Supplementary Online Content

Chauquet S, Zhu Z, O'Donovan M, Walters J, Wray N, Shah S. Association of antihypertensive drug target genes with psychiatric disorders: a mendelian randomization study. *JAMA Psychiatry.* Published online March 10, 2021. doi:10.1001/jamapsychiatry.2021.0005

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. MR Analysis

Assumptions of MR

MR analysis makes three major assumptions that need to be met. Firstly, the instrument is robustly associated with the exposure of interest. Weak instruments can lead to bias, and selection of instruments should assess instrument strength. Second, there are no unmeasured confounders between genetic instruments and outcome. Sufficient adjustment for population stratification in GWAS can help minimise confounding from ancestry, while identifying association of instruments with other covariates may help to identify and assess other potential sources for confounding. Finally, the instrument should not affect the outcome except through the biomarker of interest. The assumptions of an MR analysis hold if the associations of a genetic variant with other covariates arise because of the serial and sequential effects of the biomarker of interest on others residing more distally on the same causal pathway to disease[1] (i.e. association through vertical pleiotropy rather than horizontal pleiotropy). It is plausible for a single genetic variant to meet these conditions if the biological process linking the variant with the exposure is well understood[2]. Where mRNA or protein levels are the primary exposure of interest, any pleiotropy observed of a *cis*-SNP instrumenting its encoded mRNA or protein is more likely to be vertical than horizontal in origin[1].

SMR Analysis

Briefly, let y be the outcome of interest, x be gene expression (exposure), g be a genetic instrument, b_{gx} be the effect of g on x (estimated by eQTL studies), and b_{gy} be the effect of g on y (estimated by GWAS). The effect of exposure x on outcome y, free of any non-genetic confounders, is defined as b_{gy}/b_{gx} (Wald ratio). The SMR approach selects a single most significantly associated eQTL SNP (located near the target gene i.e. *cis*-eQTL SNP) as an instrument. The SMR tool also implements the heterogeneity in dependent instruments (HEIDI) test to assess if the observed association between gene expression and outcome is due to a linkage scenario, where rather than the SNP affecting disease via gene expression regulation, the SNP that influences expression is in linkage disequilibrium (LD) with another SNP that independently influences the outcome, which would violate one of the assumptions of MR (see **Assumptions of MR** section above). As recommended by the authors of SMR, a HEIDI test p-value < 0.01 was considered to indicate association due to a linkage scenario. SMR requires a reference dataset from which to estimate LD between SNPs. For this purpose, a random sample of 10,000 unrelated individuals of European ancestry from the UK Biobank data was used. Though we do not believe there to be sample overlap between exposure and outcome data, SMR test statistics have been shown to be robust to even complete sample overlap. All analyses were performed using SMR version 1.02.

Assessing Instrument strength

The F-statistic from the regression of the exposure (gene expression) on the instrument (eQTL SNP) is usually quoted in single sample MR studies as a measure of the strength of an instrument. By rule of thumb, instruments with an F-statistic less than 10 are considered 'weak instruments'[3]. A value of 10 indicates that bias in the estimated causal effect due to measurement error is around 10% of the true value of causal effect. In the context of two-sample MR which uses GWAS summary data, the F-statistic can be generated using the approximation described by Bowden et al[4]. F-statistic for SNP j can be approximated as $F_j = \hat{\gamma}_j^2 / \sigma_{X_j}^2$ where $\hat{\gamma}_j$ is the SNP-exposure association and σ_{X_j} is the standard error of the SNP-exposure association.

Assessing association due to linkage

In the absence of confounding factors such as population stratification and assortative mating, MR can estimate an apparent effect of an exposure on the outcome in 3 scenarios: vertical pleiotropy (causality); horizontal pleiotropy, where the same SNP influences the exposure and outcome through independent pathways (e.g. if the same SNP affects expression of two different genes, but only one gene plays a casual role with respect to the outcome); or linkage, where the SNP that influences the exposure is in linkage disequilibrium (LD) with another SNP that independently influences the outcome. We used the heterogeneity in dependent instruments (HEIDI) test implemented in the SMR tool to identify association between gene expression and outcome due to a linkage scenario[5]. Briefly, if gene expression and a trait share the same causal variant, the b_{xy} values calculated for any SNPs in LD (using the default value of $r^2 > 0.05$ and also $r^2 < 0.9$ to avoid issues of collinearity) with the causal variant should be identical. Therefore, testing against this null hypothesis of a single causal variant is equivalent to testing for heterogeneity in the b_{xy} values estimated for the SNPs in the *cis*-eQTL region[5]. Since heterogeneity estimates may not be robust if using only a small number of SNPs, the HEIDI test required a minimum of 5 SNPs for estimating heterogeneity. As recommended by the authors, a HEIDI test p-value < 0.01 was considered to indicate heterogeneity in b_{xy} values, suggesting that association between gene expression and outcome is most likely to be due to a linkage scenario. SMR requires a reference dataset from which to estimate

LD between SNPs. For this purpose, a random sample of 10,000 unrelated individuals of European ancestry from the UK Biobank data was used.

Colocalisation analysis

For any statistically significant MR associations, we applied a Bayesian colocalisation approach to estimate the posterior probability for a common causal variant for gene expression and outcome[6] using the coloc (v3.1) R package (https://cran.r-project.org/web/packages/coloc/). Since for each gene, SNP associations in blood were only available for those within a 2MB window, colocalisation analysis was restricted to using only these *cis*-SNP associations. Default priors were used for analysis.

Assessing horizontal pleiotropy

The same genetic variant could be associated with expression of more than one gene, which could invalidate the assumption that the genetic instrument is associated with outcome only via changes in gene expression of the drug target gene. To assess the presence of horizontal pleiotropy, for each genetic instrument we also extracted available associations with all other nearby genes (within a 2Mb window). For genes whose expression was associated at nominal significance ($p < 0.05$), we performed SMR analysis to determine if the expression of these genes was associated with the outcome of interest, and colocalisation analysis to determine the posterior probability of a shared causal variant.

MR analysis with brain gene expression and psychiatric outcomes

We performed SMR analysis between gene expression in brain tissue and psychiatric outcomes. We used eQTL data from the PsychENCODE resource sample size of 1,387. The effect estimate from SMR represents the effect on disease risk (odds ratio) per 1 SD change in gene expression. Only associations for SNPs located within a 1 Mb window around each gene are available in the PsychENCODE eQTL data. Given different brain regions may have greater relevance to different diseases, we also ran SMR analysis using gene expression in 13 different brain regions from the latest version of GTEx (v8), with sample size ranging between ~130 - 250. For GTEx, data is only available for SNP-gene associations within a 2 Mb window around the transcription start site (**eTable 1)**.

Assessing confounding due to ancestry

One potential confounder in MR analyses is ancestry. For both the bipolar disorder and MDD GWAS analysis was restricted to individuals with European ancestry. For schizophrenia, though a small proportion (~5%) of individuals consisted of non-European ancestry, analysis was performed using matched ancestry controls, and within-ancestry analysis prior to meta-analysis to avoid confounding. As sensitivity analyses we ran SMR analysis using blood eQTL data and case-control GWAS summary data from individuals of European ancestry only (33,640 schizophrenia cases and 43,456 controls) and East Asian ancestry only (22,778 schizophrenia cases and 35,362 controls [7], providing two independent outcome datasets of different ancestry (**eTable 2**). For the analysis with the East Asian schizophrenia GWAS, we used unrelated East Asian samples in UK Biobank as our LD reference .

MR analysis using a proxy for the ACE insertion/deletion as an instrument

The ACE indel consists of the presence or absence of a 250-bp DNA fragment in intron 16 associated with ACE enzyme activity. The SNP rs4343 is considered the best proxy for the *ACE I/D*[8], and has been associated with increased plasma ACE activity accounting for 16.2% of the total variance in ACE activity[9], as well as increased ACE protein levels in cerebral spinal fluid[10]. Linkage disequilibrium (LD) between rs4343 and the best eQTL SNP selected by the SMR method as an instrument for *ACE* expression in blood (rs4277405), is around r^2 0.35 in Europeans (LD estimation from LDlink https://ldlink.nci.nih.gov/). We performed MR analysis using rs4343 as an instrument for *ACE* expression in blood to determine if consistent effects were observed on schizophrenia risk.

MR analysis of other genes in the renin-angiotensin pathway

ACE is part of the renin-angiotensin (RAS) pathway that regulates blood pressure. To determine whether other components of the pathway show any association with schizophrenia, we performed MR analysis on other genes in the RAS pathway (*AGT*, *REN*, *AGTR1*), regardless of whether they are targets of antihypertensive medication or not. *AGT, AGTR1* and *REN* were absent in the blood eQTL summary data. All 3 genes were present either in the PsychENCODE or GTEx brain data. We therefore identified the strongest eQTL SNPs for these genes in the brain eQTL datasets and performed SMR analysis with schizophrenia as an outcome.

MR analysis with ACE protein levels in plasma and cerebral spinal fluid (CSF)

We performed SMR analysis using GWAS summary data of ACE protein levels from CSF in 544 European ancestry individuals[10] and plasma in 818 European ancestry individuals[11] (Supplementary Table 2) with schizophrenia as the outcome. These data are from Alzheimer's disease case/control samples, but disease status was adjusted for in GWA analysis. There is no known sample overlap with eQTL or psychiatric GWAS samples. Data are publicly available, but summary data is only provided for SNPs where association p-value is < 0.05. The effect estimate from SMR represents the effect on disease risk (odds ratio) per 1 SD change in protein levels. We also performed multi-SNP MR analysis using all independently associated (*cis* and *trans*) variants as instruments for ACE levels. Analyses were performed using the TwoSampleMR v0.5.2 package (https://mrcieu.github.io/TwoSampleMR/) in R CRAN. A GWAS association p-value threshold of $5x10^{-8}$ and a pairwise LD r^2 of 0.1 were used to identify multiple, independent instruments for ACE protein levels.

Reverse MR analysis of schizophrenia risk on ACE protein levels in plasma and cerebral spinal fluid (CSF)

As the published eQTL summary data only contained cis-associations, reverse MR was not possible using these datasets. The published pQTL GWAS data for ACE plasma and CSF only contained SNPs with p<0.05, and therefore most of the schizophrenia-associated SNPs and proxies were absent in this data. We contacted the authors of the ACE pQTL GWAS who provided us with full GWAS summary data from an updated (unpublished) dataset. Correlation for intersecting SNPs between betas in the published downloaded summary data versus the new unpublished data was 0.99 for both plasma and CSF data.

Simulating the effect of selection bias on causal estimates

Using the approach by Gkatzionis & Burgess[12] we use simulations to investigate the effect of selection bias on the ACE expression-schizophrenia association. It is known that that there are increased cardiovascular disease (CVD)-related deaths in individuals with schizophrenia. Given the known relationship between higher ACE and increased risk of hypertension, and therefore increased risk of CVD, we assume that selection bias is caused by increased ACE levels, either due to increased hypertension and/or increased cardiovascular disease, or other ACE-associated disease resulting in reduced survival in cases. We run simulations under 2 scenarios:

1) **Null model** ($b_{xy} = 0$), assuming no association without selection

2) **Causality model** ($b_{xy} \neq 0$), assuming an existing causal relationship without selection.

To mimic the real data, the parameters used in the simulations were close to the observed estimates.

1. Simulation under the null model

- *Simulation of ACE eQTL data*: We first randomly sampled 32,000 individuals (dataset 1) to simulate ACE gene expression using the linear model,

$$x = b_{zx}z + b_{cx}c + e_x,$$

where,

  - z is a causal variant of x;
  - allele frequency f was simulated to be that of the ACE SNPs rs4277405 in unrelated European participants of UK Biobank v3: f=0.375 for rs4277405:C:T with C as the reference allele (associated with increased ACE expression).
  - variance of x explained by z (R^2_{zx}) is 2.5×10^{-3} (based on eQTL association for rs4277405 in the eQTLGen data);
  - the effect of the variant z on gene expression x is $\mathrm{E}(b_{zx}) = \sqrt{R^2_{zx}/[2f(1-f)]}$
  - c is a confounding factor, simulated from a normal distribution $N(0, R^2_{cx})$ with R^2_{cx} being the variance of x explained by c, $R^2_{cx} = 0.3$, assuming a third of the remaining variance is attributed to c, $b_{cx} = 1$;
  - e_x is the residual, $e_x \sim N(0, \sqrt{1 - R^2_{zx} - R^2_{zc}})$, which leads to $var(x) = 1$ in the simulations.

- *Simulation of case/control data:*
  We randomly sampled another 200,000 individuals (dataset 2) to simulate x using the same approach as above, without any sample overlap with dataset 1. Given the simulated x, disease risk was simulated using the regression model,

$$y = b_{cy}c + e_y$$

where,

  - c was sampled from a normal distribution $N(0, R^2_{cy})$ with $R^2_{cy} = 0.3$ (i.e a third of the remaining variance is attributed to c), $b_{cy} = -1$;

– e_y is the residual, $e_y \sim N(0, \sqrt{1 - R_{cy}^2})$, which leads to $var(y) = 1$

We assume the population prevalence for disease to be 10%, and therefore individuals with the top 10% y were considered as cases (1), and the rest considered as controls (0). The actual population prevalence of schizophrenia is only around ~1%, but as number of cases would be low and therefore lower power, we have used 10% in our simulations.

- *Simulation of the effect of selection bias on the causal effect:*
Suppose there was selection bias in cases which was caused by ACE, either dying early from cardiovascular, hypertensive or other ACE-associated disease, or ascertainment bias of cases induced by ACE. We then simulated the probability of selection (π) using the general regression model,

$$logit(\pi) = \gamma_0 + \gamma_x x + \gamma_c c$$

where
γ_0=-1.6, which suggests ~80% of all the individuals would be included due to selection
γ_x is effect of x on π, γ_x = -2, -1, -0.5, -0.2 and 0.
γ_c is effect of c on π, $\gamma_c = \gamma_x$.
Schizophrenia cases were more likely to be excluded in the simulation, because of negative γ_x and γ_c.

The selection event (s) was then sampled from a Bernoulli distribution, $s \sim Bernoulli(\pi)$.

We estimated $\hat{b}_{zx}$ by linear regression from dataset 1 and estimated $\hat{b}_{zy}$ (effect of z on y) by logistic regression from dataset 2. MR analysis was subsequently conducted. Each simulation was replicated 1,000 times.

2. Simulation under the causality model

We carried out simulation under the causality model using the same approach as above, but simulating disease using the regression model,

$$y = b_{xy} x + b_{cy} c + e_y,$$

where
– b_{xy} is the effect size of x on y, $b_{xy} = -0.55$ (based on the observed causal effect of blood *ACE* expression on schizophrenia risk);
– the variance of y explained by x, $R_{xy}^2 = b_{xy}^2 = 0.3025$.
– Then $e_x \sim N(0, \sqrt{1 - R_{xy}^2 - R_{cy}^2 - 2b_{xy}\text{cov}(x, c)})$,
– where $\text{cov}(x, c) = R_{cx} R_{cy}$.

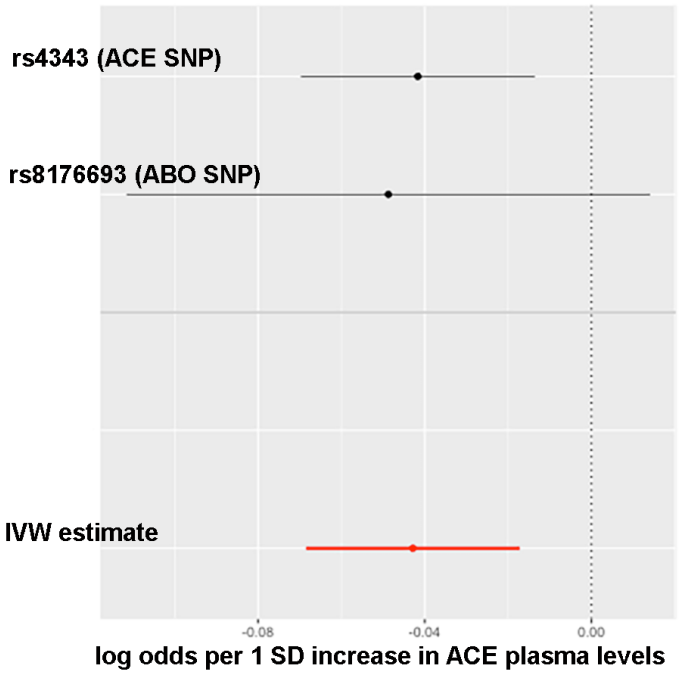
Each simulation was replicated 1,000 times.

MR Power Calculations

We have used the power calculations from Stephen Burgess (PMID: 24608958),
also implemented as an online tool (http://sb452.shinyapps.io/power/) to estimate the causal effect for which we have 80% power to detect.

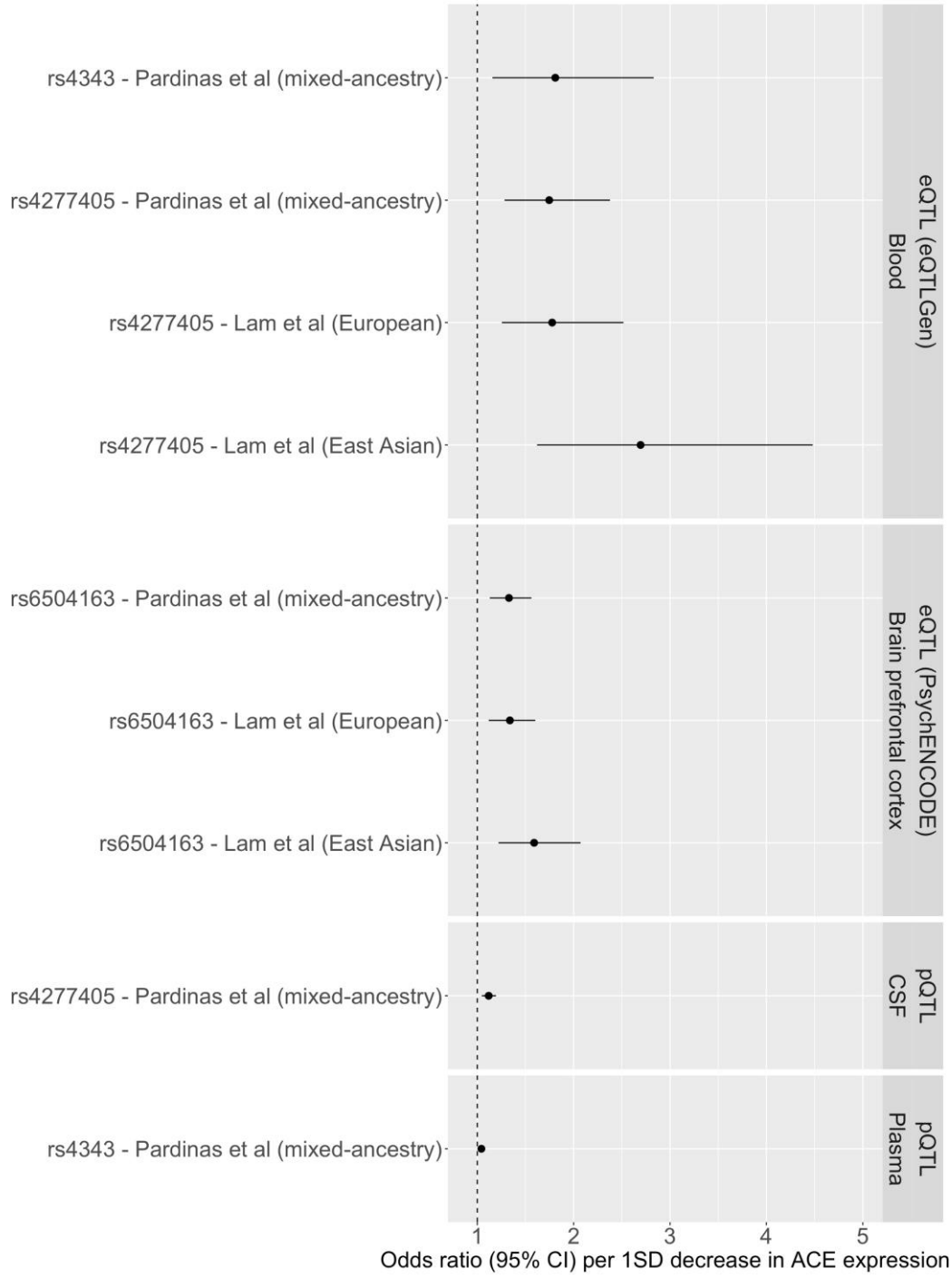
eFigure 1. Multi-SNP MR for ACE Protein Levels and Schizophrenia Risk

The figure below shows the log odds of disease per 1 SD increase in ACE plasma levels, for each independent SNP instrument, and their combined effect estimated by the IVW method (shown in red). Using GWAS summary data of ACE protein levels in cerebral spinal fluid (CSF) and plasma, after LD pruning (using an r^2 threshold of 0.1 and GWA p-value < 5e-8), only 1 ACE variant remained for ACE CSF levels, while 3 SNPs survived for plasma ACE levels: one in ACE (rs4343, proxy for the ACE indel) and 2 in/near the ABO gene (rs2519093, rs8176693). rs4343 explained ~22% of the variance in plasma ACE levels while the ABO SNPs explain around 4 % each. This is concordant with results from an independent GWAS of ACE plasma activity in a samples of East Asian ancestry[9], which also reported rs4343 and two independent variants in/near the ABO gene providing independent replication of the two *trans*-pQTLs for ACE. rs2519093 and its proxy SNPs were absent in the schizophrenia GWAS summary data. Therefore, multi-SNP MR of plasma ACE levels was only possible using 1 ACE SNP (rs4343) and 1 ABO SNP (rs8176693), and only the IVW approach was feasible. Though the association for rs8176693 was not nominally significant, the direction of effect of the ABO SNP on schizophrenia risk was concordant with that of the ACE SNP, and a concordant association between higher ACE levels and lower risk of schizophrenia was observed (OR per SD increase in plasma ACE (95% CI) = 0.96 (0.93 – 0.98); $p=1x10^{-3}$).



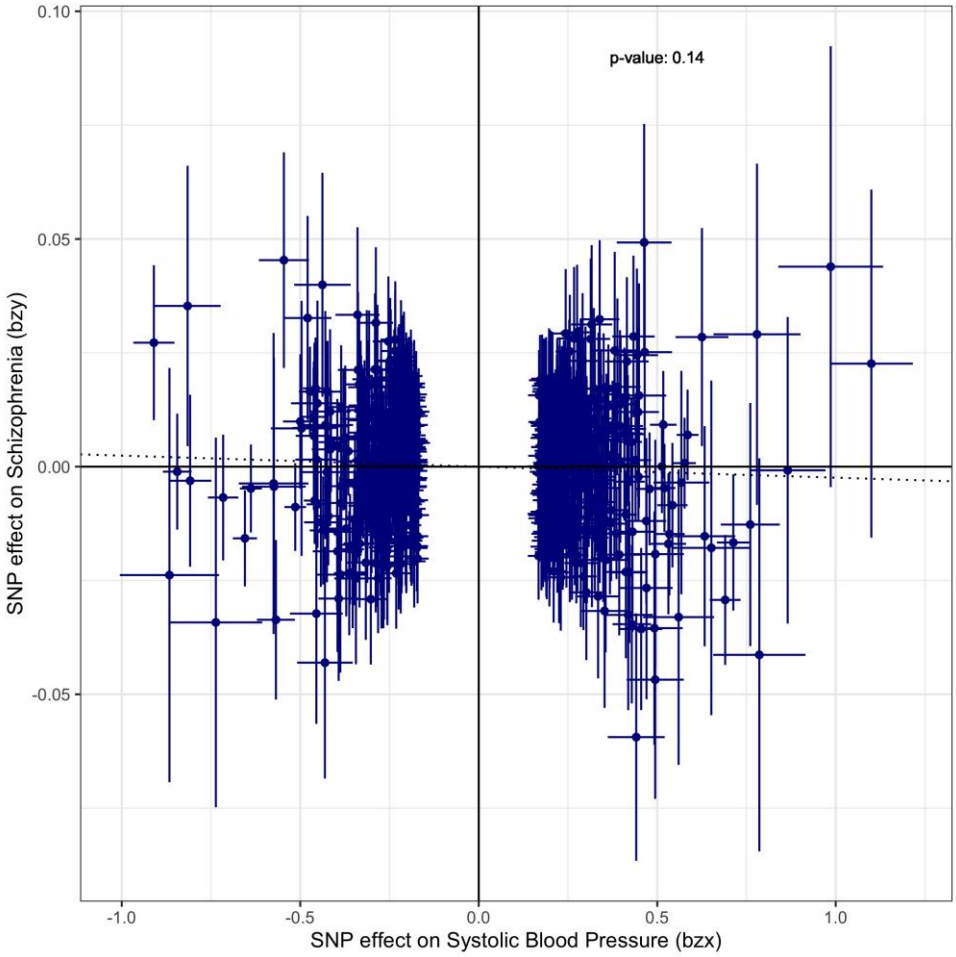
eFigure 2. Summary of Results From Sensitivity Analyses of the Effect of ACE Gene Expression or Protein Levels on Schizophrenia Risk

The plots shows the odds ratio for schizophrenia risk per 1 standard deviation (SD) increase in ACE gene expression or protein levels using a combination of different SNP instruments, quantitative trait loci datasets and outcome datasets. Abbreviations: eQTL; expression quantitative trait loci, pQTL protein; quantitative trait loci, CSF; cerebral spinal fluid, CI; confidence interval.



eFigure 3. MR Analysis of Systolic Blood Pressure as Exposure With Schizophrenia Risk

The plots show the effect sizes and standard errors (error bars) for the associations of independent SNP instruments for the exposure of interest and schizophrenia. SNP effects on the systolic blood pressure (b_{zx}; mmHg per effect allele) are shown on the x-axis, while the effects on schizophrenia risk (b_{zy}; log odds per effect allele) are shown on the y-axis. The p-value for the MR association is shown.



Simulating the effect of selection bias on causal estimates (eTable 15 -16)

1. Null model

eTable 15. Estimates of b_{xy} Under the Null With Selection Bias

SNP	γ_x or γ_c	$\hat{b}_{xy}$	χ^2	FPR	n_{cases}	p_{cases}	$n_{controls}$	$p_{controls}$
	/	0.003	0.919	0.044	20000	1.000	180000	1.000
	0	0.005	0.900	0.043	16641	0.832	149761	0.832
rs4277405	-0.2	0.013	0.887	0.036	15635	0.782	149646	0.831
	-0.5	0.063	0.986	0.043	13606	0.680	146526	0.814
	-1	0.174	1.575	0.122	10394	0.520	138103	0.767
	-2	0.316	2.471	0.235	7046	0.352	124497	0.692
	/	0.001	0.910	0.032	20000	1.000	179999	1.000
	0	-0.001	0.909	0.035	16641	0.832	149762	0.832
rs4343	-0.2	0.009	0.898	0.034	15635	0.782	149645	0.831
	-0.5	0.059	0.967	0.044	13606	0.680	146527	0.814
	-1	0.166	1.507	0.103	10394	0.520	138105	0.767
	-2	0.325	2.527	0.252	7047	0.352	124501	0.692

Notes:
1. "/" - no selection bias
2. $\hat{b}_{xy}$ – mean of $\hat{b}_{xy}$ across the 1,000 replicates;
3. χ^2 – mean χ^2 across the 1,000 replicates;
4. FPR – false positive rate. The associations with *p*-value < 0.05 were considered as false positives;
5. n – sample size;
6. p - proportion of the remaining individuals under selection bias.

Interpretation:
As fewer schizophrenia cases get included in the simulation due to selection bias, $\hat{b}_{xy}$ becomes positive with increased false positive rate (FPR). Under a moderate selective pressure, e.g. γ_x= -0.5, FPR < 0.05, despite $\hat{b}_{xy}$ is slightly larger than 0. This suggests that an association between *ACE* expression and schizophrenia is unlikely to be identified with small or moderate selection if there is no true association. We repeated simulations using a disease population prevalence of 1%, which did not change our conclusion.

2. Causality model

eTable 16. Estimates of b_{xy} Under the Causality With Selection Bias

SNP	γ_x or γ_c	$\hat{b}_{xy}$	χ^2	n_{cases}	p_{cases}	$n_{controls}$	$p_{controls}$
	/	-1.091	31.016	20000	1.000	180000	1.000
	0	-1.088	27.555	16641	0.832	149762	0.832
rs4277405	-0.2	-1.097	25.823	14613	0.731	150667	0.837
	-0.5	-1.099	20.981	10542	0.527	149590	0.831
	-1	-1.071	10.892	4589	0.229	143908	0.799
	-2	-0.990	2.318	622	0.031	130921	0.727
	/	-1.085	30.941	19998	1.000	180002	1.000
	0	-1.086	27.640	16639	0.832	149764	0.832
rs4343	-0.2	-1.095	25.914	14611	0.731	150669	0.837
	-0.5	-1.079	20.601	10542	0.527	149591	0.831

	-1	-1.043	10.567	4589	0.229	143909	0.799
	-2	-1.003	2.369	622	0.031	130926	0.727

Notes:
1. "/" - no selection bias
2. $\hat{b}_{xy}$ – mean of $\hat{b}_{xy}$ across the 1,000 replicates;
3. χ^2 – mean χ^2 across the 1,000 replicates;
4. n – sample size;
5. p - proportion of the remaining individuals under selection bias.

Interpretation

$\hat{b}_{xy}$ which was approximately -1.1 without selective pressure, but became smaller with decreased χ^2, with increased selective pressure. When γ_x= -0.5, $\hat{b}_{xy}$ was not too different from the estimate without selection, with a smaller χ^2, ~20. Therefore, the association between lower ACE expression and schizophrenia association is likely to be identified even with small/moderate selection, if there is a true association. We also repeated simulations using a disease population prevalence of 1%, which did not change our conclusion.

MR Power Calculations (eTable 17 - 19)

eTable 17. Association Between ACE Brain Expression and Schizophrenia Risk

Genes for which there was an association with at least nominal significance are marked by a single asterisk(*), while those that passed multiple testing correction (p<0.0035) are marked by a double asterisk (**)

Brain Tissue	Top-associated SNP	r-square	F-statistic	eQTL Sample Size	odds ratio with 80% power
PsychENCODE prefrontal cortex**	rs6504163	0.031	45	1387	1.11
Cerebellum**	rs4308	0.141	39	241	1.05
Anterior cingulate cortex BA24**	rs6504163	0.115	23	176	1.05
Frontal Cortex BA9**	rs4291	0.144	35	209	1.05
Cerebellar Hemisphere**	rs4292	0.108	26	215	1.06
Cortex**	rs4968783	0.102	29	255	1.06
Hippocampus*	rs4309	0.053	11	197	1.08
Putamen basal ganglia	rs4344	0.036	8	205	1.10
Nucleus accumbens basal ganglia	rs12451780	0.033	8	246	1.10
Amygdala	rs2727280	0.053	8	152	1.08
Caudate basal ganglia	rs12950385	0.024	6	246	1.12
Spinal cord cervical c-1	rs75995183	0.068	12	159	1.07
Hypothalamus	rs4968775	0.029	6	202	1.11
Substantia nigra	rs113806292	0.091	14	139	1.06

eTable 18. Association Between Blood Gene Expression and Schizophrenia Risk

For 17 out of the 22 genes we have 80% power to detect (at 5% significance level) a smaller causal effect of gene expression on schizophrenia than that observed for ACE (OR = 1.75). Genes for which there was an association with at least nominal significance are marked by a single asterisk (*), while those that passed multiple testing correction ($p<7.6x10^{-4}$) are marked by double asterisks (**)

Gene	SNP	Frequency of effect allele	F-statistic	r-square	Sample size	Ratio cases:controls	Causal effect with 80% power
SLC12A1	rs964611	0.15	37527	0.044	105318	1.59	1.09
MTOR	rs4845986	0.27	756	0.024	105318	1.59	1.12
CA4	rs34820870	0.032	496	0.016	105318	1.59	1.15
SLC12A2	rs17764730	0.25	387	0.012	105318	1.59	1.18
KCNH2	rs4725984	0.36	310	0.0098	105318	1.59	1.20
CACNB2	rs72787951	0.45	296	0.0093	105318	1.59	1.20
SLC16A1	rs7169	0.44	241	0.0076	105318	1.59	1.23
ADRB2	rs1432622	0.45	239	0.0076	105318	1.59	1.23
KCNJ11*	rs2074310	0.36	166	0.0053	105318	1.59	1.28
P4HA1	rs6480668	0.064	139	0.0044	105318	1.59	1.31
CACNA2D2	rs62260815	0.11	123	0.0039	105318	1.59	1.33
CACNA1D	rs9830632	0.27	94	0.003	105318	1.59	1.38
CYP17A1**	rs284856	0.4	91	0.0029	105318	1.59	1.39
NISCH**	rs187084	0.4	92	0.0029	105318	1.59	1.39
ACE**	rs4277405	0.38	72	0.0023	105318	1.59	1.45
SLC12A3	rs56228609	0.31	69	0.0022	105318	1.59	1.46
NR3C1*	rs2398631	0.21	54	0.0017	105318	1.59	1.54
ATP1A1	rs6704439	0.27	21	0.00067	105318	1.59	1.98
CACNA1H	rs71384658	0.29	19	6.00E-04	105318	1.59	2.06
SCNN1D	rs4970365	0.068	15	0.00049	105318	1.59	2.23
CA12	rs12909041	0.19	14	0.00047	105318	1.59	2.27
AOC3	rs513495	0.2	10	0.00033	105318	1.59	2.65

eTable 19. Association Between Blood Gene Expression and Bipolar Disorder

19 out of the 22 genes had 80% power to detect (at 5% significance level) a causal effect smaller than that observed for *CYP17A1* blood expression on bipolar disorder (OR = 1.95), which was significant in the bipolar disorder analysis.

Genes for which there was an association with at least nominal significance are marked by a single asterisk (*), while those that passed multiple testing correction ($p<7.6x10^{-4}$) are marked by double asterisks (**)

Gene	SNP	Frequency of effect allele	F-statistic	r-square	Sample size	Ratio cases:controls	Causal effect with 80% power
SLC12A1	rs964611	0.15	37527	0.044	51710	1.54	1.13
MTOR	rs4845986	0.27	756	0.024	51710	1.54	1.18
CA4	rs34820870	0.032	496	0.016	51710	1.54	1.22
SLC12A2	rs17764730	0.25	387	0.012	51710	1.54	1.26
KCNH2	rs2072412	0.27	324	0.01	51710	1.54	1.29
CACNB2	rs72787951	0.45	296	0.0093	51710	1.54	1.30
ADRB2	rs2082395	0.45	248	0.0078	51710	1.54	1.33
CACNA1H	rs117177120	0.051	239	0.0076	51710	1.54	1.34
SLC16A1	rs7169	0.44	241	0.0076	51710	1.54	1.34
KCNJ11	rs2074310	0.36	166	0.0053	51710	1.54	1.41
P4HA1	rs6480668	0.064	139	0.0044	51710	1.54	1.46
SCNN1D	rs11804831	0.19	129	0.0041	51710	1.54	1.48
CACNA2D2	rs62260815	0.11	123	0.0039	51710	1.54	1.50
CACNA1D	rs9830632	0.27	94	0.003	51710	1.54	1.58
CYP17A1**	rs284856	0.4	91	0.0029	51710	1.54	1.60
NISCH**	rs187084	0.4	92	0.0029	51710	1.54	1.60
ACE*	rs4277405	0.38	72	0.0023	51710	1.54	1.69
SLC12A3	rs56228609	0.31	69	0.0022	51710	1.54	1.71
NR3C1	rs4912908	0.21	55	0.0017	51710	1.54	1.84
ATP1A1	rs6704439	0.27	21	0.00067	51710	1.54	2.65
CA12	rs146993121	0.0016	15	0.00048	51710	1.54	2.72
AOC3	rs55751736	0.0081	14	0.00046	51710	1.54	2.72

eTable 20. Association Between Blood Gene Expression and Major Depression

Genes for which there was an association with at least nominal significance are marked by a single asterisk (*), while those that passed multiple testing correction ($p<7.6x10^{-4}$) are marked by double asterisks (**)

Gene	SNP	Frequency of effect allele	F-statistic	r-square	Sample size	Ratio cases:controls	Causal effect with 80% power
SLC12A1*	rs964611	0.15	37527	0.044	480359	2.55	1.04
MTOR	rs4845986	0.27	756	0.024	480359	2.55	1.06
CA4	rs34820870	0.032	496	0.016	480359	2.55	1.07
SLC12A2	rs17764730	0.25	387	0.012	480359	2.55	1.09
KCNH2	rs2072412	0.27	324	0.01	480359	2.55	1.09
CACNB2	rs72787951	0.45	296	0.0093	480359	2.55	1.10
ADRB2	rs2082395	0.45	248	0.0078	480359	2.55	1.11
SLC16A1	rs7169	0.44	241	0.0076	480359	2.55	1.11
CACNA1H	rs7201107	0.061	208	0.0066	480359	2.55	1.12
KCNJ11	rs2074310	0.36	166	0.0053	480359	2.55	1.13
P4HA1	rs6480668	0.064	139	0.0044	480359	2.55	1.14
SCNN1D	rs11804831	0.19	129	0.0041	480359	2.55	1.15
CACNA2D2	rs62260815	0.11	123	0.0039	480359	2.55	1.15
CACNA1D	rs9830632	0.27	94	0.003	480359	2.55	1.18
CYP17A1*	rs284856	0.4	91	0.0029	480359	2.55	1.18
NISCH*	rs187084	0.4	92	0.0029	480359	2.55	1.18
ACE*	rs4277405	0.38	72	0.0023	480359	2.55	1.21
SLC12A3	rs56228609	0.31	69	0.0022	480359	2.55	1.21
NR3C1	rs4912908	0.21	55	0.0017	480359	2.55	1.24
ATP1A1	rs6704439	0.27	21	0.00067	480359	2.55	1.41
CA12	rs12909041	0.19	14	0.00047	480359	2.55	1.51
AOC3	rs66832424	0.045	11	0.00035	480359	2.55	1.62

eReferences.

1. Swerdlow DI, Kuchenbaecker KB, Shah S, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol*. 2016;45(5):1600-1616. doi:10.1093/ije/dyw088
2. Davies NM, Holmes M V, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018;362:k601. doi:10.1136/bmj.k601
3. Staiger D, Stock JH. Instrumental Variables Regression with Weak Instruments. *Econometrica*. 1997;65(3):557. doi:10.2307/2171753
4. Bowden J, Del Greco M. F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol*. 2016;45(6):1961-1974. doi:10.1093/ije/dyw220
5. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481-487. doi:10.1038/ng.3538
6. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. Williams SM, ed. *PLoS Genet*. 2014;10(5):e1004383. doi:10.1371/journal.pgen.1004383
7. Lam M, Chen C-Y, Li Z, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet*. 2019;51(12):1670-1678. doi:10.1038/s41588-019-0512-x
8. Abdollahi MR, Huang S, Rodriguez S, et al. Homogeneous Assay of rs4343, an ACE I/D Proxy, and an Analysis in the British Women's Heart and Health Study (BWHHS). *Dis Markers*. 2008;24(1):11-17. doi:10.1155/2008/813679
9. Chung C-M, Wang R-Y, Chen J-W, et al. A genome-wide association study identifies new loci for ACE activity: potential implications for response to ACE inhibitor. *Pharmacogenomics J*. 2010;10(6):537-544. doi:10.1038/tpj.2009.70
10. Kauwe JSK, Bailey MH, Ridge PG, et al. Genome-Wide Association Study of CSF Levels of 59 Alzheimer's Disease Candidate Proteins: Significant Associations with Proteins Involved in Amyloid Processing and Inflammation. Geschwind DH, ed. *PLoS Genet*. 2014;10(10):e1004758. doi:10.1371/journal.pgen.1004758
11. Deming Y, Xia J, Cai Y, et al. Genetic studies of plasma analytes identify novel potential biomarkers for several complex traits. *Sci Rep*. 2016;6(1):18092. doi:10.1038/srep18092
12. Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *Int J Epidemiol*. 2019;48(3):691-701. doi:10.1093/ije/dyy202