Supplementary Table 1: Simulation parameters (see excel file). (a)-(e) correspond to Tables S2-S6 respectively.

	Description	$p_{\text{MR}} < \alpha$	$p_{\text{Egger}} < \alpha$	$p_{\text{Med}} < \alpha$	$p_{\text{MBE}} < \alpha$	$p_{\text{bi-MR}} < \alpha$	$p_{\text{aux}} < \alpha$	$p_{\text{LCV}} < \alpha$	$\hat{\text{gcp}}$	$\hat{\rho}_g$
a	Default	0.056	0.044	0.118	0.094	0.047	0.002	0	-0.001	0.001
b	Uncorrelated pleiotropy	0.048	0.054	0.161	0.117	0.049	0.01	0	0.005	-0.001
c	Strong uncorrelated pleiotropy	0.046	0.06	0.248	0.098	0.064	0.01	0	0	0
d	Weak uncorrelated pleiotropy	0.04	0.053	0.148	0.092	0.052	0.006	0	0	0.001
e	Nonzero genetic correlation	1	0.211	0.958	0.096	0.054	0.032	0.018	-0	0.2
f	High genetic correlation	1	0.401	1	0.06	0.053	0.054	0.022	-0.001	0.4
g	Very high genetic correlation	1	0.896	1	1	0.051	0.06	0.026	-0.002	0.8
h	Unequal polygenicity	1	0.933	0.995	0.072	0.862	0.104	0.038	-0.002	0.2
i	Very unequal polygenicity	1	0.983	1	0.999	1	0.368	0.038	-0.003	0.199
j	Slightly unequal polygenicity	1	0.527	0.972	0.104	0.187	0.034	0.029	-0.001	0.2
k	Unequal power	1	0.809	1	0.974	0.892	1	0.061	-0.001	0.2
l	Very unequal power	1	0.834	1	0.97	0.971	1	0.049	-0	0.199
m	Slightly unequal power	1	1	1	0.034	0.804	1	0.03	0	0.199

Supplementary Table 2: Comparison with existing methods in null simulations with no LD. Five types of MR methods (two-sample MR, MR-Egger, Weighted Median, Mode-based Estimator and Bidirectional MR) are used, as well as LCV and the auxiliary test. The proportion of simulations out of $n = 1000$ with $p < 0.05$ is reported. (a) Default parameters with no pleiotropy: sample size $N_1 = N_2 = 100$k; number of SNPs $M = 50$k; 5% of SNPs causal for each trait, 0 SNPs causal for both traits. (b) 1% of SNPs are pleiotropic, with uncorrelated effect sizes on each trait; 4% of SNPs are causal for each trait exclusively. Average effect size is identical for pleiotropic and non-pleiotropic SNPs. (c) 2.5% of SNPs are pleiotropic; 2.5% of SNPs are causal for each trait exclusively. (d) 0.5% of SNPs are pleiotropic; 4.5% of SNPs are causal for each trait exclusively. (e) Genetic correlation of 0.2. 1% of SNPs are pleiotropic, with identical effect sizes on both traits; 4% of SNPs are causal for each trait exclusively. Note that none of the methods are expected to perform differently when the genetic correlation is negative instead of positive. (f) Genetic correlation of 0.4. 2% of SNPs are pleiotropic; 3% of SNPs are causal for each trait exclusively. (g) Genetic correlation of 0.8. 4% of SNPs are pleiotropic; 1% of SNPs are causal for each trait exclusively. (h) Unequal polygenicity, in addition to a genetic correlation of 0.2: 1% of SNPs causal for both traits, 8% of SNPs causal for trait 1 only, 2% of SNPs causal for trait 2 only. (i) Unequal polygenicity, in addition to a genetic correlation of 0.2: 1% of SNPs causal for both traits, 16% of SNPs causal for trait 1 only, 1% of SNPs causal for trait 2 only. (j) Unequal polygenicity, in addition to a genetic correlation of 0.2: 1% of SNPs causal for both traits, 5% of SNPs causal for trait 1 only, 3% of SNPs causal for trait 2 only. (k) Unequal power, in addition to a genetic correlation of 0.2: $N = 50$k for trait 1, $N = 200$k for trait 2. In addition, polygenicity is unequal for pleiotropic and non-pleiotropic SNPs (but equal between each trait): 0.5% of SNPs are causal for both traits, and 8% of SNPs are causal for each trait exclusively. (l) Unequal power, in addition to a genetic correlation of 0.2 and unequal polygenicity between pleiotropic and non-pleiotropic SNPs: $N = 50$k for trait 1, $N = 400$k for trait 2. (m) Unequal power, in addition to a genetic correlation of 0.2 and unequal polygenicity between pleiotropic and non-pleiotropic SNPs: $N = 100$k for trait 1, $N = 200$k for trait 2. We note that the weighted median method produced slightly inflated false positive rates even in the absence of a genetic correlation; this behavior may be due to miscalibrated standard errors, which are computed using a parametric bootstrap approach.

	Description	$p_{\text{MR}} < \alpha$	$p_{\text{Egger}} < \alpha$	$p_{\text{Med}} < \alpha$	$p_{\text{MBE}} < \alpha$	$p_{\text{bi-MR}} < \alpha$	$p_{\text{aux}} < \alpha$	$p_{\text{LCV}} < \alpha$	gĉp	$\hat{\rho}_g$
a	Default parameters	0.316	0	0.052	0	0.039	0.984	0.995	0.698	0.201
b	Low N1	0.001	0.001	0	0	0	0.006	0.886	0.646	0.199
c	Very low N1	0	0	0	0	0	0.077	0.553	0.52	0.205
d	High N1	1	0.007	0.981	0.004	0.076	1	1	0.748	0.201
e	Low N2	0.127	0.003	0.054	0	0	0.975	0.789	0.607	0.2
f	Very low N2	0.037	0.002	0.058	0	0	0.535	0.302	0.39	0.202
g	High N2	0.656	0.001	0.042	0	0.268	0.252	0.999	0.744	0.202
h	Small effect size	0.029	0	0.001	0	0.008	0.267	0.081	0.466	0.099
i	Very small effect size	0.002	0.002	0.001	0	0.002	0.007	0	0.217	0.052
j	Large effect size	0.927	0.005	0.723	0.012	0.064	1	1	0.856	0.401
k	High polygenicity	0.002	0	0.001	0	0.001	0.58	0.46	0.566	0.202
l	Very high polygenicity	0	0	0	0	0	0.136	0.018	0.375	0.201
m	Low polygenicity	0.999	0.008	0.856	0	0.191	1	1	0.782	0.201
n	Pleiotropy	0.138	0.004	0.076	0	0.008	0.618	0.793	0.351	0.2
o	Strong pleiotropy	0.087	0	0.09	0	0	0.179	0.493	0.22	0.199
p	Partial causality (gcp=0.75)	0.034	0	0.014	0.001	0.01	0.774	0.809	0.607	0.2
q	Partial causality (gcp=0.5)	0.004	0	0	0	0.003	0.282	0.277	0.486	0.198
r	Partial causality (gcp=0.25)	0	0.001	0.001	0	0.002	0.04	0.029	0.296	0.2

Supplementary Table 3: Comparison with existing methods in non-null simulations with no LD. Five types of MR methods (two-sample MR, MR-Egger, Weighted Median, Mode-based Estimator and Bidirectional MR) are compared with LCV and the auxiliary test, and the proportion of simulations out of $n = 1000$ with $p < 0.001$ is reported. (a) Causal simulation with default parameters: sample size $N_1 = N_2 = 25$k; number of SNPs $M = 50$k; 5% of SNPs causal for trait 1, with a causal effect of size 0.2 on trait 2; 5% of SNPs causal for trait 2 but not trait 1; heritability of 0.2 for both traits. (b) Lower sample size for trait 1 (the causal trait): $N_1 = 12.5$k, $N_2 = 25$k. (c) Lower sample size for trait 1 (the causal trait): $N_1 = 6.2$k, $N_2 = 25$k. (d) Higher sample size for trait 1: $N_1 = 50$k, $N_2 = 25$k. (e) Lower sample size for trait 2: $N_1 = 12.5$k, $N_2 = 25$k. (f) Lower sample size for trait 2: $N_1 = 25$k, $N_2 = 6.2$k. (g) Higher sample size for trait 2: $N_1 = 25$k, $N_2 = 50$k. (h) Smaller causal effect size of 0.1. (i) Very small causal effect size of 0.05. (j) Large causal effect size of 0.4. (k) Higher trait 1 polygenicity: 10% of SNPs are causal for trait 1, and an additional 5% of SNPs are causal for trait 2. (l) Very high trait 1 polygenicity: 20% of SNPs are causal for trait 1, and an additional 5% of SNPs are causal for trait 2. Lower trait 1 polygenicity: 2.5% of SNPs are causal for trait 1, and an additional 5% of SNPs are causal for trait 2. (m) Additional pleiotropy: 2.5% of SNPs are causal for both traits with extra effects on trait 2 in addition to the component of their effects which is mediated by trait 1. 2.5% of SNPs are causal for trait 1 with mediated effects (but no additional pleiotropic effects) on trait 2, and 2.5% of SNPs are causal for trait 2 only. (o) Strong additional pleiotropy: all SNPs affecting trait 1 (5% of SNPs) have additional pleiotropic non-mediated effects on trait 2; no SNPs affect trait 2 only (or trait 1 only). (p) Partial genetic causality, with a gcp of 0.75 and a genetic correlation of 0.2. (q) Partial genetic causality, with a gcp of 0.5 and a genetic correlation of 0.2. Partial genetic causality, with a gcp of 0.25 and a genetic correlation of 0.2.

	Description	$p_{\text{MR}} < \alpha$	$p_{\text{Egger}} < \alpha$	$p_{\text{Med}} < \alpha$	$p_{\text{MBE}} < \alpha$	$p_{\text{bi-MR}} < \alpha$	$p_{\text{aux}} < \alpha$	$p_{\text{LCV}} < \alpha$	gĉp	$\hat{\rho}_g$
a	No pleiotropy	0.054	0.044	0.137	0.098	0.053	0.002	0	0.001	-0.001
b	Low genetic correlation	1	0.182	0.277	0.088	0.04	0.017	0.011	0.001	0.07
c	Nonzero genetic correlation	1	0.344	0.968	0.086	0.043	0.023	0.034	-0.001	0.142
d	High genetic correlation	1	0.612	1	0.041	0.026	0.032	0.022	0.001	0.283
e	Very high genetic correlation	1	0.979	1	0.876	0.023	0.032	0.029	-0.002	0.566
f	Slightly unequal polygenicity	1	0.049	0.973	0.075	0.13	0.037	0.033	-0	0.141
g	Unequal polygenicity	1	0.334	0.992	0.086	0.748	0.103	0.034	0.001	0.141
h	Very unequal polygenicity	1	0.979	0.999	0.075	0.999	0.414	0.045	0.001	0.141
i	Slightly unequal power	1	0.973	0.999	0.006	0.783	0.97	0.036	0.001	0.142
j	Unequal power	0.99	0.202	0.99	0.565	0.439	1	0.058	0.002	0.141
k	Very unequal power	0.997	0.203	0.986	0.577	0.773	1	0.039	0.001	0.141
l	Causal with weak pleiotropy	0.104	0.005	0.007	0	0.029	0.699	0.535	0.551	0.141
m	Causal with pleiotropy	0.109	0	0.008	0	0.031	0.652	0.577	0.554	0.143
n	Causal with very strong pleiotropy	0.074	0.001	0.015	0	0.013	0.432	0.498	0.494	0.141

Supplementary Table 4: Simulations with no LD under Gaussian mixture models. Five types of MR methods (two-sample MR, MR-Egger, Weighted Median, Mode-based Estimator and Bidirectional MR) are compared with LCV and the auxiliary test, and the proportion of simulations, out of $n = 1000$, with $p < \alpha$ is reported, where $\alpha = 0.05$ for null simulations (top) and $\alpha = 0.001$ for causal simulations (bottom). Mean estimated gcp and genetic correlation (ρ_g) are also reported. (a) Null simulation with no pleiotropy. 5% of SNPs are causal for each trait. (b) Null simulation with a low genetic correlation ($\rho_g = \sqrt{2}/20$). 1% of SNPs are causal for both traits with correlated effects ($r^2 = 0.5$), and 4% of SNPs are causal for each trait exclusively. (c) Null simulation with a medium genetic correlation ($\rho_g = \sqrt{2}/10$) (corresponds to Supplementary Figure 1a). (d) Null simulation with a high genetic correlation ($\rho_g = \sqrt{2}/5$). (e) Null simulation with a very high genetic correlation ($\rho_g = 2\sqrt{2}/5$). (f) Null simulation with slightly unequal polygenicity between the two traits. 1% of SNPs are causal for both traits with correlated effects ($r^2 = 0.5$), 3% of SNPs are causal for trait 1 exclusively, and 5% of SNPs are causal for trait 2 exclusively. (g) Null simulation with moderately unequal polygenicity between the two traits (corresponds to Supplementary Figure 1b). 2% of SNPs are causal for trait 1 exclusively, and 8% of SNPs are causal for trait 2 exclusively. (h) Null simulation with very unequal polygenicity between the two traits. 1% of SNPs are causal for trait 1 exclusively, and 16% of SNPs are causal for trait 2 exclusively. (i) Null simulation with slightly unequal power between the two traits. 0.5% of SNPs are causal for both traits with correlated effects ($r^2 = 0.5$), and 8% of SNPs are causal for each trait exclusively. Sample sizes is $N_1 = 50$k and $N_2 = 100$k. (j) Null simulation with moderately unequal power between the two traits (corresponds to Supplementary Figure 1c) ($N_1 = 25$k, $N_2 = 100$k). (k) Null simulation with very unequal power between the two traits ($N_1 = 25$k, $N_2 = 400$k). (l) Causal simulation with weak additional pleiotropic effects. 1% of SNPs were causal for both traits, with correlated effect sizes ($r^2 = 2/3$); 4% of SNPs were causal for trait 2 exclusively, and 0% of SNPs were causal for trait 1 exclusively. (m) Causal simulation with moderately strong additional pleiotropic effects ($r^2 = 0.5$ between effect sizes of correlated SNPs). (n) Causal simulation with very strong additional pleiotropic effects ($r^2 = 0.125$ between effect sizes of correlated SNPs).

	Description	$p_{\text{MR}} < \alpha$	$p_{\text{Egger}} < \alpha$	$p_{\text{Med}} < \alpha$	$p_{\text{MBE}} < \alpha$	$p_{\text{bi-MR}} < \alpha$	$p_{\text{aux}} < \alpha$	$p_{\text{LCV}} < \alpha$	gĉp	$\hat{\rho}_g$
a	Equal polygenicity	1	0.83	1	0.018	0.027	0.03	0.014	-0	0.458
b	Slightly unequal polygenicity	1	0.964	1	0.023	0.766	0.176	0.975	0.176	0.459
c	Unequal polygenicity	1	1	1	0.247	1	0.912	1	0.422	0.459
d	Very unequal polygenicity	1	1	1	0.088	1	1	1	0.514	0.459
e	Equal polygenicity, N=500k	1	1	1	0.06	0.011	0.002	0.014	0	0.458
f	Slightly unequal polygenicity, N=500k	1	1	1	0.063	0.047	0.01	0.995	0.177	0.458
g	Unequal polygenicity, N=500k	1	1	1	0.065	0.474	0.081	1	0.421	0.458
h	Very unequal polygenicity, N=500k	1	1	1	0.064	0.918	0.369	1	0.514	0.458
i	Causal with weak confounder	0.999	0.016	0.89	0	0.246	1	1	0.81	0.282
j	Causal with confounder	0.995	0.005	0.728	0.001	0.085	0.997	0.999	0.649	0.36
k	Causal with strong confounder	0.937	0	0.763	0	0.001	0.817	0.489	0.284	0.521

Supplementary Table 5: Simulations with no LD under multiple intermediary models. Five types of MR methods (two-sample MR, MR-Egger, Weighted Median, Mode-based Estimator and Bidirectional MR) are compared with LCV and the auxiliary test, and the proportion of simulations, out of $n = 1000$, with $p < \alpha$ is reported, where $\alpha = 0.05$ for null simulations (top) and $\alpha = 0.001$ for causal simulations (bottom). Mean estimated gcp and genetic correlation (ρ_g) are also reported. (a) Null simulation with two intermediaries having equal polygenicity (corresponds to Supplementary Figure 1d). The first intermediary explains 15% and 35% of heritability for trait 1 and trait 2 respectively, and the second intermediary explains 35% and 15% respectively. 2% of SNPs are causal for each intermediary. In all null simulations in this table, 4% of SNPs are causal for each trait exclusively. (b) Null simulation with two intermediaries having slightly unequal polygenicity. 2% and 4% of SNPs are causal for each intermediary respectively. (c) Null simulation with two intermediaries having unequal polygenicity (corresponds to Supplementary Figure 1e). 1% and 8% of SNPs are causal for each intermediary respectively. (d) Null simulation with two intermediaries having very unequal polygenicity. 0.5% and 16% of SNPs are causal for each intermediary respectively. (e-h) Null simulations similar to (a-d), but at higher sample size ($N = 500$k instead of $N = 100$k). (i) Causal simulation with an additional latent variable explaining a small proportion of the genetic correlation. 2% of SNPs are causal for trait 1 with proportional effects on trait 2 (causal effect size: $q_2 = 0.2$); 2% of SNPs are causal for both traits via the additional latent variable, explaining 10% of heritability for each 2 exclusively. (j) Causal simulation with an additional latent variable explaining approximately half of the genetic correlation. The additional latent variable explains 20% of heritability for each trait. (k) Causal simulation with an additional latent variable explaining a large proportion of the genetic correlation (corresponds to Supplementary Figure 1g). The additional latent variable explains 40% of heritability for each trait.

		ρ_g	$p < .05$	$p < .001$	Mean χ^2	Mean gĉp	gĉp std dev	RMS $\hat{\sigma}$	Z_h
a	Zero genetic correlation	0	0	0	0.32	-0.00	0.11	0.55	8
b	Low genetic correlation	0.1	0.009	0	0.58	0.00	0.14	0.29	8.5
c	Default parameter values	0.2	0.058	0.003	1.09	-0.00	0.07	0.08	8.6
d	High genetic correlation	0.4	0.067	0.004	1.2	-0.00	0.1	0.11	8
e	Very high genetic correlation	0.8	0.058	0.002	1.13	-0.00	0.21	0.24	5.8
f	Uncorrelated pleiotropic effects	0.2	0.054	0.001	1.06	-0.00	0.08	0.09	8.7
g	Differential polygenicity	0.2	0.062	0.002	1.1	-0.01	0.08	0.08	10
h	Very different polygenicity	0.2	0.067	0.004	1.19	-0.01	0.1	0.1	11.2
i	Low N_1	0.2	0.063	0.004	1.14	0.01	0.12	0.13	5
j	Very low N_1	0.2	0.228	0.132	11.2	0.11	0.35	0.33	1.4
k	Different heritability	0.2	0.061	0.005	1.7	0.00	0.09	0.1	6.5
l	High phenotypic correlation	0.2	0.057	0.002	1.12	0.00	0.07	0.08	8.7
m	Zero phenotypic correlation	0.2	0.057	0.005	1.1	0.00	0.07	0.08	8.6
n	Uncorrelated pleiotropic effects	0	0.001	0	0.3	0.00	0.14	0.52	8
o	Differential polygenicity	0	0	0	0.31	-0.02	0.12	0.55	9.8
p	Very different polygenicity	0	0.001	0	0.31	-0.05	0.14	0.52	11.4
q	Low N_1	0	0.001	0.001	0.31	0.00	0.14	0.52	5
r	Very low N_1	0	0.272	0.216	46.4	0.27	0.32	0.39	1.4
s	Different heritability	0	0	0	0.28	-0.00	0.11	0.55	6.3
t	Causal	0.2	0.965	0.94	258	0.76	0.12	0.16	8.6
u	Partially causal	0.2	0.706	0.347	12.9	0.56	0.15	0.24	10
v	Low N_1	0.2	0.852	0.768	66	0.65	0.17	0.2	5.1
w	Very low N_1	0.2	0.452	0.378	102	0.39	0.35	0.35	1.4
x	Low N_2	0.2	0.843	0.714	40.8	0.60	0.18	0.21	8.7
y	Weak causal effect	0.1	0.422	0.104	6.36	0.49	0.18	0.32	8.7
z	Y_1 less polygenic	0.2	0.997	0.996	7331	0.90	0.08	0.07	3.6
aa	Y_1 more polygenic	0.2	0.155	0.004	2.39	0.28	0.2	0.47	13.3
bb	Y_1 infinitessimal	0.2	0.012	0	0.7	0.07	0.2	0.5	14.2

Supplementary Table 6: Additional simulations with LD. Proportion of simulations (out of n = 5000) with LCV p-value for partial causality less than 0.05 and less than 0.001; mean χ^2 statistic; mean gĉp (in each case, standard error is less than 0.01); empirical standard deviation of gĉp; root mean squared estimated standard error; mean heritability Z-score for trait 1. Simulations a-s are null (gcp = 0), and simulations t-bb are non-null. (a-e) Different values of the genetic correlation (ρ_g). When the genetic correlation is zero or near-zero, we observe conservative p-values and overestimates of the gĉp standard error. (f) Uncorrelated pleiotropic effects: 0.3% of SNPs affect both traits with independent effect sizes. (g-h) Differential or very different polygenicity: 0.2% and 0.8% of SNPs, or 0.1% and 1.6% of SNPs respectively, have direct effects on each trait. (i-j) Low or very low sample size for trait 1: either N_1 = 20k or N_1 = 4k respectively, and N_2 = 100k. (k) Different heritability: h_1^2 = 0.1 and h_2^2 = 0.5. (l) High phenotypic correlation of 0.4, compared with ρ_g = 0.2. (m) Zero phenotypic correlation. (n) Uncorrelated pleiotropic effects: 0.3% of SNPs affect both traits with independent effect sizes. (o-p) Differential or very different polygenicity: 0.2% and 0.8% of SNPs, or 0.1% and 1.6% of SNPs respectively, have direct effects on each trait. (q-r) Low or very low sample size for trait 1: either N_1 = 20k or N_1 = 4k respectively, and N_2 = 100k. (s) Different heritability: h_1^2 = 0.1 and h_2^2 = 0.5. (t) Causal. (u) Partially causal (gcp = 0.5). (v-w) Causal, with low or very low sample size for the causal trait (N_1 = 20k or N_1 = 4k, and N_2 = 100k). (x) Causal, with low sample size in the downstream trait (N_2 = 20k, N_1 = 100k). (y) Weak causal effect (0.1 rather than 0.25). (z-bb) Varying polygenicity for the causal trait: instead of 0.5% of SNPs causal, either 0.05%, 5%, or 100% of SNPs causal for z-bb respectively.

		ρ_g	$p < .05$	$p < .001$	Mean χ^2	Mean gĉp	gĉp std dev	RMS $\hat{\sigma}$	Z_h
a	Default parameter values	0.2	0.034	0	0.9	0.00	0.05	0.06	16.7
b	Zero genetic correlation	0	0.001	0	0.31	-0.00	0.11	0.55	15.3
c	Very high genetic correlation	0.8	0.033	0.002	0.94	-0.00	0.11	0.43	10.9
d	Uncorrelated pleiotropic effects	0.2	0.032	0.000	0.87	-0.00	0.06	0.09	16.6
e	Differential polygenicity	0.2	0.034	0.002	0.86	-0.00	0.05	0.07	19.3
f	Low N_1	0.2	0.042	0.002	0.9	0.00	0.1	0.12	8.7
g	Very low N_1	0.2	0.254	0.16	19.96	0.08	0.35	0.32	2.2
h	Causal	0.2	0.968	0.943	257.17	0.76	0.11	0.16	16.5
i	Partially causal	0.2	0.765	0.369	12.54	0.57	0.15	0.23	19.2

Supplementary Table 7: Simulations with LD using constrained-intercept LD score regression to estimate the heritability. This heritability estimation method is less noisy than variable-intercept LD score regression but can produce biased estimates on real data due to population stratification and cryptic relatedness.[19] Proportion of simulations (out of n = 2000) with p-value for partial causality less than .05 and less than .001; mean χ^2 statistic for partial causality; mean gĉp; standard deviation of gcp estimates; root-mean squared estimated standard error. Simulations a-f are null (gcp = 0), and simulations g-h are non-null. (a) Realistic simulation parameters (see Methods). (b) Genetic correlation ρ_g = 0. (c) Genetic correlation ρ_g = 0.75. (d) Uncorrelated pleiotropic effects in addition to a genetic correlation: 50% of SNPs with direct (non-mediated) effects on each trait are shared between the two traits. (e) Differential polygenicity: 0.2% and 0.05% of SNPs have direct effects on each trait. (f) Different sample size: N_1 = 1000k and N_2 = 500k. (g) Different sample size: N_1 = 20k and N_2 = 500k. (h) Full genetic causality: gcp = 1, with causal effect equal to the genetic correlation (0.25). (i) Partial genetic causality: gcp = 0.5.

	Ground truth	PC corrected	$p_{\text{LCV}} < .05$	$p_{\text{LCV}} < .001$	median gĉp
a	Uncorrelated	0	0.58	0.38	-0.49
b	Non-causal correlated	0	0.34	0.13	-0.22
c	Uncorrelated	1	0.006	0	0.00
d	Non-causal correlated	1	0.056	0.006	0.02
e	Causal	0	0.22	0.078	0.05
f	Causal corrected	1	0.99	0.90	0.78

Supplementary Table 8: Confounding due to population stratification and correction for stratification using PCA. Simulations were performed using UK Biobank genotypes for chromosome 1, with environmental stratification added along PC1 explaining 1% of variance for trait 1 and 2% for trait 2. LCV was applied to summary statistics before and after correction for PC1, either when Y_1 was causal for Y_2, when Y_1 and Y_2 were genetically correlated with no partial causality, or when Y_1 and Y_2 had no genetic correlation. Based on n = 500 simulations.

	Regression coefficient (std err)	RMSE	RMPV
Ascertained simulations (43%)	0.97 (.004)	0.15	0.13
All simulations	1.00 (.005)	0.24	0.20

Supplementary Table 9: Unbiasedness of estimated gcp and standard error in simulations with random true parameter values, using real LD. We drew random values of gcp (and ρ_g) from a $Unif(-1, 1)$ distribution and compared true and estimated values of gcp, either for all $n = 10,000$ simulations or for a subset (43%) of simulations in which the genetic correlation was nominally significant $p < 0.05$ and the evidence for partial causality was strong ($p < 0.001$). We report the regression coefficient of true on estimated gcp values with standard error, as well as the root mean squared error and the root mean posterior variance estimate.

Phenotype	Reference	N (thousands)	Z_h
Anorexia	Boraska et al., 2014 Mol Psych	32	17.8
Autism Spectrum	PGC Cross-Disorder Group, 2013 Lancet	10	12.1
Bipolar Disorder	BIP Working Group of the PGC, 2011 Nat Genet	17	11.8
Breast Cancer	Amos et al., 2016 Cancer Epidemiol. Biomarkers Prev.	~ 447*	16
Celiac Disease	Dubois et al., 2010 Nat Genet	15	10.4
Crohns Disease	Jostins et al., 2012 Nature	21	12.1
Depressive symptoms	Okbay et al., 2016 Nat Genet	161	13.1
HDL	Teslovich et al., 2010 Nature	98	8.2
HbA1c	Soranzo et al., 2010 Diabetes	46	8.8
LDL	Teslovich et al., 2010 Nature	93	8.1
Lupus	Bentham et al., 2015 Nat Genet	14	10.2
Prostate Cancer	Amos et al., 2016 Cancer Epidemiol. Biomarkers Prev.	~ 447*	7.5
Schizophrenia	SCZ Working Group of the PGC, 2014 Nature	70	17.4
Triglycerides	Teslovich et al., 2010 Nature	94	9.5
Ulcerative Colitis	Jostins et al., 2012 Nature	27	8.8
Eosinophil count †	UK Biobank 27–29	~ 460**	20.8
Reticulocyte count †	UK Biobank 27–29	~ 460	19.9
Lymphocyte count †	UK Biobank 27–29	~ 460	22.7
Mean corpuscular hemoglobin †	UK Biobank 27–29	~ 460	14.3
Mean platelet volume †	UK Biobank 27–29	~ 460	15.7
Monocyte count †	UK Biobank 27–29	~ 460	15.1
Platelet count †	UK Biobank 27–29	~ 460	20.2
Platelet distribution width †	UK Biobank 27–29	~ 460	17.1
RBC distribution width †	UK Biobank 27–29	~ 460	19.7
RBC count †	UK Biobank 27–29	~ 460	17.5
White cell count †	UK Biobank 27–29	~ 460	20.7
Bone mineral density - heel †	UK Biobank 27–29	~ 460	29
Balding - male*** †	UK Biobank 27–29	~ 230	16.1
BMI †	UK Biobank 27–29	~ 460	27.5
Height †	UK Biobank 27–29	~ 460	24.7
BP - diastolic †	UK Biobank 27–29	~ 460	32.3
BP - systolic †	UK Biobank 27–29	~ 460	28.3
College †	UK Biobank 27–29	~ 460	19.1
Smoking status †	UK Biobank 27–29	~ 460	24.9
Eczema †	UK Biobank 27–29	~ 460	21.8
Asthma †	UK Biobank 27–29	~ 460	16.8

Dermatology †	UK Biobank 27–29	~ 460	9.1
Myocardial infarction****	UK Biobank 27–29	~ 460	18.6
High cholesterol †	UK Biobank 27–29	~ 460	15.6
Hypertension †	UK Biobank 27–29	~ 460	36.2
Hypothyroidism †	UK Biobank 27–29	~ 460	20.1
Type 2 Diabetes †	UK Biobank 27–29	~ 460	19.5
Basal metabolic rate †	UK Biobank 27–29	~ 460	23.4
FEV1/FVC †	UK Biobank 27–29	~ 460	17.7
FVC †	UK Biobank 27–29	~ 460	18.8
Neuroticism †	UK Biobank 27–29	~ 460	28.7
Morning person †	UK Biobank 27–29	~ 460	21.1
Age at menarche †	UK Biobank 27–29	~ 230	24
Age at menopause †	UK Biobank 27–29	~ 230	19.1
Number children - female	UK Biobank 27–29	~ 230	14.4
Number children - male	UK Biobank 27–29	~ 230	15.1

Supplementary Table 10: 52 GWAS datasets included in the analysis. Most UK Biobank summary statistics are publicly available.[29] All datasets have heritability Z-score $Z_h > 7$ and estimated genetic correlation $\hat{\rho}_g < 0.9$ with other traits. Summary statistics for $\sim 1,000,000$ HapMap3 SNPs were used, excluding the MHC region. *Total number of samples genotyped by OncoArray; actual sample size is slightly less than 447k. These numbers are excluded from average reported sample size for non-UK Biobank traits. **Actual sample size for UK Biobank analyses is slightly less than 460k (respectively 230k for sex-specific traits), owing to incomplete phenotype data. For most case control traits, effective sample size is substantially less than 460k due to the low fraction of cases. ***The balding phenotype was the "balding 4" UK Biobank category, corresponding to nearly-complete baldness. ****We confirmed that self-reported MI in UK Biobank was highly genetically correlated with CAD in CARDIoGRAM consortium data[39] ($\rho_g = 1.34(0.25)$; not significantly different from 1). †Summary statistics publicly available at https://data.broadinstitute.org/alkesgroup/UKBB/.

Trait 1	Trait 2	p_{LCV}	$\hat{\rho}_g$ (std err)	gĉp(std err)	p_{aux}	MR ref
Triglycerides	Hypertension	5×10^{-39}	0.25 (0.04)	0.95 (0.04)	0.04	
BMI	Heart attack	3×10^{-9}	0.34 (0.09)	0.94 (0.11)	0.22	32, 38
Triglycerides	Heart attack	8×10^{-32}	0.30 (0.06)	0.90 (0.08)	0.04	4
Triglycerides	BP - systolic	6×10^{-41}	0.13 (0.03)	0.89 (0.08)	8×10^{-4}	
HDL	Hypertension	6×10^{-22}	-0.29 (0.06)	0.87 (0.09)	0.15	
LDL	High cholesterol	8×10^{-7}	0.77 (0.07)	0.86 (0.11)	0.08	
Triglycerides	Mean cell volume	10×10^{-19}	-0.20 (0.04)	0.86 (0.11)	2×10^{-4}	
Triglycerides	BP - diastolic	5×10^{-39}	0.11 (0.04)	0.86 (0.10)	0.004	
Platelet volume	Platelet count	6×10^{-10}	-0.66 (0.03)	0.84 (0.10)	0.18	
BMI	Hypertension	2×10^{-16}	0.38 (0.03)	0.83 (0.11)	0.06	11, 38
Triglycerides	Platelet dist width	5×10^{-17}	0.19 (0.04)	0.81 (0.13)	7×10^{-5}	
LDL	BMD	4×10^{-34}	-0.12 (0.05)	0.80 (0.12)	0.02	
BMI	FVC	4×10^{-13}	-0.22 (0.03)	0.79 (0.17)	0.001	72
Triglycerides	Reticulocyte count	2×10^{-10}	0.33 (0.05)	0.79 (0.14)	0.02	
Triglycerides	Eosinophil count	3×10^{-17}	0.14 (0.05)	0.75 (0.16)	0.001	
Balding - male	Num children - male	2×10^{-30}	-0.16 (0.05)	0.75 (0.13)	2×10^{-4}	
HDL	Platelet dist width	8×10^{-17}	-0.14 (0.04)	0.75 (0.16)	0.004	
RBC dist width	Type 2 Diabetes	3×10^{-4}	0.11 (0.03)	0.73 (0.19)	0.21	
LDL	Heart attack	2×10^{-31}	0.17 (0.08)	0.73 (0.13)	6×10^{-4}	3, 13
Platelet dist width	Platelet count	1×10^{-7}	-0.47 (0.04)	0.73 (0.15)	0.04	
Hypothyroidism	Type 2 Diabetes	2×10^{-4}	0.22 (0.05)	0.73 (0.29)	0.2	
HDL	Type 2 Diabetes	2×10^{-7}	-0.40 (0.06)	0.72 (0.17)	0.35	
Hypothyroidism	Heart attack	6×10^{-12}	0.26 (0.05)	0.72 (0.16)	0.08	
High cholesterol	Heart attack	2×10^{-4}	0.52 (0.12)	0.71 (0.19)	0.32	
HDL	BP - diastolic	4×10^{-17}	-0.12 (0.06)	0.70 (0.18)	0.005	
Platelet dist width	Reticulocyte count	1×10^{-7}	0.13 (0.04)	0.69 (0.20)	0.005	
LDL	College	1×10^{-10}	-0.13 (0.05)	0.68 (0.30)	0.35	
Triglycerides	Monocyte count	1×10^{-4}	0.14 (0.04)	0.67 (0.21)	0.09	
Type 2 Diabetes	Ulcerative Colitis	2×10^{-5}	-0.14 (0.07)	0.65 (0.23)	0.41	
BMI	Reticulocyte count	4×10^{-5}	0.39 (0.03)	0.64 (0.25)	10×10^{-4}	
HDL	FEV1/FVC	1×10^{-13}	-0.09 (0.04)	0.56 (0.08)	0.19	
High cholesterol	Neuroticism	2×10^{-14}	0.09 (0.03)	0.55 (0.19)	0.32	
Triglycerides	Basal metab rate	2×10^{-8}	0.08 (0.04)	0.55 (0.13)	0.25	
Height	BMD	3×10^{-14}	-0.09 (0.04)	0.50 (0.14)	2×10^{-8}	
Triglycerides	Height	3×10^{-14}	-0.10 (0.03)	0.45 (0.09)	0.15	
HbA1C	High cholesterol	5×10^{-22}	0.25 (0.06)	0.44 (0.16)	0.49	
Age at menarche	Height	7×10^{-11}	0.16 (0.04)	0.43 (0.10)	2×10^{-5}	11
High cholesterol	Smoking status	5×10^{-19}	0.13 (0.03)	0.42 (0.02)	0.52	
Reticulocyte count	Hypertension	2×10^{-4}	0.27 (0.04)	0.41 (0.13)	0.75	
BMI	Asthma	4×10^{-14}	0.21 (0.03)	0.40 (0.27)	0.05	72
High cholesterol	Monocyte count	4×10^{-4}	0.09 (0.03)	0.40 (0.15)	0.2	
Height	Basal metab rate	10×10^{-9}	0.57 (0.03)	0.39 (0.07)	0.006	
Eczema	FEV1/FVC	2×10^{-15}	-0.08 (0.03)	0.36 (0.10)	2×10^{-5}	
Height	College	3×10^{-6}	0.17 (0.03)	0.33 (0.10)	0.06	71
Prostrate cancer	Hypothyroidism	10×10^{-5}	-0.12 (0.05)	0.30 (0.38)	0.19	
Crohns Disease	LDL	4×10^{-13}	-0.12 (0.06)	0.29 (0.15)	0.82	
High cholesterol	Type 2 Diabetes	4×10^{-6}	0.42 (0.05)	0.24 (0.30)	0.62	
RBC count	Monocyte count	8×10^{-7}	0.14 (0.05)	0.24 (0.46)	0.31	
HbA1C	BMI	7×10^{-17}	0.25 (0.05)	0.23 (0.35)	0.77	
Basal metab rate	Hypothyroidism	6×10^{-21}	0.11 (0.04)	0.21 (0.04)	0.04	

Platelet dist width	Corpuscular hemoglobin	5×10^{-14}	-0.06 (0.02)	0.15 (0.14)	0.08
Depressive syndrome	Asthma	4×10^{-4}	0.21 (0.05)	0.14 (0.08)	0.37
BMI	High cholesterol	2×10^{-6}	0.33 (0.06)	0.13 (0.12)	0.25
Age at menopause	Depressive syndrome	2×10^{-7}	-0.27 (0.06)	0.12 (0.32)	0.41
White cell count	BMI	7×10^{-5}	0.24 (0.03)	0.09 (0.16)	1
Asthma	Lymphocyte count	2×10^{-4}	0.09 (0.04)	0.08 (0.19)	0.57
Num children - male	Hypothyroidism	9×10^{-11}	0.18 (0.05)	0.03 (0.26)	0.87
College	High cholesterol	2×10^{-8}	-0.23 (0.03)	0.01 (0.08)	0.34
RBC dist width	High cholesterol	4×10^{-4}	0.11 (0.04)	0.00 (0.17)	0.35

Supplementary Table 11: Pairs of traits with evidence of partial genetic causality. We restricted to pairs of traits having a nominally significant genetic correlation (two-tailed $p < 0.05$; 429 trait pairs) and reported all traits with strong evidence of partial causality (1% FDR). Trait pairs are ordered so that trait 1 is genetically causal or partially genetically causal for trait 2. We have provided references for each trait pair with existing support in the MR literature that we are aware of. For some trait pairs, there was strong evidence for partial causality but low and noisy gcp estimates. This phenomenon may occur due to multiple intermediaries, which can cause the estimated mixed fourth moments to have opposite signs. When this occurs, the approximate likelihood function is sometimes bimodal, with no support for any specific value of gcp (because there is no value of gcp that produces mixed fourth moments of opposite signs). While this phenomenon appears to occur for several traits with gcp estimates close to zero, there were no trait pairs with statistically significant evidence that their mixed fourth moments had opposite signs.

Supplementary Table 12: Results for all genetically correlated pairs of phenotypes (see Excel file). We report p-values for LCV, MR, MR-Egger, bidirectional MR, and the auxiliary test for partial causality. We also report gcp estimates using LCV and genetic correlation estimates using LDSC.[16]

Trait 1	Trait 2	Traits conditioned	p_{LCV}	$\hat{\rho}_g$ (std err)	gĉp(post std err)
BMI	MI	LDL, TG	6×10^{-26}	0.28(0.09)	0.48(0.42)
Triglycerides	MI	LDL, BMI	1×10^{-20}	0.18(0.07)	0.82(0.13)
LDL	MI	TG, BMI	N/A	0.02(0.09)	N/A
Hypothyroidism	MI	LDL, TG, BMI	2×10^{-23}	0.17(0.06)	0.78(0.14)
High cholesterol	MI	LDL, TG, BMI	0.006	0.42(0.15)	0.59(0.23)
HDL	MI	LDL, TG, BMI	0.8	-0.16(0.06)	-0.15(0.48)

Supplementary Table 13: Conditional analyses of MI and potential MI risk factors. Trait 1 summary statistics were residualized on summary statistics for BMI, LDL and triglycerides, and these were analyzed in conjunction with summary statistics for MI (see Online Methods). LCV results are reported for traits whose genetic correlation with MI remained significant ($p < 0.05$) after residualizing; results are reported as N/A for other traits. BMI, LDL and triglycerides were chosen as covariates because they represent well-established causal risk factors for MI. This approach is motivated by a scenario in which the covariates have fully genetically causal effects on both trait 1 and MI. If the covariates are genetically correlated with but not causal for trait 1, then this approach could potentially introduce collider bias[35] and false positive associations. Moreover, if the effect of trait 1 on MI is mediated by one of the covariates, evidence for a causal effect may persist. Due to these limitations, we view this approach as a sensitivity analysis, and we do not recommend applying LCV to residualized summary statistics as a primary analysis.

Trait 1	Trait 2	p_{LCV}	$\hat{\rho}_g$ (std err)	gĉp(std err)
BMI	Type 2 Diabetes	0.086	0.60 (0.04)	0.45 (0.41)
Asthma	FVC	0.084	-0.17 (0.04)	0.22 (0.12)
Asthma	FEV1/FVC	0.12	-0.30 (0.04)	0.16 (0.10)
Smoking status	FVC	0.29	-0.03 (0.03)	-0.03 (0.53)
Smoking status	FEV1/FVC	0.40	0.03 (0.03)	-0.08 (0.47)
Smoking status	MI	0.17	0.23 (0.05)	0.49 (0.33)
Anorexia	BMI	0.26	-0.17 (0.04)	0.31 (0.45)

Supplementary Table 14: Plausible causal relationships not identified by LCV.

Trait 1	Trait 2	p_{LCV}	$p_{\text{bi-MR}}$	$\hat{\rho}_g$ (std err)	gĉp(std err)
Triglycerides	MI	8×10^{-32}	0.009	0.30 (0.06)	0.90 (0.08)
Triglycerides	BP - systolic	6×10^{-41}	0.002	0.13 (0.03)	0.89 (0.08)
HDL	Hypertension	6×10^{-22}	0.05	-0.29 (0.06)	0.87 (0.09)
LDL	High cholesterol	8×10^{-7}	0.48	0.77 (0.07)	0.86 (0.11)
LDL	Bone mineral density - heel	4×10^{-34}	0.06	-0.12 (0.05)	0.80 (0.12)
Triglycerides	Eosinophil count	3×10^{-17}	0.2	0.14 (0.05)	0.75 (0.16)
Balding4	Number children - male	2×10^{-30}	0.06	-0.16 (0.05)	0.75 (0.13)
HDL	Platelet distribution width	8×10^{-17}	0.004	-0.14 (0.04)	0.75 (0.16)
RBC distribution width	Type 2 Diabetes	3×10^{-4}	0.45	0.11 (0.03)	0.73 (0.19)
LDL	MI	2×10^{-31}	0.05	0.17 (0.08)	0.73 (0.13)
Platelet distribution width	Platelet count	1×10^{-7}	0.03	-0.47 (0.04)	0.73 (0.15)

Hypothyroidism	Type 2 Diabetes	2×10^{-4}	0.57	0.22 (0.05)	0.73 (0.29)
HDL	Type 2 Diabetes	2×10^{-7}	0.95	-0.40 (0.06)	0.72 (0.17)
Hypothyroidism	MI	6×10^{-12}	0.06	0.26 (0.05)	0.72 (0.16)
High cholesterol	MI	2×10^{-4}	0.15	0.52 (0.12)	0.71 (0.19)
HDL	BP - diastolic	4×10^{-17}	0.25	-0.12 (0.06)	0.70 (0.18)
Platelet distribution width	Reticulocyte count	1×10^{-7}	0.07	0.13 (0.04)	0.69 (0.20)
LDL	College	1×10^{-10}	0.13	-0.13 (0.05)	0.68 (0.30)
Triglycerides	Monocyte count	1×10^{-4}	0.56	0.14 (0.04)	0.67 (0.21)
Type 2 Diabetes	Ulcerative Colitis	2×10^{-5}	0.61	-0.14 (0.07)	0.65 (0.23)
Type 2 Diabetes	Hypertension	8×10^{-4}	0.12	0.44 (0.05)	0.56 (0.18)
HDL	FEV1/FVC	1×10^{-13}	0.21	-0.09 (0.04)	0.56 (0.08)
High cholesterol	Neuroticism	2×10^{-14}	0.06	0.09 (0.03)	0.55 (0.19)
Triglycerides	Basal metabolic rate	2×10^{-8}	0.94	0.08 (0.04)	0.55 (0.13)
Triglycerides	Height	3×10^{-14}	0.81	-0.10 (0.03)	0.45 (0.09)
HbA1C	High cholesterol	5×10^{-22}	0.46	0.25 (0.06)	0.44 (0.16)
High cholesterol	Smoking status	5×10^{-19}	0.004	0.13 (0.03)	0.42 (0.02)
Reticulocyte count	Hypertension	2×10^{-4}	0.24	0.27 (0.04)	0.41 (0.13)
BMI	Asthma	4×10^{-14}	0.002	0.21 (0.03)	0.40 (0.27)
High cholesterol	Monocyte count	4×10^{-4}	0.96	0.09 (0.03)	0.40 (0.15)
Height	College	3×10^{-6}	0.19	0.17 (0.03)	0.33 (0.10)
Prostrate cancer	Hypothyroidism	10×10^{-5}	0.07	-0.12 (0.05)	0.30 (0.38)
Crohns Disease	LDL	4×10^{-13}	0.5	-0.12 (0.06)	0.29 (0.15)
High cholesterol	Type 2 Diabetes	4×10^{-6}	0.002	0.42 (0.05)	0.24 (0.30)
RBC count	Monocyte count	8×10^{-7}	0.001	0.14 (0.05)	0.24 (0.46)
HbA1C	BMI	7×10^{-17}	0.05	0.25 (0.05)	0.23 (0.35)
Basal metabolic rate	Hypothyroidism	6×10^{-21}	0.006	0.11 (0.04)	0.21 (0.04)
Platelet distribution width	Mean corpuscular hemoglobin	5×10^{-14}	0.02	-0.06 (0.02)	0.15 (0.14)
Depressive syndrome	Asthma	4×10^{-4}	0.09	0.21 (0.05)	0.14 (0.08)
Age at menopause	Depressive syndrome	2×10^{-7}	0.43	-0.27 (0.06)	0.12 (0.32)
Asthma	Lymphocyte count	2×10^{-4}	0.73	0.09 (0.04)	0.08 (0.19)
Number children - male	Hypothyroidism	9×10^{-11}	0.13	0.18 (0.05)	0.03 (0.26)
RBC distribution width	High cholesterol	4×10^{-4}	0.22	0.11 (0.04)	0.00 (0.17)
Type 2 Diabetes	Mean cell volume	0.002	8×10^{-5}	-0.15 (0.03)	0.77 (0.20)
Height	BMI	0.002	2×10^{-4}	-0.17 (0.03)	0.65 (0.23)
Type 2 Diabetes	HbA1C	0.05	1×10^{-5}	0.47 (0.09)	0.54 (0.28)
Schizophrenia	Basal metabolic rate	0.06	3×10^{-4}	-0.09 (0.04)	0.51 (0.32)
BMI	Lymphocyte count	0.002	4×10^{-14}	0.20 (0.03)	0.50 (0.21)
BMI	Hypothyroidism	0.08	0.001	0.16 (0.03)	0.49 (0.31)
College	Reticulocyte count	0.13	6×10^{-12}	-0.19 (0.04)	0.47 (0.30)
BMI	Type 2 Diabetes	0.04	$< 10^{-20}$	0.60 (0.04)	0.45 (0.41)
Reticulocyte count	Smoking status	0.02	2×10^{-8}	0.10 (0.04)	0.42 (0.23)
Height	Platelet count	0.002	7×10^{-8}	-0.13 (0.03)	0.42 (0.17)
Age at menopause	Type 2 Diabetes	0.38	3×10^{-4}	-0.17 (0.05)	0.39 (0.35)
College	White cell count	0.03	3×10^{-8}	-0.22 (0.04)	0.38 (0.20)
College	Lymphocyte count	0.06	3×10^{-5}	-0.10 (0.05)	0.34 (0.31)
Basal metabolic rate	College	0.06	7×10^{-4}	-0.07 (0.03)	0.29 (0.38)
Smoking status	Asthma	0.19	2×10^{-4}	0.09 (0.04)	0.29 (0.45)
BMI	Monocyte count	0.35	8×10^{-7}	0.14 (0.03)	0.28 (0.40)
BMI	RBC distribution width	0.19	2×10^{-9}	0.18 (0.02)	0.26 (0.21)
Mean cell volume	Smoking status	0.14	2×10^{-6}	0.13 (0.03)	0.24 (0.25)
HDL	BMI	0.08	1×10^{-6}	-0.36 (0.05)	0.22 (0.15)

Mean corpuscular hemoglobin	Smoking status	0.22	3×10^{-7}	0.08 (0.03)	0.21 (0.35)
BMI	Basal metabolic rate	0.18	6×10^{-13}	0.66 (0.02)	0.21 (0.22)
Mean corpuscular hemoglobin	RBC count	0.27	2×10^{-5}	-0.67 (0.03)	0.19 (0.29)
BMI	FEV1/FVC	0.34	7×10^{-8}	0.17 (0.02)	0.13 (0.17)
College	BP - diastolic	0.48	3×10^{-4}	-0.09 (0.04)	0.08 (0.35)
BMI	Triglycerides	0.01	3×10^{-4}	0.27 (0.05)	0.08 (0.25)
RBC count	BMI	0.39	2×10^{-6}	0.10 (0.03)	0.08 (0.23)
Smoking status	White cell count	0.29	8×10^{-6}	0.12 (0.05)	0.07 (0.32)
College	Monocyte count	0.27	6×10^{-5}	-0.15 (0.03)	0.06 (0.49)
Triglycerides	Hypertension	5×10^{-39}	2×10^{-4}	0.25 (0.04)	0.95 (0.04)
BMI	MI	3×10^{-9}	2×10^{-7}	0.34 (0.09)	0.94 (0.11)
Triglycerides	Mean cell volume	10×10^{-19}	8×10^{-5}	-0.20 (0.04)	0.86 (0.11)
Triglycerides	BP - diastolic	5×10^{-39}	2×10^{-5}	0.11 (0.04)	0.86 (0.10)
Mean platelet volume	Platelet count	6×10^{-10}	2×10^{-4}	-0.66 (0.03)	0.84 (0.10)
BMI	Hypertension	2×10^{-16}	7×10^{-15}	0.38 (0.03)	0.83 (0.11)
Triglycerides	Platelet distribution width	5×10^{-17}	3×10^{-5}	0.19 (0.04)	0.81 (0.13)
BMI	FVC	4×10^{-13}	2×10^{-11}	-0.22 (0.03)	0.79 (0.17)
Triglycerides	Reticulocyte count	2×10^{-10}	9×10^{-5}	0.33 (0.05)	0.79 (0.14)
BMI	Reticulocyte count	4×10^{-5}	$< 10^{-20}$	0.39 (0.03)	0.64 (0.25)
Height	Bone mineral density - heel	3×10^{-14}	1×10^{-6}	-0.09 (0.04)	0.50 (0.14)
Age at menarche	Height	7×10^{-11}	6×10^{-12}	0.16 (0.04)	0.43 (0.10)
Height	Basal metabolic rate	10×10^{-9}	8×10^{-5}	0.57 (0.03)	0.39 (0.07)
Eczema	FEV1/FVC	2×10^{-15}	5×10^{-8}	-0.08 (0.03)	0.36 (0.10)
BMI	High cholesterol	2×10^{-6}	1×10^{-8}	0.33 (0.06)	0.13 (0.12)
White cell count	BMI	7×10^{-5}	2×10^{-15}	0.24 (0.03)	0.09 (0.16)
College	High cholesterol	2×10^{-8}	9×10^{-4}	-0.23 (0.03)	0.01 (0.08)

Supplementary Table 15: Pairs of traits identified by either LCV only (top), Bidirectional MR only (middle), or both methods (bottom).

Trait 1	Trait 2	$\hat{\rho}_g$ (std err)	p_{LCV}	$p_{\text{outliers removed}}$
HC	Neuroticism	0.09 (0.03)	2×10^{-14}	0.15
RBCs	Monocytes	0.14 (0.05)	8×10^{-7}	0.004
College	High cholesterol	-0.23 (0.03)	2×10^{-8}	0.18
BMI	T2D	0.61 (0.04)	0.09	9×10^{-6}
High cholesterol	RBCs	0.08 (0.03)	0.002	10×10^{-8}
Asthma	White cells	0.14 (0.04)	0.92	3×10^{-10}
RDW	High cholesterol	0.11 (0.04)	8×10^{-4}	2×10^{-7}
HDL	Age at menopause	0.10 (0.05)	0.83	5×10^{-6}

Supplementary Table 16: Trait pairs with discordant results after outlier removal. Three trait pairs were no longer significant (top; one of these pairs remained nearly significant), and five new trait pairs became significant after we applied the outlier-removal procedure (bottom; two of these pairs were previously nearly significant). We generally do not recommend removing outlier loci, because they may contain valuable information. When outlier removal causes a significant result to become non-significant, this does not imply that failure to remove outliers causes false positives, as the outlier removal procedure (which removes entire jackknife blocks) can result in reduced power, particularly when multiple jackknife blocks are removed.

X	Random variable*
Y_1, Y_2	Random variable
L	Random variable
γ_1, γ_2	Random variable
π	Random variable
α_1, α_2	Random variable
ρ_g	Fixed parameter
gcp	Fixed parameter**
q_1, q_2	Fixed parameter**

Supplementary Table 17: List of variables and fixed parameters under the LCV model. *X can be viewed as fixed if in-sample LD is used for the LD score regression steps (Y and L would still be viewed as random). **For the purpose of estimation, we impose a prior on gcp, and thus also implicitly on q_1, q_2 (conditional on ρ_g); however, this choice is made for convenience and is not integral to the LCV model.

Supplementary Note

Contents

1 Simulations

1.1 Existing Mendelian Randomization methods

Two-sample MR. As described in ref. 5, we ascertained significant SNPs ($p < 5 \times 10^{-8}$, χ^2 test) for the exposure and performed an unweighted regression, with intercept fixed at zero, of the estimated effect sizes on the outcome with the estimated effect sizes on the exposure (in practice, a MAF-weighted and LD-adjusted regression is often used; in our simulations, all SNPs had equal MAF, and there was no LD). To assess the significance of the regression coefficient, we estimated the standard error as $\text{se} = \sqrt{\frac{\frac{1}{K}\sum_{k=1}^{K}\bar{\beta}_{k2}^2}{\sum_{k=1}^{K}\hat{\beta}_{k1}^2}}$, where $\bar{\beta}_{k2}$ is the k^{th} residual, N_2 is the sample size in the outcome cohort, and K is the number of significant SNPs. This estimate of the standard error

allows the residuals to be overdispersed compared with the error that is expected from the GWAS sample size. To obtain p values, we applied a two-tailed t-test to the regression coefficient divided by its standard error, with $K-1$ degrees of freedom.

MR-Egger. As described in ref. 7, we ascertained significant SNPs for the exposure and coded them so that the alternative allele had a positive estimated effect on the exposure. We performed an unweighted regression with a fitted intercept of the estimated effect sizes on the outcome on the estimated effect sizes on the exposure. We assessed the significance of the regression using the same procedure as for two-sample MR, except that the t-test used $K-2$ rather than $K-1$ degrees of freedom.

Bidirectional MR. We implemented bidirectional mendelian randomization in a manner similar to ref. 11. Significant SNPs were ascertained for each trait. If the same SNP was significant for both traits, then it was assigned only to the trait where it ranked higher (if a SNP ranked equally high for both traits, it was excluded from both SNP sets). The Spearman correlations r_1, r_2 between the z scores for each trait was computed on each set of SNPs, and we applied a χ^2_1 test to

$$\chi^2 = \frac{1}{\frac{1}{K_1-3}+\frac{1}{K_2-3}}(\text{atanh}(r_1)-\text{atanh}(r_2))^2, \tag{1}$$

where K_j is the number of significant SNPs for trait j. In ref. 11, the statistics $\text{atanh}(r_j)$ were also used, but a relative likelihood comparing several different models was reported instead of a p-value. We chose to report p-values for Bidirectional MR in order to allow a direct comparison with other methods.

Weighted median. As described in ref. 8, we ascertained significant SNPs for the exposure and computed ratio estimates and weights for each SNP. We computed the weighted median of the ratio estimates and estimated the standard error using a parametric bootstrap (100 bootstrap runs). We assessed significance using a Z test.

Mode based estimator. We ascertained significant SNPs for the exposure and computed ratio estimates for each SNP. We fit a curve to the observed ratio estimates using the Matlab fitdist() function with a bandwidth parameter as recommended in ref. 10, with uniform SNP weights. We verified that the Matlab fitdist() function produces identical curves as the original implementation in R. We computed the mode of the smoothed distribution and estimated its standard error using a parametric bootstrap (100 bootstrap runs). We assessed significance using a Z test.

Application of MR to real data. For our applications of MR and related methods to real data, we selected genetic instruments using a greedy pruning procedure. We ranked all genome-wide significant SNPs for the exposure ($p < 5 \times 10^{-8}$) by χ^2 statistic. Iteratively, we removed all SNPs within 1cM of the first SNP in the list, obtaining a set of independent lead SNPs separated by at least 1cM. We confirmed using an LD reference panel that our 1cM window was sufficient to minimize LD among the set of retained SNPs. We applied each MR method as described above; in particular, we performed unweighted regressions for MR and MR-Egger.

Application of MR to LDL and BMD. We applied two-sample MR (see above) to 8 curated SNPs that were previously used to show that LDL has a causal effect on CAD in ref. 3. 10 SNPs were used in ref. 3, of which summary statistics were available for 8 SNPs: rs646776, rs6511720, rs11206510, rs562338, rs6544713, rs7953249, rs10402271 and rs3846663.

1.2 Simulations involving LCV model violations

In order to investigate potential limitations of our approach, we performed null and causal simulations under genetic architectures that violate LCV model assumptions. As noted above, partial genetic causality is well-defined without making LCV (or other) model assumption (see Online Methods). There are two classes of LCV model violations: *independence violations* and *proportionality violations*. Roughly, independence violations involve a violation of the independence assumption between (1) mediated effects (π) and (2) direct effects (γ) while still satisfying a key proportionality condition related to the mixed fourth moments; as a result, independence violations are not expected to cause LCV to produce false positives (see Online Methods). Proportionality violations, on the other hand, violate this proportionality condition and are potentially more problematic. A representative example of an independence violation is a bivariate Gaussian mixture model where one of the mixture components generates imperfectly correlated effect sizes on the two traits. These SNPs underlying this mixture component can be viewed as having both an effect on L and also a residual effect on the two traits directly, in violation of the independence assumption. First, we performed null simulations under a Gaussian mixture model with a nonzero genetic correlation. These simulations were similar to the simulations reported in Figure 2b, except that the correlated SNP effect sizes (1% of SNPs) were drawn from a bivariate normal distribution with correlation 0.5 (explaining 20% of heritability for each trait; in Figure 2b, these effects were perfectly correlated). Similar to Figure 2b, LCV and bidirectional MR produced p-values that were well-calibrated, while MR and MR-Egger produced inflated p-values (Supplementary Figure 1a, Supplementary Table 4a-d). Second, similar to Figure 2c, we included differential polygenicity between the two traits, finding that differential polygenicity caused all existing methods including bidirectional MR, but not LCV, to produce false positives (Supplementary Figure 1b, Supplementary Table 4f-h). Third, similar to Figure 2d, we included differential power between the two traits, again finding that LCV produced well-calibrated p-values while existing methods produced false positives (Supplementary Figure 1c, Supplementary Table 4i-k).

A representative example of a proportionality violation is a model in which two intermediaries L_1 and L_2 have different effect sizes on the two traits, and L_1 and L_2 also have unequal polygenicity. First, for comparison purposes, we considered a model with two intermediaries with equal polygenicity; 2% of SNPs were causal for each intermediary, and 4% of SNPs were causal for each trait exclusively. Because this model implies only an independence violation (see Online Methods), we expected that LCV would not produce false positives. Indeed, LCV produced well-calibrated p-values (Supplementary Figure 1d, Supplementary Table 5a). Similar to Figure 2b and Supplementary Figure 1a, Bidirectional MR also produced well-calibrated p-values, while MR and MR-Egger produced false positives. Second, we shifted the polygenicity of the two intermediaries in opposite directions: 1% of SNPs were causal for L_1 and 8% of SNPs were causal for L_2, resulting in a proportionality violation. We expected that LCV would produce false positives, as the intermediary with lower polygenicity would disproportionately affect the mixed fourth moments. Indeed, LCV (as well as other methods) produced false positives, indicating that proportionality violations cause LCV to produce false positives (Supplementary Figure 1e, Supplementary Table 5b-d). We investigated the gcp estimates produced by LCV in these simulations, finding that LCV produced low gcp estimates ($\widehat{\text{gcp}} \approx 0.5$; Supplementary Figure 2a). We varied the difference in polygenicity as well as the difference in the relative effect sizes of the two intermediaries, finding that extreme parameter settings (e.g., a 32× difference in polygenicity in conjunction with a 25× difference in the relative effect sizes of L_1 and L_2) were required to cause LCV to produce high gcp estimates

(gcp > 0.6; Supplementary Figure 2a). Thus, proportionality violations of LCV model assumptions can cause LCV (and other methods) to produce false positives, but genetic causality remains the most parsimonious explanation for high gcp estimates.

Finally, we performed (fully) causal simulations under LCV model violations. First, we simulated an independence violation by specifying a Gaussian mixture model where every SNP affecting trait 1 also affected trait 2, but the relative effect sizes were noisy (Supplementary Figure 1f, Supplementary Table 4l-n). Sample size and polygenicity were similar to Figure 3a (4× lower sample size than Supplementary Figure 1a). As expected, LCV had lower power to detect a causal effect than in Figure 3a, although it still had moderately high power. Second, we simulated a proportionality violation by specifying both a causal effect (corresponding to L_1) and an additional genetic confounder (corresponding to L_2) (Supplementary Figure 1g, Supplementary Table 5i-k). LCV had lower power to detect a causal effect than in Figure 3a, although it still had high power. We investigated the gcp estimates produced by LCV in these simulations, finding that they were substantially lower than 1 (Supplementary Table 5i-k and Supplementary Figure 2b). Therefore, gcp estimates lower than 1 should not be viewed as conclusive evidence against a fully causal effect; an alternative explanation is that model violations cause LCV to underestimate the gcp.

In summary, we determined in null simulations that independence violations do not cause LCV to produce false positives; in addition, these simulations recapitulated the limitations of existing methods that we observed in simulations under the LCV model (Figure 2). Proportionality violations caused LCV (as well as existing methods) to produce false positives; however, extreme values of the simulation parameters were required in order for LCV to produce high gcp estimates. In causal simulations, we determined that both independence and proportionality violations lead to reduced power for LCV (and other methods), as well as downwardly biased gcp estimates for LCV.

1.3 Simulations with LD

We performed simulations with LD to assess the robustness of LCV; we note that LD can potentially affect the performance of our method, which uses a modified version of LD score regression[16,19] to normalize effect size estimates and to estimate genetic correlations. LD was computed using $M = 596k$ common SNPs in $N = 145k$ samples of European ancestry from the UK Biobank interim release.[27] Unlike our simulations with no LD, these simulations also included sample overlap. Because existing methods exhibited major limitations in simulations with no LD (Figure 2), we restricted these simulations to the LCV method.

First, we performed null simulations to assess calibration. We chose a set of default parameters similar to Figure 2b and varied each parameter in turn. In particular, similar to Figure 2, these simulations included uncorrelated pleiotropy, genetic correlations, differential polygenicity between the two traits, and differential power between the two traits (Supplementary Table 6a-m). LCV produced approximately well-calibrated or conservative false positive rates. Slight inflation was observed due to noise in our heritability estimates (Supplementary Table 6c-m); proper calibration was restored by using constrained-intercept LD score regression[19] (resulting in more precise heritability estimates) (Supplementary Table 7a-f). To avoid problems with noisy heritability estimates, we restrict our analyses of real traits to data sets with highly significant heritability estimates (Z score for nonzero h^2 = $Z_h > 7$). We also determined that uncorrected population stratification led to false positives (Supplementary Table 8).

Second, we performed causal simulations to assess power. We chose a set of default parameters

similar to our null simulations, finding that LCV was well-powered (Supplementary Table 6t), although its power was lower than in simulations with no LD (Figure 3a). We varied each parameter in turn, finding that power was reduced when we reduced the sample size, increased the polygenicity of the causal trait, reduced the causal effect size, or simulated a partially causal rather than fully causal genetic architecture (Supplementary Table 6u-bb), similar to simulations with no LD (Figure 3b-f). These simulations indicate that LCV is well-powered to detect a causal effect for large GWAS under most realistic parameter settings, although its power does depend on genetic parameters that are difficult to predict.

Third, to assess the unbiasedness of gcp posterior mean (and variance) estimates, we performed simulations in which the true value of gcp was drawn uniformly from $[-1, 1]$ (corresponding to the prior that LCV uses to compute its posterior mean estimates, see Online Methods). We expected posterior-mean estimates to be unbiased in the Bayesian sense that $E(\text{gcp}|\hat{\text{gcp}}) = \hat{\text{gcp}}$ (which differs from the usual definition of unbiasedness, that $E(\hat{\text{gcp}}|\text{gcp}) = \text{gcp}$).[26] Thus, we binned these simulations by $\hat{\text{gcp}}$ and plotted the mean value of gcp within each bin (Supplementary Figure 3). We determined that mean gcp within each bin was concordant with $\hat{\text{gcp}}$. In addition, the root mean squared error was 0.15, approximately consistent with the root mean posterior variance estimate of 0.13 (Supplementary Table 9).

In summary, we confirmed using simulations with LD that LCV produces well-calibrated false positive rates under a wide range of realistic genetic architectures; some p-value inflation was observed when heritability estimates were noisy, but false positives can be avoided in analyses of real traits by restricting to traits with highly significant heritability ($Z_h > 7$). We also confirmed that LCV is well-powered to detect a causal effect under a wide range of realistic genetic architectures, and produces unbiased posterior mean estimates of the gcp.

1.4 Simulation details

In order to simulate summary statistics with no LD, first, we chose causal effect sizes for each SNP on each trait according to the LCV model. For all simulations except for Supplementary Table 4, the causal effect size vector for trait k was

$$\beta_k = \frac{h_k^2}{M}(q_k\pi + \gamma_k), \tag{2}$$

where in all simulations except for Supplementary Table 5, q_k was a scalar, and π and γ_k were $1 \times M$ vectors. In Supplementary Table 5, q_k was a 1×2 vector and π was a $2 \times M$ matrix. Entries of π were drawn from i.i.d. point-normal distribution with mean zero, variance 1, and expected proportion of causal SNPs equal to p_π. Entries of γ_k were drawn from i.i.d. point-normal distributions with expected proportion of causal SNPs equal to p_{γ_k}; we modeled colocalization between non-mediated effects by fixing some expected proportion of SNPs $p_{\gamma_{1,2}} < \min(p_{\gamma_1}, p_{\gamma_2})$ as having nonzero values of both γ_1 and γ_2. Then, we centered and re-scaled the nonzero entries of π and γ_k, so that they had mean 0 and variance 1 and $1 - q_k^2$, respectively.

For simulations in Supplementary Table 4, effect sizes were drawn from a mixture of Normal distributions: there was a point mass at (0,0); a component with $\sigma_1^2 = 0, \sigma_2^2 \neq 0$; a component with $\sigma_1^2 \neq 0, \sigma_2^2 = 0$; and a component with $\sigma_1^2 \neq 0, \sigma_2^2 \neq 0, \sigma_{12} = \sqrt{\sigma_1^2\sigma_2^2}$. Values of $M, N_k, N_{\text{shared}}, \rho_{\text{total}}, p_{\gamma_k}, p_{\gamma_{1,2}}, h_k^2, p_\pi, q_k$ for each simulation can be found in Supplementary Table 1.

Second, we simulated summary statistics as

$$\hat{\beta}_k \sim N(\beta_k, \frac{1}{N_k} I), \tag{3}$$

where β_k is the vector of true causal effect sizes for trait k and N_k is the sample size for trait k. When we ran LCV on these summary statistics, we used constrained-intercept LD score regression rather than variable-intercept LD score regression both to normalize the effect estimates[19] and to estimate the genetic correlation,[16] with LD scores equal to one for every SNP.

In simulations with LD, we first simulated causal effect sizes for each trait in the same manner as simulations with no LD. Then, we obtained summary statistics in one of two ways, either using real genotypes or using real LD only.

For simulations with real genotypes modeling population stratification (Supplementary Table 8), we chose effect sizes for each SNP and each trait from the LCV model with various parameters and multiplied these effect size vectors by real genotype vectors from UK Biobank,[27] adding noise to obtain simulated phenotypes. For computational efficiency, we restricted these genotypes to chromosome 1 (M = 43k). We added stratification directly to the phenotype values along PC1 (computed on 43k SNPs and $N_1 + N_2$ individuals), with effect sizes $\sqrt{0.01}$ and $\sqrt{0.02}$ for trait 1 and trait 2, respectively. We then re-normalized phenotypes to have variance 1; afterwards, ~1% and ~2% of variance were explained by PC1 for each trait respectively. We estimated SNP effect sizes for each trait by correlating each SNP with the phenotypic values in N_k individuals. In corrected simulations (Supplementary Table 8b,d,f), we residualized the PC1 SNP loadings (computed on all $N_1 + N_2$ individuals) from the SNP effect estimates, a procedure which is effectively equivalent to correction of the individual-level data.[25]

For other simulations, we simulated summary statistics without first simulating phenotypic values, using the fact that the sampling distribution of Z-scores is approximately:[23]

$$Z \sim N(\sqrt{N} R\beta, R), \tag{4}$$

where R is the LD matrix and β is the vector of true effect sizes. We estimated R from the N = 145k UK Biobank cohort using plink with an LD window size of 2Mb (M = 596k), which we converted into a block diagonal matrix with 1001 blocks. The number 1001 was chosen instead of the number 1000 so that the boundaries of these blocks would not align with the boundaries of our 100 jackknife blocks; the use of blocks allowed us to avoid diagonalizing a matrix of size 596k, while not significantly changing overall LD patterns (there are ~$50,000$ independent SNPs in the genome, and $1001 << 50,000$). Because the use of a 2Mb window causes the estimated LD matrix to be non-positive semidefinite (even after converting it into a block diagonal matrix), each block was converted into a positive semidefinite matrix by diagonalizing it and removing its negative eigenvalues: that is, we replaced each block $A = V\Sigma V^T$ with the matrix B, where $B = V \max(0, \Sigma) V^T$. Then, because the removal of negative eigenvalues causes B' to have entries slightly different from one, we re-normalized each block: $C = D^{-1/2} B D^{-1/2}$, where D is the diagonal matrix corresponding to the diagonal of B. Even though the diagonal elements of B are close to 1 (mostly between 0.99 and 1.01), this step is important to obtain reliable heritability estimates using LD score regression because otherwise the diagonal elements of the LD matrix will be strongly correlated with the LD scores ($r^2 \approx 0.5$) and the heritability estimates will be upwardly biased, especially at low sample sizes.

We concatenated the blocks $C_1, ..., C_{1001}$ to obtain a positive semi-definite block-diagonal matrix R'. We also computed and concatenated the matrix square root of each block. In order to obtain

samples from a Normal distribution with mean $R'\beta$ and variance $\frac{1}{N}R'$, we multiplied a vector having independent standard normal entries by the matrix square root of R' and added this noise vector to the vector of true marginal effect sizes, $R'\beta$. We computed LD scores directly from R. For simulations with sample overlap, the summary statistics were correlated between the two GWAS: the correlation between the noise term in the estimated effect of SNP i on trait 1 and the estimated effect of SNP j on trait 2 was $R'_{ij}\rho_{\text{total}}N_{\text{shared}}/\sqrt{N_1 N_2}$, which is the amount of correlation that would be expected if the total (genetic plus environmental) correlation between the traits is ρ_{total}.[16]

2 Characterization of LCV model violations

In this section, we define partial genetic causality without making LCV (or other) model assumptions and characterize the type of LCV model violation that causes LCV to produce false positives and bias. There are two classes of LCV model violations: *independence violations* and *proportionality violations*. Roughly, independence violations involve a violation of the independence assumption between mediated effects (π) and direct effects (γ) while still satisfying a key proportionality condition related to the mixed fourth moments; as a result, independence violations are not expected to cause LCV to produce false positives (see Online Methods). Proportionality violations, on the other hand, violate this proportionality condition and are potentially more problematic. In order to make this characterization, it is necessary to define partial genetic causality in a more general setting, without assuming the LCV model. Partial genetic causality is defined in terms of the correlated genetic component of the bivariate SNP effect size distribution, which generalizes the shared genetic component modeled by LCV; unlike the shared genetic component, the correlated genetic component does not have proportional effects on both traits (but merely correlated effects).

2.1 Definition of partial genetic causality without LCV model assumptions

Let $A = (\alpha_1, \alpha_2)$ be the bivariate distribution of marginal effect sizes, normalized to have zero mean and unit variance. First, we define an *even genetic component* of A as a distribution $T = (t_1, t_2)$ that is independent of its complement $A - T$ and that satisfies a mirror symmetry condition:

$$(t_1, t_2) \sim (-t_1, t_2) \sim (t_1, -t_2). \tag{5}$$

Equivalently, the density function of T is an even function of both variables. Note that an even genetic component does not contribute to the genetic correlation. In order to define the "correlated genetic component," we would like to define a maximal even component, i.e. an even component that explains the largest possible amount of heritability for both traits. However, if A follows a Gaussian distribution, then there is no maximal even component: instead, the even genetic component that maximizes the proportion of trait 1 heritability explained fails to maximize the proportion of trait 2 heritability explained. This fact is related to the observation that the LCV model is non-identifiable when the effect size distribution for L follows a Gaussian distribution, and *only* when it follows a Gaussian distribution (see Identifiability). Generalizing this result, we conjecture that there exists an even component that is maximal up to a Gaussian term. More precisely, there exists a maximal even component $T^* = (t_1^*, t_2^*)$ such that for any even component $T = (t_1, t_2)$, there exists a (possibly degenerate) Gaussian random variable $Z = (z_1^*, z_2^*)$ independent of T^* such that $T^* + Z$ is an even component and $E((t_1^* + z_1)^2) \geq E(t_1^2)$ and $E((t_2^* + z_2)^2) \geq E(t_2^2)$.

We define the *correlated genetic component* $S = (s_1, s_2)$ as the complement of the maximal even component and the Gaussian term. Trait 1 is defined as *partially genetically causal* for trait 2 if $E(s_1^2) > E(s_2^2)$, and vice versa. We may also define the genetic causality proportion using main text equation (1), substituting $E(s_k^2)$ for q_k^2. However, the interpretation of the gcp is not as clear in this more general setting. Note that the correlated genetic component may be identically 0, for example if A is bivariate Gaussian or if A itself is an even component; in both cases, there is no partial causality, and the genetic causality proportion is undefined. In practice, if the correlated genetic component is 0 or nearly 0, LCV will produce null p-values and low, noisy gcp estimates.

2.2 Independence violations and proportionality violations

The LCV model assumption is equivalent to the statement that the correlated genetic component resembles a line through the origin (and there is no Gaussian term): $S = (q_1\pi, q_2\pi)$, for some random variable π and fixed parameters q_1, q_2 such that $\rho_g = q_1 q_2$. Under the LCV model we refer to this distribution as the *shared genetic component* because its effects are fully shared (rather than merely correlated) between the two traits. This assumption enables an inference approach based on mixed fourth moments because it implies that the mixed fourth moments of the correlated component are proportional to the respective variances:

$$E(s_1 s_2 s_k^2) \propto E(s_k^2), \tag{6}$$

where under the LCV model, the proportionality constant is $q_1 q_2 E(\pi^4)$. However, the interpretation of the gcp is not as clear in this more general setting; in particular, a gcp of 1 implies that every SNP affecting trait 1 also affects trait 2, but not proportionally. Note that the correlated genetic component may be identically 0, for example if A is bivariate Gaussian or if A itself is an even genetic component; in both cases, there is no partial causality, and the genetic causality proportion is undefined. In practice, if the correlated genetic component is 0 or nearly 0, LCV will produce null p-values and low, noisy gcp estimates.

Intuitively, this type of violation arises as a result of non-independence between mediated effects (π) and direct effects (γ), causing "noise" from the direct effects to be incorporated into the correlated component. For this reason, we call such violations *independence violations*; genetic architectures that violate the proportionality condition we call *proportionality violations*. In the presence of an independence violation, we obtain the following moment condition, generalizing main text equation (2):

$$E(\alpha_1 \alpha_2 \alpha_k^2) = cE(s_k^2) + 3\rho_g \tag{7}$$

where c is a proportionality constant. In particular, if $E(s_1^2) = E(s_2^2)$ (no partial causality), then $E(\alpha_1 \alpha_2^3) = E(\alpha_2 \alpha_1^3)$, and LCV is expected to produce well-calibrated p-values. Conversely, under a proportionality violation, LCV is expected to produce inflated p-values under the null.

3 Discussion of additional trait pairs

3.1 Positive results

We briefly discuss several other trait pairs with significant evidence of partial genetic causality, including novel results and results that have previously been reported (Supplementary Table 11).

- We identified four traits with evidence for a fully or partially genetically causal effect on hypertension (Supplementary Table 11), which is genetically correlated with MI ($\hat{\rho}_g = 0.49(0.10)$). These included genetically causal effects of BMI, consistent with the published literature,[11,38] as well as triglycerides and HDL. The genetically causal effect of HDL indicates that there exist major metabolic pathways affecting hypertension with little or no corresponding effect on MI. The positive partially genetically causal effect of reticulocyte count, which had a low gcp estimate ($\hat{\text{gcp}} = 0.41(0.13)$), is likely related to the substantial genetic correlation of reticulocyte count with triglycerides ($\hat{\rho}_g = 0.33(0.05)$) and BMI ($\hat{\rho}_g = 0.39(0.03)$).

- We detected evidence for a fully or partially genetically causal effect of triglycerides on five cell blood traits: mean cell volume, platelet distribution width, reticulocyte count, eosinophil count and monocyte count (Table 1). These results highlight the pervasive effects of metabolic pathways, which can induce genetic correlations with cardiovascular phenotypes. For example, shared metabolic pathways may explain the high genetic correlation of reticulocyte count with MI ($\hat{\rho}_g = 0.31(0.06)$) and hypertension ($\hat{\rho}_g = 0.27(0.04)$).

- There was evidence for a negative fully or partially genetically causal effect of BMI on FVC, consistent with a longitudinal association between increased BMI and decreased FVC.[69] Similarly, there was evidence for partially genetically causal effects of fasting glucose on FVC and of HDL on FEV1/FVC; these trait pairs had lower gcp estimates and genetic correlations, possibly consistent with mediation of the respective genetic correlations by BMI. There was also evidence for partially causal effects of eczema on FEV1/FVC and of BMI on asthma, with low gcp estimates.

- There was evidence for a negative fully or partially genetically causal effect of balding on number of children in males. Two possible explanations are shared pathways involving androgens[70] and sexual selection against early balding.

- There was evidence for a fully or partially genetically causal effect of HDL and red blood cell distribution width on T2D, with a much higher genetic correlation for HDL ($\rho_g = -0.40(0.06)$). A published MR study provided no strong evidence for an effect of HDL on T2D, despite being well powered to detect a fully genetically causal effect, given the high genetic correlation.[73] It is possible that there is a partially genetically causal effect that MR may have lower power to detect, as it is expected by chance that some trait pairs having gcp estimates of around 0.7 would have the true gcp values below 0.5 (HDL and T2D: $\hat{\text{gcp}} = 0.72(0.17)$). A gcp of ~ 0.5 for these traits would be less surprising, if lipid traits more broadly have a causal effect on T2D.

- There was evidence for a positive fully or partially genetically causal effect of BMI on triglycerides, consistent with results using MR[38] and bidirectional MR.[11] There was also evidence for a positive genetically causal effect of LDL on the self-reported high cholesterol phenotype, consistent with LDL cholesterol representing one component of this compound phenotype.

- There was evidence for fully or partially genetically causal effects of several traits on various platelet phenotypes: large negative effects on platelet count for platelet distribution width and platelet volume, and effects of triglycerides and HDL on platelet distribution width.

- It has been suggested that height has a causal effect on educational attainment.[71] While our results support a partially genetically causal effect, the low gcp estimate ($\hat{\text{gcp}} = 0.33(0.10)$)

suggests shared developmental pathways rather than direct causality, highlighting the benefit of our non dichotomous approach to causal inference. There was a similar result for age at menarche and height, which was previously reported using Bidirectional MR.[11]

3.2 Negative results

Several causal or plausibly causal relationships were not identified by LCV (Supplementary Table 14). We note that non-significant LCV p-values do not constitute evidence against a causal effect. (Confidently low gcp estimates do constitute evidence against a causal effect, but LCV did not produce confidently low gcp estimates for most trait pairs discussed below; Supplementary Table 14). First, LCV did not identify a causal effect of BMI on T2D, due to two outlier loci that do not support a causal effect. After applying an outlier removal procedure to remove these loci (see Online Methods), LCV provides convincing evidence for a fully or partially genetically causal effect ($p = 9 \times 10^{-6}$). Pleiotropic outlier loci can cause LCV to produce false negatives (but not false positives, as our use of a block-jackknife to estimate statistical significance ensures that significant evidence of partial genetic causality will never be based on a single large-effect locus); however, this phenomenon appears to be uncommon (Supplementary Table 16), and we generally do not recommend removing outlier loci because they may contain valuable information. Second, LCV did not identify a causal effect of asthma on pulmonary function (FVC or FEV1/FVC). A possible explanation is diagnosis bias: if individuals with low pulmonary function (for reasons unrelated to asthma) are more likely to be diagnosed with asthma, then this bias would mask the causal effect of asthma on pulmonary function. Third, LCV did not identify a causal effect of smoking status on pulmonary function or MI. A possible explanation is that many SNPs affect smoking status only indirectly, with a primary effect on smoking heaviness or deepness of inhalation.[59] Such SNPs would have much larger effects on cardiopulmonary traits than would be expected based on their effect on smoking status. This type of pleiotropy causes LCV to have lower power (Supplementary Figure 1f). Fourth, LCV did not identify a causal effect of anorexia on BMI. A possible explanation is the high polygenicity of anorexia, as LCV has lower power when the polygenicity of the causal trait is high (Supplementary Table 3e). We note that for most of the trait pairs described above, Bidirectional MR also did not detect a causal effect (Supplementary Table 15).

4 Auxiliary test for partial genetic causality

4.1 Overview of method

Partial genetic causality is well-defined without any type of model assumption (see Online Methods). In order to test for partial genetic causality without using model assumptions, we use an auxiliary test that directly estimates the correlated mixture component of the bivariate distribution of SNP effect sizes and compares the proportion of heritability explained by this correlated component for each trait. This estimate is transformed into an estimate of the distribution of the correlated genetic component using a heuristic, and we compute the difference between the variances of this bivariate distribution. If the estimated correlated component explains a greater proportion of variance (heritability) for trait 1 than for trait 2, it suggests that trait 1 is partially genetically causal for trait 2. We use a block jackknife to determine whether this difference is significantly different from zero.

In order to estimate the bivariate SNP effect size distribution $f(x_1, x_2)$, we utilize a kernel estimator, which can be thought of as a smoothed scatter plot. Each SNP is replaced with a bivariate Gaussian centered at the estimated effect sizes for that SNP:

$$\hat{f}(x_1, x_2) = \frac{1}{Z} \sum_{1}^{M} w_i \phi\left(\frac{x_1 - \hat{\alpha}_{i1}}{\sigma_1}, \frac{x_2 - \hat{\alpha}_{i2}}{\sigma_2}\right), \tag{8}$$

where $\phi(x, y)$ is the bivariate Gaussian density with mean 0 and variance I. The variance of these Gaussians can be thought of as a smoothing parameter, with larger variance corresponding to more smoothing. We choose $\sigma_1^2 = 1/N_2$ and $\sigma_2^2 = 1/N_1$, or substituting the LD score intercept divided by N instead of $1/N$. By using the sampling variance for trait 1 as the smoothing parameter for trait 2 and vice versa, we attempt to reduce the amount of bias that results from unequal sample size between the two traits (however, this approach is not fully successful; see below). These Gaussians are weighted using the same weights as LCV, and we evaluate the density on a grid.

In order to transform the estimated effect size distribution f into an estimate of the distribution of the correlated genetic component, g, we use the following heuristic:

$$g(x_1, y) \approx f(x_1, x_2) + f(-x_1, -x_2) - f(-x_1, x_2) - f(x_1, -x_2). \tag{9}$$

This transformation corresponds to a decomposition of f into two mixture components: g, and an even mixture component. In contrast, the correlated genetic component is defined as an additive component of the effect size distribution; in general, the mixture of two distributions is not the same as their sum. For two independent sparse distributions, however, their sum can be approximated by their mixture, and this fact motivates equation (9).

After obtaining an estimate $\hat{g}$ of the distribution of the correlated genetic component, we compute its variances: $\hat{V}_k = \int x_k^2 \hat{g}(x_1, x_2) dx_1 dx_2$. We compute the significance of the statistic $V_1 - V_2$ using a block jackknife with 100 blocks to estimate its standard error, together with a single-tailed Z test.

4.2 Performance in Simulations

We evaluated the auxiliary test in simulations without LD (Tables 2-5). We found that it has significant limitations. First, it produces false positives in null simulations with unequal power between the two traits (Supplementary Table 2k-m), and also to a lesser extent in simulations with unequal polygenicity between the traits (Supplementary Table 2h-j). Second, while it generally had power comparable to LCV, in some simulations it had substantially lower power (Supplementary Table 3b-c,n-o); these simulations appear to understate the difference in power between the auxiliary test and LCV, as the auxiliary test had much lower power when applied to real data.

We evaluated the auxiliary test in challenging simulations involving multiple intermediaries with unequal polygenicity; these simulations, which constitute proportionality violations, caused LCV to produce false positives (see Simulations with no LD: LCV model violations). While the auxiliary test did have an inflated false positive rate, it was less inflated than LCV (Supplementary Table 5b-d). When we increased the sample size from N = 100k to N = 500k (many of our real datasets have N = 460k), we found that the auxiliary test was far less likely to produce false positives, while LCV was no less likely.

Given the limitations of the auxiliary test, we strongly recommend against using it as a standalone test for partial causality, and we also do not recommend using it as a sensitivity analysis

for LCV. However, it can be used to provide some aggregate replication of LCV results, as at large sample size it produces false positives under orthogonal conditions as LCV does. Results of the auxiliary test on trait pairs that were significant using LCV are listed in Supplementary Table 11, and results on all 429 genetically correlated trait pairs are provided in Supplementary Table 12.

4.3 Application to real data

We applied the auxiliary test to the 30 trait pairs with high gcp estimates, finding that the estimated direction of effect was concordant with LCV for 30/30 trait pairs (Supplementary Table 11). While the auxiliary test replicated the LCV result at a nominal significance level (single-tailed $p < 0.05$) for only 17/30 trait pairs, the fraction 17/30 is expected to be an underestimate of the true positive rate, due to limited power. Indeed, when we applied the auxiliary test to the remaining 394 trait pairs, it produced positive results at the corresponding significance level (two-tailed $p < 0.10$) for only 41/394 trait pairs (39 expected under the null; includes 7/29 trait pairs that LCV reported as significant with $\hat{\text{gcp}} < 0.6$; Supplementary Table 12). This analysis confirms that the 30 trait pairs reported in Table 1 are extremely unlikely to be false positives.

5 Limitations

In addition to the two limitations listed in the Discussion section, this study has several other limitations. First, LCV can be susceptible to false negatives due to outlier loci, bias in disease diagnosis, strong pleiotropic effects, or a highly polygenic causal trait (Supplementary Table 14). However, LCV is well-powered to detect a causal effect in most simulations (Figure 3), and it detects many established causal relationships among real traits with very high statistical significance (Table 1). Second, LCV is not currently applicable to traits with small sample size and/or heritability, due to low power as well as incorrect calibration. However, GWAS summary statistics at large sample sizes have become publicly available for increasing numbers of diseases and traits, including UK Biobank traits.[29] Third, the LCV model can be confounded by shared population stratification, so it is critical for association statistics to be corrected for stratification. Fourth, while many trait pairs have high gcp estimates ($\hat{\text{gcp}} > 0.6$), it is not clear whether most of these trait pairs reflect fully or partially genetically causal relationships. A gcp of 1 and a gcp of ~0.6 would be interpreted differently, as a gcp of ~0.6 suggests that only some interventions on trait 1 will modify trait 2, depending on their mechanism of action. This type of uncertainty can be reduced at higher sample size, but not eliminated entirely. Fifth, even full genetic causality must be interpreted with caution before designing disease interventions, as interventions may fail to mimic genetic perturbations. For example, factors affecting a developmental phenotype such as height might need to be modified at the correct developmental time point in order to have any effect; this limitation broadly applies to all methods for inferring causality using genetic data. Sixth, LCV does not model LD explicitly (unlike cross-trait LD score regression[16]), and consequently it models the marginal, rather than the causal, effect size distribution. Modeling the causal effect size distribution while explicitly accounting for LD would enable LCV analyses to be conditioned on various functional annotations, enabling models involving different shared genetic components such as SNPs linked to gene regulation in different cell types. Seventh, power might also be increased by including rare and low-frequency variants; even though these SNPs explain less complex trait heritability than common SNPs,[20,61] they may contribute significantly to power if the genetic architecture among these SNPs is more

sparse than among common SNPs. Eighth, we cannot infer whether inferred causal effects are linear. For example, it is plausible that BMI would have a small effect on MI risk for low-BMI individuals and a large effect for high-BMI individuals, but this type of nonlinearity cannot be gleaned from summary statistics (unless MI summary statistics were stratified by BMI). Ninth, MR-style analyses have been applied to gene expression,[62–64] and the potential for confounding due to pleiotropy in these studies could possibly motivate the use of LCV in this setting, but LCV is not applicable to molecular traits, which may be insufficiently polygenic for the LCV random-effects model to be well-powered. Finally, we have not exhaustively benchmarked LCV against every published MR method, but have restricted our simulations to the most widely used MR methods.[5,7–11] We note that there exist additional methods that aim to improve robustness by excluding or effectively down-weighting variants whose causal effect estimates appear to be outliers,[6,12] conceptually similar to the weighted median[8] and mode-based estimator;[10] however, we believe that any method that relies on genome-wide significant SNPs for a single one trait is likely to be confounded by genetic correlations (Figure 2). We further note that MR should ideally be applied to carefully curated sets of genetic variants that aim to exclude pleiotropic effects (MR with curation), but that curated sets of genetic variants are unavailable for most complex traits; in particular, it is difficult to compare LCV to MR with curation, as the performance of MR with curation will strongly depend on the quality of information used for curation, which can vary in practice.

6 Identifiability

We ask when q^2 is identifiable: under what conditions is there only one value (q_1^2, q_2^2) that produces the joint distribution A of (α_1, α_2), for any choice of the distribution B of $(\pi, \gamma_1, \gamma_2)$? It is possible that q_1^2 and q_2^2 are not identifiable: for example, if A is multivariate Normal, then the relationship between α_1 and α_2 is fully parameterized by their correlation, and there is no asymmetry that can be exploited in order to separate q_1^2 from q_2^2. In main text equation (2), $\kappa = 0$ and no information is gleaned from the mixed fourth moments.

Interestingly, the Gaussian case is the only non-identifiable case. The following proposition asserts that the LCV model is identifiable under an independence assumption if and only if π does not follow a normal distribution. It does not matter what the marginal distributions of γ_1 and γ_2 are. This result echoes similar results in Independent Components Analysis,[68] which separates independent, additive signals exploiting non-Gaussianity. We note that there exist identifiable cases under which our method will not be able to estimate q^2: our moments-based estimator makes assumptions about the joint distribution of $(\pi, \gamma_1, \gamma_2)$ that are weaker than independence, and as a result, our estimator requires a slightly stronger identifiability assumption than non-Gaussianity, namely that $E(\pi^4) - 3 \neq 0$. If π does follow a Gaussian distribution, then LCV will have no power to estimate gcp or to identify a causal effect, but it will not lead to false positives or to confident false negatives (see Supplementary Table 6bb).

Proposition 1. *Assume that γ_1, γ_2, π are independently distributed, with joint distribution B. Let $A(B, q)$ be the joint distribution on α for some choice of q. Then q is uniquely determined, up to sign flipping, by A if and only if the marginal distribution of π is non-Gaussian.*

Proof. The characteristic functions for B and A are:

$$\begin{aligned}\phi_B(s_1, s_2, s_3) =& E(\exp(i(s_1\gamma_1 + s_2\gamma_2 + s_3\pi))),\\ \phi_A(s_1, s_2) =& E(\exp(i(s_1\beta_1 + s_2\beta_2))).\end{aligned}$$

Because $\alpha_k = q_k\pi + \gamma_k$,

$$\phi_A(s_1, s_2) = \phi_B(s_1, s_2, q_1s_1 + q_2s_2).$$

By the independence assumption, ϕ_B factors:

$$\phi_D(s_1, s_2, q_1s_1 + q_2s_2) = a_1(s_1)a_2(s_2)b(q_1s_1 + q_2s_2).$$

Now, suppose that there is some other q_1', q_2' and some $\phi_{B'}$ (which also factors) such that:

$$\phi_D(s_1, s_2, q_1s_1 + q_2s_2) = \phi_{B'}(s_1, s_2, q_1's_1 + q_2's_2).$$

Without loss of generality, $q_1' = rq_1$ and $q_2' = q_2/r$, since q_1q_2 is the genetic correlation. Factoring $\phi_{B'}$, there exists b' such that

$$\forall s_1,\, s_2,\, b(q_1s_1 + q_2s_2) \propto b'(q_1s_1r + q_2s_2/r),$$

where $\propto$ hides factors of the form $a(s_1)$ and $a(s_2)$. Now, either $r = \pm 1$, or for some imaginary scalar z,

$$b(q_1s_1 + q_2s_2) \propto \exp(z(q_1s_1 + q_2s_2)^2) \propto \exp(zq_1s_1q_2s_2) \propto \exp(z(q_1s_1r + q_2s_2/r)^2).$$

(z must be imaginary in order to have a valid characteristic function). This is precisely the form of the Normal characteristic function:

$$\phi_{N(\mu,\sigma^2)}(s) = \exp(i\mu s)\exp(i(\sigma s)^2/2)$$

so π must be Normally distributed. □

References

[1] Davey Smith, George, and Shah Ebrahim. "Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease?" International journal of epidemiology 32.1 (2003): 1-22.

[2] Davey Smith, George, and Gibran Hemani. "Mendelian randomization: genetic anchors for causal inference in epidemiological studies." Human molecular genetics 23.R1 (2014): R89-R98.

[3] Voight, Benjamin F., et al. "Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study." The Lancet 380.9841 (2012): 572-580.

[4] Do, Ron, et al. "Common variants associated with plasma triglycerides and risk for coronary artery disease." Nature genetics 45.11 (2013): 1345-1352.

[5] Burgess, Stephen, Adam Butterworth, and Simon G. Thompson. "Mendelian randomization analysis with multiple genetic variants using summarized data." Genetic epidemiology 37.7 (2013): 658-635.

[6] Kang, Hyunseung, et al. "Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization." Journal of the American Statistical Association 111.513 (2016): 132-144.

[7] Bowden, Jack, George Davey Smith, and Stephen Burgess. "Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression." International journal of epidemiology 44.2 (2015): 512-525.

[8] Bowden, Jack, et al. "Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator." Genetic epidemiology 40.4 (2016): 304-314.

[9] Hemani, Gibran, et al. "MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations." BioRxiv (2016): 078972.

[10] Hartwig, Fernando Pires, George Davey Smith, and Jack Bowden. "Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption." International journal of epidemiology 46.6 (2017): 1985-1998.

[11] Pickrell, Joseph K., et al. "Detection and interpretation of shared genetic influences on 42 human traits." Nature genetics 48.7 (2016): 709.

[12] Verbanck, Marie, et al. "Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases." Nature genetics 50.5 (2018): 693.

[13] Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. "Sequence variations in PCSK9, low LDL, and protection against coronary heart disease." New England Journal of Medicine 354 (2006): 1264-72.

[14] Paaby, Annalise B., and Matthew V. Rockman. "The many faces of pleiotropy." Trends in Genetics 29.2 (2013): 63-73.

[15] VanderWeele, Tyler J., et al. "Methodological challenges in Mendelian randomization." Epidemiology 25.3 (2014): 427.

[16] Bulik-Sullivan, Brendan, et al. "An atlas of genetic correlations across human diseases and traits." Nature genetics 47.11 (2015): 1236-1241.

[17] Welsh, Paul, et al. "Unraveling the directional link between adiposity and inflammation: a bidirectional Mendelian randomization approach." The Journal of Clinical Endocrinology & Metabolism 95.1 (2010): 93-99.

[18] Vimaleswaran, Karani S., et al. "Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts." PLoS Med 10.2 (2013): e1001383.

[19] Bulik-Sullivan, Brendan K., et al. "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies." Nature genetics 47.3 (2015): 291-295.

[20] Yang, Jian, et al. "Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index." Nature genetics 47.10 (2015): 1114.

[21] Kolesar, Michal, et al. "Identification and inference with many invalid instruments." Journal of Business & Economic Statistics 33.4 (2015): 474-484.

[22] Burgess, Stephen, and Simon G. Thompson. "Interpreting findings from Mendelian randomization using the MR-Egger method." European Journal of Epidemiology (2017): 1-13.

[23] Conneely, Karen N., and Michael Boehnke. "So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests." The American Journal of Human Genetics 81.6 (2007): 1158-1168.

[24] Galinsky, Kevin J., et al. "Population structure of UK Biobank and ancient Eurasians reveals adaptation at genes influencing blood pressure." The American Journal of Human Genetics 99.5 (2016): 1130-1139.

[25] Bhatia, Gaurav, et al. "Correcting subtle stratification in summary association statistics." bioRxiv (2016): 076133.

[26] Goddard, Michael E., et al. "Estimating effects and making predictions from genome-wide marker data." Statistical Science 24.4 (2009): 517-529.

[27] Sudlow, Cathie, et al. "UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age." PLoS medicine 12.3 (2015): e1001779.

[28] Bycroft, Clare, et al. "Genome-wide genetic data on 500,000 UK Biobank participants." bioRxiv (2017): 163298.

[29] Loh, Po-Ru, et al. "Mixed model association for biobank-scale data sets." bioRxiv (2017): 194944.

[30] Holmes, Michael V., Mika Ala-Korpela, and George Davey Smith. "Mendelian randomization in cardiometabolic disease: challenges in evaluating causality." Nature Reviews Cardiology (2017): 577-590.

[31] Smith, George Davey, et al. "The association between BMI and mortality using offspring BMI as an indicator of own BMI: large intergenerational mortality study." Bmj 339 (2009): b5043.

[32] Nordestgaard, Brge G., et al. "The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach." PLoS Med 9.5 (2012): e1001212.

[33] Hgg, Sara, et al. "Adiposity as a cause of cardiovascular disease: a Mendelian randomization study." International journal of epidemiology 44.2 (2015): 578-586.

[34] Holmes, Michael V., et al. "Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis." The American Journal of Human Genetics 94.2 (2014): 198-208.

[35] Cole, Stephen R., et al. "Illustrating bias due to conditioning on a collider." International journal of epidemiology 39.2 (2009): 417-420.

[36] Aschard, Hugues, et al. "Adjusting for heritable covariates can bias effect estimates in genome-wide association studies." The American Journal of Human Genetics 96.2 (2015): 329-339.

[37] Ross, Stephanie, et al. "Mendelian randomization analysis supports the causal role of dysglycaemia and diabetes in the risk of coronary artery disease." European heart journal 36.23 (2015): 1454-1462.

[38] Lyall, Donald M., et al. "Association of body mass index with cardiometabolic disease in the UK Biobank: a Mendelian randomization study." JAMA cardiology 2.8 (2017): 882-889.

[39] Schunkert, Heribert, et al. "Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease." Nature genetics 43.4 (2011): 333-338.

[40] Klein, Irwin, and Kaie Ojamaa. "Thyroid hormone and the cardiovascular system." New England Journal of Medicine 344.7 (2001): 501-509.

[41] Grais, Ira Martin, and James R. Sowers. "Thyroid and the heart." The American journal of medicine 127.8 (2014): 691-698.

[42] Zhao, Jie V., and C. Mary Schooling. "Thyroid function and ischemic heart disease: a Mendelian randomization study." Scientific reports 7:8515 (2017): 8515.

[43] Monzani, F. et al. "Effect of levothyroxine on cardiac function and structure in subclinical hypothyroidism: a double blind, placebo-controlled study." J. Clin. Endocrinol. Metab. 86 (2001): 1110-1115.

[44] Meier, C. et al. "TSH-controlled L-thyroxine therapy reduces cholesterol levels and clinical symptoms in subclinical hypothyroidism: a double blind, placebo-controlled trial (Basel Thyroid Study)." J. Clin. Endocrinol. Metab. 86 (2001): 4430-4863.

[45] Monzani, F. et al. "Effect of levothyroxine replacement on lipid profile and intima-media thickness in subclinical hypothyroidism: a double-blind, placebo- controlled study." J. Clin. Endocrinol. Metab. 89 (2004): 2099-2106.

[46] Razvi, S. et al. "The beneficial effect of L-thyroxine on cardiovascular risk factors, endothelial function, and quality of life in subclinical hypothyroidism: randomized, crossover trial." J. Clin. Endocrinol. Metab. 92 (2007): 1715-1723.

[47] Nagasaki, T. et al. "Decrease of brachial-ankle pulse wave velocity in female subclinical hypothyroid patients during normalization of thyroid function: a double-blind, placebo-controlled study." Eur. J. Endocrinol. 160 (2009): 409-415.

[48] Chaker, Layal, et al. "Thyroid function and risk of type 2 diabetes: a population-based prospective cohort study." BMC medicine 14.1 (2016): 150.

[49] Brenta, Gabriela, et al. "Acute thyroid hormone withdrawal in athyreotic patients results in a state of insulin resistance." Thyroid 19.6 (2009): 665-669.

[50] Wang, Zongze, et al. "Effects of Statins on Bone Mineral Density and Fracture Risk: A PRISMA-compliant Systematic Review and Meta-Analysis." Medicine 95.22 (2016): e3042.

[51] Yerges, Laura M., et al. "Decreased bone mineral density in subjects carrying familial defective apolipoprotein B-100." The Journal of Clinical Endocrinology & Metabolism 98.12 (2013): E1999-E2005.

[52] Sanjak, Jaleal S., et al. "Evidence of directional and stabilizing selection in contemporary humans." Proceedings of the National Academy of Sciences (2017): 201707227.

[53] Price, George R. "Selection and covariance." Nature 227 (1970): 520-521.

[54] Clarke, T. K., et al. "Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population." Molecular psychiatry 21.3 (2016): 419-425.

[55] Keller, Matthew C., and Geoffrey Miller. "Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best?" Behavioral and Brain Sciences 29.4 (2006): 385-404.

[56] Mullins, Niamh, et al. "Reproductive fitness and genetic risk of psychiatric disorders in the general population." Nature communications 8 (2017): 15833.

[57] Davies, Gail, et al. "Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112,151)." Molecular psychiatry 21.6 (2016): 758.

[58] UK10K Consortium. "The UK10K project identifies rare variants in health and disease." Nature 526.7571 (2015): 82.

[59] Ware, Jennifer J., et al. "Genome-wide meta-analysis of cotinine levels in cigarette smokers identifies locus at 4q13. 2." Scientific reports 6 (2016): 20092.

[60] Burgess, Stephen, et al. "Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways." International journal of epidemiology 44.2 (2014): 484-495.

[61] Schoech, Armin, et al. "Quantification of frequency-dependent genetic architectures and action of negative selection in 25 UK Biobank traits." bioRxiv (2017): 188086.

[62] Gamazon, Eric R., et al. "A gene-based association method for mapping traits using reference transcriptome data." Nature genetics 47.9 (2015): 1091-1098.

[63] Gusev, Alexander, et al. "Integrative approaches for large-scale transcriptome-wide association studies." Nature genetics 48 (2016): 245-252.

[64] Zhu, Zhihong, et al. "Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets." Nature genetics 48 (2016):481:487.

[65] The GTEx consortium, et al. "Genetic effects on gene expression across human tissues." Nature 550.7675 (2017): 204.

[66] Mokry, Lauren E., et al. "Vitamin D and risk of multiple sclerosis: a Mendelian randomization study." PLoS medicine 12.8 (2015): e1001866.

[67] Child, Dennis. "The essentials of factor analysis." A&C Black (2006).

[68] Comon, Pierre. "Independent component analysis, a new concept?" Signal processing 36.3 (1994): 287-314.

[69] Thyagarajan, Bharat, et al. "Longitudinal association of body mass index with lung function: the CARDIA study." Respiratory research 9.1 (2008): 31.

[70] Ellis, Justine A., Margaret Stebbing, and Stephen B. Harrap. "Polymorphism of the androgen receptor gene is associated with male pattern baldness." Journal of investigative dermatology 116.3 (2001): 452-455.

[71] Tyrrell, Jessica, et al. "Height, body mass index, and socioeconomic status: mendelian randomization study in UK Biobank." BMJ 352 (2016): i582.

[72] Skaaby, Tea, et al. "Estimating the causal effect of body mass index on hay fever, asthma, and lung function using Mendelian randomization." Allergy (2017).

[73] Haase, Christiane L., et al. "High-density lipoprotein cholesterol and risk of type 2 diabetes: a Mendelian randomization study." Diabetes (2015): db141603.