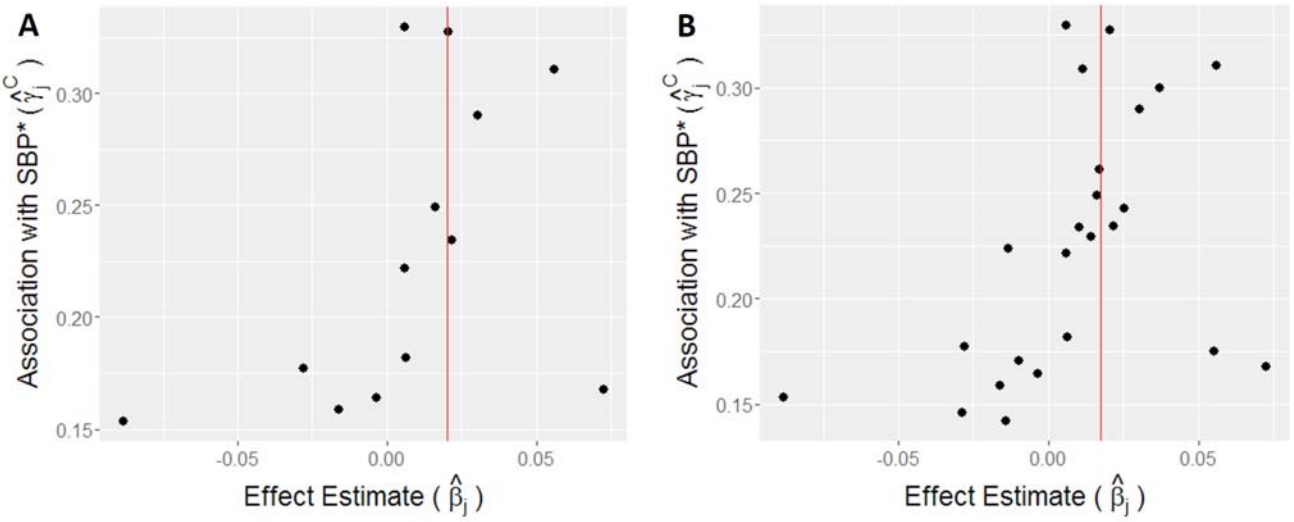
SUPPLEMENTARY DATA

Systolic blood pressure and risk of type 2 diabetes: a Mendelian Randomization study

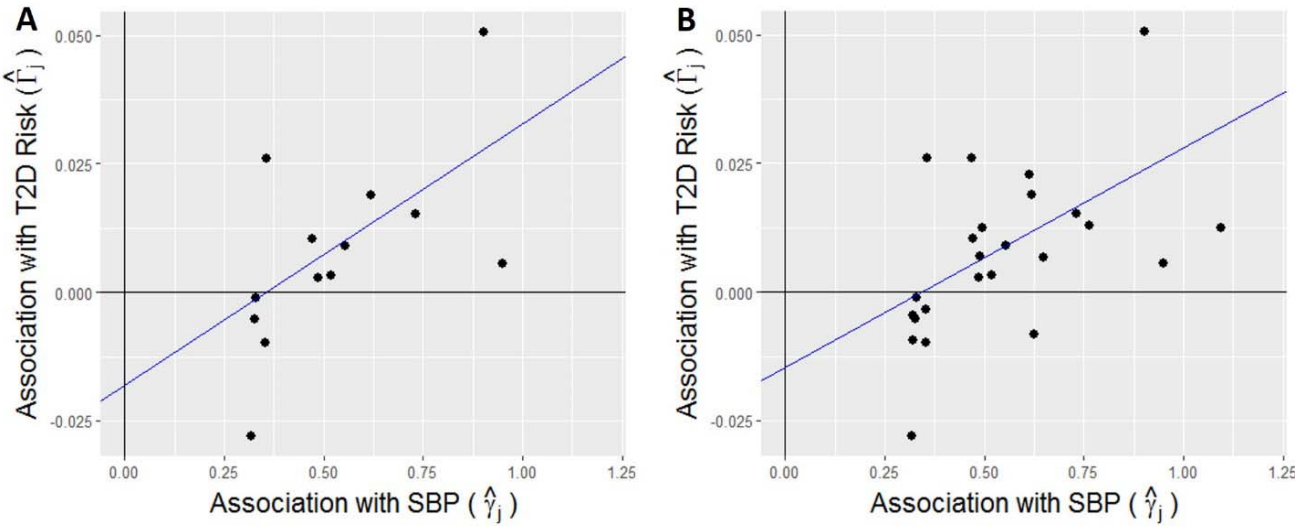
Rachael C. Aikens, Wei Zhao, Danish Saleheen, Muredach P. Reilly, Stephen E. Epstein, Emmi Tikkanen, Veikko Salomaa, and Benjamin F. Voight

Supplementary Figure 1. Funnel plots of all SNPs from (A) the conservative and (B) the expanded instrument set. SBP associations have been corrected by effect allele frequency as described by Bowden et al. previously.1 Red vertical line denotes combined GRS effect estimate from all SNPs.



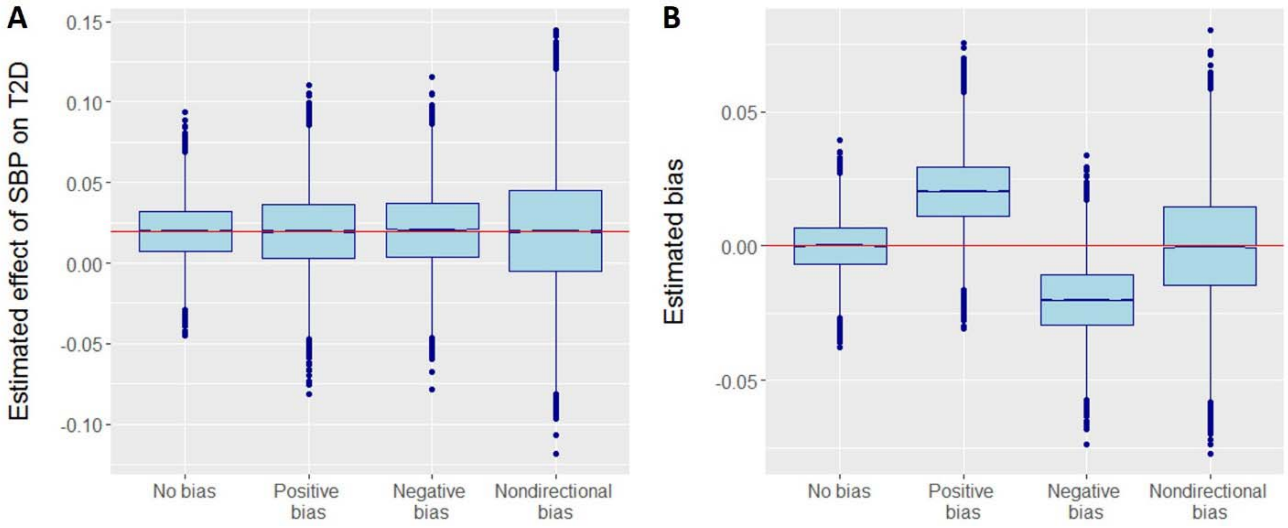
SUPPLEMENTARY DATA

Supplementary Figure 2. Egger regression plots for the conservative and expanded genetic instruments. The slope of the regression line is the egger regression estimate for the effect of SBP on T2D risk (in log odds per mmHg). The y-intercept of the regression an estimate of the level and direction of bias present in the typical GRS or inverse-variance weighted estimate due to pleiotropy. A negative y-intercept in this case indicates that any bias present in this analysis will result in underestimation, rather than overestimation, of the causal effect size.



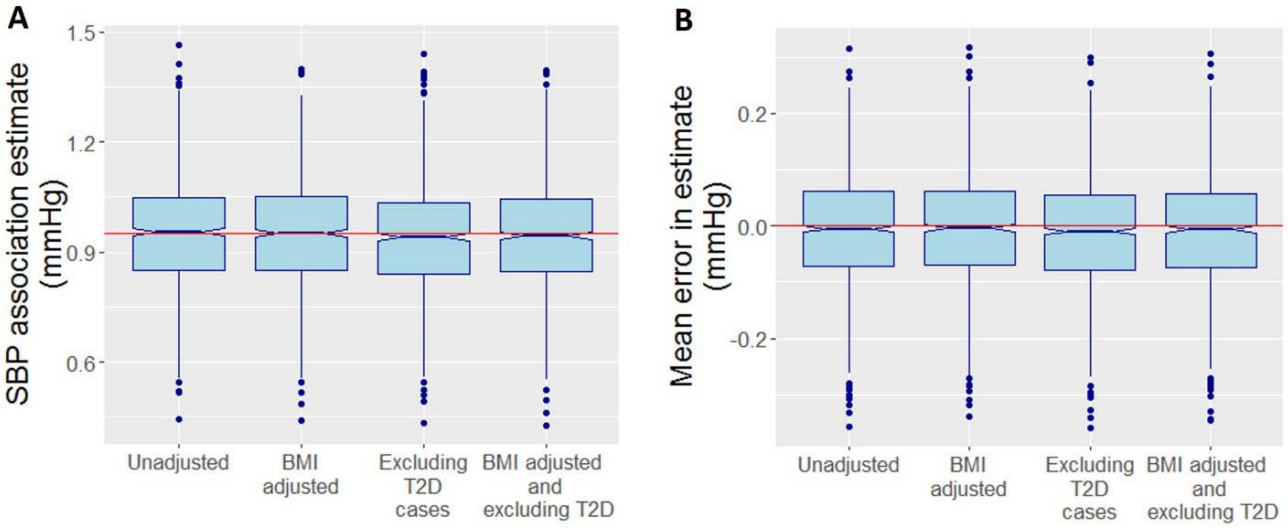
SUPPLEMENTARY DATA

Supplementary Figure 3. Egger Regression results from n=10000 simulated datasets of SNP association data based on our conservative instrument. All figures are in log-odds per millimeter of mercury. (A) Distribution of Egger Regression effect estimates. Horizontal line marks the true effect of SBP on T2D as set in the simulation. (B) Distributions of Egger Regression bias estimates. Horizontal line marks zero bias.



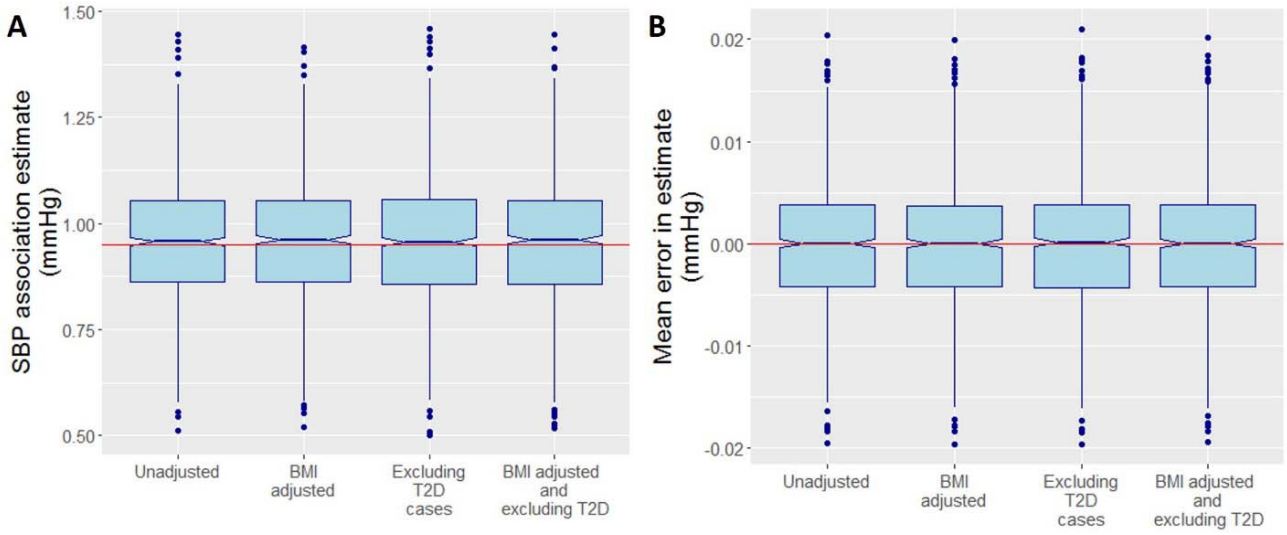
SUPPLEMENTARY DATA

Supplementary Figure 4. Simulation analysis for sample ascertainment and analysis conditions from SBP genome-wide association studies. Results from n=1000 simulated GWAS of 150,000 individuals after adjusting for BMI, excluding diabetes cases, or both, under a model in which both BMI and SBP effect type 2 diabetes risk. (A) SBP association estimates for a representative SNP in our conservative instrument (rs6015450). The red horizontal line denotes the true effect size. (B) Mean error in effect estimates over n=13 SNPs used in conservative instrument.



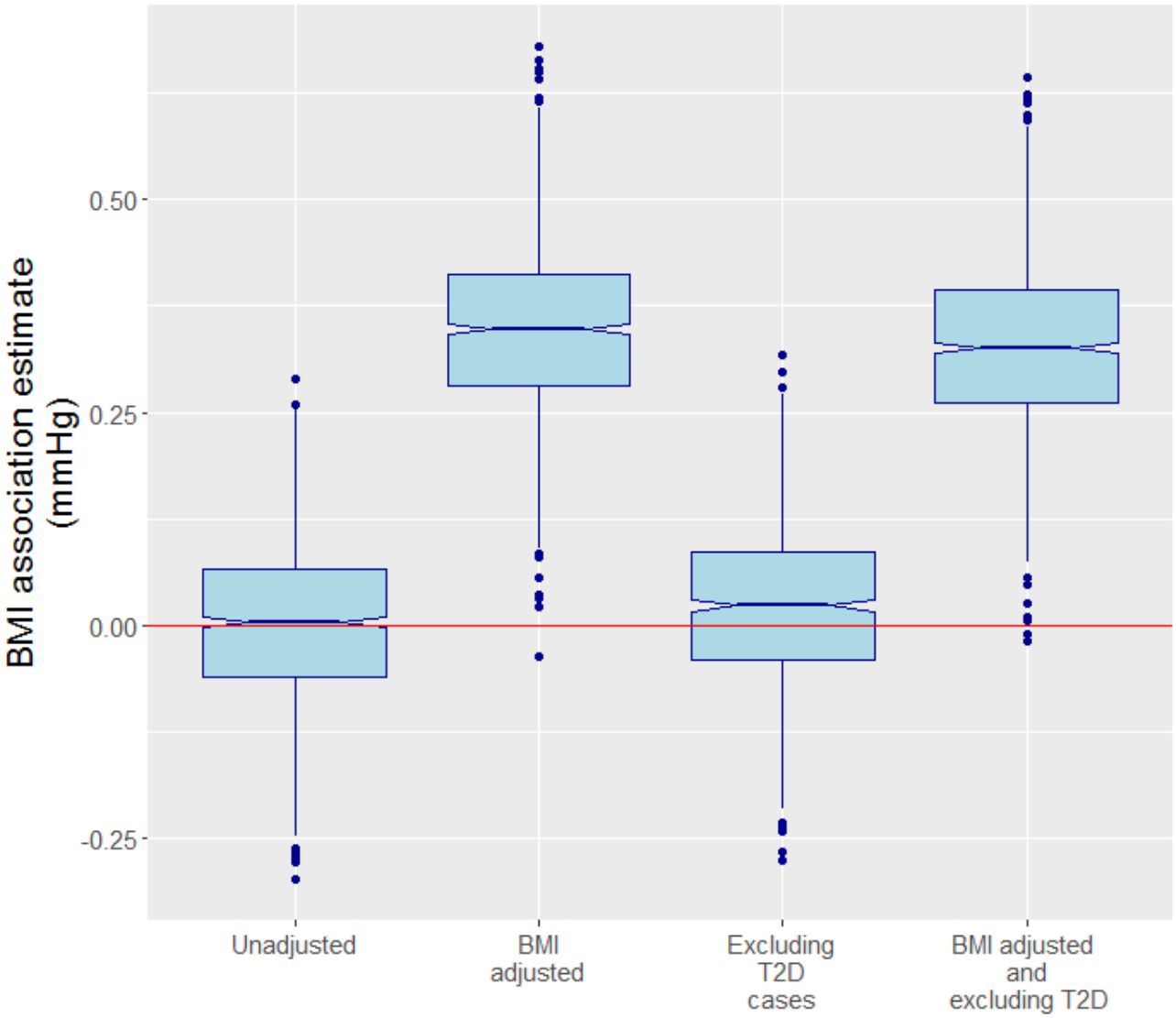
SUPPLEMENTARY DATA

Supplementary Figure 5. Simulation results from n=1000 simulated GWAS generated as in Supplementary figure 4, but under a model in which only BMI effects T2D risk and SBP does not. (A) SBP association estimates for a representative SNP in our conservative instrument (rs6015450). The red horizontal line denotes the true effect size. (B) Mean error in effect estimates over n=13 SNPs used in conservative instrument.



SUPPLEMENTARY DATA

Supplementary Figure 6. Positive control experiment example of confounding due to Collider bias for BMI, using a representative SNP exclusively associated with SBP. Data shown are effect estimates (in standard deviations of BMI) from n=1000 simulated datasets of 150,000 individuals, based on a model in which both SBP and BMI have an effect on T2D risk.



SUPPLEMENTARY DATA

Supplementary Table 1. Summary of SNP Sets

SNPID	EXPANDED-SET	CONSERVATIVE-SET
rs932764	1	1
rs805303	1	1
rs7129220	1	1
rs633185	1	1
rs6015450	1	1
rs4373814	1	1
rs419076	1	1
rs381815	1	1
rs2521501	1	1
rs17367504	1	1
rs1458038	1	1
rs1327235	1	1
rs11953630	1	1
rs4590817	1	0
rs2932538	1	0
rs2014912	1	0
rs1813353	1	0
rs1799945	1	0
rs17608766	1	0
rs17249754	1	0
rs1563788	1	0
rs1378942	1	0
rs13359291	1	0
rs12946454	1	0
rs12940887	1	0
rs1173771	1	0
rs11191548	1	0
rs10850411	1	0

SUPPLEMENTARY DATA

Supplementary Table 2A. Summary for Genotype Risk Score analysis

		Cumulative GRS Calculations							Unweighted GRS Calculations							Weighted median estimator			
	Number of SNPs	OR	á, effect estimate	SE	95% CI: lower bound (normal)	95% CI: upper bound (normal)	Chi squared statistic	p (Chi squared)	OR	á, effect estimate	SE	95% CI: lower bound (normal)	95% CI: upper bound (normal)	Chi squared statistic	p (Chi squared)	á	SE	95% CI: lower bound (bootstrap)	95% CI: upper bound (bootstrap)
Expanded	28	1.018	0.0181	0.00461	0.0091	0.0271	15.32	9.05E-05	1.0078	0.00780	0.00253	0.00284	0.01276	9.52	2.04E-03	0.0175	0.0067	0.0060	0.0325
Conservative	13	1.021	0.0203	0.00639	0.0078	0.0328	10.10	1.48E-03	1.0084	0.00835	0.00353	0.00143	0.01527	5.59	1.80E-02	0.0193	0.0090	0.0031	0.0386
Expanded-excluding-rs2521501	27	1.018	0.0174	0.00474	0.0081	0.0267	13.45	2.45E-04	1.0074	0.00733	0.00258	0.00227	0.01239	8.06	4.52E-03	0.0165	0.0068	0.0048	0.0315
Conservative-excluding-rs2521501	12	1.019	0.0192	0.00674	0.0060	0.0324	8.12	4.39E-03	1.0075	0.00744	0.00368	0.00023	0.01465	4.09	4.32E-02	0.0175	0.0096	0.0005	0.0384

SUPPLEMENTARY DATA

Supplementary Table 2B. Summary of results heterogeneity analysis

		heterogeneity test		GRS calculations by SNP			
	Number of SNPs	Cochran's Q	p (heterogeneity test)	SNPs reporting positive $\hat{\alpha}$	p (two-sided binomial test)	Reporting significant $\hat{\alpha}$	p (two-sided binomial test)
Expanded	28	24.7	0.59	19 of 28	0.087	3 of 28	0.16
Conservative	13	15.78	0.202	9 of 13	0.27	2 of 13	0.14
Expanded-excluding-rs2521501	27	24.3	0.56	18 of 27	0.12	3 of 27	0.15
Conservative-excluding-rs2521501	12	15.5	0.16	8 of 12	0.39	2 of 12	0.12

SUPPLEMENTARY DATA

Supplementary Table 2C. Summary of results for Regression-based causal inference analysis

		Inverse Variance Weighted Regression								Egger Regression											
	Number of SNPs	á, effect estimate	SE	95% CI: lower bound (bootstrap)	95% CI: upper bound (bootstrap)	95% CI: lower bound (student's t)	95% CI: upper bound (student's t)	t statistic	p (Student's t)	OR	á, effect estimate	SE	95% CI: lower bound (bootstrap)	95% CI: upper bound (bootstrap)	95% CI: lower bound (student's t)	95% CI: upper bound (student's t)	t statistic	p (Student's t)	Bias estimate	SE	t statistic	p (students t)
Expanded	28	0.0181	0.0046	0.0084	0.0269	0.0086	0.0275	3.914	5.54E-04	1.043	0.0422	0.0139	0.0059	0.0623	0.0136	0.0708	3.032	5.45E-03	-0.0140	0.0076	-1.839	0.077
Conservative	13	0.0203	0.0064	0.0073	0.0323	0.0064	0.0342	3.179	7.94E-03	1.052	0.0508	0.0184	0.0041	0.0827	0.0104	0.0912	2.765	1.84E-02	-0.0180	0.0102	-1.770	0.104
Expanded-excluding-rs2521501	27	0.0174	0.0047	0.0077	0.0263	0.0076	0.0271	3.668	1.10E-03	1.042	0.0413	0.0140	0.0060	0.0621	0.0124	0.0702	2.947	6.86E-03	-0.0139	0.0076	-1.814	0.082
Conservative-excluding-rs2521501	12	0.0192	0.0067	0.0055	0.0324	0.0044	0.0340	2.849	1.58E-02	1.051	0.0495	0.0186	0.0028	0.0829	0.0081	0.0909	2.662	2.38E-02	-0.0178	0.0102	-1.749	0.111

SUPPLEMENTARY DATA

Supplementary Table 3. GRS calculations and raw data for expanded instrument of n = 28 SNPs

General Info							SBP Association, $\hat{\gamma}_j$				T2D Association, $\hat{\Gamma}_j$				Genetic Risk Score Calculations by SNP			
SNP rsid	Locus/Nearby Gene	Chr	Position	Trait Effect Allele	Non-Effect Allele	Effect Allele Frequency	SBP Association, $\hat{\gamma}_j$ (mmHg incr.)	SBP Association Standard Error, SE_γ	SBP Association P-Value (Wald assumption)	SBP Association Reference (Pubmed ID)	OR	T2D Association, $\hat{\Gamma}_j$	T2D Association Standard Error, σ_j	T2D Association P-Value	GRS for Individual SNP	GRS Standard Error	Chi Squared Statistic	p value (Chi Squared test)
rs10850411	TBX3	12	115387796	T	C	0.7111	0.322	0.069	3.1E-06	21909115	0.991	-0.0093	0.0115	0.42	-0.0289	0.0357	0.65	0.42
rs11191548	NT5C2	10	104846428	T	C	0.08707	1.095	0.1041	7.1E-26	21909115	1.013	0.0125	0.0178	0.48	0.0114	0.0163	0.49	0.48
rs1173771	NPR3	5	32815028	G	A	0.5976	0.495	0.0781	2.3E-10	21909115	1.013	0.0125	0.0104	0.23	0.0253	0.0210	1.44	0.23
rs11953630	EBF1	5	157845402	C	T	0.3298	0.357	0.0789	6.0E-06	21909115	1.026	0.0259	0.0128	0.043	0.0725	0.0359	4.09	0.043
rs12940887	ZNF652	17	47402807	T	C	0.6346	0.354	0.0621	1.2E-08	21909115	0.997	-0.0035	0.0107	0.74	-0.0099	0.0302	0.11	0.74
rs12946454	PLCD3	17	43208121	T	A	0.7203	0.57	0.15	1.0E-08	19430483	1.015	0.0146	0.0121	0.23	0.0256	0.0212	1.46	0.23
rs1327235	JAG1	20	10969280	G	A	0.4657	0.329	0.0774	2.1E-05	21909115	0.999	-0.0011	0.0103	0.92	-0.0033	0.0313	0.01	0.92
rs13359291	PRDM6	5	122476457	A	G	0.31	0.530	0.07	8.9E-16	26390057	1.013	0.0129	0.0215	0.55	0.0243	0.0406	0.36	0.55
rs1378942	CSK	15	75077117	C	A	0.6055	0.613	0.0621	5.6E-23	21909115	1.023	0.0227	0.0108	0.035	0.0370	0.0176	4.42	0.035
rs1458038	FGF5	4	81164973	T	C	0.277	0.732	0.0792	2.4E-20	21909115	1.015	0.0152	0.0115	0.19	0.0208	0.0157	1.75	0.19
rs1563788	SLC22A7	6	43308363	T	C	0.31	0.510	0.06	2.2E-16	26390057	1.020	0.0197	0.018	0.27	0.0386	0.0353	1.20	0.27
rs17249754	ATP2B1	12	90060836	G	A	0.1359	0.763	0.119	1.4E-10	21909115	1.013	0.0129	0.0232	0.58	0.0169	0.0304	0.31	0.58
rs17367504	MTHFR	1	11862778	A	G	0.1372	0.903	0.179	4.5E-07	21909115	1.052	0.0506	0.0142	0.00037	0.0560	0.0157	12.70	0.00037
rs17608766	GOSR2	17	45013521	C	T	0.1662	0.470	0.1084	1.5E-05	21909115	1.026	0.026	0.0153	0.089	0.0553	0.0326	2.89	0.089
rs1799945	HFE	6	26091429	G	C	0.847	0.649	0.1214	9.0E-08	21909115	1.007	0.0067	0.0165	0.69	0.0103	0.0254	0.16	0.69
rs1813353	CACNB2	10	18907698	T	C	0.6728	0.489	0.0895	4.7E-08	21909115	1.007	0.007	0.0163	0.67	0.0143	0.0333	0.18	0.67
rs2014912	ARHGAP24	4	86715670	T	C	0.16	0.620	0.08	5.4E-17	26390057	0.997	-0.0028	0.0222	0.90	-0.0045	0.0358	0.016	0.90
rs2932538	MOV10	1	113216543	G	A	0.2652	0.321	0.0768	2.9E-05	21909115	0.996	-0.0045	0.0124	0.72	-0.0140	0.0386	0.13	0.72
rs2521501	FES	15	91437388	T	A	0.3232	0.620	0.0862	6.4E-13	21909115	1.019	0.0188	0.0125	0.13	0.0303	0.0202	2.26	0.13
rs381815	PLEKHA7	11	16902018	T	C	0.2968	0.485	0.0972	6.0E-07	21909115	1.003	0.0029	0.0126	0.82	0.0060	0.0260	0.053	0.82
rs419076	MECOM	3	169100636	T	C	0.4802	0.355	0.0678	1.6E-07	21909115	0.990	-0.0099	0.0164	0.55	-0.0279	0.0462	0.36	0.55
rs4373814	CACNB2	10	18419972	C	G	0.367	0.318	0.0692	4.3E-06	21909115	0.972	-0.028	0.0169	0.097	-0.0881	0.0531	2.75	0.097
rs4590817	ARID5B	10	63467803	G	C	0.1504	0.626	0.1066	4.3E-09	21909115	0.992	-0.0083	0.0147	0.57	-0.0133	0.0235	0.32	0.57
rs6015450	ZNF831	20	57750867	G	A	0.8602	0.951	0.1134	0.0E+00	21909115	1.006	0.0056	0.0157	0.72	0.0059	0.0165	0.13	0.72
rs633185	ARHGAP42	11	100593788	C	G	0.7177	0.553	0.0838	4.1E-11	21909115	1.009	0.0089	0.0117	0.44	0.0161	0.0212	0.58	0.44
rs7129220	ADM	11	10350788	A	G	0.1425	0.520	0.1079	1.4E-06	21909115	1.003	0.0033	0.0155	0.83	0.0063	0.0298	0.045	0.83
rs805303	BAG6	6	31616616	G	A	0.6214	0.327	0.0671	1.1E-06	21909115	0.995	-0.0052	0.0113	0.65	-0.0159	0.0346	0.21	0.65
rs932764	PLCE1	10	95895940	G	A	0.4551	0.471	0.0759	5.5E-10	21909115	1.010	0.0103	0.0106	0.33	0.0219	0.0225	0.94	0.33

SUPPLEMENTARY DATA

Supplementary Table 4. GRS calculations and raw data for conservative instrument of n = 13 SNPs

General Info							SBP Association, $\hat{\gamma}_j$				T2D Association, $\hat{\Gamma}_j$				Genetic Risk Score Calculations by SNP			
SNP rsid	Locus/Nearby Gene	Chr	Position	Trait Effect Allele	Non-Effect Allele	Effect Allele Frequency	SBP Association, $\hat{\gamma}_j$ (mmHg incr.)	SBP Association Standard Error, SE_γ	SBP Association P-Value (Wald assumption)	SBP Association Reference (Pubmed ID)	OR	T2D Association, $\hat{\Gamma}_j$	T2D Association Standard Error, σ_j	T2D Association P-Value	GRS for Individual SNP	GRS Standard Error	Chi Squared Statistic	p value (Chi Squared test)
rs11953630	EBF1	5	157845402	C	T	0.3298	0.357	0.0789	6.0E-06	21909115	1.026	0.0259	0.0128	0.043	0.0725	0.0359	4.09	0.043
rs1327235	JAG1	20	10969280	G	A	0.4657	0.329	0.0774	2.1E-05	21909115	0.999	-0.0011	0.0103	0.92	-0.0033	0.0313	0.01	0.92
rs1458038	FGF5	4	81164973	T	C	0.277	0.732	0.0792	2.4E-20	21909115	1.015	0.0152	0.0115	0.19	0.0208	0.0157	1.75	0.19
rs17367504	MTHFR	1	11862778	A	G	0.1372	0.903	0.1790	4.5E-07	21909115	1.052	0.0506	0.0142	0.00037	0.0560	0.0157	12.70	0.00037
rs2521501	FES	15	91437388	T	A	0.3232	0.620	0.0862	6.4E-13	21909115	1.019	0.0188	0.0125	0.13	0.0303	0.0202	2.26	0.13
rs381815	PLEKHA7	11	16902018	T	C	0.2968	0.485	0.0972	6.0E-07	21909115	1.003	0.0029	0.0126	0.82	0.0060	0.0260	0.05	0.82
rs419076	MECOM	3	169100636	T	C	0.4802	0.355	0.0678	1.6E-07	21909115	0.990	-0.0099	0.0164	0.55	-0.0279	0.0462	0.36	0.55
rs4373814	CACNB2	10	18419972	C	G	0.367	0.318	0.0692	4.3E-06	21909115	0.972	-0.028	0.0169	0.097	-0.0881	0.0531	2.75	0.097
rs6015450	ZNF831	20	57750867	G	A	0.8602	0.951	0.1134	5.0E-17	21909115	1.006	0.0056	0.0157	0.72	0.0059	0.0165	0.13	0.72
rs633185	ARHGAP42	11	100593788	C	G	0.7177	0.553	0.0838	4.1E-11	21909115	1.009	0.0089	0.0117	0.44	0.0161	0.0212	0.58	0.44
rs7129220	ADM	11	10350788	A	G	0.1425	0.520	0.1079	1.4E-06	21909115	1.003	0.0033	0.0155	0.83	0.0063	0.0298	0.05	0.83
rs805303	BAG6	6	31616616	G	A	0.6214	0.327	0.0671	1.1E-06	21909115	0.995	-0.0052	0.0113	0.65	-0.0159	0.0346	0.21	0.65
rs932764	PLCE1	10	95895940	G	A	0.4551	0.471	0.0759	5.5E-10	21909115	1.010	0.0103	0.0106	0.33	0.0219	0.0225	0.94	0.33

SUPPLEMENTARY DATA

Supplementary Table 5. Parameters used to generate bias in Egger Regression simulations.

Type of bias	***Distribution for Simulation***
No bias	*No bias added*
Positive Bias	*Uniform(0, 0.04)*
Negative Bias	*Uniform(-0.04, 0)*
Nondirectional bias	*Uniform(-0.04, 0.04)*

SUPPLEMENTARY DATA

Supplementary Table 6. Percent of simulations under each bias distribution which reported that bias was present in the analysis (two-tailed t-test with *p* threshold < 0.05).

Type of bias	***Power to Detect Bias***
Positive Bias	*0.439*
Negative Bias	*0.427*
Nondirectional bias	*0.307*

SUPPLEMENTARY DATA

Supplementary Table 7. Simulation results for casual effect estimate via instrumental variable weighted regression (IVW) and Egger Regression (ER) modeling a pro-diabetic drug use among hypertensive subjects. 95% CI represents the error on the mean. *Based on 10,000 simulations. Results reported in units of change in log odds of T2D risk per SD change in SBP. Analysis performed is Egger Regression.

	*IVW Effect Estimate**		*ER Effect Estimate (Slope)**		*ER Bias Estimate (Intercept)**	
% DRUG USE	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
0	0.539	0.537-0.541	0.513	0.508 - 0.519	0.00082	0.00065 - 0.00099
20	0.544	0.542-0.546	0.495	0.489 - 0.501	0.00156	0.00138 - 0.00173
40	0.548	0.546-0.550	0.484	0.478 - 0.490	0.00201	0.00183 - 0.00219
60	0.550	0.547-0.552	0.461	0.455 - 0.467	0.00279	0.00261 - 0.00296
80	0.554	0.552-0.556	0.452	0.446 - 0.458	0.00322	0.00305 - 0.00340
100	0.555	0.553-0.558	0.427	0.421 - 0.433	0.00404	0.00387 - 0.00421

SUPPLEMENTARY DATA

Analyzing Egger Regression performance through simulation

Since the use of Egger Regression for Mendelian Randomization studies is a relatively novel technique[1], we sought to better understand the behavior of this analytical tool through simulation. Using our conservative instrument as a baseline, simulated n=10,000 datasets for analysis by adding noise and bias to the T2D association of each SNP:

$$\begin{aligned} \tilde{\Gamma}_j &= \alpha\hat{\gamma}_j + noise + bias \\ noise &\sim \text{Normal}(0, \hat{S}_j) \\ bias &\sim \text{Uniform} \end{aligned} \qquad (1)$$

Where α is the true effect of SBP on T2D risk (set in these simulations at 0.02 log-odds increase per mmHg), $\tilde{\Gamma}_j$ is the T2D association estimate for the jth SNP generated for a given simulation, $\hat{\gamma}_j$ is the actual estimated SBP association, and $\hat{S}_j$ is the true standard error in T2D association for that SNP. By adjusting the upper and lower limits of the uniform distribution for the added bias, we were able to generate datasets affected by different levels and directions of pleotropic bias. The parameters for bias added are listed in **Supplementary Table 5.**

From each these simulated datasets, we ran Egger regression analyses and estimated the power to detect bias. The complete R code (v3.3.0) used to simulate and analyze these datasets is available at https://github.com/raikens1/T2D_MR/. Using this set-up, we found that Egger regression power to detect negative bias is limited (**Supplementary Table 6**). Egger Regression effect estimates have a higher variance when bias is at play (6×10^{-4} under negative bias compared with 3×10^{-4} with no bias, F-test $p = 2 \times 10^{-16}$). However, the effect estimates from this test are still correct on average (**Supplementary Figure 3**, two-tailed t-test for significant error under negative bias: p = 0.45).

***In Silico* Test for bias due to adjustment in the primary scan**

In order to understand whether adjustment for adiposity in the primary scan resulted in bias in our GRS, we simulated GWAS under two different causal models:

A. Both BMI and SBP affect T2D risk
B. BMI, but not SBP, affects T2D risk

Under each causal model, we used the MR_predictor simulation engine described previously[2] to generate n=1000 sets of genotype and phenotype from 150,000 individuals. To construct our simulations, association of SBP with BMI[3], and BMI-associated T2D risk[4] were drawn from the literature, and the T2D prevalence was tuned to give a realistic case/control ratio in simulation (targeting a 9.8% diabetes prevalence, in agreement with estimates by Cowie et al.[5]).We then used the Plink analysis toolset (v1.07)[6,7] to generate linear SBP association estimates for each of the 13 SNPs in our conservative instrument set over n=1000 simulations. As a summary statistic, we observed the distribution of the mean error over all SNPs, as:

SUPPLEMENTARY DATA

$$\text{mean error for } i^{th} \text{ simulation} = \frac{1}{13}\sum_{j=1}^{13}(\gamma_j - \hat{\gamma_{ij}}) \qquad (2)$$

where that γ_j and γ_{ij}-hat respectively represent the actual SBP association of the j^{th} SNP and the estimate for that association generated from the i^{th} simulation (where the association for a given SNP is always relative to the blood-pressure-increasing allele). Mean errors were tested for significance using a twosided t-test in R (v3.3.0). The code used to run plink and MR_predictor and the relevant MR_predictor input files are available at https://github.com/raikens1/T2D_MR/, and the MR_predictor simulation toolset and documentation are additionally available online (http://coruscant.itmat.upenn.edu/mr_predictor/).

We found in both scenarios that, even when certain corrections in the primary scan did result in statistically significant bias, the magnitude of this change was sufficiently small compared to our estimated SBP associations that it could not be expected to have any notable effect on our final result (**Supplementary Figures 4 and 5**).

As an additional check, we sought to demonstrate that our simulation framework was sufficient to detect strong collider bias if it did indeed arise (collider bias has been illustrated previously[8]). Since high BMI is known to cause high blood pressure[9–11], adjusting for SBP in a linear association analysis will cause SBP-related SNPs to falsely associate with BMI. We used the PLINK toolset to perform these association analyses for n=1000 simulated datasets under the model that both SBP and BMI increase type two diabetes risk. When SBP was used as a covariate in these analyses, we found that simulations tended to report a false BMI association for SNPs related exclusively to SBP (**Supplementary Figure 6**).

***In Silico* Test for bias due to pro-diabetic antihypertensive use in GWAS cohorts**

Since evidence suggests that various antihypertensive medications (namely beta-blockers and thiazide diuretics) are linked to increased type 2 diabetes risk, we considered the possibility that the putative link between SBP and type 2 diabetes risk can be explained by the use of diabetogenic antihypertensive use by the subjects of our GWAS cohorts. If hypertensive subjects used an anti-hypertensive medication that increased diabetes risk, we would expect our risk score to be positively biased. This is because we expect, based on genotype, a log-additive increase in drug use on a liability scale (with respect to blood pressure). Put another way: each genetic variant increases the chance of crossing the hypertension liability threshold by a small amount. Each variant thus increases drug use amount proportional to the SBP effect. This applies to each SNP: weaker-effect SBP SNPs have lower chance for antihypertensive use, while stronger SBP SNPs will contribute a greater chance. Individuals will carry a random collection of these variants. However, the impact of drug use on type 2 diabetes risk is the same (the magnitude of the effect does not change by genotype). This is analogous to systemic, positive bias from unmeasured confounding, which can be measured and subsequently accounted for by Egger Regression.

We performed a simulation experiment to verify this intuition. We generated 33K cases and 33K controls, the equivalently powered effective symmetric sample size of our T2D study. Among simulated subjects, we assumed 60% of T2D cases were hypertensive (> 140 mmHg SBP), 30% of controls as hypertensive. These rough estimates were obtained from recent literature[12,13]. Then, we varied (from 0-100%) the percentage of hypertensive subjects that take a drug that increases T2D risk, and assumed that

this drug use increases T2D by 1.4-fold (according to a literature estimate for beta blockers[14]). This boils down to T2D subjects having a higher prevalence of an exposure (i.e., drug use) that increases the baseline risk for a subset of participants (i.e. hypertension).

Simulations demonstrate a positive bias that grows in magnitude as the percent of pro-diabetic antihypertensive drug use among subjects who are hypertensive increases (see **Supplementary Table 7**, below). This effect also resulted in a corresponding reduction in the casual effect from the Egger Regression analysis, as one would expect in the presence of positive, directional confounding (**Supplementary Table 7**). This effect also does slightly increase the casual effect estimate from the GRS method (0.540 for no drug use to 0.556 for 100% drug use, **Supplementary Table 7**). Based on this analysis, we did not observe significant evidence of bias for either of our risk scores. Moreover, the direction of that term trended toward negative, rather than positive, contrary to what would be expected from this drug-confounding effect. While assumptions made here are unlikely to perfectly match the specifics of the contributing T2D cohort(s) to our study, the results support our intuition above: (i) that the direction of this type of bias should be positive, (ii) that Egger regression can identify (and adjust) for this effects, at least under this specific model, and (iii) that in the real data, we observed a trend in the opposite direction of this putative effect: negative rather than positive bias.

SUPPLEMENTARY DATA

CITATIONS

1. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44,** 512–25 (2015).

2. Voight, B. F. MR_predictor: a simulation engine for Mendelian Randomization studies. *Bioinformatics* **30,** 3432–4 (2014).

3. Dudina, A. *et al.* Relationships between body mass index, cardiovascular mortality, and risk factors: a report from the SCORE investigators. *Eur. J. Cardiovasc. Prev. Rehabil.* **18,** 731–42 (2011).

4. Sheikh, M. A., Lund, E. & Braaten, T. The predictive effect of body mass index on type 2 diabetes in the Norwegian women and cancer study. *Lipids Health Dis.* **13,** 164 (2014).

5. Cowie, C. C. *et al.* Prevalence of diabetes and impaired fasting glucose in adults in the U.S. population: National Health And Nutrition Examination Survey 1999-2002. *Diabetes Care* **29,** 1263–8 (2006).

6. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–75 (2007).

7. Purcell, S. PLINK (version 1.07). (2009). at <http://pngu.mgh.harvard.edu/purcell/plink/>

8. Cole, S. R. *et al.* Illustrating bias due to conditioning on a collider. *Int. J. Epidemiol.* **39,** 417–20 (2010).

9. Drøyvold, W. B., Midthjell, K., Nilsen, T. I. L. & Holmen, J. Change in body mass index and its impact on blood pressure: a prospective population study. *Int. J. Obes. (Lond).* **29,** 650–5 (2005).

10. Dudina, A. *et al.* Relationships between body mass index, cardiovascular mortality, and risk factors: a report from the SCORE investigators. *Eur. J. Cardiovasc. Prev. Rehabil.* **18,** 731–42 (2011).

11. Dua, S., Bhuker, M., Sharma, P., Dhall, M. & Kapoor, S. Body mass index relates to blood pressure among adults. *N. Am. J. Med. Sci.* **6,** 89–95 (2014).

12. Colosia, A. D., Palencia, R. & Khan, S. Prevalence of hypertension and obesity in patients with type 2 diabetes mellitus in observational studies: a systematic literature review. *Diabetes. Metab. Syndr. Obes.* **6,** 327–38 (2013).

13. Egan, B. M., Li, J., Hutchison, F. N. & Ferdinand, K. C. Hypertension in the United States 1999-2012: Progress Toward Healthy People 2020 Goals. *Circulation* (2014).

14. Lam, S. K. H. *et al.* Incident diabetes in clinical trials of antihypertensive drugs. *Lancet (London, England)* **369,** 1513-4-5 (2007).