

Supplementary Material

The effects of testosterone and SHBG on lung function in males and females: A multivariable MR study

Diana A. van der Plaat¹, Alexandra Lenoir²⁻³, Shyamali C. Dharmage⁴, James F. Potts¹, Francisco Gómez-Real⁵⁻⁶, Seif O. Shaheen⁷, Deborah Jarvis¹, Cosetta Minelli^{1}, Bénédicte Leynaert^{8*}*

¹ National Heart and Lung Institute (NHLI), Imperial College London, London, United Kingdom

² Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

³ Gesundheitsamt Fürstentfeldbruck, Fürstentfeldbruck, Germany

⁴ Allergy and Lung Health Unit, School of Population and Global Health, University of Melbourne, Melbourne, Australia.

⁵ Department of Clinical Science, University of Bergen, Bergen, Norway

⁶ Department of Gynecology and Obstetrics, Haukeland University Hospital, Bergen, Norway

⁷ Wolfson Institute of Population Health, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

⁸ Université Paris-Saclay, UVSQ, Université Paris-Sud, Inserm, Équipe d'Épidémiologie Respiratoire Intégrative, CESP, 94807, Villejuif, France

* shared last author

Correspondence to: Diana A. van der Plaat, PhD, National Heart and Lung Institute (NHLI), Imperial College London, London, United Kingdom (e-mail: d.van-der-plaat@imperial.ac.uk).

Index

Index.....	2
Supplementary Methods	3
UK Biobank data.....	3
GWA analysis for selection of genetic instrument.....	4
Mendelian Randomisation (MR).....	5
MR assumptions.....	5
Pleiotropy.....	5
G-X and G-Y estimation.....	6
Multivariable Mendelian Randomisation (MVMR) methods	6
Pleiotropy detection	8
MVMR – Secondary analyses.....	8
Subgroup analyses	8
Longitudinal MVMR analyses.....	9
Observational analyses	9
Results presentation	9
Appendix 1. Variable names and UKB variable numbers used in this study.	10
References	11
Supplementary Figures	13
Figure S1. The workflow of the study divided over four main phases.	13
Figure S2. Manhattan plots – total testosterone (TT)	14
Figure S3. Manhattan plots – sex hormone-binding globulin (SHBG)	15
Figure S4. Leave-One-Out RE-IVW analysis plots in males.	16
Figure S5. Leave-One-Out RE-IVW analysis plots in females.	17
Supplementary Tables	18
Table S1. Overview of identification and replication of number of SNPs associated with hormone levels in UK biobank.....	18
Table S2: MVMR G-X and G-Y estimates.....	18
Table S3: Results of the MVMR analysis of the effect of SHBG and testosterone on FVC, FEV ₁ , and FEV ₁ /FVC separately in males and females.....	18
Table S4: Results of the MVMR analysis excluding “outlier” SNPs.....	18
Table S5: Results from PhenoScanner for the pleiotropic SNPs identified by MVMR-PRESSO.	18
Table S6: Results of subgroup MVMR analyses by obesity, physical activity, menopausal status in females, and age (>50 years) in males.....	18
Table S7: Results of the MVMR analysis of the effect of SHBG and testosterone on FVC, FEV ₁ , and FEV ₁ /FVC decline separately in males and females.....	18
Table S8: Results of the observational analyses.....	18

Supplementary Methods

In UK Biobank (UKB), we first performed a Genome-Wide Association (GWA) analysis to identify SNPs associated with TT and SHBG to be used as genetic instruments for our MVMR analyses. Subsequently, the SNP-hormone and SNP-lung function estimates were used for our one-sample MVMR to investigate the effect of TT and SHBG on lung function. We applied different MVMR analyses to test the robustness of the findings, investigated possible effect modification in population subgroups, and assessed the effects of TT and SHBG on lung function decline using MVMR. Finally, for comparison, we performed conventional linear regression analyses of TT and SHBG on lung function. All analyses were performed using UKB data and were stratified by sex. Figure S1 describes this workflow with some further details.

UK Biobank data

We included subjects from the publicly available UKB, which is a large, multicentre, population-based prospective study across 22 centres in the United Kingdom [1]. At baseline (2006-2010), over 500,000 participants aged 39-70 were recruited, completed online questionnaires and underwent physical measurements as well as blood sampling. UK Biobank has received ethics approval from the National Health Service National Research Ethics Service (Ref 11/NW/0382) and all participants provided written consent.

Mid 2019, data became available on a wide range of biomarkers from blood and urine collected at baseline, including serum levels of Total Testosterone (TT) and sex hormone-binding globulin (SHBG). Serum TT and SHBG levels were measured using a chemiluminescent immunoassay (Beckman Coulter Ltd, UK, UniCel Dcl 800) with results expressed in nanomole per litre (nmol/L).

Three continuous lung function outcomes were investigated: FEV₁ (mL), FVC (mL), and FEV₁/FVC (%). Only white subjects with spirometry data that passed the “best measure” criteria as developed by the UK BiLEVE study provided by Professor Martin Tobin at the University of Leicester (UKB data-fields 20150 and 20151) were included.[2] The highest measure for FEV₁ and FVC (Fields 3062 and 3063) was selected if it was “acceptable” according to corresponding acceptability field (Field 3061). A blow was deemed acceptable if recorded as no problems or “USER_ACCEPTED”.

Genotyping was performed in three batches using the UK BiLEVE Axiom Array (first 50,000 subjects) and the UK Biobank Axiom Array (data released at two time points). Details on genotyping and imputation methods are available elsewhere [3].

Of the 502,536 participants, we excluded subjects with missing genetic (N=6,678) or biomarker (N=11,142) data (Figure 1). The final dataset (N=484,716) was used to create two datasets, which were subsequently split based on sex: the hormones and the lung function datasets.

The hormone dataset was subsequently randomly divided into a “discovery hormone dataset” (2/3 of the sample: N=323,144) used in the GWA analyses and a “validation hormone dataset” (1/3 of the sample: N=161,572) used for the replication of GWA findings and for the multivariable Mendelian Randomisation (MVMR) analyses to estimate the effect of the genetic variants on TT and SHBG.

For the “lung function dataset”, another 124,890 subjects were excluded because of missing good quality lung function data (see above), leading to a final dataset with 341,826 subjects. This dataset was used for the MVMR analyses to estimate the effect of the genetic variants on lung function and for the observational analyses. The “best measure” variables were limited to white participants, and therefore only these were included in the “lung function dataset”, while the hormone datasets also included a small proportion of participants from other ethnicities (5.4%). In addition, there is about a 70% overlap between the sex hormone and lung function datasets.

See appendix 1 for a list of variables and their UKB variable number that were used in this study.

GWA analysis for selection of genetic instrument

To identify genetic instruments for TT and SHBG, we performed GWA analyses in UKB, separately for males and females, on natural log-transformed TT and SHBG levels, as the residuals were not normally distributed. We used linear mixed-models implemented within BOLT-LMM [4], which account for relatedness and fine-scale population structure, and we adjusted for age, genotyping batch/array, and centre.

Independent genome-wide significant SNPs were selected based on the “discovery hormone dataset” using the online platform FUMA GWAS [5], with the following settings: p-value lead SNP = 5×10^{-8} , p-value cut-off = 0.05, r^2 threshold = 0.05, reference population = UKB/release2b 10k European, minimum MAF = 0, distance between linkage disequilibrium (LD) blocks > 10,000 kb. These independent SNPs were considered replicated if in the “validation hormone dataset” the direction of effect was the same and their Bonferroni-corrected one-side p-value was statistically significant ($p < 0.05/\text{number of SNPs per outcome and sex}$).

In total, we identified 674 independent SNPs and were able to replicate 450 SNPs (Supplementary Table S1 and see Supplementary Figures S2 and S3 for the Manhattan plots). For TT, we replicated 92 SNPs in males and 63 SNPs in females, which explained about 8.1% and 3.6% of the variance in TT levels, respectively. For SHBG, we replicated 213 SNPs in males and 153 SNPs in females, which explained about 20.5% and 13.0% of the variance in SHBG levels, respectively. See Supplementary Table S2 for details on which SNPs were associated with which trait and the annotated genes. Of the replicated 450 SNPs, 43 were associated with both TT and SHBG and there was at least one IV associated with both traits in all male and female analyses.

Interestingly, a larger proportion of SNPs was shared between TT and SHBG in males compared to females, and this could both have a statistical and a biological reason.

First, fewer signals might have been detected for TT in females due lower statistical power, as about 20% of females had TT levels below the detection limit, compared with only 5% of males (for SHBG, 12-14% of data were missing for both). Moreover, the variances of the two hormones in the study population were larger in females than males, particularly for TT (TT: 0.22 in females vs. 0.11 in males; SHBG: 0.25 vs. 0.18), and this would lower the power of the GWAS in females compared with males thus again reducing the number of detectable signals.

Second, a plausible biological reason might be that the sources of TT production and the levels of TT and SHBG are different in males and females. In females about 25% of TT is produced in the ovary, another 25% in the adrenal gland and the remaining 50% in peripheral tissues. The main precursors of TT are androstenedione in ovaries and Dehydroepiandrosterone (DHEA) in the adrenal glands. In addition, in post-menopausal women, the production of steroid hormones by the ovaries is severely decreased (and >60% of UKB females are post-menopausal). In males, the majority of TT is produced in the testes. The major source of SHBG in males and females is the liver. These differences could indicate different underlying genetic determinants and a lower overlap between the two in the different sexes.

Lastly, in general, power to detect an effect in MR improves with a higher percentage of the variance explained (R^2) and a larger sample size.[6] As a reference, SNPs selected as genetic instruments in our previous MR studies on age at menarche and age at menopause explained 7.4% and 6%, respectively [7,8].

Mendelian Randomisation (MR)

MR assumptions

While Mendelian Randomisation (MR) is not affected by traditional confounding or reverse causation, its validity relies on assumptions (Figure E1) [9]. The three key assumptions of MR are:

- 1) Relevance assumption: SNPs are associated with exposure or factor of interest (TT and SHBG levels). In the current study we identified SNPs associated with TT and SHBG levels.
- 2) Independent assumption: the association between SNP and outcome is NOT confounded.
- 3) Exclusion restriction assumption: SNPs affect outcome (lung function) only through the factor of interest and not through independent pathways (i.e. absence of pleiotropy).

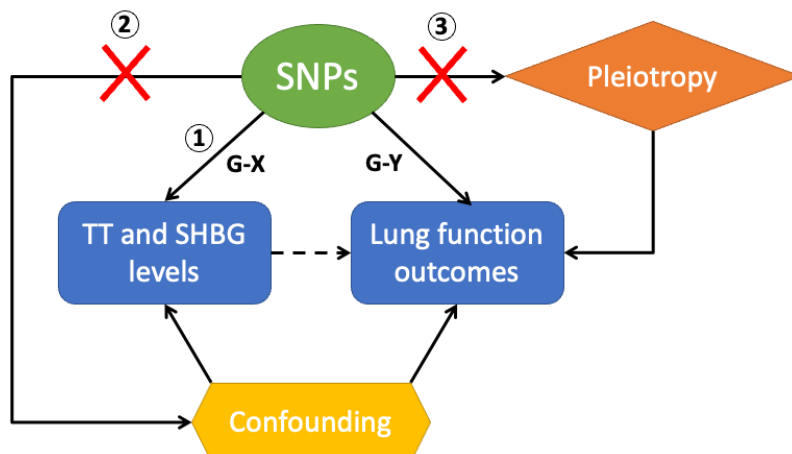


Figure E1: Schematic overview of Mendelian randomisation (MR). For each genetic variant (SNP), we estimated the G-X (association of the SNP with Total Testosterone (TT) and SHBG) and the G-Y (association of the SNP with lung function outcomes) associations. MR assumptions: 1) relevance assumption; 2) independent assumption; and 3) exclusion restriction assumption.

Pleiotropy

The most problematic assumption is the absence of ("horizontal") pleiotropy [10]. There are two types of pleiotropy, vertical and horizontal pleiotropy (Figure E2).

Vertical pleiotropy occurs when a genetic variant has a phenotypic effect on a trait (the factor of interest), which in turn has its own effect on another trait which increases the risk of disease (cascade of events). For example, a SNP in the gene *FTO* influences BMI, which is associated with increased blood pressure and which in turn increases the risk for coronary artery disease. In the case of vertical pleiotropy, there is a causal pathway between the factor of interest and the outcome.

Horizontal pleiotropy occurs when a genetic instrument (SNP) affects multiple traits and thereby affect the outcome through pathways that are independent from the factor of interest. For example, if a SNP associates both with the factor of interest and with an independent risk factor that both independently affect the outcome of interest, the MR estimate is the combined effect of these risk factors, rather than of the factor of interest alone. This type of pleiotropy can therefore bias the MR estimate. For example, smoking, a typical confounder in observational studies, could affect MR through pleiotropy if (but only if) the genetic instruments used for TT and SHBG were also associated with smoking.

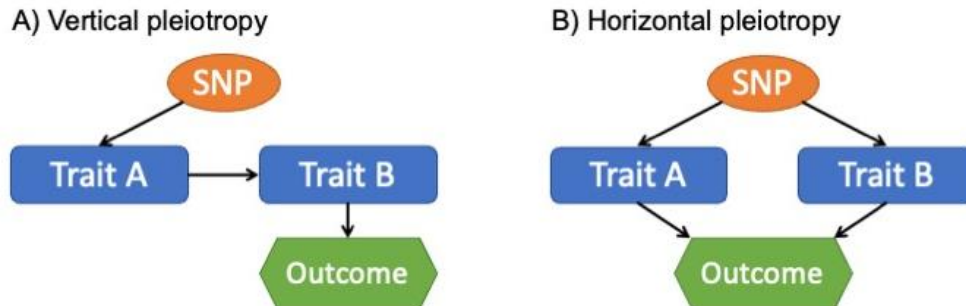


Figure E2: Depiction of vertical and horizontal Pleiotropy. A) Vertical pleiotropy: valid MR analysis as genetic instrument (SNP) affects the outcome via risk factor of interests (Trait A) in a single pathway. B) Horizontal pleiotropy: can cause bias in MR as genetic instrument affects the outcome via independent pathways (traits).

G-X and G-Y estimation

For the MR analyses, we estimated the causal effects of TT and SHBG on lung function derived from the SNP-hormone (G-X) and the SNP-lung function (G-Y) association estimates, both derived from UKB data.

As G-X estimates (Figure 1), separately by sex, we used the effect estimates of the SNPs on SHBG and TT from the GWA replication analysis in UKB and removed SNPs that were in LD ($r^2 > 0.1$) between the SNP lists for TT and SHBG as they were jointly modelled. In total, 178 independent SNPs (67 for TT and 138 for SHBG) were included in the male MVMR analyses, and 174 (55 for TT and 120 for SHBG) in the female analyses, all with F-statistic > 10 to avoid weak instrument bias.[11]

To obtain the G-Y estimates (Figure 1), again separately in males and females, we estimated the effect of each SNP on FEV₁, FVC and FEV₁/FVC in UKB using BOLT-LMM and adjusting for age, age², height, genotyping batch/array, and centre in the “lung function dataset”. We adjusted for height as height is a strong predictor of lung function, particularly FVC and FEV₁, with values standardised for height (as well as age and sex) being those of clinical interest.

The G-X and G-Y estimates for all SNPs separately by sex can be found in Supplementary Table S2.

Multivariable Mendelian Randomisation (MVMR) methods

As levels of SHBG and TT are partly determined by the same genes (independence assumption), we used multivariable MR (MVMR), where TT and SHBG are modelled together to estimate their causal effects on lung function (FEV₁, FVC and FEV₁/FVC), independent of each other. We derived the MVMR effect estimate from the SNP-hormone (G-X) and the SNP-lung function (G-Y) association estimates.

As G-X estimates, we used the effect estimates of our GWA replication analysis and removed SNPs that were in LD ($r^2 > 0.1$) between the SNP lists of TT and SHBG. In total, 178 independent SNPs (67 for TT and 138 for SHBG) were included in the male MVMR analyses, and 174 (55 for TT and 120 for SHBG) in the female analyses, all with F-statistic > 10 . To obtain the G-Y estimates, again stratified by sex, we estimated the effect of each SNP on FEV₁, FVC and FEV₁/FVC in UKB using BOLT-LMM and adjusting for age, age², height, genotyping batch/array, and centre. G-X and G-Y estimates for all SNPs and by sex are reported in Supplementary Table S2.

Several methods have been developed to detect and control for (horizontal) pleiotropy [12]. We applied six MVMR methods that make different assumptions about the pleiotropic effects:

- Methods assuming no pleiotropy:

- 1) Two-Stage Least Squares (2SLS): This is the only method we used that requires individual-level data, and is the gold standard for a one-sample MR in the absence of pleiotropy. It consists of two regression stages: a first-stage regression of the exposures on the instrumental variables (SNPs), and a second-stage regression of the outcome on the fitted values of each of the exposures from the first stage [13]. In the context of a multivariable MR, the first-stage regression is a multivariate multiple regression (multiple dependent variables and multiple independent variables), while the second-stage regression is a univariate multiple regression (a single dependent variable and multiple independent variables) [14]. All 2SLS analyses were performed in the “lung function dataset” which has complete data on TT/SHBG, lung function and the genetic variants.
- 2) Fixed-effect inverse-variance weighted (FE-IVW) meta-analysis: This is the most powerful two-sample MR method, but, similarly to the 2SLS, assumes no pleiotropy [15]. Here G-X and G-Y estimates are obtained for each SNP, and SNP-specific MR estimates are then derived using the Wald estimator (ratio of G-Y over G-X), with standard error calculated using the Delta method [16]. The estimate of the causal effect of interest is obtained by pooling the SNP-specific MR estimates using a fixed-effect meta-analysis model [15]. The IVW approach, both FE-IVW and RE-IVW (see next paragraph), has been extended to multivariable MR [14,17].

- Methods that account for pleiotropy but make different assumptions:

- 3) Random-effects inverse-variance weighted (RE-IVW) meta-analysis: This method is performed similarly to the FE-IVW, but with a random-effects, instead of a fixed-effect, meta-analysis model to allow for heterogeneity across SNP-specific MR estimates, which represents a statistical measure of pleiotropy. RE-IVW assumes that pleiotropic effects across SNPs are random (balanced pleiotropy), and that their magnitude is independent of the magnitude of the corresponding G-X effects (InSIDE assumption) [15]. This method, which has been extended to multivariable MR [17], is sensitive to outliers and directional pleiotropy.
- 4) Robust regression (MM-estimation): It is a natural extension of IVW performing robust regression using MM-estimation along with Tukey’s bisquare objective function [18]. It provides robustness to outliers, when there are a relatively small number of invalid instruments, by effectively capping residuals of a certain magnitude. This method is easily extended to multivariable MR, as described by Koller and Stahel [19].
- 5) Weighted median based estimation: It assumes that more than 50% of the information contributing to the analysis comes from genetic variants that are valid (i.e. not pleiotropic) [20]. It has been extended to multivariable MR, as described Grant *et al.* [17].
- 6) MR-Egger regression: G-Y estimates for the individual SNPs are regressed on their G-X estimates, with the intercept representing the overall pleiotropy and the slope the adjusted MR estimate [21]. This method has been extended for use in multivariable MR by Rees *et al.* [22], who show that this multivariable version has advantages over the original (univariable) MR-Egger in terms of both plausibility of the underlying assumptions and statistical power.

All MVMR analyses were performed using R (version 4.0.2). In particular, we used: the *AER* package [23] for method 1 (2SLS); the *MendelianRandomization* package [24] for methods 2, 3 and 6 (FE-IVW, RE-IVW and MR-Egger); and specific R codes provided by Grant *et al.* [17] for methods 4 and 5.

Pleiotropy detection

To detect possible pleiotropy, we assessed the between-instrument heterogeneity as a marker of pleiotropy using the Q test p-value in IVW, the Sargan p-value in 2SLS, and the intercept p-value in MR-Egger.

To assess the validity of our genetic instruments, took two approaches to identify and remove potential pleiotropic SNPs which were subsequently excluded in sensitivity analyses;

- 1) a statistical approach using MVMR-PRESSO [25] to detect “outliers”. The MVMR-PRESSO outlier test evaluates the presence of specific horizontal pleiotropic outlier variants by using the observed and expected distributions of the tested variant. It computes an empirical p-value adjusted for multiple testing using the Bonferroni correction. Outliers were detected for each sex and each of the three lung function outcomes. Per sex, we combined all the outliers detected per outcome and excluded these in follow-up analyses.
- 2) using PhenoScanner [26] to identify SNPs associated with potential sources of pleiotropy identified a priori: smoking, weight-related traits, diabetes, insulin, leptin and adiponectin (n= 36 SNPs).

To better understand the relevance of potential pleiotropic pathways, for all detected outliers, we searched for association of these SNPs with any other traits in previous GWA studies at genome-wide significance ($p < 5 \times 10^{-8}$) using PhenoScanner [26]. We also performed the Leave-One-Out RE-IVW analysis to identify the most influential pleiotropic SNPs [27]. In a Leave-One-Out analysis, one variant is removed from the MVMR analysis, and the causal effect is re-estimated. This is done one-by-one for each of the identified outlier SNPs by MVMR-PRESSO. If there is one genetic variant that is particularly strongly associated with the exposure, it can have a particularly large horizontal pleiotropic effect, thereby dominating the estimate of the causal effect.

MVMR – Secondary analyses

Subgroup analyses

We investigated possible effect modification for both TT and SHBG by:

- 1) Obesity, as several studies have reported interactions between sex-hormones and obesity on asthma [28] and lung function [8]. Obesity was defined as a body mass index (BMI) $\geq 30 \text{ kg/m}^2$.
- 2) Physical activity, as randomised controlled trials in COPD patients have suggested that testosterone supplementation might be more effective when combined with exercise training [29–31]. Moderate physical activity was defined as ≥ 10 minutes of moderate physical activity for ≥ 4 days/week based on the question: “In a typical week, on how many days did you do 10 minutes or more of moderate physical activities like carrying light loads, cycling at normal pace? (Do not include walking)”.
- 3) Menopausal status in females and age (below/above 50) in males, as the effects of the SHBG and testosterone might vary with age and age-related changes in the hormonal-milieu. Menopausal status was based on the question: “Have you had your menopause (periods stopped)”. Females who responded “yes” were categorised as post-menopausal and who responded “No” and were aged below 60 years were categorised as pre-menopausal. Menopausal status broadly classified females in two age groups, ≤ 50 (median age at menopause in UKB) and > 50 ; we therefore performed a similar subgroup analysis in males by age, below and above 50 years.

All subgroup analyses were performed separately by sex; "weak" SNPs (F-statistic ≤ 10) were excluded, and pleiotropy was dealt with as in the main analyses.

Longitudinal MVMR analyses

We also performed MVMR to assess the effect of TT and SHBG on lung function decline using longitudinal data from UKB. We included 6,812 males and 7,470 females who had data at baseline and at second follow-up until March 2019, with a mean time between the two visits of 8.3 years. Please see Lenoir *et al.*[29] for more details on how high-quality longitudinal spirometry was obtained in UKB. We obtained the G-Y estimates stratified by sex by estimating the effect of each SNP on yearly decline of FEV₁, FVC and FEV₁/FVC in UKB using the same model as in the main analysis. Yearly decline was calculated by subtracting the second lung function reading from the reading at baseline and dividing it by the time (years) between the readings. Not all the 450 SNPs passed QC (minor allele frequency >1%, missingness per marker <1.5% and Hardy-Weinberg equilibrium $p < 1 \times 10^{-6}$) and therefore less SNPs were included in these MVMR analyses (see Supplementary Table S2 for the G-Y estimates of the included SNPs).

Observational analyses

To compare our MVMR results with finding from observational analyses, we performed multiple linear regression to estimate the independent association of natural log-transformed TT and SHBG levels (TT and SHBG included in the same model) on the three lung function outcomes, separately per sex. The model was adjusted for the same variables as in the UKB study by Lenoir *et al.*[29]; age, age², height, BMI, smoking status (never/former/current), pack years, Townsend Deprivation Index, number of days per week of moderate physical activity, and fresh fruit intake (0-1, 2-3, and ≥ 4 portions), as well as time of day of the appointment (before or after 12am) to control for diurnal variability of hormone levels.

Results presentation

To help interpretation of the results and figures (based on log-transformed hormone values, i.e. linear-log models), we calculated the average regression coefficient (beta) across MVMR methods and multiplied it by $\ln(1.1)$ to display the effect of a 10% increase in SHBG or TT on lung function. We did not do this for the effect estimates reported in the Supplementary Tables, which represent the effect on lung function of a 1-unit increase in log-transformed SHBG and TT.

Appendix 1. Variable names and UKB variable numbers used in this study.

Variable Name	UKB Variable number
Sex	31
UK Biobank assessment centre	54
Townsend Deprivation Index	189
Age Completing Full Time Education	845
Moderate physical activity	884
Fresh fruit intake	1309
Menopause	2724
Bilateral oophorectomy	2834
Time of blow measurement	3060
Ever had hysterectomy	3591
Qualifications	6138
Height	12144
Smoking Status	20116
FEV ₁ best measure	20150
FVC best measure	20151
Pack years	20161
Body mass index (BMI)	21001
Weight	21002
Age when attended assessment centre	21003
Array	22000
Principle Components (PCs)	22009
Recommended genomic analysis exclusions	22010
Genetic relatedness pairing	22011
Genetic relatedness factor	22012
Doctor diagnosed asthma	22127

The UK Biobank showcase of resources including more information on these variables can be found online at: <https://biobank.ctsu.ox.ac.uk/crystal/index.cgi>

References

1. Sudlow C, Gallacher J, Allen N, 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 Med* 2015;12:e1001779. doi:10.1371/journal.pmed.1001779
2. Wain L V, Shrine N, Miller S, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* 2015;3:769–81. doi:10.1016/S2213-2600(15)00283-0
3. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–9. doi:10.1038/s41586-018-0579-z
4. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015;47:284–90. doi:10.1038/ng.3190
5. Watanabe K, Taskesen E, van Bochoven A, et al. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017;8:1826. doi:10.1038/s41467-017-01261-5
6. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol* 2014;43:922–9. doi:10.1093/ije/dyu005
7. Minelli C, van der Plaat DA, Leynaert B, et al. Age at puberty and risk of asthma: A Mendelian randomisation study. *PLoS Med* 2018;15:e1002634. doi:10.1371/journal.pmed.1002634
8. van der Plaat DA, Pereira M, Pesce G, et al. Age at menopause and lung function: a Mendelian randomisation study. *Eur Respir J* 2019;54. doi:10.1183/13993003.02421-2018
9. Burgess S, Thompson SG. *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation*. CRC Press 2015.
10. Davies NM, Holmes M v., Davey Smith G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ (Online)* 2018;362:k601. doi:10.1136/bmj.k601
11. Lawlor DA, Harbord RM, Sterne JAC, et al. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;27:1133–63. doi:10.1002/sim.3034
12. Sheehan NA, Didelez V, Burton PR, et al. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med* 2008;5:1205–10. doi:10.1371/journal.pmed.0050177
13. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res* 2017;26:2333–55. doi:10.1177/0962280215597579
14. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015;181:251–60. doi:10.1093/aje/kwu283
15. Bowden J, Del Greco M F, Minelli C, et al. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 2017;36:1783–802. doi:10.1002/sim.7221
16. Thompson JR, Minelli C, del Greco M F. Mendelian Randomization using Public Data from Genetic Consortia. *Int J Biostat* 2016;12. doi:10.1515/ijb-2015-0074
17. Grant AJ, Burgess S. Pleiotropy robust methods for multivariable Mendelian randomization. *Stat Med* 2021;40:5813–30. doi:https://doi.org/10.1002/sim.9156
18. Rees JMB, Wood AM, Dudbridge F, et al. Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates. *PLoS One* 2019;14:e0222362. doi:10.1371/journal.pone.0222362

19. Koller M, Stahel WA. Sharpening Wald-type inference in robust regression for small samples. *Comput Stat Data Anal* 2011;55:2504–15. doi:10.1016/j.csda.2011.02.014
20. Bowden J, Davey Smith G, Haycock PC, et al. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* 2016;40:304–14. doi:10.1002/gepi.21965
21. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512–25. doi:10.1093/ije/dyv080
22. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat Med* 2017;36:4705–18. doi:10.1002/sim.7492
23. Kleiber C, Zeileis A. *Applied Econometrics with R*. Springer New York 2008.
24. Yavorska OO, Burgess S. MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017;46:1734–9. doi:10.1093/ije/dyx034
25. Verbanck M, Chen CY, Neale B, et al. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018;50:693–698. doi:10.1038/s41588-018-0099-7
26. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics* 2016;32:3207–9. doi:10.1093/bioinformatics/btw373
27. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res* 2020;4:186. doi:10.12688/wellcomeopenres.15555.2
28. Han Y-Y, Forno E, Celedón JC. Sex Steroid Hormones and Asthma in a Nationwide Study of U.S. Adults. *Am J Respir Crit Care Med* 2020;201:158–66. doi:10.1164/rccm.201905-0996OC
29. Lenoir A, Fuertes E, Gómez-Real F, et al. Lung function changes over 8 years and testosterone markers in both sexes: UK Biobank. *ERJ Open Res* 2020;6. doi:10.1183/23120541.00070-2020
30. Baillargeon J, Urban RJ, Zhang W, et al. Testosterone replacement therapy and hospitalization rates in men with COPD. *Chron Respir Dis* 2019;16:1479972318793004. doi:10.1177/1479972318793004
31. Pan L, Wang M, Xie X, et al. Effects of anabolic steroids on chronic obstructive pulmonary disease: a meta-analysis of randomised controlled trials. *PLoS One* 2014;9:e84855. doi:10.1371/journal.pone.0084855

Supplementary Figures

Figure S1. The workflow of the study divided over four main phases.

In UK Biobank (UKB), we first performed a Genome-Wide Association (GWA) analysis to identify SNPs associated with TT and SHBG to be used as instruments in our MVMR analyses. Estimates of the SNP-hormone effects (G-X) from the replication stage of the GWA, together with estimates of the SNP-lung function effects (G-Y) obtained in UKB for the identified SNPs, were used for our one-sample MVMR to investigate the effect of TT and SHBG on lung function. We applied different MVMR analyses to test the robustness of the findings, and we investigated possible effect modification in population subgroups. As a secondary analysis, we also performed MVMR to investigate the effects of TT and SHBG on lung function decline, using longitudinal data available for a small subsample of UKB. Finally, we performed conventional linear regression analysis of lung function on TT and SHBG in UKB for comparison with the results of our main MVMR analysis. All analyses were performed stratified by sex.

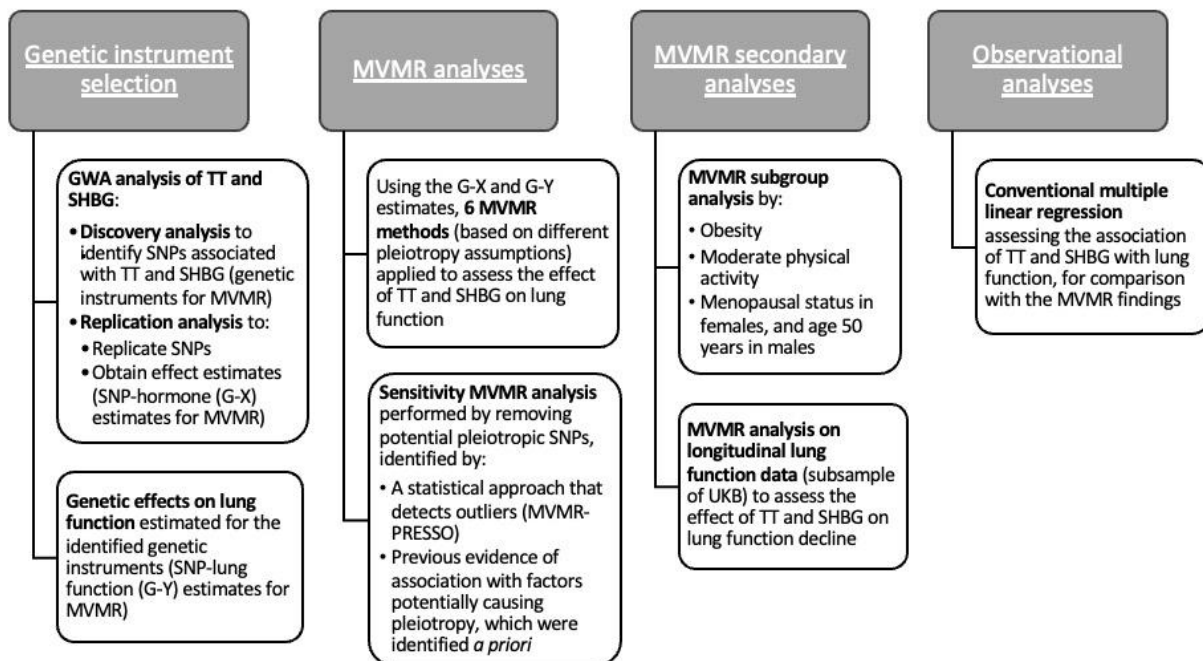
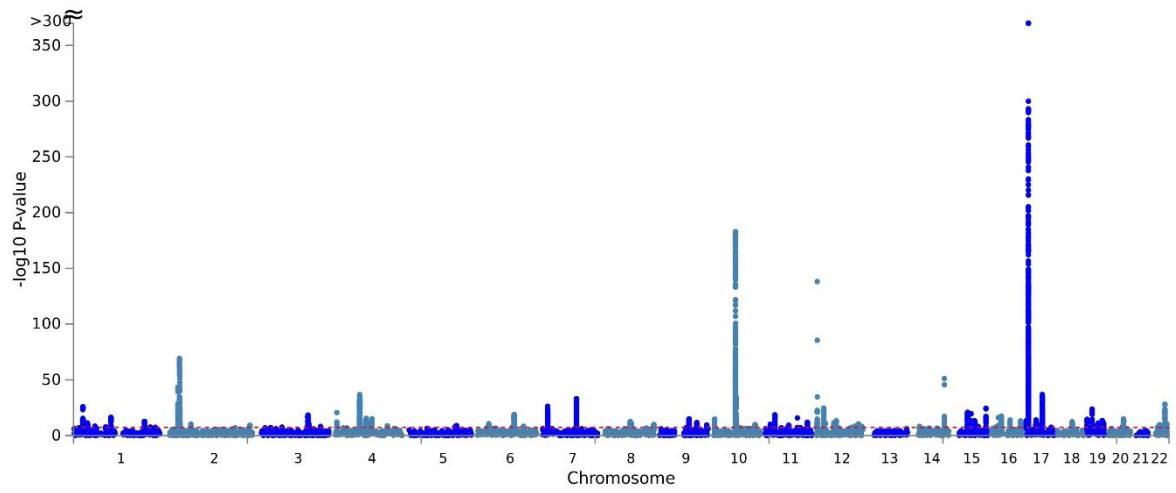


Figure S2. Manhattan plots – total testosterone (TT)

A) TT Male



B) TT Female

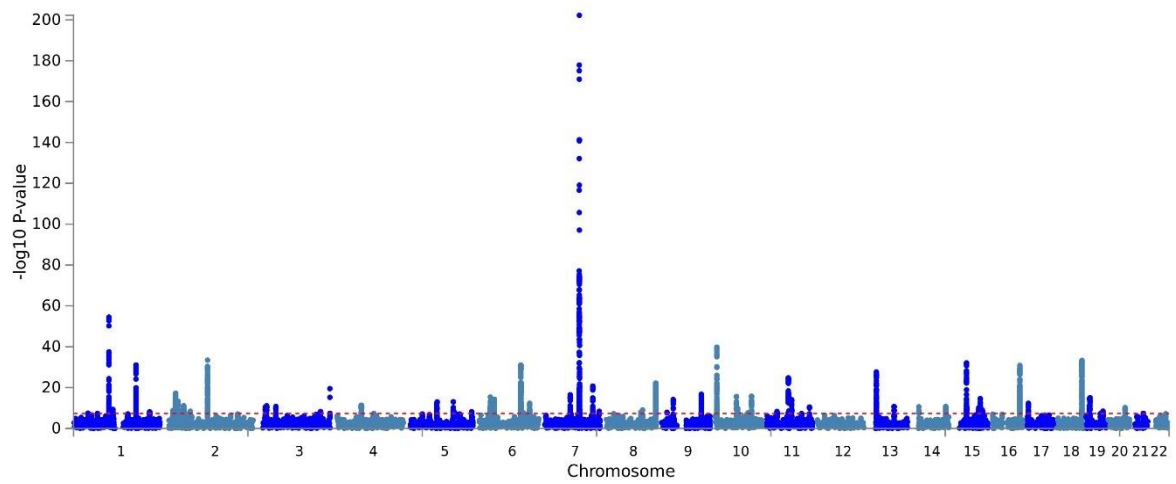
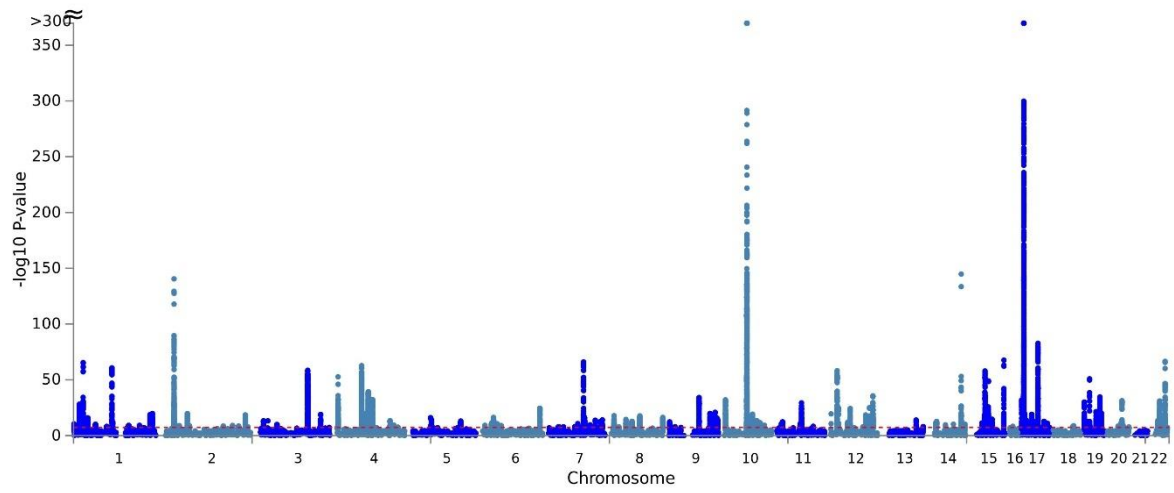


Figure S3. Manhattan plots – sex hormone-binding globulin (SHBG)

A) SHBG Male



B) SHBG Female

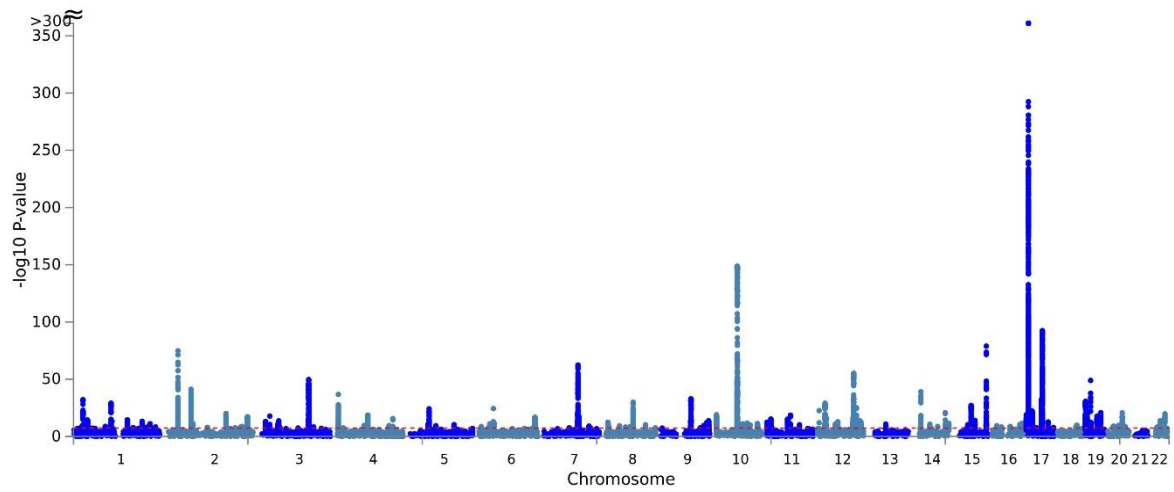


Figure S4. Leave-One-Out RE-IVW analysis plots in males.

Possible pleiotropic SNPs identified by MVMR-PRESSO were removed from the analysis one by one.
 * all = original analysis with all SNPs; rs... = analysis without this SNP; all_out = analysis with all possible pleiotropic SNPs excluded.

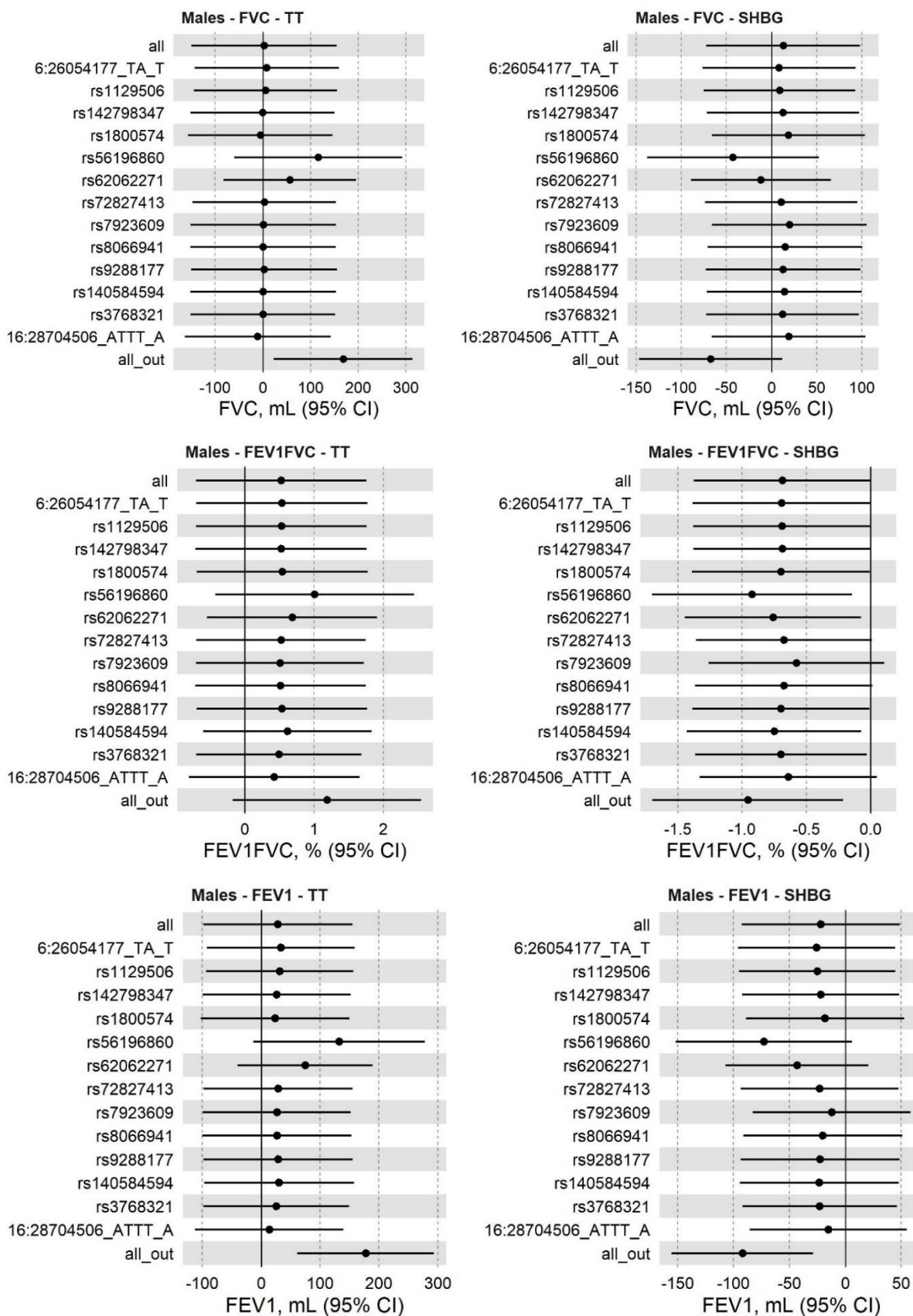
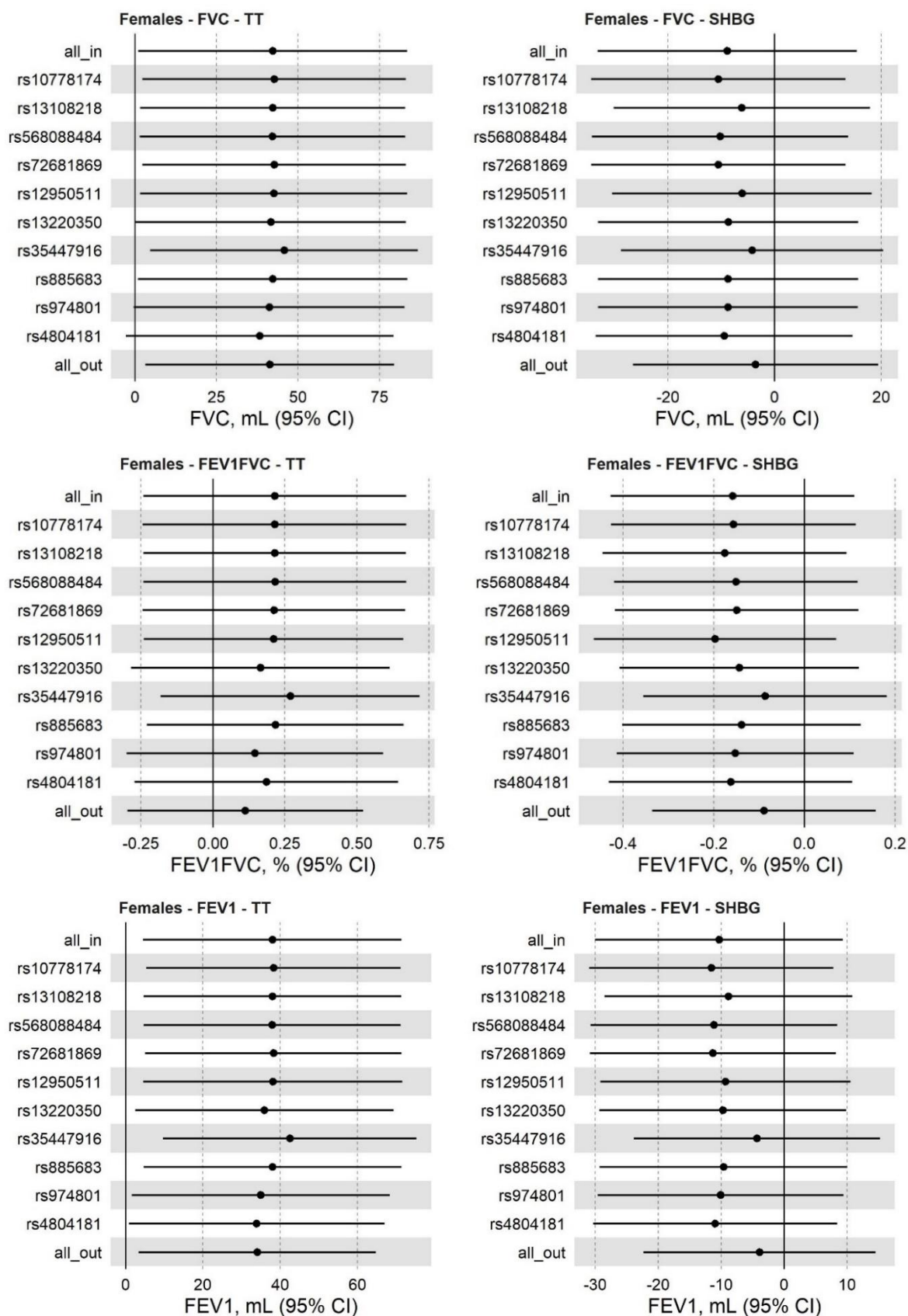


Figure S5. Leave-One-Out RE-IVW analysis plots in females.

Possible pleiotropic SNPs identified by MVMR-PRESSO were removed from the analysis one by one. * all = original analysis with all SNPs; rs... = analysis without this SNP; all_out = analysis with all possible pleiotropic SNPs excluded.



Supplementary Tables

Overview of Supplementary Tables which are provided as separate documents.

Table S1. Overview of identification and replication of number of SNPs associated with hormone levels in UK biobank.

Table S2: MVMR G-X and G-Y estimates.

Table S3: Results of the MVMR analysis of the effect of SHBG and testosterone on FVC, FEV₁, and FEV₁/FVC separately in males and females.

Table S4: Results of the MVMR analysis excluding “outlier” SNPs.

Table S5: Results from PhenoScanner for the pleiotropic SNPs identified by MVMR-PRESSO.

Table S6: Results of subgroup MVMR analyses by obesity, physical activity, menopausal status in females, and age (>50 years) in males.

Table S7: Results of the MVMR analysis of the effect of SHBG and testosterone on FVC, FEV₁, and FEV₁/FVC decline separately in males and females.

Table S8: Results of the observational analyses.