



Sex-specific genetic effects across biomarkers

Emily Flynn¹ · Yosuke Tanigawa¹ · Fatima Rodriguez² · Russ B. Altman^{1,3,4,5} · Nasa Sinnott-Armstrong⁵ · Manuel A. Rivas¹

Received: 18 February 2020 / Revised: 28 July 2020 / Accepted: 4 August 2020 / Published online: 1 September 2020
© The Author(s), under exclusive licence to European Society of Human Genetics 2020

Abstract

Sex differences have been shown in laboratory biomarkers; however, the extent to which this is due to genetics is unknown. In this study, we infer sex-specific genetic parameters (heritability and genetic correlation) across 33 quantitative biomarker traits in 181,064 females and 156,135 males from the UK Biobank study. We apply a Bayesian Mixture Model, Sex Effects Mixture Model (SEMM), to Genome-wide Association Study summary statistics in order to (1) estimate the contributions of sex to the genetic variance of these biomarkers and (2) identify variants whose statistical association with these traits is sex-specific. We find that the genetics of most biomarker traits are shared between males and females, with the notable exception of testosterone, where we identify 119 female and 445 male-specific variants. These include protein-altering variants in steroid hormone production genes (*POR*, *UGT2B7*). Using the sex-specific variants as genetic instruments for Mendelian randomization, we find evidence for causal links between testosterone levels and height, body mass index, waist and hip circumference, and type 2 diabetes. We also show that sex-specific polygenic risk score models for testosterone outperform a combined model. Overall, these results demonstrate that while sex has a limited role in the genetics of most biomarker traits, sex plays an important role in testosterone genetics.

Introduction

Sex differences have been documented in many phenotypes and diseases [1]. Across quantitative traits, men and women typically have overlapping distributions with different means, examples of these traits include height and body mass index

(BMI) [2]. Previous studies have demonstrated that some of this difference is due to sex-specific genetic factors [1, 3]. Genome-wide Association Studies (GWAS) are increasingly used to identify variants that contribute to sex differences, and recently, gene-by-sex interactions have been identified across many phenotypes, including in anthropometric traits [4], irritable bowel syndrome [5], and glioma [6].

Examination of blood and urine laboratory biomarker levels reveal sex differences [7]; however, it is unknown to what extent these sex differences are related to underlying differences in the genetic architecture versus environmental differences. Heritability, or the fraction of phenotypic variability explained by genetic variance, was initially estimated from family studies; but now, with the increasing availability of genome-wide data, common genetic variants (such as single nucleotide polymorphisms (SNP)) are used for this estimation [8]. Methods for estimating SNP-based heritability include LD-score regression, restricted maximum likelihood estimation, Haseman–Elston regression, and the moment-matching approach [9–12]. These methods are applied to a sample of unrelated individuals in order to quantify the proportion of phenotypic variance explained by all genetic variants in the GWAS.

At a trait level, we can use the sex-specific heritabilities and the between-sex genetic correlation to examine what

Supplementary information The online version of this article (<https://doi.org/10.1038/s41431-020-00712-w>) contains supplementary material, which is available to authorized users.

✉ Emily Flynn
erflynn@stanford.edu

✉ Manuel A. Rivas
mrivas@stanford.edu

¹ Biomedical Informatics Training Program, Stanford University, Stanford, CA, USA

² Department of Medicine, Division of Cardiovascular Medicine and the Cardiovascular Institute, Stanford University, Stanford, CA, USA

³ Department of Medicine, Stanford University, Stanford, CA, USA

⁴ Department of Bioengineering, Stanford University, Stanford, CA, USA

⁵ Department of Genetics, Stanford University, Stanford, CA, USA

fraction of the genetics of that trait is shared. The UK Biobank is a prospective population-based study of 500,000 individuals that includes both genetic and phenotypic data, allowing for rich SNP-based estimation of heritability [13]. While most traits do not show sex effects on heritability [14], previous studies have documented these differences in a subset of traits, including fat distribution and other anthropometric traits [4, 15]. However, there has yet to be an analysis of these sex differences across biomarkers.

Here we present an approach for estimating the extent to which genetic effects are correlated between sexes and identifying the proportion of relevant variants that have shared effects versus effects that are specific to each sex. We apply this approach to blood and urine biomarker data from the UK Biobank to examine sex differences in genetic effects, and find differences primarily in the genetic determinants of testosterone level. Furthermore, we use these identified sex differences to provide hypotheses about biological mechanisms including (1) examination of protein-altering variants and tissues where these genes are selectively expressed, (2) causal inference using Mendelian randomization (MR) to assess relationships between testosterone and other traits, and (3) improved genetic risk prediction models for testosterone.

Materials and methods

Genotype data

We used genotype data from the UK Biobank dataset release version 2 and the hg19 human genome reference for all analyses in the study [16]. To minimize the variability due to population structure in our dataset, we restricted our analyses to unrelated White British individuals (as indicated by self-reported ethnicity, UKBB field ID 21000) without missing data (see Supplementary methods for details). Variant annotations, filtering, and LD pruning were performed as previously described [17, 18]. We additionally filtered for variants with Hardy–Weinberg equilibrium $<10^{-7}$ and $<1\%$ missingness, and used `plink --xchr-model 2`.

Anthropometric traits

To demonstrate the utility of our method, we applied SEMM to previously examined anthropometric traits [4, 15] (field IDs in Table S1; Supplementary methods).

Selection and processing of biomarker traits

We focused on 33 of 38 biomarkers previously described and covariate-adjusted [7] (field IDs in Table S2; Supplementary methods).

Menopause phenotype definition

We used a stringent definition for dividing individuals into pre- vs. postmenopause to separate into clear categories and avoid including peri-menopause (see Supplementary methods for the definition and relevant field IDs).

Summary statistic generation

Genome-wide association summary statistics were generated separately for males and females using PLINK v2.00aLM (2 April 2019). Age, genotyping array used, and the first four PCs were included as covariates. Variants with missing standard errors or standard errors >0.2 in either sex were also removed.

Sex Effects Mixture Models (SEMM)

We used a two-component mixture model consisting of a point mass centered at zero and a multivariate normal distribution to estimate the variance–covariance matrix, from which we estimate the genetic correlation and heritability. We extended this model to a four-component model with male- and female-specific components in order to identify variants with sex-specific genetic effects (see Supplementary methods).

Sex-specific multivariate polygenic prediction

To construct sex-stratified polygenic risk score (PRS) models using multivariate penalized regression, we created a random split dataset of White British individuals in UK Biobank into 70% training, 10% validation, and 20% test sets. We used covariate adjustment testosterone residual values as described previously [7]; however, we did so in both sex-separated and combined cohorts (see Table S14; Supplementary methods).

Mendelian randomization (MR)

We used MR-Base to test for evidence for causal associations between testosterone and ten outcomes of interest using the sets of female- and male-specific testosterone variants [19]. Variants were pruned for LD with clumping and the analysis was performed with MR Egger, inverse variance weighted (IVW), and IVW with fixed effects. For each of the outcomes, we used summary statistics from both a UK Biobank and non-UK Biobank source and sex-divided outcomes where available. The traits include: waist circumference (WC) [20], hip circumference (HC) [20], height [21, 22], BMI [23], age at menarche [24], age at menopause [25], prostate cancer [26], heart disease [27], type 2 diabetes (T2D) [28], and stroke [29] (Table S12). A Bonferroni

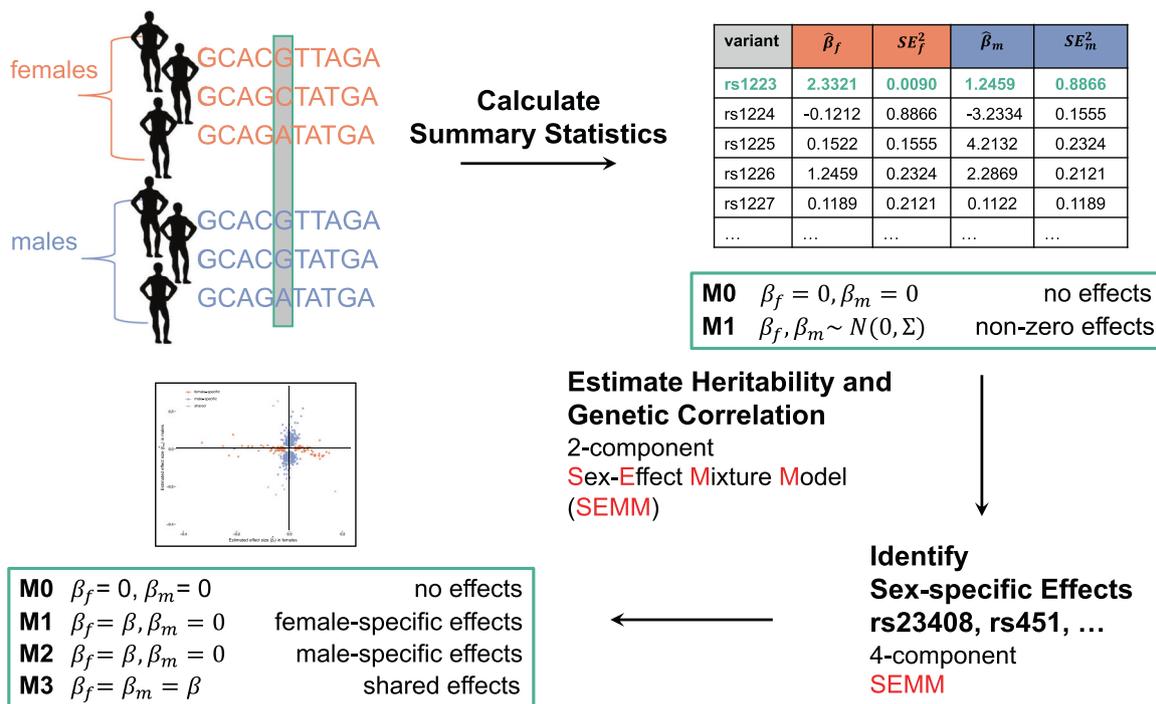


Fig. 1 Schematic overview of Sex-Effect Mixture Model. We prepared a dataset of 33 serum and urine biomarkers from 337199 individuals in UK Biobank. We calculated GWAS summary statistics from males (light purple) and females (orange) separately, so that for every trait, we had an effect estimate (and) and standard error for each variant in each sex. We use a two-component Bayesian Sex-Effect Mixture Model (SEMM), with no-effect and nonzero effect components, to estimate SNP-based heritability and the genetic correlation

between males and females for each biomarker. A four-component extension of the SEMM contains two additional components for separate male and female effects. This model allows us to distinguish between four cases: genetic variants that have no effect (illustrated as M0), genetic variants that have a stronger association with the trait in females or males (M1, orange, or M2, light purple), and genetic variants that have similar effects in females and males (M3, gray).

correction was used to account for multiple tests (p value $< 0.05/168 = 2.98 \times 10^{-4}$).

Results

Sex-Effect Mixture Models

We built a two-component Bayesian SEMM for estimating the contributions of sex to genetic variance using GWAS summary statistics (Fig. 1). The model contains a null component, for variants that do not contribute to the trait, and a non-null component, for variants that represent the genetic contribution to that trait. Variants driving male and female traits in the non-null component are modeled as two-dimensional vectors drawn from a multivariate normal distribution with a variance–covariance matrix that can be used to estimate the genetic correlation between sexes. To assess whether our approach obtains reliable estimates in real data, we applied SEMM to traits from Rawlik et al. [15] and obtained overlapping genetic correlations and similar but not identical heritability estimates (Table S1 and Figs. S1, S2).

We extended our two-component SEMM to a four-component model to identify genetic variants with different effects in males and females (Fig. 1). To do so, we add two components for detecting genetic variants that have stronger effects in one sex. Similar to the two-component model, the four-component SEMM also contains a no-effect and shared-effect component. Through fitting this model, we are able to separate genetic variants that have “shared” effects, where the variant or set of variants have the same effect in males and females, and “sex-specific” effects, where the variants have different effects in males than in females (e.g., this variant is associated with higher lab values in females but not in males).

To demonstrate the efficacy of the four-component SEMM, we applied the model to four traits, waist–hip ratio, arm-fat ratio, leg-fat ratio, and trunk-fat ratio with previously identified sex-specific genetic effects. We identified 367, 560, 832, and 1158 genetic variants that had significantly stronger associations in females in waist–hip ratio, arm-fat ratio, leg-fat ratio, and trunk-fat ratio, respectively. In males, only 12 variants were found in arm-fat ratio (estimated false discovery rate 4.9–6.8% across all traits, see Table S3a–c and Fig. S3). Included in the female-

specific waist–hip ratio variants were genetic variants proximal to four of six previously reported genes (*COBL1/GRB14*, *VEGFA*, *PPARG*, and *HSD17B4*). Fat ratio variants were proximal to one of the male and 48 of the female-specific genes previously identified, indicating that we capture known sex-specific signal (see Table S3d for the overlap and Table S4 for the full lists). In addition, we validate these sex-specific variants by showing they have similar effect sizes in a held-out cohort (see Supplementary methods, Fig. S6 and Table S11).

Sex-differential heritability

We applied the two-component SEMM to 33 UK Biobank biomarkers [7] in order to estimate the sex-specific heritability and genetic correlation for each trait (see Table S2 for a full list of these traits). While a large fraction of biomarkers had overlapping heritability estimates, we found sex differences in the heritability of 17 of 33 biomarkers, including testosterone, IGF-1, non-albumin protein, SHBG, total protein (higher in males), apolipoprotein B, C-reactive protein, cholesterol, creatinine, cystatin C, eGFR, gamma glutamyltransferase, HDL-C, LDL-C, potassium in urine, sodium in urine, and urate (higher in females, Fig. 2a). Of these, cholesterol, creatinine, and sodium in urine, LDL, testosterone, and urate showed >1.3-fold differences. For the majority of traits, the between-sex genetic correlations were close to 1.0, indicating shared additive genetic effects between males and females (Fig. 2b). By contrast, for testosterone, we estimated a genetic correlation

of only 0.120 (2.5–97.5 percentile interval: 0.0805–0.163), indicating largely nonoverlapping genetic effects between males and females (see Table S5; these estimates are consistent across priors: Fig. S4a and Table S6a, b).

The heritability of a particular trait can vary across the lifetime, as genetics may explain more or less of the variation in that particular trait. Previous studies have found that pre- and postmenopausal women have different heritability for BMI, waist and hip measures, and lipid biomarkers [30]. To examine this across biomarkers in the UK Biobank population, we applied our two-component SEMM to summary statistics for pre- and postmenopausal women. We found that genetic correlations between pre- and postmenopausal women close to 1.0, and all traits had higher or equivalent within-sex (between pre- and postmenopausal women) than between-sex (between either group and men) genetic correlations (Fig. 2c and Table S7).

Identification of genetic variants with sex-specific effects

We applied our four-component SEMM to all 33 biomarkers to identify genetic variants with sex-specific and shared effects. In total, our analysis found 26,561 variants with effects on the traits of interest (Table S8). As expected, the majority (25,950) of these variants showed shared effects between sexes, and most traits had few or no sex-specific variants. We identified 148 and 463 genetic variants with sex-specific effects in females and males, the bulk of them associated with testosterone (80.4% and 96.1%,

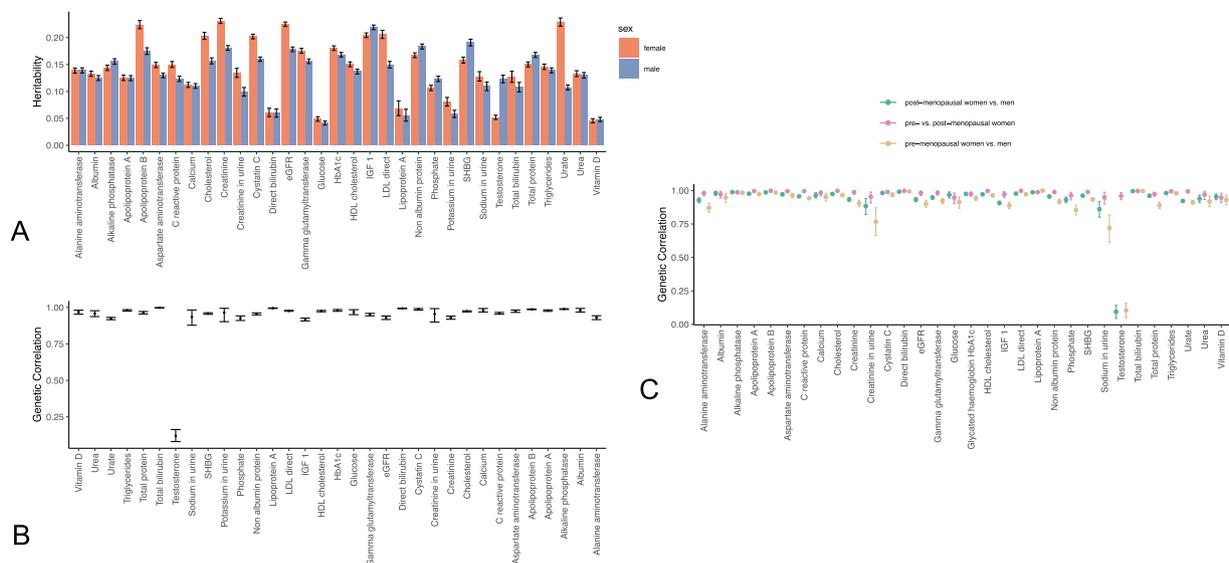


Fig. 2 Heritability and genetic correlations of biomarkers between females and males and related to menopausal status. a SNP-based heritability estimates of 33 biomarkers for females (orange) and males (light purple). **b** Correlation of genetic effects between males and females for 33 biomarkers. **c** The genetic correlation within women

(pre- vs. postmenopausal, in pink) was higher or equal to than either that between both postmenopausal women and men (green) and premenopausal women and men (tan). Error bars in all three plots indicate the 2.5–97.5 percentile interval from STAN sampling.

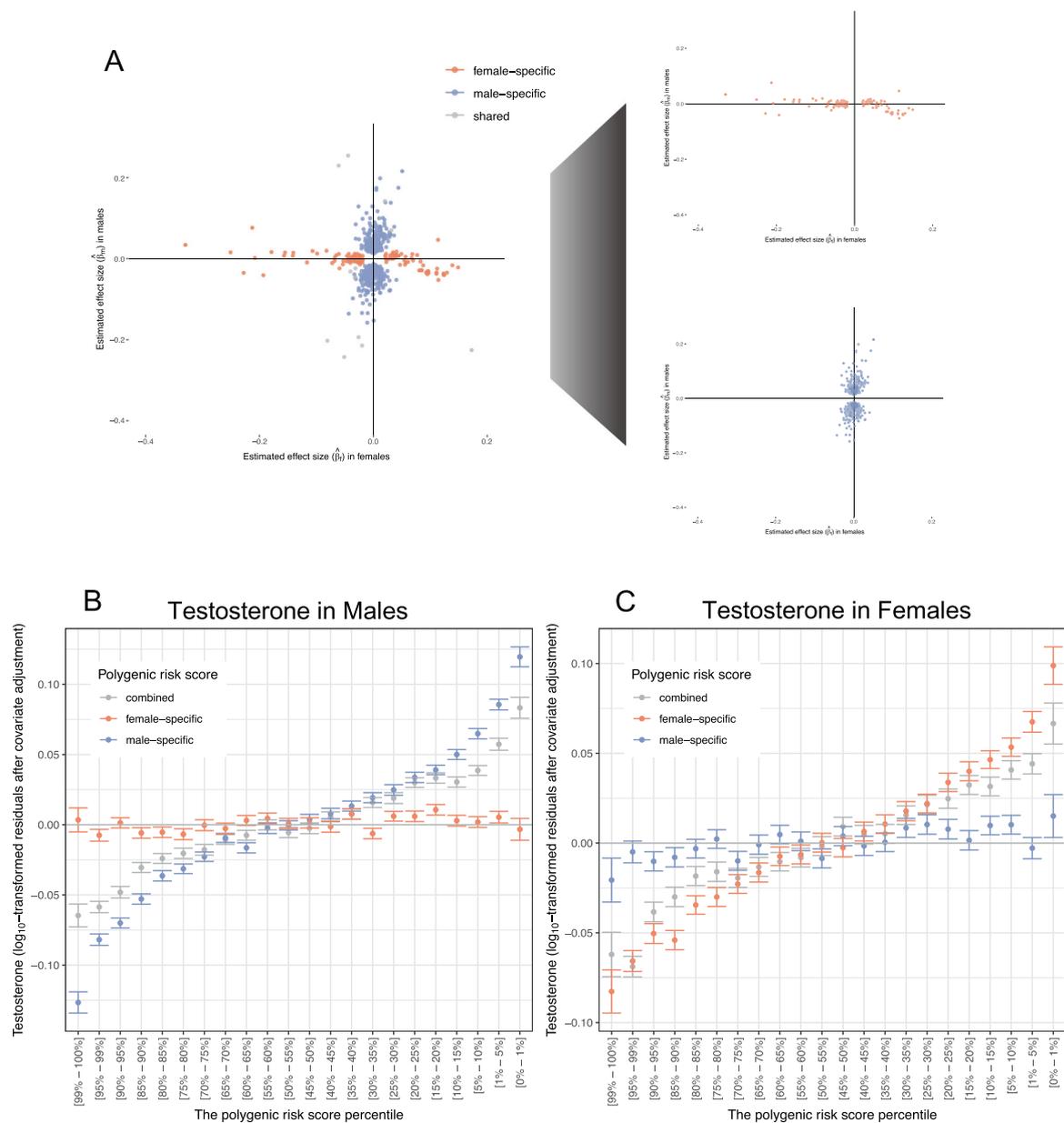


Fig. 3 Identification of genetic variants with sex-specific effects on testosterone levels. **a** Estimated effect sizes for genetic variants with nonzero effects on testosterone are shown (x-axis, estimated effect size in females; y-axis, estimated effect size in males). Light purple dots correspond to variants that belong to the “male-specific” effect component; orange corresponds to the “female-specific” effect component; and gray dots correspond to genetic variants that belong to the “shared” effect component of SEMM. Polygenic risk score (PRS)

respectively; see Figs. 3a, S5 and Tables S9, S10). Of the testosterone variants, 54 male-specific, one female-specific, and one shared variant are located on the X chromosome, indicating enrichment of X chromosomal variants in male testosterone genetics, consistent with previous reports [31]. In addition, using tissue-specific enrichment analysis [32], we find enrichment of liver genes in the genes proximal to

predictions for testosterone in male (**b**) and female (**c**) individuals for male-specific (light purple), female-specific (orange), and combined (gray) PRS models. Plots show predicted stratified risk bins for testosterone levels (x-axis) versus mean covariate-adjusted testosterone for those individuals (y-axis). These values were calculated on a held-out test set of unrelated White British individuals. The error bars represent standard errors.

male- but not female-specific variants ($p = 6.21 \times 10^{-7}$ and $p > 0.1$, respectively; Fig. S5, Supplementary methods).

Previous testosterone genetics studies have focused on males. In our study, in addition to identifying male-specific variants in known testosterone-related genes (*AR*, *JMJ1DC*, and *FAM9B*), we also identify multiple female-specific variants with strong positive or negative effects on

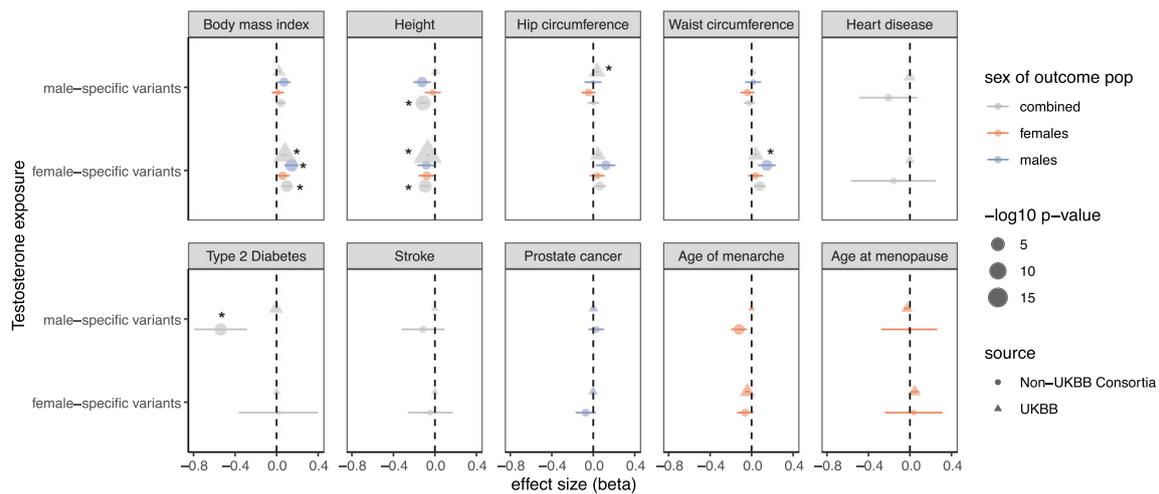


Fig. 4 Results of Mendelian randomization tests with sex-specific testosterone variants as instruments. This plot shows the Mendelian randomization results for all traits, with each trait shown separately. Effect sizes (betas) are estimated using either the female or male-specific variants as instrumental variables for testosterone exposure. Ninety-five percent confidence intervals are shown for each estimate. Points are colored by the sex of the outcome population (light purple

for males, orange for females, and gray for combined), with size indicating the $-\log_{10} p$ value, and shape showing the source of the GWAS statistics for the trait (triangles for UKBB = UK Biobank, circles for all others). Trait/exposure pairs that are significant after multiple hypothesis correction are indicated with an asterisk. The list of non-UK Biobank consortia traits is included in Supplementary methods and Table S10.

testosterone. In females, these include missense variants in *STAG3* (*rs149048452*, $\beta = -0.33$ and $p = 8.97 \times 10^{-9}$), a meiosis cohesion complex protein containing variants associated with premature ovarian failure [33], and *POR* (*rs17853284*, $\beta = -0.23$ and $p = 8.98 \times 10^{-15}$), a cytochrome p450 oxidoreductase where deficiencies associated with amenorrhea, disordered steroidogenesis, and congenital adrenal hyperplasia [34] (Table S10a). Many female-specific missense variants are located in genes associated with steroid hormone production (*LIPE*, *POR*, *UGT2B7*) or gamete formation (*STAG3*, *MCM9*, *TSBP1*, *ZAN*); although *ZAN* and *TSBP1* encode the sperm zonadhesin protein and testis-expressed protein 1, respectively. To our knowledge, these associations are novel and may help with understanding testosterone genetics in women.

Mendelian Randomization of sex-specific genetic effects

After identifying genetic variants with sex-specific effects on testosterone levels, we used Mendelian Randomization (MR) to examine whether these biomarkers are causally related to disease outcomes or other commonly measured traits. The intuition is that if a genetic variant is associated with differing levels of a biomarker, this provides a natural experiment, and we can examine whether the predicted variance in that biomarker based on the genetic variant is associated with the outcome variance, which indicates a causal effect. Recently, MR studies have found evidence for causal links between testosterone and cardiovascular disease [35] but not cognition [36] or BMI [37]. We

aggregated a total of ten outcomes (Table S12), including anthropometric traits (height, BMI, WC, and HC), disease outcomes (heart disease, stroke, and type 2 diabetes), and sex-specific traits (ages at menarche and menopause, prostate cancer), and used the IVW method [38] to assess the causal effects of the sex-specific variants identified in our analysis (Fig. 4 and Table S13). We found that testosterone levels showed evidence of a causal association with BMI and WC using female-specific variants as instruments and HC using male-specific variants, with estimated effects consistent with higher testosterone increasing BMI, WC, and HC ($p = 1.3 \times 10^{-12}$, 1.1×10^{-4} , 2.6×10^{-5} ; $\beta = 0.081$, 0.04 , 0.036 ; $SE = 0.011$, 0.01 , 0.0086 respectively, Fig. S8). A previous MR study [35] examined testosterone for causal effects on HC, WC, and BMI, but did not find evidence of an association; however, it is possible we are able to find these associations because we used sex-specific genetic instruments. Both female and male variants showed evidence of a causal association with height ($p = 6.1 \times 10^{-6}$ and 9.8×10^{-9}), with higher testosterone associated with decreased height ($\beta = -0.093$ and -0.11 , $SE = 0.021$, 0.020). This is in contrast to evidence of a positive relationship between height and testosterone levels at a population level and in a previous MR study [35, 39]. For all of these associations, we observe similar effects in the UK Biobank and GIANT datasets, with MR Egger and IVW (Table S13).

Male-specific testosterone levels show evidence of an association with type 2 diabetes (T2D) ($p = 3.1 \times 10^{-5}$); higher testosterone is related to T2D risk reduction ($\beta = -0.54$, $SE = 0.13$) using data from the combined

DIAGRAM and MetaboChip study [28]. This association was found using the IVW method; MR Egger estimates indicate that the relationship is in the reverse direction and is not significant ($\beta = 1.1$, $SE = 0.61$, $p = 0.10$). Several longitudinal studies have shown that low levels of testosterone predict the later development of T2D or metabolic syndrome [40].

Sex-specific multivariate polygenic risk prediction

Motivated by the sex differences in testosterone genetics, we tested whether sex-specific PRS would have better predict testosterone levels than a sex-combined model. We applied batch screening iterative lasso [41] to train multivariate penalized regression models for males and females. While the two sex-specific and combined models are consistent on a held-out test set ($\rho = 0.59$ and 0.60 , $p < 2.2 \times 10^{-16}$; Fig. S9), the sex-specific models have improved performance in the sexes they were trained on over the combined model, and low performance in the opposite sex ($R^2 = 0.31$ vs. 0.21 vs. 0.020 and 0.18 vs. 0.13 vs. 0.023 for male and female vs. combined vs. opposite sex). Overall, these results highlight the benefits of sex-specific polygenic prediction for testosterone (Fig. 3b, c and S10).

Discussion

We sought to examine how genetics relates to sex differences in biomarker levels. To answer this question, we studied the genetics of 33 biomarkers in UK Biobank males and females using SEMM, a two- and four-component Bayesian Mixture Model. SEMM has the benefit of both estimating the underlying genetic architecture and identifying genetic variants with shared and sex-specific effects. For the majority of the traits we analyzed, we do not see strong sex differences in genetic effects, which is expected and previously documented in the literature [14]. Namely, the genetics of these traits are shared (as indicated by genetic correlations close to one and similar heritabilities), and the traits have few or no variants with sex-specific effects.

By contrast, we found little overlap between males and females in the genetics of testosterone levels. In addition to finding significant sex differences in genetic architecture, we also identified over five hundred genetic variants with male- or female-specific effects. Because of the male-female differences in testosterone genetics, we examined the subset of these sex-specific variants that are protein-altering and the tissue-specific expression patterns of proximal genes. The protein-altering variants associated with female-specific effects testosterone include variants in genes associated

with steroid hormone production and gamete production. Tissue-specific enrichment analysis reveals that the genes proximal to these sex-specific variants are enriched in liver in males but not females. It is hypothesized that the relationship between testosterone and liver disease may have different etiology in men and women [42]. In addition, we built sex-specific polygenic risk models, which showed improved predictive performance over a sex-combined model.

We used MR to assess whether testosterone may be causally implicated in a broad range of diseases and phenotype measurements, and found associations with BMI, WC, HC, height, and T2D. The relationships with BMI, WC, and HC are novel MR associations, and it is possible that we are able to identify them because we are using a novel set of genetic instruments and most previous MR studies in testosterone excluded women. Our analysis shows a decreasing effect of testosterone on height, which is surprising, as previous studies indicated that higher testosterone is associated with higher stature [35]. However, testosterone is sometimes used as a therapy for tall males with delayed puberty, and results in accelerated initial growth but overall stunting of stature [43]. Further work is required to understand this association. Finally, the potential causal relationship with T2D supports the hypothesis that testosterone treatment for reducing diabetes risk in men may be a worthwhile approach, and matches the MR recent findings of Ruth et al. [44], that higher testosterone levels reduce T2D risk in men and increase risk in women. While we did not use sex-separated T2D data in our MR analysis, the associations between T2D and testosterone levels were only found with male-specific testosterone variants as instruments, which are consistent with this effect.

SEMM is made publicly available as an R package. In addition, the inference results are available for visualization as a web application in the Global Biobank Engine [45], and all the sex-specific genetic variants we identified are included in the Supplementary Tables (Tables S4 and S10). We focused on identifying sex-specific effects in biomarkers; however, there are many other traits and diseases that also show phenotypic sex differences and may have underlying sex-specific genetic effects, so future work may involve application of our model to those traits. While we applied this method to examine sex differences in genetic effects, SEMM could also be used to for any other type of binary gene-covariate analysis.

Testosterone is frequently thought of as a male sex hormone because of its higher levels in men and its involvement in the development of the male reproductive tract and secondary sex characteristics. However, females also produce testosterone, albeit at lower levels, and elevated testosterone is associated with polycystic ovarian syndrome

and metabolic disorders [46, 47]. Previous work examining the genetics of testosterone in females did not find associations [48] and previous MR studies have been limited by the lack of known testosterone variants in women [35]. Our analysis expands on and addresses this by using a larger population, carefully adjusted biomarkers, and our SEMM method to identify variants. Our results demonstrate that the genetics of testosterone levels is complex and highly polygenic in both males and females. Further, our work highlights the importance of also examining female variability in testosterone levels, and of considering sex as a variable. It is important to note that our analysis was limited to White British individuals; and as such may not generalize to other populations. In addition, our analysis binarized sex and did not consider gender, so our results do not consider gender-related effects or include intersex, transgender, and non-binary individuals [49]. Future analyses should include more diverse population cohorts, an expanded sex-gender spectrum, rare genetic effects, and additional reproductive health outcomes; we anticipate that this will lead to better understanding of the translational impact of these findings.

Data availability

The sex-specific biomarker variants are in the supplement and in the Global Biobank Engine: <https://biobankengine.stanford.edu/sex-effects> [45]. The polygenic risk score coefficients are available at figshare (<https://doi.org/10.6084/m9.figshare.12793490.v1>).

Code availability

SEMM is publicly available as an R package on GitHub at <https://github.com/rivas-lab/semm>; additional scripts used in the analysis are included at <https://github.com/rivas-lab/sex-diff-biomarker-genetics>.

Acknowledgements This research has been conducted using the UK Biobank resource. We thank all the participants in the UK Biobank study. The primary and processed data used to generate the analyses presented here are available in the UK Biobank access management system (<https://amsportal.ukbiobank.ac.uk/>) for application 24983, “generating effective therapeutic hypotheses from genomic and hospital linkage data” (<http://www.ukbiobank.ac.uk/wp-content/uploads/2017/06/24983-Dr-Manuel-Rivas.pdf>). We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results.

Funding EF is supported by the NIH NLM F31 Fellowship F31LM013053 and the Stanford Data Science program. RBA is supported by the NIH GM 102365, LM 005652, and the Chan Zuckerberg Biohub. YT is supported by a Funai Overseas Scholarship from the Funai Foundation for Information Technology and School of Medicine at Stanford University. FR is supported by a career development award from the National Heart, Lung, and Blood Institute, National Institutes

of Health (1K01HL144607). NS-A is supported by a Stanford Graduate Fellowship. MAR is supported by National Human Genome Research Institute of the National Institutes of Health (R01HG010140) and the National Institute of Health Center for Multi- and Trans-Ethnic Mapping of Mendelian and Complex Diseases Grant (5U01 HG009080).

Author contributions MAR conceived and designed the study. EF and MAR designed and carried out the statistical and computational analyses. YT performed PRS analysis. EF, YT, NS-A, and MAR carried out quality control of the data. The manuscript was written by EF and MAR. All authors provided feedback. MAR supervised all aspects of the study.

Compliance with ethical standards

Conflict of interest The authors filed a provisional patent application No. 62/925,133 titled “Genetic Determination of Hormone Levels and Applications Thereof.” The authors declare that they have no conflict of interest.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Ober C, Loisel DA, Gilad Y. Sex-specific genetic architecture of human disease. *Nat Rev Genet.* 2008;9:911–22.
2. Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, et al. The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. *PLoS Genet.* 2015;11:e1005378.
3. Khramtsova EA, Davis LK, Stranger BE. The role of sex in the genomics of human complex traits. *Nat Rev Genet.* 2019;20:173–90.
4. Rask-Andersen M, Karlsson T, Ek WE, Johansson Å. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat Commun.* 2019;10:339.
5. Bonfiglio F, Zheng T, Garcia-Etxebarria K, Hadizadeh F, Bujanda L, Bresso F, et al. Female-specific association between variants on chromosome 9 and self-reported diagnosis of irritable bowel syndrome. *Gastroenterology.* 2018;155:168–79.
6. Ostrom QT, Kinnersley B, Wrensch MR, Eckel-Passow JE, Armstrong G, Rice T, et al. Sex-specific genome-wide association study in glioma identifies new risk locus at 3p21.31 in females, and finds sex-differences in risk at 8q24.21. *Scientific Reports.* 2017;229112. <https://doi.org/10.1101/229112>.
7. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars NJ, Aguirre M, Venkataraman GR, et al. Genetics of 38 blood and urine biomarkers in the UK Biobank. *bioRxiv.* 2019; <https://doi.org/10.1101/660506>.
8. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. *Nat Genet.* 2017;49:1304–10.
9. Bulik-Sullivan B. Relationship between LD score and Haseman-Elston regression. *bioRxiv.* 2015. <https://doi.org/10.1101/018283>.
10. Hill WG. Estimation of heritability by regression using collateral relatives: linear heritability estimation. *Genetical Res.* 1978;32:265–74.
11. Ni G, Moser G, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Wray NR, Lee SH. Estimation of genetic correlation via linkage disequilibrium score regression and genomic restricted maximum likelihood. *Am J Hum Genet.* 2018;102:1185–94.

12. Speed D, Balding DJ. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat Genet.* 2019;51:277–84.
13. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, 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.
14. Stringer S, Polderman TJC, Posthuma D. Author correction: majority of human traits do not show evidence for sex-specific genetic and environmental effects. *Sci Rep.* 2018;8:18060.
15. Rawlik K, Canela-Xandri O, Tenesa A. Evidence for sex-specific genetic architectures across a spectrum of human complex traits. *Genome Biol.* 2016;17:166.
16. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562:203–9.
17. DeBoever C, Tanigawa Y, Lindholm ME, McInnes G, Lavertu A, Ingelsson E, et al. Medical relevance of protein-truncating variants across 337,205 individuals in the UK Biobank study. *Nat Commun.* 2018;9:1612.
18. Tanigawa Y, Li J, Justesen JM, Horn H, Aguirre M, DeBoever C, et al. Components of genetic associations across 2,138 phenotypes in the UK Biobank highlight novel adipocyte biology. *Nat Commun.* 2019;10:4064.
19. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife.* 2018;7:e34408.
20. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature.* 2015;518:187–96.
21. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet.* 2014;46:1173–86.
22. Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, et al. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* 2013;9:e1003500.
23. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518:197–206.
24. Perry JR, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature.* 2014;514:92–97.
25. Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet.* 2015; 47:1294–303.
26. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet.* 2018;50:928–36.
27. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet.* 2013;45:25–33.
28. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet.* 2012;44:981–90.
29. Malik R, Traylor M, Pulit SL, Bevan S, Hopewell JC, Holliday EG, et al. Low-frequency and common genetic variation in ischemic stroke: the METASTROKE collaboration. *Neurology.* 2016;86:1217–26.
30. Kelemen LE, Atkinson EJ, de Andrade M, Shane Pankratz V, Cunningham JM, Wang A, et al. Linkage analysis of obesity phenotypes in pre- and post-menopausal women from a United States mid-western population. *BMC Med Genet.* 2010;11:156.
31. Ohlsson C, Wallaschofski H, Lunetta KL, Stolk L, Perry JRB, Koster A, et al. Genetic determinants of serum testosterone concentrations in men. *PLoS Genet.* 2011;7:e1002313.
32. Xu X, Wells AB, O'Brien DR, Nehorai A, Dougherty JD. Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J Neurosci.* 2014;34:1420–31.
33. Stabej LeQuesne, Williams P, James HJ, Tekman C, Stanescu M, Kleta HC, et al. STAG3 truncating variant as the cause of primary ovarian insufficiency. *Eur J Hum Genet.* 2016;24:135–8.
34. Idkowiak J, Cragun D, Hopkin RJ, Arlt W. Cytochrome P450 oxidoreductase deficiency. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, et al., editors. *GeneReviews®*. Seattle: University of Washington; 2005.
35. Luo S, Au Yeung SL, Zhao JV, Burgess S, Schooling CM. Association of genetically predicted testosterone with thromboembolism, heart failure, and myocardial infarction: mendelian randomisation study in UK Biobank. *BMJ.* 2019;364:1476.
36. Zhao JV, Lam TH, Jiang C, Cherny SS, Liu B, Cheng KK, et al. A Mendelian randomization study of testosterone and cognition in men. *Sci Rep.* 2016;6:21306.
37. Eriksson J, Haring R, Grarup N, Vandenput L, Wallaschofski H, Lorentzen E, et al. Causal relationship between obesity and serum testosterone status in men: a bi-directional mendelian randomization analysis. *PLoS One.* 2017;12:e0176277.
38. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med.* 2017;36:1783–802.
39. Handelsman DJ, Yeap B, Flicker L, Martin S, Wittert GA, Ly LP. Age-specific population centiles for androgen status in men. *Eur J Endocrinol.* 2015;173:809–17.
40. Antonio L, Wu FCW, O'Neill TW, Pye SR, Carter EL, Finn JD, et al. Associations between sex steroids and the development of metabolic syndrome: a longitudinal study in European men. *J Clin Endocrinol Metab.* 2015;100:1396–404.
41. Qian J, Du W, Tanigawa Y, Aguirre M, Tibshirani R, Rivas MA, et al. A fast and flexible algorithm for solving the Lasso in large-scale and ultrahigh-dimensional problems. *bioRxiv.* 2019; <https://doi.org/10.1101/630079>.
42. Yim JY, Kim J, Kim D, Ahmed A. Serum testosterone and non-alcoholic fatty liver disease in men and women in the US. *Liver Int.* 2018;38:2051–9.
43. Zachmann M, Ferrandez A, Mürset G, Gnehm HE, Prader A. Testosterone treatment of excessively tall boys. *J Pediatr.* 1976;88:116–23.
44. Ruth KS, Day FR, Tyrrell J, Thompson DJ, Wood AR, Mahajan A, et al. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat Med.* 2020;26:252–8.
45. McInnes G, Tanigawa Y, DeBoever C, Lavertu A, Olivieri JE, Aguirre M, et al. Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics.* 2019;35:2495–7.
46. Haring R, Baumeister SE, Völzke H, Dörr M, Felix SB, Kroemer HK, et al. Prospective association of low total testosterone concentrations with an adverse lipid profile and increased incident dyslipidemia. *Eur J Cardiovasc Prev Rehabil.* 2011;18:86–96.
47. Kim JJ, Kim D, Yim JY, Kang JH, Han KH. Polycystic ovary syndrome with hyperandrogenism as a risk factor for non-obese non-alcoholic fatty liver disease. *Aliment Pharmacol Ther.* 2017;45:1403–12.

48. Prescott J, Thompson DJ, Kraft P, Chanock SJ, Audley T, Brown J, et al. Genome-wide association study of circulating estradiol, testosterone, and sex hormone-binding globulin in postmenopausal women. *PLoS One*. 2012;7:e37815.
49. Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S, et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *npj Digit Med*. 2020;3:81.