Rehnberg, E., Gustafsson, S., et al. (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nature genetics 44, 991-1005.
4. Stergiakouli, E., Gaillard, R., Tavare, J.M., Balthasar, N., Loos, R.J., Taal, H.R., Evans, D.M., Rivadeneira, F., St Pourcain, B., Uitterlinden, A.G., et al. (2014). Genome-wide association study of height-adjusted BMI in childhood identifies functional variant in ADCY3. Obesity.
5. Loth, D.W., Artigas, M.S., Gharib, S.A., Wain, L.V., Franceschini, N., Koch, B., Pottinger, T.D., Smith, A.V., Duan, Q., Oldmeadow, C., et al. (2014). Genome-wide association analysis identifies six new loci associated with forced vital capacity. Nature genetics 46, 669-677.
6. Hancock, D.B., Eijgelsheim, M., Wilk, J.B., Gharib, S.A., Loehr, L.R., Marciante, K.D., Franceschini, N., van Durme, Y.M., Chen, T.H., Barr, R.G., et al. (2010). Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nature genetics 42, 45-52.
7. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Magi, R., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature genetics 42, 937-948.