THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Sex differences in genetic architecture in the UK Biobank

Citation for published version:
Bernabeu, E, Canela-Xandri, O, Rawlik, K, Talenti, A, Prendergast, J & Tenesa, A 2021, 'Sex differences in genetic architecture in the UK Biobank', *Nature Genetics*, vol. 53, no. 9, pp. 1283-1289. https://doi.org/10.1038/s41588-021-00912-0

Digital Object Identifier (DOI):
10.1038/s41588-021-00912-0

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Nature Genetics

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 21. May. 2024

Title:

Sex differences in genetic architecture in the UK Biobank

Authors:

Elena Bernabeu[1], Oriol Canela-Xandri[2], Konrad Rawlik[1], Andrea Talenti[1], James Prendergast[1], Albert Tenesa[1,2]

Affiliations:

[1] The Roslin Institute, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG, UK

[2] Institute of Genetics and Cancer (IGC), University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh, EH4 2XU, UK

[3] Corresponding author
Dr Albert Tenesa
The Roslin Institute
The University of Edinburgh Easter Bush
Roslin, Midlothian
EH25 9RG
UK
Tel: 0044 (0)131 651 9100
Fax: 0044 (0)131 651 9220
Email: Albert.Tenesa@ed.ac.uk

Males and females present differences in complex traits and in the risk of a wide array of diseases. Gene by sex interactions (GxS) are thought to account for some of these differences. However, the extent and basis of these interactions are poorly understood. Here we provide insights into both the scope and mechanism of GxS across the genome of circa 450,000 individuals of European ancestry and 530 complex traits in the UK Biobank. We found small yet widespread differences in genetic architecture across traits. We also found that, in some cases, sex-agnostic efforts might be missing loci of interest, and looked into possible improvements in the prediction of high-level phenotypes. Finally, we studied the potential functional role of the differences observed through sex-biased gene expression and gene-level analyses. Our results suggest the need to consider sex-aware analyses for future studies in order to shed light into possible sex-specific molecular mechanisms.

INTRODUCTION

In recent years, there has been growing evidence of common genetic variation having different effects on males and females[1,2]. This, along with sex-biases observed in the human transcriptome[3–10], the presence of a distinct hormone milieu in each sex, and differential environmental pressures arising from gender societal roles[2,11], has led to an increased study of the potential importance of GxS interactions to understand the underlying biology of complex traits, including the estimation of disease risk. Previous studies have investigated differences in heritability between the sexes (h^2)[12–15], departure of genetic correlations from 1 (r_g)[12,14,16–19], and performed sex-stratified genome-wide association studies (GWAS) to directly assess differences in the effects of genetic variants between the sexes[6,20–27]. These studies have however been limited with regards to the number of traits studied or statistical power. Furthermore, insights into how differences in genetic architecture translate into differences in complex traits have been lacking. Accounting for sex differences is of great importance as sex-agnostic analyses could potentially be masking sex-specific effects, and which could - if better understood - lead to better personalized treatment and an improved understanding of the biological mechanisms driving these differences[17,28].

The objective of the current study was to assess both the existence and scope of GxS interactions in the human genome by estimating sex-specific heritability and genetic correlations, as well as performing sex-stratified GWAS analyses. To this end, we analysed 530 traits using 450K individuals of European ancestry from the UK Biobank. Furthermore, we evaluated the potential of improving trait predictions using sex-stratified polygenic scores (PGS), as well as looked into the possibility of missing loci of interest in non-sex stratified studies. Finally, to shed light on the downstream effects of sex-differences in genetic architecture, we performed a functional *in silico* analysis.

RESULTS

Data overview

Using the July 2017 release of the UK Biobank dataset, we performed sex-stratified GWASs and partition of variance analyses for 530 traits (446 binary and 84 non-binary, Supplementary Table 1) within 452,264 individuals of European ancestry (245,494 females and 206,770 males) using DISSECT[29]. Linear Mixed Models (LMMs) were fitted for each phenotype by sex. We then tested the

association of 9,072,751 autosomal and 17,364 X-chromosome genetic variants, obtaining estimates of the genetic effects of each variant in each sex. In our quality control (QC) stage we excluded genetic variants with a minimum allele frequency (MAF) < 10% in the analysis of binary traits due to the general limited number of cases available, thus reducing the number of variants considered for these traits to 4,229,346 autosomal and 7,227 genotyped X-chromosome genetic variants (see Methods). The results of the autosomal analyses were used to estimate sex-stratified genetic parameters (such as heritability) and genetic correlations. We then tested for differences in these genetic parameters and between the effects of genetic variants estimated within sexes (see Methods).

Because of the different QC treatment of binary and non-binary traits, as well as the difference in phenotype characteristics, results are presented separately for both throughout this work, not as a means to compare the results between the two but so as to contain them both to their own separate categories.

Heritability differences between the sexes

Heritability (here referring to SNP heritability[30]) is defined as the fraction of the variation of a trait that can be explained by the additive effects of genetic variation. A difference in heritability between the sexes would entail a difference in the fraction of the variance of a trait that is accounted for by the genotype, and thus a possible difference in the underlying genetic mechanisms of said trait. Out of the 530 traits studied, 41/84 (48.88%) non-binary traits and 30/446 (6.73%) binary traits showed significant differences in their heritability between the sexes (FDR corrected p, termed q, < 0.05, Figure 1, Supplementary Table 1). Of these, a total of 25/41 (60.98%) non-binary and 14/30 (46.67%) binary traits had a larger heritability in males than in females. Non-binary traits with the largest significant difference in heritability between the sexes included body mass traits, while binary traits included ankylosing spondylitis, disorders of mineral metabolism and soft tissue disorders (Supplementary Table 1).

Although differences in the heritability of traits between the sexes can offer potential insights into genetic differences, one must also consider that these could arise due to differences in environmental variances. Hence, we looked for differences in genetic variance between the sexes ($q < 0.05$). We found that 65/84 (77.38%) non-binary and 136/446 (30.49%) binary traits showed significant differences in the amount of genetic variance estimated in each sex, including impedance-related traits, and diseases of the thyroid and heart (Extended Data Figure 1, Extended Data Figure 2, Supplementary Table 1). A total of 6/84 (7.14%) non-binary traits and 4/446 (0.90%) binary traits presented a significant difference in heritability but no significant difference in genetic variance between the sexes, indicating differences in environmental variance. These included numerous blood phenotypes (such as lymphocyte percentage and mean reticulocyte volume), ease of skin tanning, venous thrombosis disease, anaphylaxis/allergy and diseases of the digestive system.

Finally, we observed significant differences in evolvability (a measure of the ability to undergo adaptation) between sexes for 56/84 (66.67%) non-binary and 35/446 (7.85%) binary traits ($q < 0.05$, Extended Data Figure 2, Supplementary Table 1). These included binary traits like ankylosing spondylitis and malabsorption/coeliac disease, as well as non-binary traits like current tobacco

smoking. These estimates offer further evidence for differences in the underlying genetic architecture of the traits considered, paralleling previous reports at smaller scales for traits including height, waist-hip circumference ratio and weight[12–14].

Genetic correlations indicate genotype by sex interactions

Genetic correlations between two sub-groups of the population are usually interpreted as a measure of shared underlying genetics, and are a means to estimate the size of putative genotype by group interactions. Genetic correlations between the sexes can thus offer insights into the common genetic control of complex traits and diseases of males and females.

We obtained genetic correlations between the sexes for a total of 83 non-binary and 77 binary traits with over 5,000 cases using LD score regression (LDSC, Supplementary Table 1)[31] which met our QC criteria (see Methods). Genetic correlations ranged from 0.716 to 0.996 for non-binary traits and from 0.226 to 1.099 for binary traits (note that with heritability close to zero application of LDSC can result in r_g exceeding the theoretically valid range [-1, 1][32]). A total of 58/83 (69.88%) non-binary traits and 11/77 (14.29%) binary traits had a r_g significantly different from 1 ($q < 0.05$, Figure 2A). These included binary traits like hernia ($r_g = 0.59$, $q = 4.04 \times 10^{-10}$), eczema ($r_g = 0.61$, $q = 0.04$) and gastric reflux ($r_g = 0.67$, $q = 0.02$), and non-binary traits like waist-hip circumference ratio ($r_g = 0.72$, $q = 8.43 \times 10^{-37}$) and alcohol intake frequency ($r_g = 0.85$, $q = 7.49 \times 10^{-9}$, Figure 2B). Our r_g estimates for several non-binary traits were in line with what has previously been published, noting that one of the compared studies also used UK Biobank data (Supplementary Table 2, Extended Data Figure 3)[12,14,16].

Genome-wide genetic effect comparison across traits

We directly assessed whether each genetic variant in the genome had different effects in males and females through sex-stratified GWASs. Our genome and trait-wide genetic effect comparison between the sexes (see Methods) yielded a total of 61/84 (72.62%) non-binary and 42/446 (9.42%) binary traits with at least one autosomal genetic variant presenting a significantly different effect at a $p < 1 \times 10^{-8}$ threshold (Supplementary Table 3), hereon termed as a sex-dimorphic SNP or sdSNP (Table 1 shows traits with the largest number of independent autosomal sdSNPs, i.e. lead sdSNPs, found). The distribution of sdSNPs across the genome for the traits with most autosomal lead sdSNPs is shown in Figure 3. When testing the X-chromosome variants, we found 28/84 (33.33%) non-binary traits with at least one sdSNP. Considering the autosomal genome, the trait with the largest number of sdSNPs was waist-hip circumference ratio, a complex trait that has appeared frequently in analyses of sex differences in genetic architecture[12,20,21]. A total of 2,421 sdSNPs were found for this trait, which represent 100 unique loci after linkage disequilibrium (LD) clumping (see Supplementary Note). The trait with the most sdSNPs in the X-chromosome was hematocrit percentage, with a total of 12 that mapped to 5 unique loci after LD clumping (Supplementary Table 1 and 3). Our results include replications of several previously reported loci for traits including anthropometric measurements or diseases like gout[20–24,33].

A total of 4,179 (4,179/9,072,751 = 0.046%) and 4,196 (4,196/4,229,346 = 0.099%) autosomal genetic variants showed evidence of GxS in at least one non-binary or binary trait respectively ($p < 1 \times 10^{-8}$),

which mapped to 264 and 88 independent loci respectively. A total of 37 (37/7,227 = 0.213%) X-chromosome variants showed evidence of sex differences in at least one non-binary trait ($p < 1 \times 10^{-8}$), which mapped to 8 unique loci. The sdSNP associated to the highest number of traits (a total of 17) was rs115775278, an imputed intergenic variant found on chromosome 16. The closest genes to this variant include LOC105371341 (an uncharacterized non-protein coding RNA gene, with transcription start site, TSS, ~40kb downstream), LOC390739 (MYC-binding protein pseudogene, TSS ~50kb upstream), PMFBP1 (Polyamine Modulated Factor 1 Binding Protein 1, TSS ~60kb downstream), and LINC01572 (a long intergenic non-protein coding RNA gene, TSS ~470kb upstream). PMFBP1 has been linked to spermatogenesis function[34]. The distribution of hits across the autosomal and X chromosome genome is shown in Extended Data Figure 4.

Several sanity checks were performed to support these results (see Supplementary Note), which included fitting alternative models (including Logistic Mixed Models[35]), comparison to the Genetic Investigation of Anthropometric Traits (GIANT) cohort[20,21], and a randomization scheme. These checks suggested results for the nucleated red blood cell percentage trait likely represent false positives (see Supplementary Note), thus its exclusion from future discussion.

Phenotype prediction with sex-stratified effect estimates

We studied whether genetic prediction could potentially be improved using sex-stratified models. To this end, we estimated genetic effects in a training population of 300,000 UK Biobank white British individuals in two different ways: (1) including both sexes in the model (obtaining sex-agnostic effects) and (2) using each sex in a separate model (obtaining sex-specific effects, see Methods). We then used a testing population consisting of 43,884 white British individuals to compare the performance of these two models in three different ways using PGSs: (1) obtaining predictions from the sex-agnostic effects (*agnostic* PGS), (2) obtaining predictions using the female effects applied to females and the male effects applied to males (*same* PGS), and (3) obtaining predictions using the female effects to predict in males and vice versa (*opposite* PGS). Prediction accuracy was measured as the correlation (r) between or the area under the curve (AUC) for our prediction and the true phenotype value for non-binary and binary traits respectively. Only lead sdSNPs were used in our PGS calculation. Due to the general low number of sdSNPs across traits, we focused our comparison on phenotypes with at least 10 lead sdSNPs. These included 7 non-binary traits (waist-hip circumference ratio, standing height, trunk fat percentage, hip circumference, whole body water mass, trunk predicted mass and trunk fat-free mass) and 3 binary traits (ankylosing spondylitis, gout and hypothyroidism/myxoedema).

Although of the 7 non-binary traits tested only waist-hip circumference ratio showed a moderately significant difference in prediction accuracy (correlation comparison $p = 0.059$, see Methods) between *same* PGS and *agnostic* PGS in males (Supplementary Table 11), all 7 traits consistently presented a larger prediction accuracy when comparing the sex-stratified model with the agnostic model, thus suggesting that the stratified model captures the effect sizes better than the agnostic model. On the other hand, we consistently observed smaller prediction accuracies when the stratified model was used to perform predictions on the opposite sex (*opposite* PGS). We did not observe any consistent prediction improvements for the 3 binary traits considered (Supplementary Table 11).

A limitation of our approach is the overlap between our discovery data set (used to establish sdSNPs) and our training and testing data sets in our prediction analysis (see Methods). We repeated our analysis with independent data sets (see Methods) for waist-hip circumference ratio, and we found that the *same* PGS and *agnostic* PGS had similar predictive ability in females (r = 0.132 with $p = 1.85 \times 10^{-98}$ and r = 0.133 with $p = 4.02 \times 10^{-99}$ respectively), the *same* PGS surpassing the *agnostic* for males (r = 0.038 with $p = 7.91 \times 10^{-8}$ and r = 0.024 with $p = 9.97 \times 10^{-4}$ respectively), however the differences in predictive power were not significantly different in either case (correlation comparison $p > 0.05$).

A possible explanation for the modest increase in predictive power found when using our sex-stratified models, when taking observed differences in heritability into account, is the potential existence of large numbers of SNPs of small dimorphic effect across the genome. These small effects remain undetected in a GWAS, and as such, are not being included in our predictions. This reasoning parallels the missing heritability problem[36], where the predicted heritability of traits can't be explained by the detected GWAS associations, a hypothesis for which is the existence of large amounts of variants of small effect that are yet to be found. Consistent with this theory, we found that our sdSNPs generally accounted for a very low percentage of the sex-specific heritability for the considered traits (Supplementary Table 12, Methods), which ranged from 0.18% to 0.65%. Waist-hip circumference ratio was the exception, for which our sdSNPs accounted for 12.10% and 1.70% of the female and male specific heritability, respectively, which could be due to the substantially larger number of sdSNPs identified. This could also be, however, due to sdSNPs having a generally small effect on the phenotypes considered.

Potential masking of loci in sex-agnostic studies

Currently, many GWASs fit non-sex-stratified models. However, a situation could arise in which (1) a locus possesses a differentially signed genetic effect in each sex or (2) a genetic variant shows a larger effect in one of the sexes and a small or no effect in the other. In any of these situations, the power of detecting the variant will be reduced in a non-stratified analysis, and the variant effect size misestimated in both sexes. This phenomenon we term as "masking" of a genetic effect.

To assess masking effects in the UK Biobank, we evaluated the total number of genetic variants that were found to be significantly associated with a trait in a sex-stratified GWAS (i.e. associated to a trait in males and/or females), but were not significantly associated in a sex-agnostic model. We performed this analysis on the 530 traits in our study, considering a genetic variant as potentially masked if it is significantly associated in females and/or males but not for the mixed population at a $p < 1 \times 10^{-8}$ threshold (see Methods).

We found that 98/446 (21.97%) binary and 78/84 (92.86%) non-binary traits had at least one genetic variant that showed potential masking across the autosomal genome (Supplementary Table 14, Extended Data Figure 5A). On average, the percentage of these variants that presented opposite signs in each sex was 20.03% (SD 36.25%) in binary traits and 5.34% (SD 9%) in non-binary traits (Extended Data Figure 5B). This may indicate that for a small percentage of traits opposite signed genetic effects are leading to masking. However, this could also be the result of smaller sample sizes leading to false positives in one sex but not the other.

A total of 93 traits (33 binary and 60 non-binary) presented at least one sdSNP and one potentially masked variant. We also found a significant correlation between the number of potentially masked variants and the number of sdSNPs (r = 0.624, p = 9.793 x 10^{-8}) for non-binary traits, as shown in Extended Data Figure 6. On average, the percentage of masked variants that presented sex differences in binary traits was 3.83% (SD 18.97%), and 1.39% (SD 2.22%) in non-binary traits (Extended Data Figure 5C). These low percentages could indicate that masked variants may have different effects on the two sexes, just not surpassing our significance threshold to be considered sdSNPs. On the other hand, 42 of our 103 traits with at least one sdSNP had one of these sdSNPs potentially masked, and, on average, the percentage of sdSNPs that were potentially masked in binary traits was 12.30% (SD 30.59%) and 18.44% (SD 21.99%) in non-binary traits (Extended Data Figure 5D). This could suggest a large number of potentially interesting variants that present a difference in genetic effect between the sexes could be being missed in sex-agnostic studies.

Gene-level analyses

To gain insight into the biological meaning of these results, gene enrichment analyses were carried out for all 103 phenotypes with at least one sdSNP. To do this, the two-tailed sex-comparison p-values for the sdSNPs found were converted to two one-tailed p-values (p_F and p_M) according to the sex which presented the largest genetic effect (see Methods). Using these two sets of p-values in combination with MAGMA we then estimated the degree of sex dimorphism of each gene, thus obtaining dimorphic gene lists, which were dominant in females or males (i.e. which presented a significantly larger effect in one sex versus the other, see Methods, Supplementary Table 15 and 16, Supplementary Note). The GENE2FUNC tool in FUMA was then used to investigate any functional enrichments among these dimorphic gene lists (Supplementary Table 17, 18 and 19, Supplementary Note) for the 10 traits with the largest number of sexually different genes. These were largely all of the anthropometric class and were: standing height, waist-hip circumference ratio, trunk predicted mass, trunk fat-free mass, trunk fat percentage, whole body fat-free mass, basal metabolic rate, impedance of arm (left), body fat percentage and hip circumference. As a background to compare our results to, this procedure was repeated using sex-agnostic GWAS results, obtaining gene sets enriched in genes associated to each of our 10 phenotypes (see Methods).

A total of 4,840 gene sets were found to be enriched in either male or female dominant genes across the 10 traits considered (q < 0.05, Supplementary Table 18 and 19). Genes dominant in one sex or the other were found to be enriched in sets in an exclusive manner (i.e. sets would not show a larger amount of both male and female dominant genes than what would be expected by random), with the average percentage of shared enriched sets across traits being 3.3%, SD = 3.34% (Supplementary Table 17). A total of 383/4,840 gene sets were found to be significantly differentially enriched between male and female dominant genes in at least one of the traits considered (Fisher's exact test q < 0.05), with an average of 12.88% (SD 12.96%) of gene sets being differentially enriched across traits (Supplementary Table 17). Furthermore, 251/383 gene sets were found to also show a significant difference in enrichment when comparing with the results of our background of sex-agnostic associated genes (Fisher's exact test q < 0.05, see Methods), the mean percentage of gene sets presenting this behaviour across traits being 67.30% (SD 18.87%, Supplementary Table 17).

Heatmaps were produced considering the aforementioned 251 gene sets, with hierarchical clustering both by gene set and by trait (Extended Data Figure 7). Most notably, we find clusters of sets pertaining to small non-coding RNA (sncRNA) biogenesis and RNA-mediated silencing, enriched in female dominant genes for body mass-related traits. It has previously been postulated that miRNA may play a role in the regulation of phenotypic sex differences due to its ability to regulate large numbers of genes with a high degree of specificity, with intervention of the sex chromosomes and/or gonadal hormones[37].

Sex differences in gene expression regulation

Differences between the sexes in complex traits could be partially explained by sex-specific gene expression regulation, which could lead to differences downstream across biological pathways and traits, and thus to the detection of GxS interactions in GWASs. Although studies have been carried out searching for differential gene expression between the sexes (sex-DE) in a variety of tissues of interest, studies linking sex to differences in gene expression regulation (sex-specific or sex-biased eQTLs) are few, with often contradictory results[5,38–41]. These mixed results could be due to the contribution of GxS to gene expression being tissue specific, a lack of sufficient statistical power, or the fact that this contribution occurs only on a small number of genes[40]. Overall, a system-wide analysis (i.e. across a wide variety of tissues) is yet to be carried out in order to determine whether there is evidence of sex-biased eQTLs and whether some tissues are more prone to sex-specific regulation than others.

In order to bring light to potential intermediary mechanisms underlying differences in genetic architecture between the sexes we investigated whether our lead sdSNPs could also be acting as sex-biased eQTLs. To do this, we performed an eQTL analysis, looking for GxS interactions in gene expression, considering genes within a 1Mb window to our lead sdSNPs (Supplementary Table 3, see Methods). This was done for a total of 39 tissues from the Genotype-Tissue Expression (GTEx) consortium v6, originating from up to 450 individuals.

A total of 8 sex-biased eQTLs were found at a $q < 0.05$ threshold (Supplementary Table 20, 21, 22 and 23). We also checked for enrichment of GxS in gene-variant pairs for variants that presented evidence of sex dimorphism (genetic effect comparison between the sexes $p < 1 \times 10^{-8}$) versus those that did not (genetic effect comparison between the sexes $p > 0.5$), using contingency tables (see Methods). We found enrichment for a small number of the tissues considered (see Supplementary Note, Supplementary Table 24).

The variant rs56705452 from chromosome 6 was found to be a sex-biased eQTL for the transcript ENSG00000204520.8, which corresponds to the gene MICA (Extended Data Figure 8), in muscle skeletal tissue. This gene encodes the highly polymorphic major histocompatibility complex class I chain-related protein A, and variations of this gene have been associated to susceptibility to psoriasis, psoriatic arthritis and ankylosing spondylitis, amongst others[42]. Interestingly, this sdSNP has been shown to bind FOXA1 through ChIP-Seq experiments[43], a protein that dictates the binding location of androgen and oestrogen receptors, and that has been found to play a role in the sexually dimorphic

presentation of various cancers[44,45]. Furthermore, this sdSNP was found to act in a sexually different manner in regards to its association to ankylosing spondylitis in our genome-wide sdSNP analysis. This result would be consistent with a hypothesis where this sdSNP is regulating MICA in a sex-dependent manner in the muscle tissue, thus leading to differences in ankylosing spondylitis presentation between the sexes when one possesses a particular variant.

The small number of significant sex-biased eQTLs found parallels the findings of recent study by Porcu et al[41], where they conclude that millions of samples would be necessary to observe sex-specific trait associations that are fully driven by sex-biased eQTLs. Overall, this type of pipeline could help in the future when short-listing biomarkers for risk susceptibility in males and females, help develop precision medicine strategies for each of the sexes, and bring light into the underlying mechanisms of the disease/trait of interest as well as possible underlying sexually different molecular networks, once larger sample sizes become available.

DISCUSSION

In this study, we have delved into the differences in genetic architecture between the sexes in the UK Biobank for a total of 530 traits from around half a million individuals. This has enabled us to assess the genetics of sex dimorphism at a depth and breadth not previously achieved.

Overall, we have found evidence of sex differences for a large number of the traits considered, though be it of generally modest magnitude, through a thorough investigation of sex-specific genomic parameters. A total of 71 traits were found to present significantly different heritability estimates between the sexes, while a total of 69 presented genetic correlations between the sexes that significantly differed from one, indicating the presence of genetic heterogeneity across these complex traits. In order to dissect this heterogeneity and pin-point genetic sites that could be differentially associated to these traits, sex-stratified GWASs were performed, yielding over 100 traits with at least one sdSNP. These traits included those of the anthropometric class as well as diseases like gout, ankylosing spondylitis or hypothyroidism. These results reinforce the need for future studies to account for genetic sex heterogeneity to fully understand the genetic underpinnings of disease and ultimately shed light on potential sex-specific biological mechanisms.

Having found evidence of GxS across the genome, we investigated whether sex-specific genetic models could improve phenotypic prediction. While no statistically significant improvement in prediction was found for the traits considered, a consistent trend of increased predictive accuracy was seen when comparing the results of sex-specific models to those of a sex-agnostic model. Putting our results in context with the heritability differences found, we postulate the potential existence of large numbers of loci presenting small amounts of dimorphism with insufficient statistical power to be detected in our analysis that could account for both this missing heritability difference as well as the absence of increased predictive power. We also investigated whether sex-agnostic models could potentially be missing loci of interest, and found indications of potential masking for 176 traits, with further investigation being needed to replicate these results.

Finally, gene set enrichment and eQTL analyses were performed in an effort to translate our GWAS results to function and bring light to potential mechanisms underlying the observed dimorphism across traits. Our eQTL analysis found a total of 8 sex-biased eQTLs, but our results parallel previous reports on the need for larger sample sizes to truly uncover potential links between sex-biased eQTLs and sdSNPs. Our gene-set enrichment analysis suggests a link to miRNA regulation, which has been hypothesized in the past to underlie sex differences. Further studies are needed to truly understand what underlies sdSNPs, moving beyond gene expression regulation mechanisms and looking at other biological regulatory mechanisms and omics data sets.

ACKNOWLEDGEMENTS

This research was funded by the BBSRC through programme grants BBS/E/D/10002070 (AT), BBS/E/D/30002275 (AT), BBS/E/D/30002276 (AT, JP, KR, AT), MRC research grants MR/P015514/1 (AT) and MR/R025851/1 (AT, OCX), HDR-UK award HDR-9004 (AT), and The Roslin Foundation's Steve Bishop PhD Fellowship (EB). This research has been conducted using the UK Biobank Resource under project 788.

AUTHOR CONTRIBUTION

A. Tenesa conceived the study. A. Tenesa, O. Canela-Xandri, K. Rawlik and E. Bernabeu designed the genetic architecture, prediction, masking and eQTL analyses. A. Talenti and J. Prendergast designed the gene-level analyses. O. Canela-Xandri, K. Rawlik and E. Bernabeu pre-processed the data and conducted modelling. E. Bernabeu conducted the statistical analyses and prepared the initial manuscript. All authors contributed and commented on the development of the manuscript.

ETHICS OVERSIGHT

The UK Biobank project was approved by the National Research Ethics Service Committee North West-Haydock (REC reference: 11/NW/0382). An electronic signed consent was obtained from the participants (more information on UK Biobank participant consent can be found here: https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/Consent.pdf). This research has been conducted using the UK Biobank Resource under project 788.

COMPETING INTEREST STATEMENT

The authors declare no competing interests.

REFERENCES (MAIN TEXT)

1. Ober, C., Loisel, D. A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nat Rev Genet* **9**, 911–922 (2008).
2. Weiss, L. A., Pan, L., Abney, M. & Ober, C. The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet* **38**, 218–222 (2006).
3. Mayne, B. T. *et al.* Large Scale Gene Expression Meta-Analysis Reveals Tissue-Specific, Sex-Biased Gene Expression in Humans. *Front Genet* **7**, 183 (2016).
4. Jansen, R. *et al.* Sex differences in the human peripheral blood transcriptome. *BMC Genomics* **15**, 33 (2014).
5. Trabzuni, D. *et al.* Widespread sex differences in gene expression and splicing in the adult human brain. *Nat Commun* **4**, 2771 (2013).
6. Kottgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* **45**, 145–154 (2013).
7. InanlooRahatloo, K. *et al.* Sex-based differences in myocardial gene expression in recently deceased organ donors with no prior cardiovascular disease. *PLoS One* **12**, e0183874 (2017).
8. Mele, M. *et al.* Human genomics. The human transcriptome across tissues and individuals. *Science (80-.).* **348**, 660–665 (2015).
9. Gershoni, M. & Pietrokovski, S. The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol* **15**, 7 (2017).
10. Guo, S. *et al.* Identification and analysis of the human sex-biased genes. *Br. Bioinform* **19**, 188–198 (2018).
11. Khramtsova, E. A., Davis, L. K. & Stranger, B. E. The role of sex in the genomics of human complex traits. *Nat. Rev. Genet.* **20**, 173–190 (2019).
12. Rawlik, K., Canela-Xandri, O. & Tenesa, A. Evidence for sex-specific genetic architectures across a spectrum of human complex traits. *Genome Biol* **17**, 166 (2016).
13. Gilks, W. P., Abbott, J. K. & Morrow, E. H. Sex differences in disease genetics: evidence, evolution, and detection. *Trends Genet* **30**, 453–463 (2014).
14. Traglia, M. *et al.* Genetic Mechanisms Leading to Sex Differences Across Common Diseases and Anthropometric Traits. *Genetics* **205**, 979–992 (2017).
15. Ge, T., Chen, C.-Y., Neale, B. M., Sabuncu, M. R. & Smoller, J. W. Phenome-wide heritability analysis of the UK Biobank. *PLOS Genet.* **13**, e1006711 (2017).
16. Yang, J. *et al.* Genome-wide genetic homogeneity between sexes and populations for human height and body mass index. *Hum Mol Genet* **24**, 7445–7449 (2015).
17. Stringer, S., Polderman, T. & Posthuma, D. Majority of human traits do not show evidence for sex-specific genetic and environmental effects. *Sci Rep* **7**, 8688 (2017).
18. Vink, J. M. *et al.* Sex differences in genetic architecture of complex phenotypes? *PLoS One* **7**, e47371 (2012).
19. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
20. Winkler, T. W. *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* **11**, e1005378 (2015).
21. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2018).
22. Doring, A. *et al.* SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat Genet* **40**, 430–436 (2008).
23. Randall, J. C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* **9**, e1003500 (2013).
24. Rask-Andersen, M., Karlsson, T., Ek, W. E. & Johansson, Å. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat. Commun.* **10**, 1–10 (2019).
25. Dumitrescu, L. *et al.* Sex differences in the genetic predictors of Alzheimer's pathology. doi:10.1093/brain/awz206
26. Graham, S. E. *et al.* Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis. *Nat. Commun.* **10**, 1–9 (2019).
27. Hartiala, J. A. *et al.* Genome-wide association study and targeted metabolomics identifies sex-specific

association of CPS1 with coronary artery disease. *Nat. Commun.* **7**, (2016).
28. Beery, A. K. & Zucker, I. Sex bias in neuroscience and biomedical research. *Neurosci Biobehav Rev* **35**, 565–572 (2011).
29. Canela-Xandri, O., Law, A., Gray, A., Woolliams, J. A. & Tenesa, A. A new tool called DISSECT for analysing large genomic data sets using a Big Data approach. *Nat Commun* **6**, 10162 (2015).
30. Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49**, 1304–1310 (2017).
31. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236–1241 (2015).
32. Ni, G., Moser, G., Schizophrenia Working Group of the Psychiatric Genomics Consortium, S. W. G. of the P. G., Wray, N. R. & Lee, S. H. Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. *Am. J. Hum. Genet.* **102**, 1185–1194 (2018).
33. Kolz, M. *et al.* Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet* **5**, e1000504 (2009).
34. Zhu, F. *et al.* Mutations in PMFBP1 Cause Acephalic Spermatozoa Syndrome. *Am. J. Hum. Genet.* **103**, 188–199 (2018).
35. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* 1–7 (2021). doi:10.1038/s41588-021-00870-7
36. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
37. Morgan, C. P. & Bale, T. L. Sex differences in microRNA regulation of gene expression: No smoke, just miRs. *Biology of Sex Differences* **3**, 22 (2012).
38. Dimas, A. S. *et al.* Sex-biased genetic effects on gene regulation in humans. *Genome Res* **22**, 2368–2375 (2012).
39. Yao, C. *et al.* Sex- and age-interacting eQTLs in human complex diseases. *Hum. Mol. Genet.* **23**, 1947–1956 (2014).
40. Kassam, I. *et al.* Autosomal genetic control of human gene expression does not differ across the sexes. *Genome Biol* **17**, 248 (2016).
41. Porcu, E. *et al.* The role of gene expression on human sexual dimorphism: too early to call. *bioRxiv* 2020.04.15.042986 (2020). doi:10.1101/2020.04.15.042986
42. Bahram, S., Bresnahan, M., Geraghty, D. E. & Spies, T. A second lineage of mammalian major histocompatibility complex class I genes. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 6259–6263 (1994).
43. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
44. Robinson, J. L. L. & Carroll, J. S. FoxA1 is a Key Mediator of Hormonal Response in Breast and Prostate Cancer. *Front. Endocrinol. (Lausanne).* **3**, 68 (2012).
45. Li, Z., Tuteja, G., Schug, J. & Kaestner, K. H. Foxa1 and Foxa2 Are Essential for Sexual Dimorphism in Liver Cancer. (2012). doi:10.1016/j.cell.2011.11.026

FIGURES

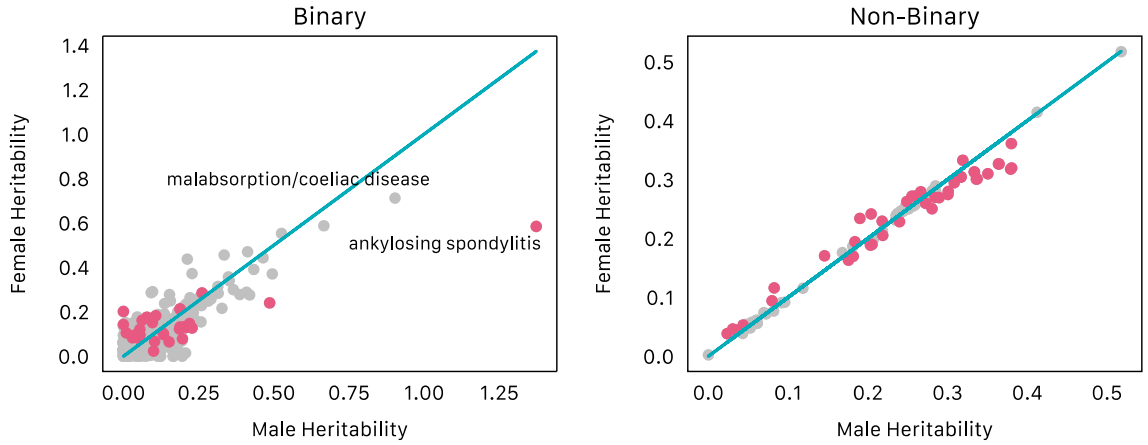


Figure 1. Sex-specific heritability estimate comparison. Scatterplots of male heritability estimates against female heritability estimates for binary traits (on the left) and non-binary traits (on the right) are shown. Each point represents a trait, which is marked in pink when heritability between the sexes is significantly different (see Methods) at a threshold of $q < 0.05$. Note that for the binary traits, heritability estimates were considered on the liability scale, which led to some estimates over 1 (see Methods). Blue line corresponds to x = y, and traits straying the furthest away from said line are annotated.

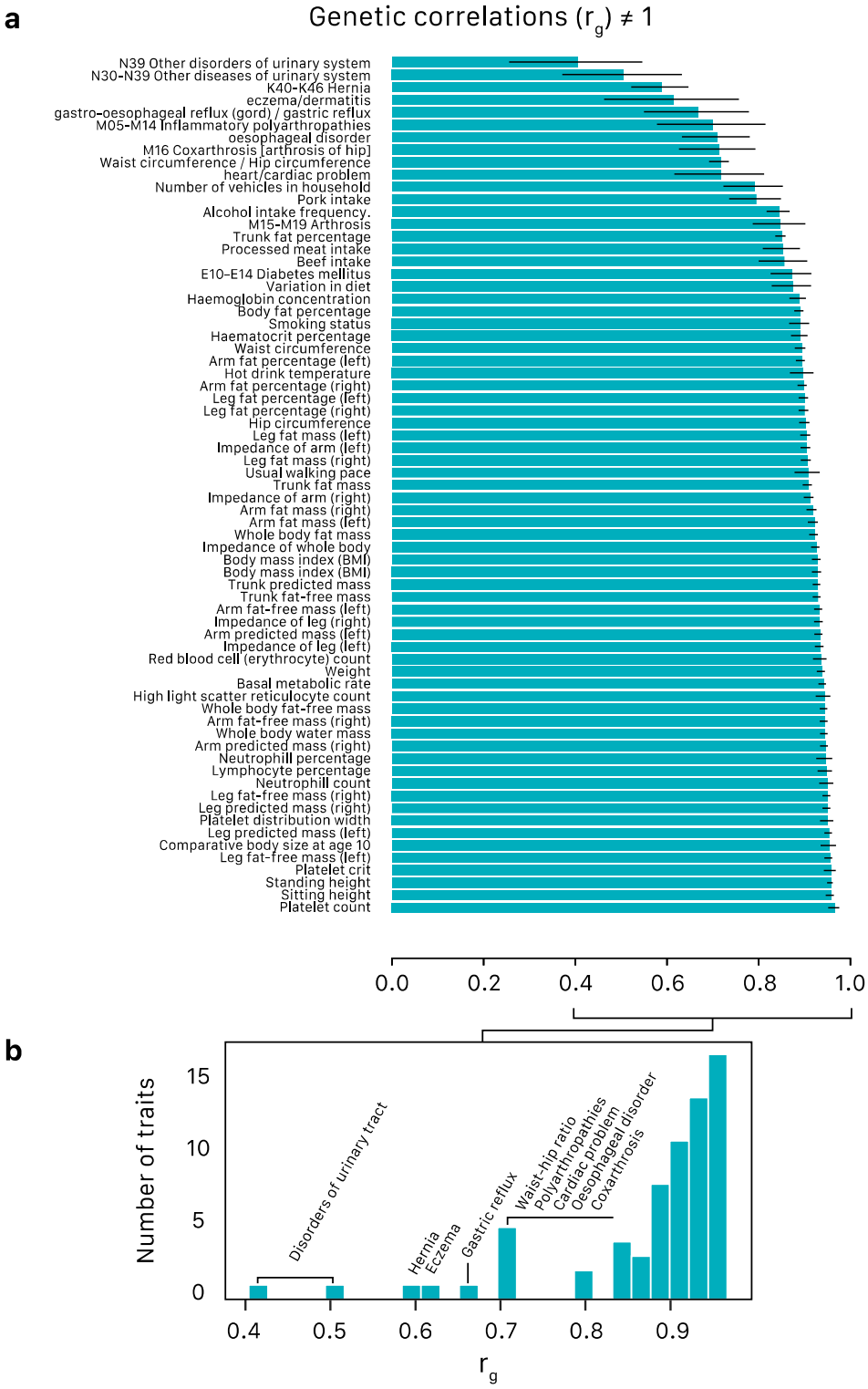


Figure 2. Genetic correlations between the sexes across traits. (a) Barplot of genetic correlations (r_g) between the sexes for traits that were found to have a r_g significantly different from one ($q < 0.05$, see Methods). Black bars indicate the standard errors of the r_g estimates (r_g +/- SE_{r_g}). (b) Histogram of genetic correlations that were found to be significantly different from one ($q < 0.05$) for both binary and non-binary traits.

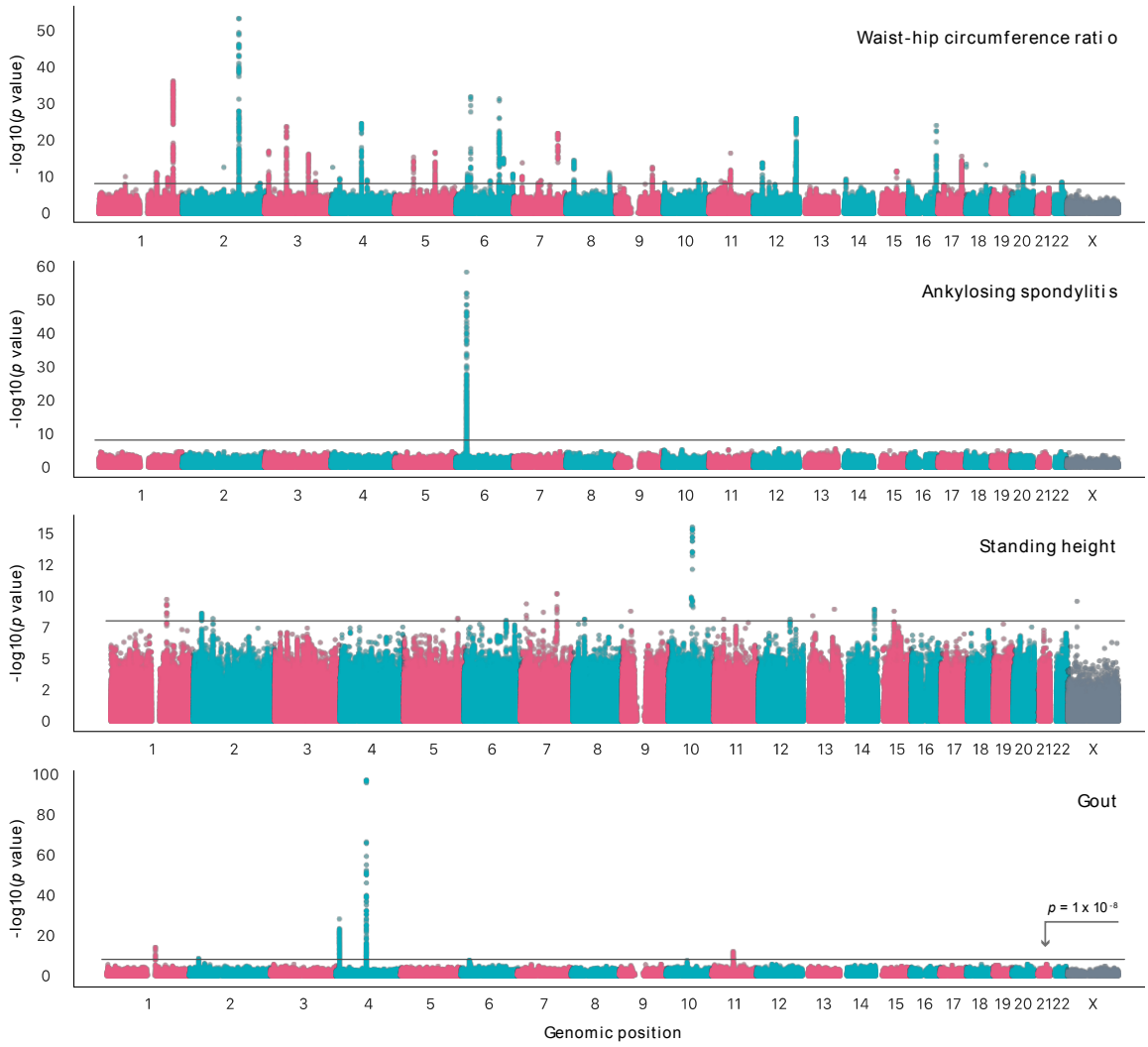
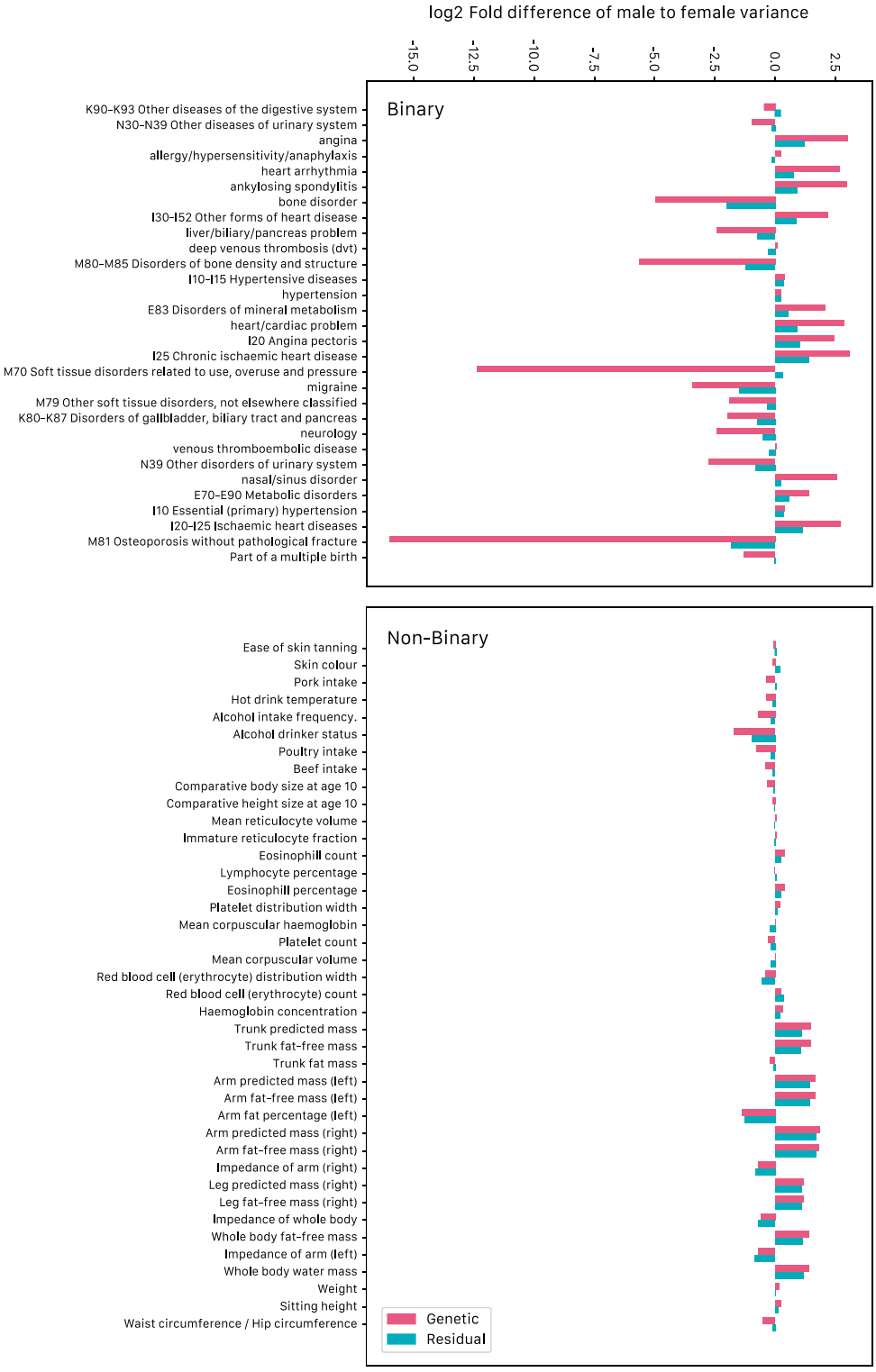
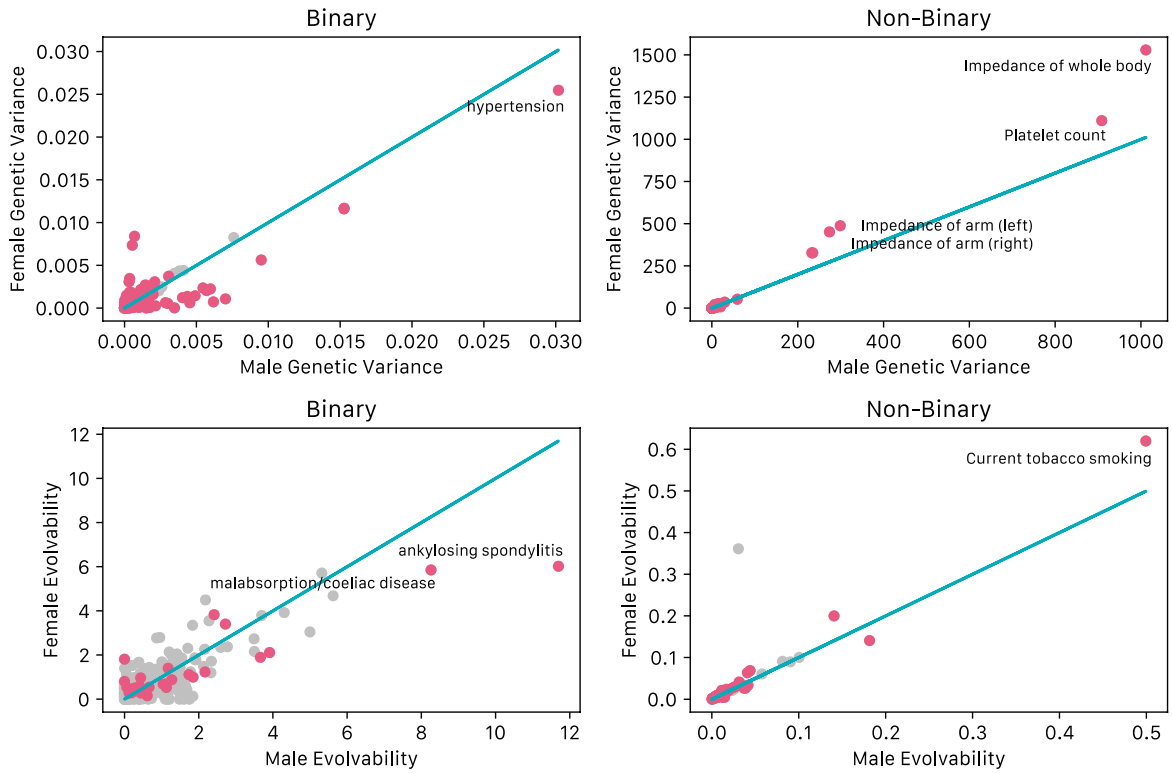


Figure 3. Manhattan plots for traits with most lead sdSNPs. The x axis corresponds to the genomic position in the genome, and the y axis to the $-\log_{10}$ *p*-value of the two-sided t-test (see Methods) for which the null hypothesis is that no difference between the sexes exists. Each point corresponds to a genetic variant. Points that go above the statistical significance line at $-\log_{10} p = 1 \times 10^{-8}$ are considered to be sdSNPs. Traits represented: waist-hip circumference ratio, ankylosing spondylitis, standing height, and gout.

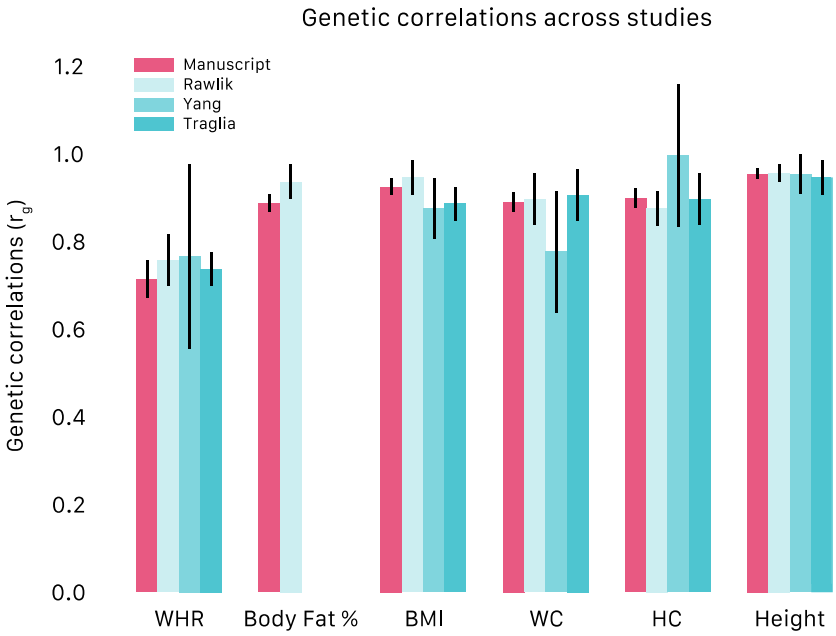
EXTENDED DATA FIGURES



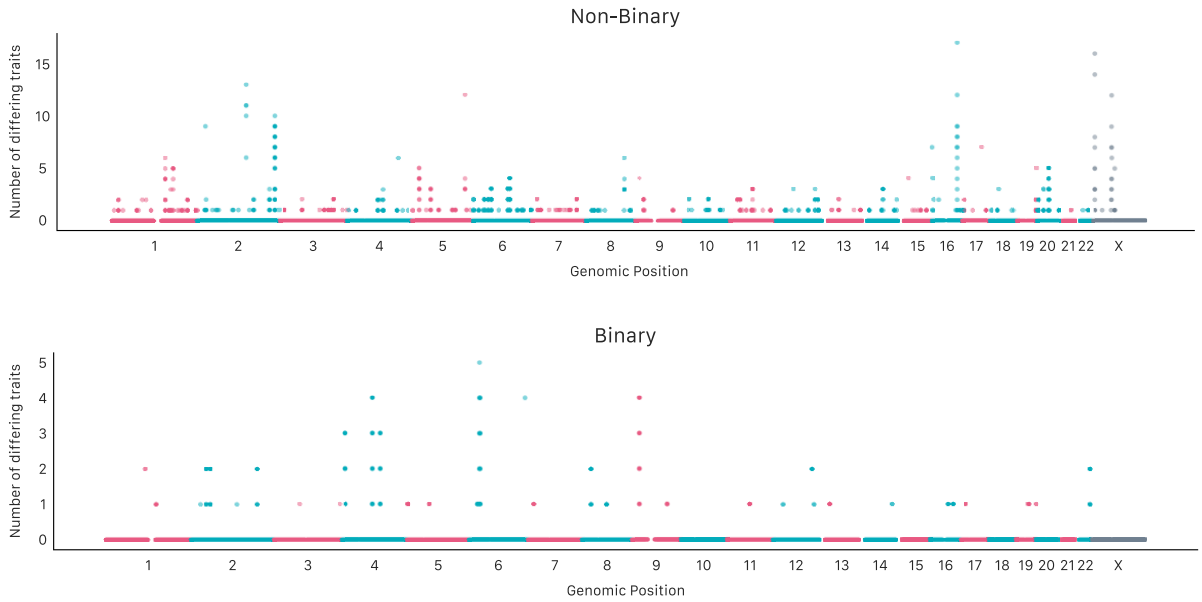
Extended Data Figure 1. Barplot of log2 fold difference in variance between males and females for binary (top) and non-binary (bottom) traits with a significantly different heritability between the sexes at a $q < 0.05$ threshold. Pink bars represent fold change between the sexes in genetic variance, and blue bars represent fold change between the sexes in residual variance. Fold change is calculated as log2(male variance/female variance), thus positive fold change = larger variance in males, negative fold change = larger variance in females.



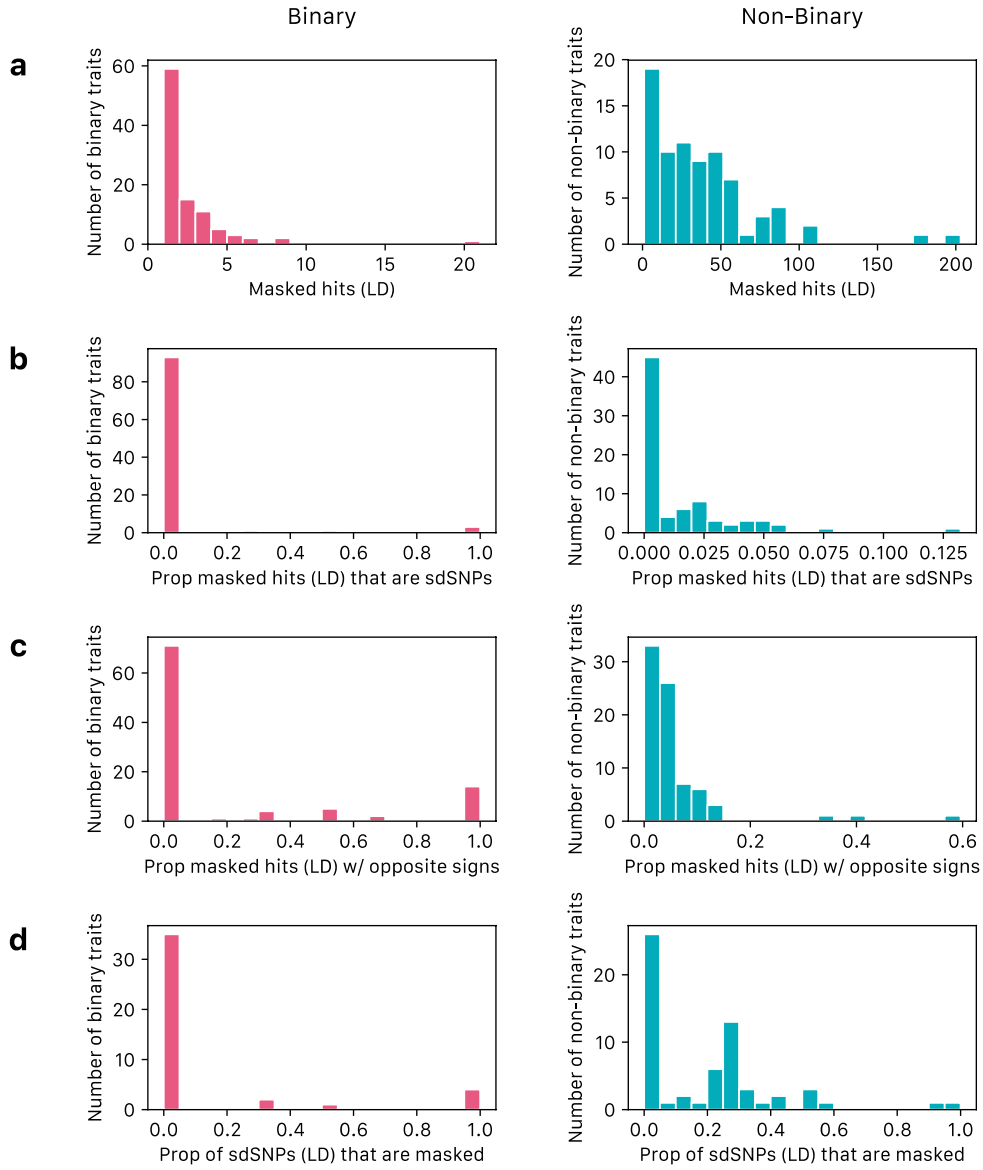
Extended Data Figure 2. Top plots: scatterplots comparing male genetic variance to female genetic variance for binary (left) and non-binary (right) traits. Bottom plots: scatterplots comparing male evolvability to female evolvability for binary (left) and non-binary (right) traits. Each point represents a trait, and pink points indicate traits for which the genomic parameter considered (genetic variance, evolvability) between the sexes is significantly different ($q < 0.05$, see Methods). Basal metabolic rate was removed as an outlier.



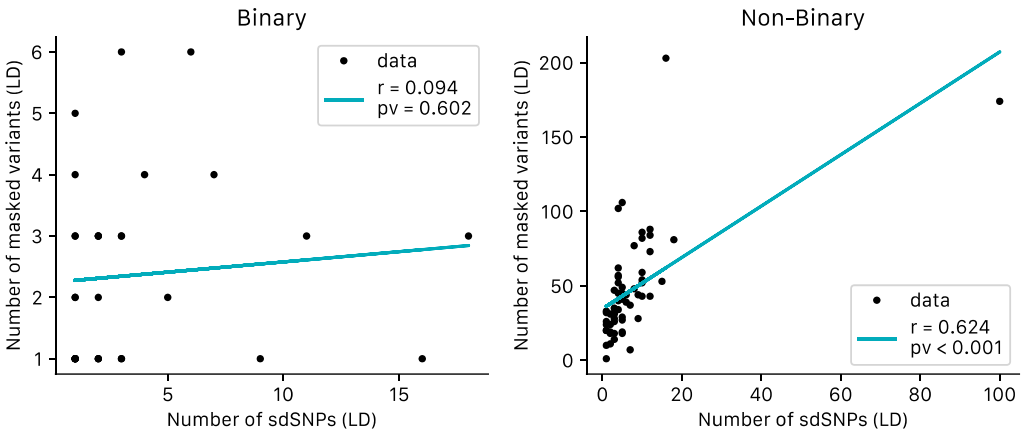
Extended Data Figure 3. Barplot comparing genetic correlation estimates from this effort to the literature for several non-binary traits ("WHR": waist-hip circumference ratio, "Body Fat %": body fat percentage, "BMI": body-mass index, "WC": waist circumference, "HC": hip circumference, and "Height": standing height). Height of bar indicates genetic correlation estimate (r_g), different colors corresponding to different publications, as represented in legend by surname of first author[12,14,16] and the values of which, are shown in Supplementary Table 2. Error bars represent 95% confidence intervals of r_g estimates (r_g +/- 1.96SE_{rg}).



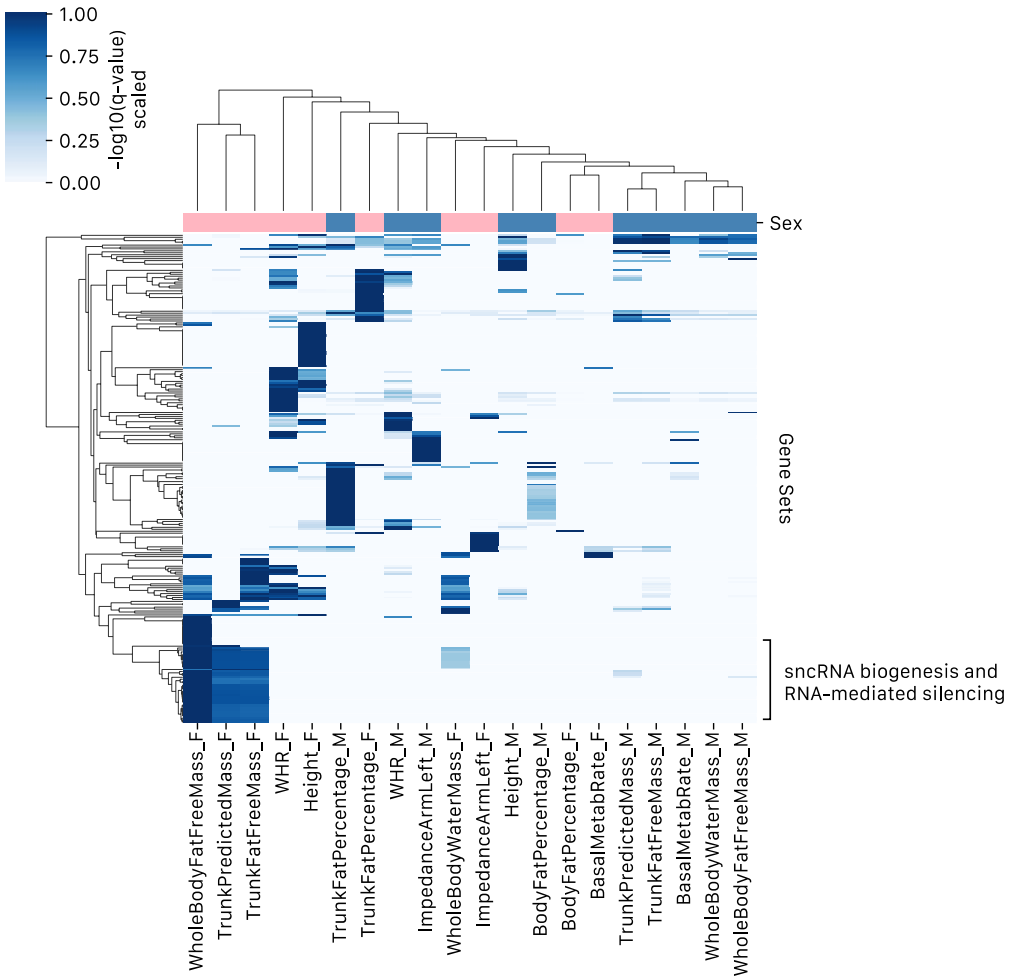
Extended Data Figure 4. Manhattan plot of number of sdSNPs ($p < 1 \times 10^{-8}$, two-sided t-test, see Methods) per genomic position, each point representing a genetic variant, and its height the number of traits it affects in a sexually different manner, for (a) non-binary traits and (b) binary traits.



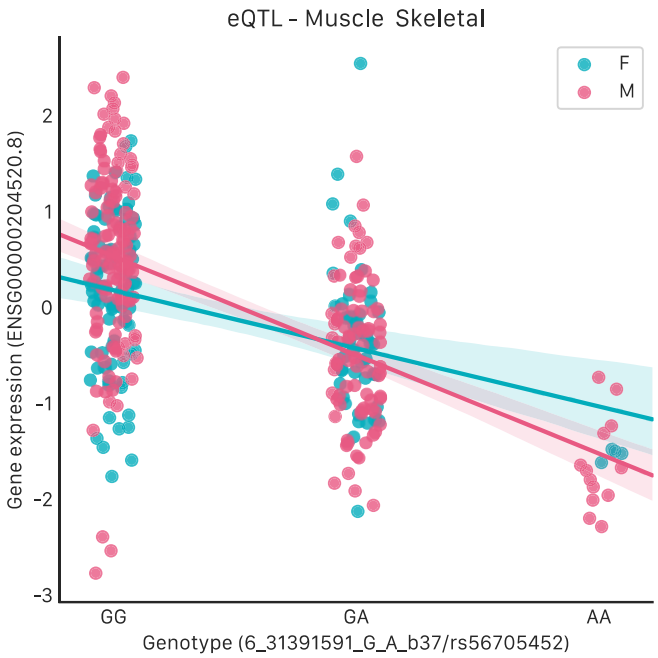
Extended Data Figure 5. For binary (left) and non-binary (right) traits with at least one potentially masked variant: (a) histogram of number of masked genetic variants, (b) histogram of proportion of masked SNPs that presented opposite sign effects between the sexes, (c) histogram of proportion of masked SNPs that were found to possess significantly different genetic effects between the sexes (two-sided t-test, $p < 1 \times 10^{-8}$ threshold, see Methods) and (d) histogram of the proportion of lead sdSNPs that were found to be masked.



Extended Data Figure 6. For binary (left) and non-binary (right) traits with at least one sdSNP and one potentially masked variant, shown is a scatterplot of number of sdSNPs against number of potentially masked variants ($p < 1 \times 10^{-8}$, and LD clumped within sex, see Methods).



Extended Data Figure 7. Heatmap with hierarchical clustering of FUMA set enrichment -log10 *q*-values for 251 gene sets that were found to be significantly differentially enriched between males and females (Fisher's exact test $q < 0.05$), as well as significantly differentially enriched to GWAS background genes (Fisher's exact test $q < 0.05$). x-axis corresponds to male or female dominant genes (see Methods) for different traits, and y axis to gene sets. Color within plot represents scaled -log10 *q*-value of enrichment, and "Sex" horizontal column indicates the sex in which the genes considered are dominant (pink for females, and blue for males).



Extended Data Figure 8. Relationship of genotype at variant rs56705452 with the expression of the transcript ENSG00000204520.8 in muscle skeletal tissue, for males (pink) and females (blue). Each point represents an individual, the x-axis its genotype at the considered variant, and the y-axis gene expression levels for the considered transcript.

TABLES

Table 1. Number of lead sdSNPs found across traits. Traits with the largest number of autosomal sdSNPs, that is, SNPs which presented a statistically significant difference between male and female genetic effect estimates ($p < 1 \times 10^{-8}$, two-sided t-test, see Methods), are shown. For each trait, total number of sdSNPs found across the genome is shown, along with the total number of independent loci presenting sex differences post-LD clumping (lead sdSNPs). Traits are sorted by number of lead sdSNPs. (*) Our analyses point to nucleated red blood cell percentage likely being a false positive (see Supplementary Note).

Trait	$p < 1 \times 10^{-8}$ (sdSNPs)	$p < 1 \times 10^{-8}$ LD clumped (lead sdSNPs)
Waist circumference / Hip circumference	2421	100
ankylosing spondylitis	626	18
Standing height	86	18
gout	708	16
*Nucleated red blood cell percentage**	*26*	*16*
Trunk fat percentage	432	15
Hip circumference	124	12
Whole body water mass	39	12
Trunk predicted mass	163	12
Trunk fat-free mass	141	12
hypothyroidism/myxoedema	168	11
Arm predicted mass (left)	52	10
Impedance of arm (left)	149	10
Impedance of whole body	233	10
Impedance of arm (right)	155	10

METHODS

UK Biobank data

UK Biobank is a large population-based prospective study with participants aged 40 to 69 years at recruitment, with extensive matching phenotypic and genomic data[46]. In this study, of the circa 490,000 individuals whose data was released in July 2017, we considered data pertaining to a total of 452,264 white European individuals. Of these, 245,494 were females and 206,770 were males, here referring to individuals whose self-proclaimed sex coincided with a XX or XY chromosomal content for females and males respectively, thus excluding individuals whose self-reported sex did not coincide with said genotypes. We also excluded individuals that were identified by UK Biobank as outliers based on genotyping missingness rate or heterogeneity, and individuals whose first or second genomic principal component differed by over 5 standard deviations from the mean of self-reported white Europeans. Finally, we removed individuals with a missingness rate > 5% for the genetic variants that passed quality control (described in Genotypes section), arriving at the aforementioned number of individuals.

Genotypes

UK Biobank's participants were genotyped using either of two arrays, the Affymetrix UK BiLEVE Axiom or the Affymetrix UK Biobank Axiom array, and later augmented by imputation of over 90 million genetic variants from the Haplotype Reference Consortium, the 1000 Genomes project, and the UK 10K project.

We excluded variants which did not pass UK Biobank quality control procedures in any of the genotyping batches and retained only bi-allelic variants with $p > 1 \times 10^{-50}$ for departure from Hardy-Weinberg and MAF > 1×10^{-4}, computed on a subset of 344,057 unrelated (Kinship coefficient < 0.0442) individuals of White-British descent with missingness rate > 2% in the study cohort, paralleling the QC procedure followed by Canela-Xandri et al[47].

A second round of QC was done prior to performing everything that followed the sex-stratified GWAS and genetic parameter estimation (described below), including calculation of genetic correlations and genetic effect comparisons between the sexes. This was done to account for the smaller sample sizes (versus those of [47]) as well as to be more stringent as we are comparing between groups. Variants were further filtered if they possessed $p < 1 \times 10^{-6}$ for departure from Hardy-Weinberg and MAF < 1×10^{-3}, computed on the aforementioned subset of unrelated individuals. A stricter MAF threshold (MAF < 1×10^{-1}) was set for binary traits due to the general limited number of cases available which can lead to an inflation of type I error rates in association tests[48]. Briefly, this is based on simulations from Loh et al's work[48], where, when the case/control ratio is 0.001, a MAF filter of 10% shows no significant inflation of Type I error rates for the sample sizes considered here. Furthermore, variants with a significant effect ($p < 1 \times 10^{-8}$) when running a GWAS on sex for the aforementioned subset were also excluded, as these could arise due to sampling bias[49]. Finally, only imputed variants with no genotyped counterpart and with an imputation score > 0.9 were retained. As a result, a total of 9,072,751 (602,984 genotyped

and 8,469,767 imputed) autosomal genetic variants and 17,364 X-chromosome genetic variants remained for our analysis of non-binary traits, and 4,229,346 autosomal (244,743 genotyped and 3,984,603 imputed) and 7,227 genotyped X-chromosome genetic variants for our analysis of binary traits.

Phenotypes

In total we analyzed 530 non-sex specific traits. These included 446 binary traits, which had at least 400 cases in each of the sexes, relating to self-reported disease status, ICD10 codes from hospitalization events, and ICD10 codes from cancer registries, as well as 84 non-binary traits comprising non-scale transformed continuous and ordered integral measures.

Our starting point were the 778 traits considered in the GeneATLAS study[47]. These were in turn extracted from the UK Biobank June 2017 release. From these original 778 traits, we further removed binary traits that had fewer than 400 traits in either sex (this thus excluded traits that were sex-specific from our study). We further removed several blood measurement related traits as well as others that did not pass our QC stage. The remaining 530 traits were considered in this study. For more information on the treatment and filtering of the phenotype data see [47].

Sex-stratified parameter estimation

To obtain genetic and environmental variance estimates for each of the sexes we used DISSECT[29] following the same methodology as described in [47]. Briefly, a partition of variance analysis was run using linear mixed models (LMMs), which were fitted for each trait with a Genetic Relationship Matrix (GRM) containing all common autosomal genetic variants (MAF > 5%) which passed QC.

Heritability was then calculated as

$$h_g^2 = \sigma_g^2 / \left(\sigma_g^2 + \sigma_\epsilon^2\right),$$

where σ_g^2 and σ_ϵ^2 are the estimates of the genetic and residual variance. Heritability for binary traits was transformed from the observed scale to the liability scale using the sex-specific population prevalence of the trait, under the assumption that there was an underlying normal distribution of liability to the considered trait, as described in [50].

Wanting to see if the sexes differ in regards to their ability to undergo adaptation, evolvability was calculated for males and females separately. Evolvability, defined as the expected evolutionary response to selection per unit of selection[51], or the ability of populations to respond to natural or sex selection[52], was calculated as

$$e = \sigma_g / m^2$$

where σ_g is the additive genetic variance of the trait and m is the trait mean.

To establish differences between heritability, genetic variance, and evolvability between the sexes, we used the t statistic

$$t = \frac{X_{males} - X_{females}}{\sqrt{SE_{males}^2 + SE_{females}^2}}$$

where X represents either heritability, genetic variance or evolvability, assumed to be independent between the sexes, and SE the standard error of the aforementioned, for males and females respectively. Two-tailed *p*-values were then FDR corrected (using the Benjamini-Hochberg procedure) to account for multiple testing.

Sex-stratified GWAS

To test the genetic effect of each variant for each of the sexes on the 530 chosen traits we ran a sex-stratified GWAS. The procedure followed for each of the sexes is that described in [47] using the DISSECT software[29]**.** Briefly, a Linear Mixed Model (LMM) was fitted,

$$y = X\beta + g + \epsilon,$$

y being the vector of phenotypes, X the matrix of fixed effects, β the effect size of these effects, g the polygenic effect that captures the population genetic structure, and ϵ the residual effect not accounted for by the fixed and random effects. Following the procedure described in [47], our curated genetic variants were regressed against the residuals of the LMM to assess association.

Genetic correlations

Genetic correlations between the sexes were calculated using the bivariate linkage disequilibrium score (LDSC) regression analysis software[31], which works directly on GWAS summary data and can thus be applied to very large sample sizes. As we were using data of European origin, we used the LD scores provided by the LDSC software and limited our genetic correlation calculation to the genetic effect estimates of SNPs for which such scores were available (1,189,831 total genetic variants, of which 1,169,868 passed LDSC's QC filters and were used in the computation). These scores were computed using the 1000 Genomes European data. We furthermore restricted our binary traits to those that had at least 5000 cases in each of the sexes, as was recommended in the documentation. Note that for traits with very low heritability this computation was unsuccessful. In total, genetic correlations were obtained for 83 non-binary and 77 binary traits.

To establish which correlations differed from one, we used the t statistic

$$t = \frac{r_g - 1}{SE_{rg}}$$

where r_g is the genetic correlation, and SE_{rg} is the standard error of the genetic correlation. Two-tailed *p*-values were FDR corrected to account for multiple testing.

Sex differences in genetic effects

To compare genetic effects across the genome between the sexes for all traits we considered the following two-tailed t test, as used in previous sex-stratified GWAS comparison studies[20,23]:

$$t = \frac{b_{males} - b_{females}}{\sqrt{SE_{males}^2 + SE_{females}^2 - 2r \cdot SE_{males} \cdot SE_{females}}}$$

where b is the estimated effect of the genetic variant considered for a given trait for males and females, SE is the standard error of the effect, and r is the Spearman rank correlation between the sexes across all genetic variants for a given trait. Both the SE and b were adjusted by the standard deviation of the trait, for each sex, to correct for scale effects that could act as confounders in the study. Some studies have opted to ignore the third term in the denominator, which estimates the covariance of the error terms, assuming r to be equal to 0[6]. However, for the traits considered this correlation ranged from -0.00335 (cervical spondylosis) to 0.34173 (standing height), thus our decision to include it.

As this test was done for all variants and all phenotypes, this effort resulted in a total of 4,808,558,030 statistical tests (9,072,751 x 530). Binary traits were then filtered further as stated in the Genotypes section. To account for multiple testing, we considered the commonly used genome wide significance cut-off of $p < 1 \times 10^{-8}$.

In order to cluster our results into independent lead variants, we used the --clump option in PLINK 1.9[53]. For each individual trait, variants found to be genome-wide significant with regards to difference between the sexes were clustered into lead variants, assigning them variants in LD within 10Mb, with an $r^2 > 0.2$ with the lead variant. To obtain the total number of independent loci across all traits, the same clustering method was used but for all variants found to be leads across traits, choosing the variant with the lowest *p*-value if variants were found in more than one trait.

Analysis checks

Our replication/sanity check methods can be divided into technical (different models, sex randomization, logistic mixed models using REGENIE[35], simulations of effect of case-control imbalance on heritability estimates) and biological (comparison to the GIANT cohort). Detailed information and results of these analyses can be found in the Supplementary Note.

Calculation of Polygenic Scores (PGS)

Using the subset of unrelated White British individuals from UK Biobank, we calculated PGS for our traits with most evidence of GxS found using our lead sdSNPs.

PGS are calculated as:

$$PGS_i = \sum_j^n b_j \cdot dosage_{ij}$$

Where PGS_i is the polygenic risk score estimate for individual i, b_j is the genetic effect estimate for variant j, $dosage_{ij}$ is the effect allele count (0, 1 or 2) of variant j in individual i, and n is the total number of genetic variants considered in the PGS calculation (here just the lead sdSNPs).

The genetic effects used in our PGS calculations were obtained re-running our original model using UK Biobank's unrelated White British participants, randomly selecting 150,000 women to obtain female specific genetic effects, 150,000 males to obtain male specific genetic effects, and 75,000 men and 75,000 females to obtain sex-agnostic genetic effects. This was done to match sample sizes.

We then proceeded to calculate PGS for a total of 34,928 females and 8,956 males from UK Biobank and which were self-reported white European, which had not been considered in the calculation of the genetic effects. We did this in three ways: using the genetic effects corresponding to the sex of the individual (*same* PGS), using the genetic effects corresponding to the opposite sex of the individual (*opposite* PGS) and using the genetic effects for the whole population (*agnostic* PGS). Thus, in total each of our circa 44K individuals had 3 PGS calculated.

In order to assess predictive power, the phenotypes of our 44K individuals (corrected by all the covariates in our original model to account for population structure and other effects) were then either regressed on the *same*, *opposite* and *agnostic* PGSs respectively in the case of non-binary traits, and in the case of binary traits the area under the curve (AUC) was calculated for the ROC curve for the three PGS groups. Statistical significance of AUCs was assessed using the two-tailed Mann Whitney U test[54]. To assess differences between PGS-phenotype correlations, these were transformed using Fisher's Z transformation[55], and then compared using the resulting Z scores (two-tailed Z test). AUCs were compared using the roc.test function from the pROC R-package[56]. Correlation, AUC and comparison statistics and *p*-values are reported in Supplementary Table 11.

An important caveat of this methodology is that there is an overlap between the discovery (the population used to declare variants as sexually different) and the replication (for which genetic effects were re-calculated and/or PGS were obtained) populations. As a way of checking whether the overlap could potentially be influencing our results, we repeated this step by obtaining the genetic effects for the circa 408K individuals of White British ethnicity in UK Biobank, repeating the steps described in the "Sex differences in genetic effects" section, to establish genetic effect differences across the genome. Then we calculated the different PGSs for the remaining circa 42K individuals of white ethnicity, again regressing on the phenotype, for our most representative trait: waist-hip circumference ratio. This, however, only serves as a validation for a single trait, meaning that caution should be taken when interpreting the predictive power of the PGSs calculated.

Heritability explained by sdSNPs

In order to put our prediction analysis into context, we obtained the proportion of sex-specific heritability explained by the sex-specific genetic effect estimates of the sdSNPs found for each trait. To do this, for each trait with i sdSNPs, the heritability of sdSNPs was calculated, for males and females separately, as

$$h^2_{sdSNP} = \frac{\Sigma_i 2p_i(1 - p_i)b_i^2}{\sigma_g^2 + \sigma_e^2}$$

where b is the sex-specific genetic effect for each of the i lead sdSNPs for a given trait, p is the frequency of the reference allele, and σ_g^2 and σ_e^2 are the sex-specific genetic and residual variance estimates for each trait respectively. The proportion of the sex-specific heritability explained by sdSNPs was then calculated for each sex as h^2_{sdSNP} / h^2.

Masking analysis

To evaluate potential masking across traits in the UK Biobank, we compared the results from sex-stratified models to those from a non-stratified model, across the 530 traits in our study. Genetic effects across the genome were assessed for significant association in males and females, and were deemed to be potentially masked if significance was reached in either at a $p < 1 \times 10^{-8}$ threshold, but not in a sex-agnostic model. 209 traits were found to have at least one potentially masked genetic variant, with a total of 127/446 (28.48%) binary and 82/84 (97.62%) non-binary traits (Supplementary Table 13).

We also performed this analysis by LD-clumping the results from our sex-stratified models prior to assessing significance in a sex-agnostic model across traits in order to account for the effects of random fluctuation (Supplementary Table 14). Using this methodology, a total of 176 traits, 98/446 (21.97%) binary and 78/84 (92.86%) non-binary, were found to present at least one potentially masked genetic variant. While we found a high correlation with our original results ($r = 0.96$, $p = 4.59 \times 10^{-299}$, Supplementary Table 14, Figure S8), numbers of potentially masked variants found decreased by around half for each trait (regression slope = 0.49, Figure S8). To be conservative, these are the results considered in the main text.

Gene level analysis

We performed a gene-level analysis to translate our SNP data into a more manageable and interpretable form. This was done for the subset of traits that presented at least one sdSNP for a total of 103 traits.

As we wanted to obtain genes relevant to each of the sexes, we began by partitioning our two-tailed *p*-values (p_{2T}) from the genetic effect (b) comparison between the sexes into two one-tailed *p*-values. For genetic variants where b_F was greater than b_M, one tailed *p*-values were calculated as:

$$p_F = \frac{p_{2T}}{2} \text{ and } p_M = 1 - (\frac{p_{2T}}{2})$$

On the other hand, when the genetic effect was greater in males, the *p*-values were calculated as

$$p_M = \frac{p_{2T}}{2} \text{ and } p_F = 1 - (\frac{p_{2T}}{2}).$$

This process led to the creation of two additional distinct sets of *p*-values for each phenotype, corresponding to sites where the genetic effect was significantly greater in males or females.

Each of these sets of *p*-values (p_{2T}, p_M and p_F) was subsequently used to identify gene level associations using MAGMA[57]. First, we annotated every gene (i.e. defined which SNPs were in the gene region) considering a range of 1kb upstream and downstream. MAGMA was then run for each phenotype and each set of one tailed *p*-values separately, considering two distinct SNP-wise models (SNP mean and top SNP), using a random sample of 1,000 unrelated white British individuals from the UK Biobank, 500 males and 500 females, as a base population for LD and MAF correction. The analysis provides three distinct *p*-values for each gene, one for the SNP mean model, one for the SNP top model, and a combined *p*-value. For subsequent analyses, we considered the combined *p*-value for each gene. Genes were declared significantly dimorphic and female or male dominant if an FDR corrected combined *p*-value (q) < 0.01 was obtained when considering p_F or p_M respectively.

The set of genes that reached our threshold were then used in the gene set and enrichment analyses using the GENE2FUNC tool in FUMA[58]. The GENE2FUNC tool in FUMA was run for the top 10 traits with the largest number of significant genes when considering a two-tailed *p*-value. As a result, FUMA was run for 10 traits for both male and female dominant genes.

Briefly, FUMA takes our list of candidate genes and checks for enrichment across: (1) differentially expressed genes (DEG) for different tissues (tissue enrichment analysis was conducted considering the GTEx V6 database[59]), and (2) biological pathways/functional categories (considering MSigDB v7, WikiPathways and GWAS Catalog[60]). Enrichment is assessed using a hypergeometric test. In this study we have focused on just biological pathways/functional categories. Significantly differentially enriched gene sets between female and male dominant genes were assessed using a Fisher's exact test, and *p*-values were FDR corrected.

As a background, the same procedure as is described in this section was followed using sex-agnostic GWAS results for the 10 traits of interest. Using MAGMA, genes associated to each of the traits were obtained, and sets enriched for genes associated to each trait were obtained using FUMA. Significantly differentially enriched gene sets between our background and our female and male dominant genes were obtained using a Fisher's exact test, and *p*-values were FDR corrected.

Heatmaps of the scaled -log10 *q*-values per set were created for each phenotype and sex. Heatmaps for gene sets were limited to those both significantly differentially enriched in at least one sex versus our background (Fisher's exact test $q < 0.05$), as well as significantly differently enriched between female and male dominant genes (Fisher exact test $q < 0.05$).

GTEx data

In order to assess whether differences in genetic architecture, as established by our analysis of the UK Biobank data, lead to differences in gene expression, we proceeded to complete an expression quantitative trait loci (eQTL) analysis looking for GxS interactions in gene expression. The data from the Genotype-Tissue Expression Project (GTEx) V6p release was used, which consists of samples and genotypes from 449 human donors (292 males and 158 females) and 39 non-sex specific and non-diseased tissue types. Each tissue type holds a different number of samples (minimum of 70, median of 149), with a male-bias present in all of them (the percentage of female samples ranging from 25% to 44%).

Processed, filtered and normalized RNA-seq data was downloaded from the GTEx portal for both the autosomal genome as well as for the X chromosome, the number of transcripts varying across tissues due to tissue-specific expression, with a median of 23,538 autosomal transcripts and 793 X-linked transcripts per tissue. Covariates were also downloaded for each of the considered tissues. Further information regarding the treatment of the data can be found in the GTEx v6p analysis methods documentation. Genotype data was also obtained for all the donors from dbGaP, and which included a total of 11,607,846 genetic variants, from genotyping efforts with Illumina OMNI 5M and 2.5M SNP Arrays and imputation from the 1000 Genomes Project Phase I version 3 reference panel. For more information regarding the processing of the genotype data see the GTEx v6p analysis methods documentation.

eQTL analysis

To assess whether our sdSNPs had different effects on the expression of nearby genes depending on sex, we assigned our lead sdSNPs nearby genes (1Mb window) using *Granges*[61] and the Biomart resource[62]. The total number of sdSNP-gene pairs were 6,591 when considering our autosomal hits for non-binary traits, 4,533 when considering our autosomal hits for binary traits, and 95 when considering our X-chromosome hits for non-binary traits.

For each of the 39 tissues, we tested for GxS in gene expression for each variant-gene pair using a linear regression model in PLINK 1.9, which was adjusted for three genotyping principal components (PCs) and PEER factors, the number of which included in the model depended on the sample size (sample sizes < 150 had 15 PEER factors, sample sizes between 150 and 250 had 30 PEER factors, and sample sizes over 250 had 35 PEER factors), as indicated in [63]**.** In the end, our gene expression model for each gene within 1Mb of a sexually dimorphic variant, and each tissue was formulated as

$$y = \mu + \beta_1 \cdot sex + \beta_2 \cdot SNP + \beta_3 \cdot SNP \cdot sex + PC_{1\text{-}3} + PEER + \varepsilon$$

where y is the gene expression of the given gene in a given tissue, μ is the mean expression levels, β_1 and β_2 are the regression coefficients for sex and genotype of the sexually dimorphic variant respectively, β_3 is the regression coefficient for the interaction of the genotype with sex, $PC_{1\text{-}3}$ and PEER

are the principal components and PEER factor covariates, and ε is the residual. FDR correction was applied to account for multiple testing.

When running our eQTL analysis we found that two tissues returned missing values across all tests performed with PLINK, the brain anterior cingulate cortex BA24 and the small intestine terminal ileum tissues. These are the two tissues with the smallest number of samples, therefore this absence of results is likely due to not enough variation being present in the phenotype.

eQTL enrichment

To assess whether our sexually dimorphic genetic variants were enriched for GxS interactions versus those not presenting sex differences, we proceeded to re-run our eQTL model using genetic variants significant for the whole population ($p < 1 \times 10^{-8}$) but that had no evidence of being sexually dimorphic (t-statistic comparing genetic effects between the sexes with $p > 0.5$). Using contingency tables and Fisher's exact test, we considered whether the number of significant variant-gene GxS terms (at a $p < 1 \times 10^{-3}$ threshold) was enriched for our sexually dimorphic variants for each of the tissues.

DATA AVAILABILITY

This research has been conducted using the UK Biobank Resource under project 788. Data from the Genotype Tissue Expression (GTEx) project V6p release was also employed, protected genotype data accessed via dbGaP, and processed gene expression data downloaded from the GTEx portal, which is openly available: https://gtexportal.org. The GIANT cohort's summary statistics were employed to compare findings to (openly available: https://portals.broadinstitute.org/collaboration/giant/), as well as Pulit et al's (Human Molecular Genetics, 2019) summary statistics pertaining to a GIANT-UKB meta-analysis (openly available: https://github.com/lindgrengroup/fatdistnGWAS). The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. The GWAS summary statistics of both autosomal and X-chromosome variants from sex-stratified models are openly available from the University of Edinburgh DataShare repository within the following collection: https://datashare.ed.ac.uk/handle/10283/3908 (clinical binary traits DOI: 10.7488/ds/3046, non-clinical binary traits DOI: 10.7488/ds/3047, non-binary traits: 10.7488/ds/3048, and logistic mixed model results DOI: 10.7488/ds/3049).

CODE AVAILABILITY

We used DISSECT (v1.15.2c, May 24, 2018), which is publicly available at http://www.dissect.ed.ac.uk/ under GNU Lesser General Public License v3), PLINK (v1.9 and v2.0, freely available online at https://www.cog-genomics.org/plink2/), BGENIX (v1.0 freely available online at https://bitbucket.org/gavinband/bgen), LD Score Regression (v.1.0.1, freely available online at https://github.com/bulik/ldsc), MAGMA (v1.06, freely available online at https://ctg.cncr.nl/software/magma), FUMA (freely available online at https://fuma.ctglab.nl/), GCTA (v. 1.91.4, freely available at https://cnsgenomics.com/software/gcta/) and REGENIE (v.1.0.7, freely available online at https://github.com/rgcgithub/regenie/tree/v1.0.7-latest). Custom code created to

perform the analysis is openly available (https://zenodo.org/record/4844680), with DOI: 10.5281/zenodo.4844680.

REFERENCES (METHODS ONLY)

46. Sudlow, C. *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.* **12**, e1001779 (2015).
47. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).
48. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
49. Pirastu, N. *et al.* Genetic analyses identify widespread sex-differential participation bias. *Nat. Genet.* **53**, 663–671 (2021).
50. Munoz, M. *et al.* Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat Genet* **48**, 980–983 (2016).
51. Hansen, T. F., Pélabon, C. & Houle, D. Heritability is not Evolvability. *Evol. Biol.* **38**, 258 (2011).
52. Garcia-Gonzalez, F., Simmons, L. W., Tomkins, J. L., Kotiaho, J. S. & Evans, J. P. Comparing evolvabilities: common errors surrounding the calculation and use of coefficients of additive genetic variation. *Evolution (N. Y).* **66**, 2341–2349 (2012).
53. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).
54. Mason, S. J. & Graham, N. E. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Q. J. R. Meteorol. Soc.* **128**, 2145–2166 (2002).
55. Fisher, R. A. Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika* **10**, 507 (1915).
56. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
57. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
58. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
59. Carithers, L. J. & Moore, H. M. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv. Biobank.* (2015). doi:10.1038/ng.2653
60. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
61. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118 (2013).
62. Kinsella, R. J. *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011**, bar030 (2011).
63. Consortium, G. Te. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
64. Bernabeu, E. Sex differences in genetic architecture in UK Biobank - Code. (2021). doi:10.5281/ZENODO.4844680