

The role of gene expression on human sexual dimorphism: too early to call

Eleonora Porcu^{1,2,§}, Annique Claringbould³, Kaido Lepik^{4,5}, BIOS Consortium†, Tom G. Richardson⁶, Federico A. Santoni⁷, Lude Franke³, Alexandre Reymond^{1,*§}, Zoltán Katalik^{2,4,8,*§}

¹ Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland

² Swiss Institute of Bioinformatics, Lausanne Switzerland

³ University Medical Centre Groningen, Groningen, the Netherlands

⁴ University Center for Primary Care and Public Health, Lausanne, Switzerland

⁵ Institute of Computer Science, University of Tartu, Tartu, Estonia

⁶ MRC Integrative Epidemiology Unit (IEU), Population Health Sciences, Bristol Medical School, University of Bristol, United Kingdom

⁷ Endocrine, Diabetes, and Metabolism Service, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland

⁸ Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

* These authors jointly supervised this work: Alexandre Reymond and Zoltán Katalik

†A full list of consortium members appears at the end of the paper.

§ Corresponding authors (eleonora.porcu@unil.ch, alexandre.reymond@unil.ch and zoltan.katalik@unil.ch)

Abstract

The genetic underpinning of sexual dimorphism is very poorly understood. The prevalence of many diseases differs between men and women, which could be in part caused by sex-specific genetic effects. Nevertheless, only a few published genome-wide association studies (GWAS) were performed separately in each sex. The reported enrichment of expression quantitative trait loci (eQTLs) among GWAS-associated SNPs suggests a potential role of sex-specific eQTLs in the sex-specific genetic mechanism underlying complex traits.

To explore this scenario, we performed a genome-wide analysis of sex-specific whole blood RNA-seq eQTLs from 3,447 individuals. Among 9 million SNP-gene pairs showing sex-combined associations, we found 18 genes with significant sex-specific *cis*-eQTLs (FDR 5%). Our genome-wide association study of the 18 top sex-specific eQTLs on >700 traits unraveled that these eQTLs do not systematically translate into detectable sex-specific trait-associations. Power analyses using real eQTL- and causal effect sizes showed that millions of samples would be necessary to observe sex-specific trait associations that are fully driven by sex-specific *cis*-eQTLs. Compensatory effects may further hamper their detection. In line with this observation, we confirmed that the sex-specific trait-associations detected so far are not driven by sex-specific *cis*-eQTLs.

Introduction

Men and women exhibit sexual dimorphism. Clear examples of sex-biased traits are anthropometric features. However, biological differences between sexes are not limited to physical traits: sex differences are also evident in incidence, prevalence and severity across diseases. For example, women are much more likely to develop autoimmune [1], while men are more likely to develop cardiovascular diseases [2].

Despite the widespread nature of these sexual differences and their noteworthy implications for medical research and treatments, little is known about their underlying biology in complex traits. While the sex chromosomes play key roles in sexual dimorphism, GWAS have identified dozens of autosomal genetic variants showing sex-specific effects [3-10], suggesting that part of the phenotypic differences might be due to accumulation of genetic variants present in both sexes at the same frequency [11], but acting in a different manner in males and females.

The strong enrichment of expression quantitative trait loci (eQTLs) among complex trait-associated loci [12-15] suggests that gene expression might be an appealing intermediate phenotype for the understanding of the biological mechanism behind SNP-trait associations. Towards this goal, several transcriptome-wide association studies (TWASs) integrating GWAS and eQTLs were proposed to identify genes whose expression is significantly associated to complex traits [16-18]. As these studies pointed to many genetic loci where variants exert their effect on phenotypes through gene expression, it is reasonable to think that sex-specific associations found by GWAS could be driven by sexual dimorphism in gene expression regulation, meaning that sex differences in eQTL effects might underlie the sex-specific GWAS associations. To explore this hypothesis, we performed a genome-wide investigation of sex biases in *cis*-eQTL effects in human autosomal genes and assessed their potential contribution to sex differences in complex traits.

Results

Sex-specific eQTLs analyses

First, we performed a genome-wide analysis of whole blood RNA-seq eQTLs to identify autosomal sex-specific eQTLs, i.e. SNPs whose effect on expression differs in magnitude between men and women. We analyzed eQTLs separately for 1,519 men and 1,928 women collected by the BIOS Consortium (<http://www.bbmri.nl/acquisition-use-analyze/bios/>). To reduce the number of tests, we restricted our analyses to autosomal variants previously detected as *cis*-eQTLs (FDR 5%) by the sex-combined analysis of the eQTLGen Consortium [19] and included in the UK10K reference panel [20].

To test the reliability of the BIOS data, we combined the results from the two sexes in a meta-analysis and compared them with those obtained by the eQTLGen Consortium (N=31,684, sex-combined results). We observed a high correlation between the betas obtained in the two studies ($r^2=0.9$).

In total, we tested 8,739,806 SNP-gene pairs (involving 3,142,796 SNPs and 16,874 genes) for sex-interaction. We found 10,285 eGenes (genes with at least one *cis*-eQTL at Bonferroni adjusted P -value < 0.05) shared between the two sexes, while 542 and 1,419 eGenes were detected only in men or women, respectively.

Identification of sex-specific *cis*-eQTLs

To identify sex-specific eQTLs, we tested the difference in the effects calculated for the two sexes separately (see Methods).

We identified 462 SNP-gene associations showing significantly different effects in men and women (FDR 5%, $P_{\text{diff}} < 2.6 \times 10^{-6}$). These sex-specific eQTLs cluster in 18 sex-specific eGenes (**Figure 1** and **Supplementary Table 1**). By analyzing only *cis*-eQTLs with a significant sex-combined effect (in the eQTLGen Consortium) we favored eQTLs that show the same direction but different magnitude of effect in men and women, as we posit that significant eQTLs showing opposite direction of effects in the two sexes would not have been detected in sex-combined analyses. Consistent with this hypothesis all 462 sex-specific eQTLs show the same direction but different magnitude of effects in both sexes.

Among the 18 sex-specific eGenes, 5 and 12 are men- and women-specific, respectively. One gene, *ZNF718*, contains eQTLs differently biased in both sexes (**Supplementary Figure 1**).

Next, to assess whether sex-specific eGenes are enriched for sex-specific biological processes, we performed enrichment analyses using 139 sex-specific eGenes, selected using a less stringent threshold ($P_{\text{diff}} < 1 \times 10^{-4}$). Although no GO term showed significant enrichment after FDR correction, we found response to glucagon (GO:0033762, $P=1.59 \times 10^{-3}$) and negative regulation of cytokine production (GO:1900016, $P=2.43 \times 10^{-3}$) as the two most enriched (top 50 GO terms are reported in **Supplementary Table 2**). These pathways are involved in two sexually dimorphic processes: glucose metabolism and inflammatory response, respectively [21] [22].

To determine if the difference of eQTL effects observed between men and women is driven by sex differences in gene expression distribution, we then compared the expression means and variances between sexes for the 18 sex-specific eGenes. While no eGene showed a significant difference in mean, we found 8 eGenes with sex-specific variance (based on $P_{\text{diff}} < 0.05/18$) (**Supplementary Table 3**).

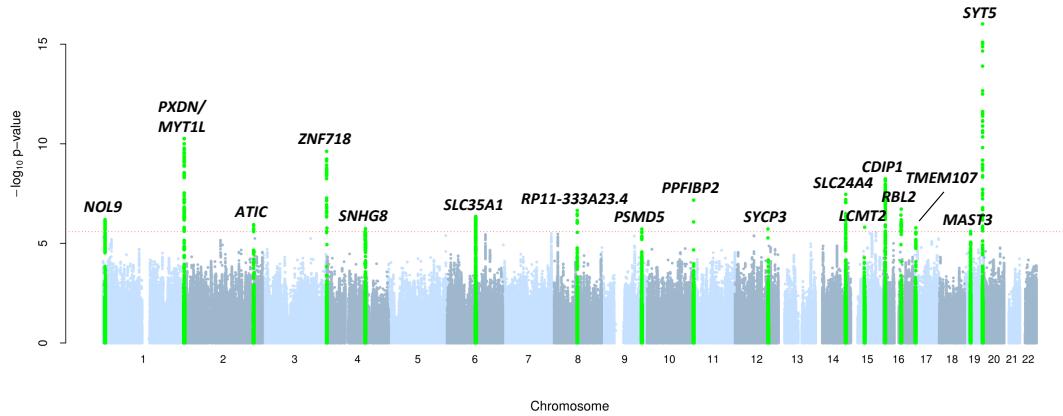


Figure1. Manhattan plot of sex-specific eQTLs.

The figure summarizes the results of our sex-specific eQTL discovery scan. SNP-gene pairs are plotted on the x-axis according to the SNP position on each autosomal chromosome in alternating light and dark blue against the P -values obtained upon testing for sex difference between effects in men and women (shown as $-\log_{10}(P_{\text{diff}})$). The red dotted line marks the 5% FDR threshold significance level ($P_{\text{diff}}=2.6\times 10^{-6}$), and SNPs in loci exceeding this threshold are highlighted in green.

Sex-specific cis-eQTLs do not translate into sex-specific trait associations

We then performed a genome-wide association study (PheWAS) to test if eQTL SNPs with sex-specific effect on expression levels have an effect on human phenotypes and if so, whether sex biases in gene expression regulation translate to sex-specific effects on complex traits. For each sex-specific eGene we selected one representative eQTL with the strongest difference in effect between men and women and ran PheWAS analyses on more than 700 phenotypes from UK Biobank (UKBB) [23].

We found that 7 of the 18 lead eQTLs were associated with 39 traits at genome-wide significant level (**Supplementary Table 4**). Interestingly all associated traits belong to two categories: either morphological (e.g.: height, weight and trunk fat mass) or hematological traits (e.g.: platelets, eosinophils and whole blood cells).

We then asked if the other 290 non-lead sex-specific eQTLs of the 7 eGenes found by PheWAS showed also a sex-specific effect on the 39 pre-selected traits (we set a Bonferroni threshold of $P_{\text{diff}} < 0.05/(20*7)$, where 20 is the effective number of independent phenotypes among the 39). For each trait, we used the sex-specific UKBB-GWAS summary statistics and observed no enrichment for the sex-specific eQTLs among the sex-specific GWAS signals and found only one eGene, *PSMD5*, for which sex-specific eQTLs are likely to translate to sex-specific associations with several obesity traits (**Figure 2**).

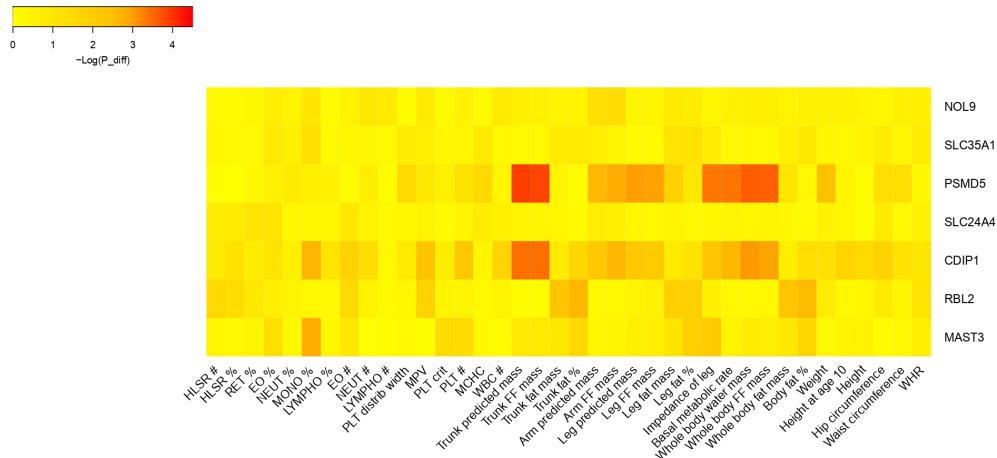


Figure2. Heatmap of sex-specific trait-associated SNPs.

The figure summarizes the results of sex-specific trait-associations driven by sex-specific eQTLs. For each sex-specific eGenes and for each phenotype, we plotted the minimum P_{diff} obtained testing for sex-difference between GWAS-effects in men and women (shown as $-\log_{10}(P_{\text{diff}})$).

Sex-specific complex trait associations are not driven by sex-specific eQTLs

Next we tested whether sex-specific SNP-trait associations are driven by sex-specific gene expression regulation. For this we looked at the two most sexually dimorphic traits, waist-to-hip ratio (WHR) and testosterone levels. We identified 803 and 266 independent SNPs showing a P -value $< 1 \times 10^{-5}$ in the sex-combined GWAS for WHR and testosterone, respectively. Among those, 121 (32 for WHR and 89 for testosterone) have a significant sex-difference in the effect on the two sexes ($P_{\text{diff}} < 0.05/(803+266)$), but none of the SNPs included in the BIOS Consortium dataset (58/121) show any sex-specific eQTL effect on genes in *cis* ($P_{\text{diff}} < 0.05/58$) (**Supplementary Tables 5 and 6**).

Sex-specific causal effects

Next, to explore the presence of sex-specific causal effect of gene expression on complex traits, we performed sex-specific transcriptome-wide Mendelian Randomization (TWMR) analyses [18] combining sex-specific eQTLs and sex-specific GWAS results. We applied such approach – as above – to testosterone levels and WHR.

We found 17 and 6 genes showing a significantly different causal effect between the two sexes in testosterone and WHR, respectively (**Supplementary Figure 2 and 3 and Supplementary Table 7**). While the sex-specific associations with WHR were all female-specific, among the 17 sex-specific genes associated with testosterone, 8 were female- and 9 male-specific, respectively.

Of note, the positive association of *IFT27* with testosterone levels observed only in men ($P_{\text{TWMR-men}} = 2.55 \times 10^{-8}$) and not in women ($P_{\text{TWMR-women}} = 0.99$) is supported by its association with Bardet-Biedl syndrome (OMIM: # 615996), for

which it has been reported that most affected males produce reduced amounts of sex hormones (hypogonadism). Among the genes associated with WHR, *CCDC92* has a significant effect only in women ($P_{TWMR\text{-women}}=2.70\times10^{-23}$) and not in men ($P_{TWMR\text{-men}}=0.09$). Interestingly, diseases associated with *CCDC92* include lipodystrophy [24], which is known to affect more women than men [25].

As a negative control, we applied TWMR to a trait not showing sexual dimorphism, such as educational attainment, and did not observe any sex-specific gene association (**Supplementary Figure 4**).

Finally, we tested whether sex-specific causal effects are driven by sex-specific gene expression regulation and observed that none of the SNPs used as instrumental variables in TWMR showed a sex-specific effect on gene expression. In addition, we ran TWMR using sex-specific WHR GWAS and sex-combined eQTLs data and found a high correlation between the causal effects ($r^2=0.79$ in females and $r^2=0.77$ in males) (**Supplementary Figure 5**), which suggests that different effects observed by TWMR are driven by sex-specific SNP-trait associations.

Power to detect sex-specific trait-associations

Since the observed sex-specific SNP-trait associations do not seem to be driven by sex-specific eQTL effects in our data, we performed power analyses using eQTL-effect differences observed in the data of the BIOS Consortium, which represents the largest biologically plausible sex-differences.

If the SNP-trait association is fully mediated by gene expression, then in females (F) and males (M)

$$\beta_{GWAS(F)} = \alpha_{TWMR(F)} * \beta_{eQTLs(F)}$$

and

$$\beta_{GWAS(M)} = \alpha_{TWMR(M)} * \beta_{eQTLs(M)}$$

where β_{GWAS} and β_{eQTLs} indicate the effect of the SNP on the trait and on gene expression, respectively, and α_{TWMR} is the causal effect of the gene expression on the trait.

Assuming the same effect of the gene expression on the trait in both sexes (i.e.: $\alpha_{TWMR(M)} = \alpha_{TWMR(F)} = \alpha_{TWMR}$), the difference of SNP effect on the trait should be

$$\beta_{GWAS(F)} - \beta_{GWAS(M)} = \alpha_{TWMR} * (\beta_{eQTLs(F)} - \beta_{eQTLs(M)})$$

Using the differences observed by the BIOS Consortium and the causal effects estimated for a large number of complex traits by TWMR [18], we observed that the power to detect sex-specific trait-associations driven by sex-specific eQTLs is null, with an exception in the case when the largest causal effect of the gene expression on the trait observed in TWMR (100th percentile) is coupled with the largest differences in eQTL-effects observed in BIOS Consortium (**Figure 3a**). We also estimated the GWAS sample size required to reach 80% power to detect differences in sex-specific GWAS. We found that, using the subset of unrelated British individuals from UKBB (N=380K) we do not have the power to detect sex-differences driven by SNPs being eQTLs for causal. Indeed, our results show that even when we used the 80th percentile of the distribution of the significant causal effects of the gene expression on the trait, we need one to five million individuals to detect the different effect driven by sex-specific eQTLs (**Figure 3b**).

Figure 3a. Heatmap showing the power to detect significantly different effect in sex-specific GWAS collecting 190,000 females and 170,000 males

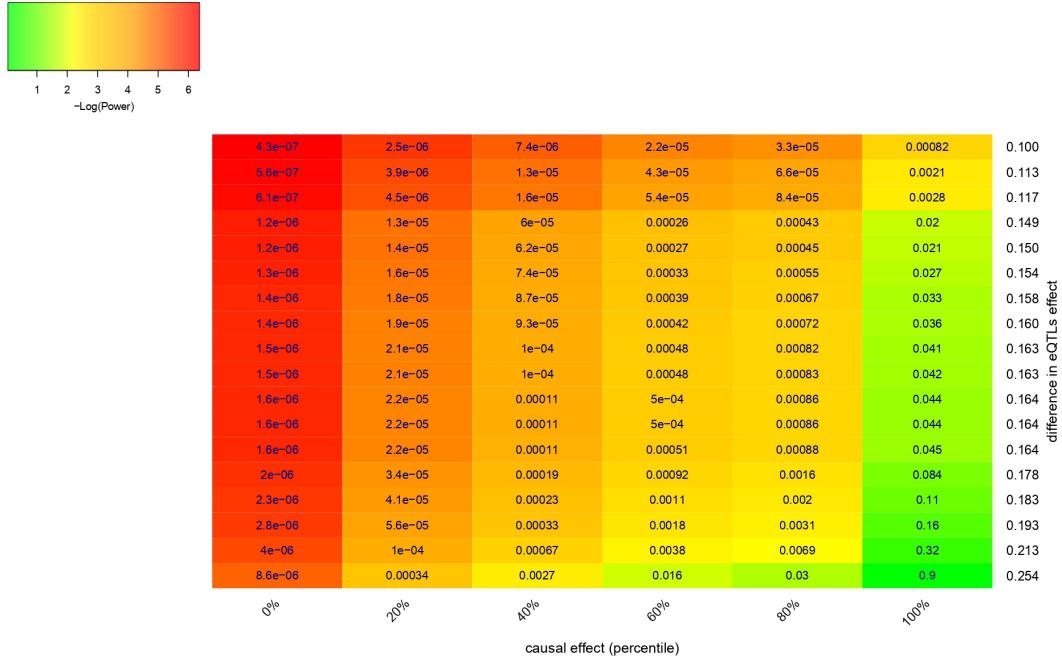
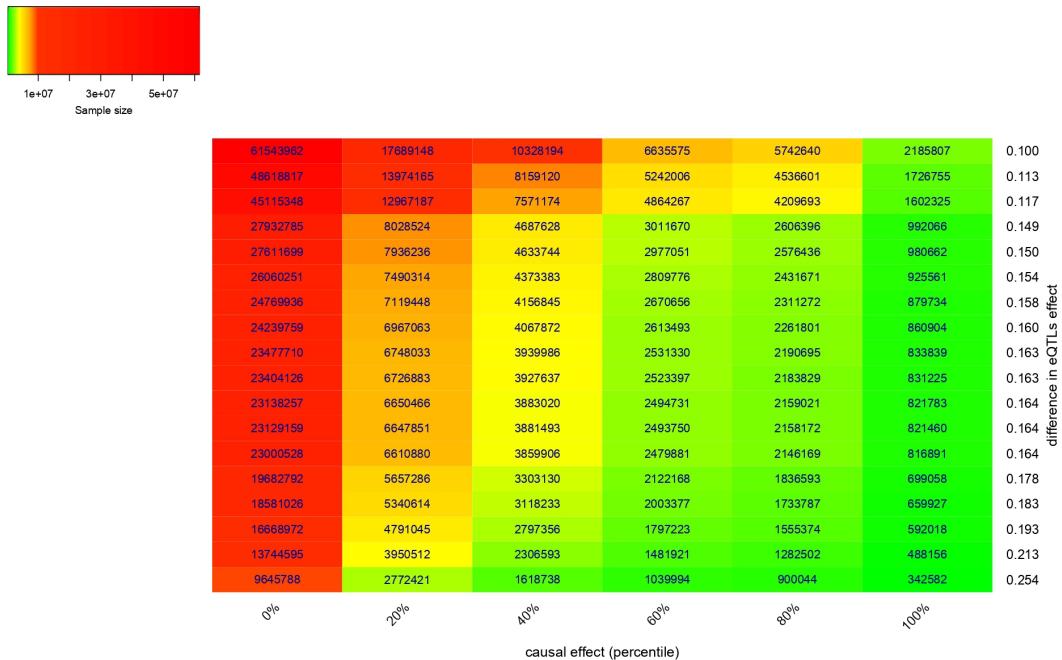


Figure 3b. Heatmap showing the total sample size (females + males) needed to detect significantly different effect in sex-specific GWAS at power > 80%.



Discussion

Notwithstanding the prominent differences in traits observed between men and women, there is little known about the role of sex-specific genetic effects. Several sex-stratified GWASs identified sex-specific genetic variants on autosomal

chromosomes, which highlights the fact that not all differences are located on the sex chromosomes [3-10]. Since many genetic variants exert their effect on complex traits through gene expression [17, 18], sex differences in eQTL effects might underlie such sex-specific GWAS associations.

In this study, we used large sex-specific eQTL data and sex-specific GWAS results to investigate the role of gene expression on the sexual dimorphism of several human phenotypes.

We identified 18 sex-specific eGenes and 7 of them were associated with traits known to exhibit sex differences. Furthermore we found that the sex-specific eGenes are nominally enriched in GO biological processes linked to sex-specific processes.

Our results suggest that sex-specific eQTLs in whole blood do not translate to detectable sex-specific trait associations, and vice versa that the observed sex-specific trait associations cannot be explained by sex-specific eQTLs. While a recent work revealed that ~11% of trait heritability could be explained by *cis*-eQTL regulation [26], our findings show that the sex-specific heritability is not detectably mediated by sex-specific gene expression regulation. Our extensive power analyses, performed using a range of realistic effect sizes, confirmed these observations. Indeed we proved that with the current sample size used in sex-specific GWAS, we do not have the power to detect differences in sex-specific trait-associations driven by sex-specific eQTLs. Our results suggest that we will be able to explore how sex-specific gene expression regulation translates to complex traits only when GWAS will be performed on millions of individuals. It is only then that we will be able to test the existence of potential compensatory mechanisms via negative feedback loops dampening such signal propagation.

There are some limitations to this study. Firstly, although the BIOS RNA-seq dataset is relatively large (including 1,918 women and 1,519 men), it is limited to whole blood. Since it is known that the effect of causal genes on diseases typically act in a tissue-specific manner [27-29], the investigation of other, more relevant, tissues could be crucial to estimate larger causal effects and unravel the sex-specific associations found by the GWAS. Moreover, although we used the biggest sex-specific GWAS results, we convincingly show that the currently available sample size is too small to reach the statistical power necessary for detecting sex-specific trait-association mediated by sex-specific blood eQTLs.

As several studies have shown that eQTL effects can be cell type-specific [30, 31], upcoming single-cell eQTL datasets [32] might be essential in identifying sex- and cell-type specific effects and unravel the biological mechanism behind sexual dimorphism. Alternatively, if sexual dimorphism of complex traits is not driven by gene expression changes we might need to explore other types of omics data to gain a deeper insight into the molecular underpinnings of sex-differences in complex diseases.

Methods

Study sample

The Biobank-based Integrative Omics Study (BIOS, <http://www.bbmri.nl/acquisition-use-analyze/bios/>) Consortium has been set up in an effort of several Dutch biobanks to create a homogenized dataset with different levels

of ‘omics’ data layers. Genotyping was performed in each cohort separately, as described before: LifeLines DEEP [33], Leiden Longevity Study [34, 35], Netherlands Twin Registry [36]; Rotterdam Study [37, 38], Prospective ALS Study Netherlands [39]. All genotypes were imputed to the Haplotype Reference Consortium [40] using the Michigan imputation server [41].

RNA-seq gene expression data was generated in The Human Genotyping facility (HugeF, Erasmus MC, Rotterdam, the Netherlands, <http://www.blimdna.org>). RNA-seq extraction and processing has been described before for a subset of the data [42]. Briefly, RNA was extracted from whole blood and paired-end sequenced using Illumina HiSeq 2000. Reads were aligned using STAR 2.3.0e [43] while masking common (MAF > 0.01) SNPs from the Genome of the Netherlands [44]. Gene-level expression was quantified using HTseq [45]. FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to check quality metrics, and we removed individuals with < 70% of reads mapping to exons (exon mapped / genome). We included only unrelated individuals in this analysis and removed population outliers by filtering out samples with >3 standard deviations from the average heterogeneity score. We removed 25 PCs, from the expression matrix with all cohorts combined, to account for unmeasured variation.

We stratified the samples by sex and performed the *cis*-eQTL mapping using a pipeline described previously [46]. In brief, the pipeline takes a window of 1Mb upstream and 1Mb downstream around each SNP to select genes or expression probes to test, based on the center position of the gene or probe. The association between these SNP-gene combinations was calculated using a Spearman correlation in each sex separately.

Differential gene expression and variance analysis

Differential expression analysis was preformed without considering genotypes, to identify genes with sex-specific expression (FDR 5%). For each gene, to test the difference in mean in the two sexes – calculated before any normalization but after quality check on samples – we used the t statistics

$$t = \frac{\mu_F - \mu_M}{\sqrt{\sigma_F^2 + \sigma_M^2}}$$

where μ and σ are the mean and the standard error, respectively.

To detect genes with sex-specific expression variance, we used normalized data and we applied two different strategies. First, we tested for difference in variance between females and males using an F-test, i.e.:

$$F = \frac{\sigma_F^2}{\sigma_M^2} \sim F(N_F, N_M)$$

Second we used the t statistic

$$t = \frac{\sigma_F^2 - \sigma_M^2}{\sqrt{var(\sigma_F^2) + var(\sigma_M^2)}}$$

Genes showing significant P -values (FDR 5%) in both tests, were used for pathways analyses.

Sex-specific eQTLs effects

To identify SNP-gene pairs with sex-difference, we computed P -values (P_{diff}) testing for difference between the standardized men-specific and women-specific β_{eQTLs} -estimates, with corresponding standard errors and using the t statistic

$$t = \frac{\beta_{\text{eQTLs}(F)} - \beta_{\text{eQTLs}(M)}}{\sqrt{\frac{1}{N_{\text{eQTLs}(F)}} + \frac{1}{N_{\text{eQTLs}(M)}}}}$$

We selected 462 sex-specific SNP-gene pairs at an FDR of 5% across all the pairs tested.

Sex-specific GWAS effects

To identify SNPs with sex-difference in the 39 associated phenotypes, we downloaded the summary statistics of the sex-stratified GWAS available at <http://www.nealelab.is/uk-biobank/>

We computed P -values (P_{diff}) testing for a difference between the men-specific and women-specific β_{GWAS} -estimates, with corresponding standard errors and using the t statistic

$$t = \frac{\beta_{\text{GWAS}(F)} - \beta_{\text{GWAS}(M)}}{\sqrt{SE_{(F)}^2 + SE_{(M)}^2}}$$

Pathway analyses

In order to explore whether certain pathways are enriched among the genes containing eQTLs with evidence for sex-difference, we applied topGO. We tested the enrichment of GO terms with regard to the list of sex-specific eGenes at $P_{\text{diff}} < 1 \times 10^{-4}$ (list of genes including at least one eQTL showing $P_{\text{diff}} < 1 \times 10^{-4}$).

Power analyses

We performed power analyses to calculate the probability that sex-specific SNPs found by GWAS are driven by sex-specific eQTLs. Using real observed data, we tested the power to detect a significant difference in β_{GWAS} in males and females starting from the difference observed in β_{eQTLs} and the causal effect of the gene expression on the phenotype calculated by TWMR (α_{TWMR}).

If the association of a SNP in a given phenotype is driven by eQTLs, then we have

$$\beta_{\text{GWAS}(\text{Females})} = \alpha_{\text{TWMR}(\text{Females})} \times \beta_{\text{eQTLs}(\text{Females})}$$

$$\beta_{GWAS(Males)} = \alpha_{TWMR(Males)} \times \beta_{eQTLs(Males)}$$

Let's suppose that the effect of the gene expression on the phenotype is the same in the two sexes

$$\alpha_{TWMR(F)} = \alpha_{TWMR(M)} = \alpha_{TWMR}$$

Then

$$\beta_{GWAS(F)} - \beta_{GWAS(M)} = \alpha_{TWMR} \times (\beta_{eQTLs(F)} - \beta_{eQTLs(M)}).$$

Since $\hat{\beta}_{GWAS(F)} \sim N(\beta_{GWAS(F)}, \frac{1}{N_{GWAS(F)}})$ and $\hat{\beta}_{GWAS(M)} \sim N(\beta_{GWAS(M)}, \frac{1}{N_{GWAS(M)}})$, our statistics t follows a normal distribution when N is large,

$$t = \frac{\hat{\beta}_{GWAS(F)} - \hat{\beta}_{GWAS(M)}}{\sqrt{\frac{1}{N_{GWAS(F)}} + \frac{1}{N_{GWAS(M)}}}} \sim N\left(\frac{\beta_{GWAS(F)} - \beta_{GWAS(M)}}{\sqrt{\frac{1}{N_{GWAS(F)}} + \frac{1}{N_{GWAS(M)}}}}, 1\right)$$

We tested the hypothesis $H_0: \vartheta = \beta_{GWAS(F)} - \beta_{GWAS(M)} = 0$ against $H_1: \vartheta = \beta_{GWAS(F)} - \beta_{GWAS(M)} \neq 0$.

Using the genome-wide significance threshold, H_0 will be rejected if

$$|t| > \Phi^{-1}(2.5 \times 10^{-8}).$$

Then, the power to detect $\beta_{GWAS(F)} - \beta_{GWAS(M)} = \vartheta$ is

$$\begin{aligned} P & \left(\left| \frac{\alpha_{TWMR} \times (\beta_{eQTLs(F)} - \beta_{eQTLs(M)})}{\sqrt{\frac{1}{N_{GWAS(F)}} + \frac{1}{N_{GWAS(M)}}}} \right| > 5.43 \right) \\ &= 2 \times \left(1 - \Phi \left(5.43 - \frac{\alpha_{TWMR} \times (\beta_{eQTLs(F)} - \beta_{eQTLs(M)})}{\sqrt{\frac{1}{N_{GWAS(F)}} + \frac{1}{N_{GWAS(M)}}}} \right) \right) \end{aligned}$$

We performed the power analyses using the difference of β_{eQTLs} observed in the 18 top sex-specific eQTLs and in 6 quantiles extracted from the distribution of significant causal effects estimated by TWMR for WHR.

Using the same statistics, we calculated the sample size to observe $\beta_{GWAS(F)} - \beta_{GWAS(M)} = \vartheta$ with power >0.8 .

Acknowledgments

This work was supported by grants from the Swiss National Science Foundation (31003A-143914 to ZK and 31003A_160203 to AR) and the Horizon2020 Twinning project ePerMed (692145 to AR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We would like to thank Chiara Auwerx, Liza Darroux and Ninon Mounier for their valuable feedback and comments on this manuscript.

Author Contributions

E.P., Z.K. and A.R. conceived and designed the study. E.P. performed statistical analyses. A.C. performed statistical analyses on BIOS Consortium dataset. L.F. supervised all statistical analyses performed in the BIOS Consortium. K.L. contributed by providing statistical support. E.P., T.G.R., F.A.S., L.F. A.R. and Z.K. contributed by providing advice on interpretation of results. E.P., A.R. and Z.K. wrote the manuscript with the participation of all authors. BIOS Consortium provided eQTLs data.

BIOS Consortium (Biobank-based Integrative Omics Study) – Author information

Management Team Bastiaan T. Heijmans (chair)¹, Peter A.C. 't Hoen², Joyce van Meurs³, Aaron Isaacs⁴, Rick Jansen⁵, Lude Franke⁶.

Cohort collection Dorret I. Boomsma⁷, René Pool⁷, Jenny van Dongen⁷, Jouke J. Hottenga⁷ (Netherlands Twin Register); Marleen MJ van Greevenbroek⁸, Coen D.A. Stehouwer⁸, Carla J.H. van der Kallen⁸, Casper G. Schalkwijk⁸ (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga⁶, Lude Franke⁶, Sasha Zhernakova⁶, Ettje F. Tigchelaar⁶ (LifeLines Deep); P. Eline Slagboom¹, Marian Beekman¹, Joris Deelen¹, Diana van Heemst⁹ (Leiden Longevity Study); Jan H. Veldink¹⁰, Leonard H. van den Berg¹⁰ (Prospective ALS Study Netherlands); Cornelia M. van Duijn⁴, Bert A. Hofman¹¹, Aaron Isaacs⁴, André G. Uitterlinden³ (Rotterdam Study).

Data Generation Joyce van Meurs (Chair)³, P. Mila Jhamai³, Michael Verbiest³, H. Eka D. Suchiman¹, Marijn Verkerk³, Ruud van der Breggen¹, Jeroen van Rooij³, Nico Lakenberg¹.

Data management and computational infrastructure Hailiang Mei (Chair)¹², Maarten van Iterson¹, Michiel van Galen², Jan Bot¹³, Dasha V. Zhernakova⁶, Rick Jansen⁵, Peter van 't Hof¹, Patrick Deelen⁶, Irene Nooren¹³, Peter A.C. 't Hoen², Bastiaan T. Heijmans¹, Matthijs Moed¹.

Data Analysis Group Lude Franke (Co-Chair)⁶, Martijn Vermaat², Dasha V. Zhernakova⁶, René Luijk¹, Marc Jan Bonder⁶, Maarten van Iterson¹, Patrick Deelen⁶, Freerk van Dijk¹⁴, Michiel van Galen², Wibowo Arindrarto¹², Szymon M. Kielbasa¹⁵, Morris A. Swertz¹⁴, Erik. W van Zwet¹⁵, Rick Jansen⁵, Peter-Bram 't Hoen (Co-Chair)², Bastiaan T. Heijmans (Co-Chair)¹.

¹ Molecular Epidemiology Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

² Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

³ Department of Internal Medicine, ErasmusMC, Rotterdam, The Netherlands

⁴ Department of Genetic Epidemiology, ErasmusMC, Rotterdam, The Netherlands

⁵ Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands

⁶ Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

⁷ Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands

⁸ Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, The Netherlands

⁹ Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands

¹⁰ Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

¹¹ Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands

¹² Sequence Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands 13. SURFsara, Amsterdam, the Netherlands

¹⁴ Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

¹⁵ Medical Statistics Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

Bibliography

1. Whitacre, C.C., S.C. Reingold, and P.A. O'Looney, *A gender gap in autoimmunity*. Science, 1999. **283**(5406): p. 1277-8.
2. Lerner, D.J. and W.B. Kannel, *Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population*. Am Heart J, 1986. **111**(2): p. 383-90.
3. Dumitrescu, L., et al., *Sex differences in the genetic predictors of Alzheimer's pathology*. Brain, 2019. **142**(9): p. 2581-2589.
4. Graham, S.E., et al., *Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis*. Nat Commun, 2019. **10**(1): p. 1847.
5. Hartiala, J.A., et al., *Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease*. Nat Commun, 2016. **7**: p. 10558.
6. Martin, J., et al., *A Genetic Investigation of Sex Bias in the Prevalence of Attention-Deficit/Hyperactivity Disorder*. Biol Psychiatry, 2018. **83**(12): p. 1044-1053.
7. 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, 2019. **28**(1): p. 166-174.
8. 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, 2013. **9**(6): p. e1003500.
9. Rask-Andersen, M., et al., *Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects*. Nat Commun, 2019. **10**(1): p. 339.
10. Zeng, Y., et al., *Sex Differences in Genetic Associations With Longevity*. JAMA Netw Open, 2018. **1**(4): p. e181670.
11. Pirastu, N.C., M.; Priyanka Nandakumar, P.; Mignogna, G.; Abdellaoui, A.; Hollis, B.; Kanai, M.; Rajagopal, V.M.; Della Briotta Parolo, P.; Baya, N.;

Carey, C.; Karjalainen, J.; Als, T.D.; Van der Zee, M.D.; Day, F.R.; Ong, K.K.; Finngen Study; Me Research Team; Consortium iPSYCH; Morisaki, T.; de Geus, E.; Bellocchio, R.; Okada, Y.; Børglum, A.D.; Joshi, P.; Auton, A.; Hinds, D.; Neale, B.M.; Walters, R.K.; Nivard, M.G.; Perry, J.R.B.; Ganna A., *Genetic analyses identify widespread sex-differential participation bias*. biorxiv, 2020.

12. Fehrman, R.S., et al., *Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA*. PLoS Genet, 2011. **7**(8): p. e1002197.

13. Hernandez, D.G., et al., *Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain*. Neurobiol Dis, 2012. **47**(1): p. 20-8.

14. Nica, A.C., et al., *Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations*. PLoS Genet, 2010. **6**(4): p. e1000895.

15. Nicolae, D.L., et al., *Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS*. PLoS Genet, 2010. **6**(4): p. e1000888.

16. Gusev, A., et al., *Integrative approaches for large-scale transcriptome-wide association studies*. Nat Genet, 2016. **48**(3): p. 245-52.

17. Zhu, Z., et al., *Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets*. Nat Genet, 2016. **48**(5): p. 481-7.

18. Porcu, E., et al., *Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits*. Nat Commun, 2019. **10**(1): p. 3300.

19. Vosa, U.C., A.; Westra, H.J.; Bonder, M.J.; Deelen, P., *Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis*. bioRxiv, 2018.

20. Huang, J., et al., *Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel*. Nat Commun, 2015. **6**: p. 8111.

21. Mauvais-Jarvis, F., *Sex differences in metabolic homeostasis, diabetes, and obesity*. Biol Sex Differ, 2015. **6**: p. 14.

22. Spolarics, Z., et al., *Inherent X-Linked Genetic Variability and Cellular Mosaicism Unique to Females Contribute to Sex-Related Differences in the Innate Immune Response*. Front Immunol, 2017. **8**: p. 1455.

23. Canela-Xandri, O., K. Rawlik, and A. Tenesa, *An atlas of genetic associations in UK Biobank*. Nat Genet, 2018. **50**(11): p. 1593-1599.

24. Lotta, L.A., et al., *Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance*. Nat Genet, 2017. **49**(1): p. 17-26.

25. Herranz, P., et al., *Lipodystrophy syndromes*. Dermatol Clin, 2008. **26**(4): p. 569-78, ix.

26. Yao, D.W.O.C., L.J.; Price, A.L.; Gusev A, *Quantifying genetic effects on disease mediated by assayed gene expression levels*. bioRxiv, 2020.

27. Musunuru, K., et al., *From noncoding variant to phenotype via SORT1 at the *Ip13* cholesterol locus*. Nature, 2010. **466**(7307): p. 714-9.

28. Ongen, H., et al., *Estimating the causal tissues for complex traits and diseases*. Nat Genet, 2017. **49**(12): p. 1676-1683.

29. Claussnitzer, M., et al., *FTO Obesity Variant Circuitry and Adipocyte Browning in Humans*. N Engl J Med, 2015. **373**(10): p. 895-907.

30. Brown, C.D., L.M. Mangravite, and B.E. Engelhardt, *Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs*. PLoS Genet, 2013. **9**(8): p. e1003649.
31. Fairfax, B.P., et al., *Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles*. Nat Genet, 2012. **44**(5): p. 502-10.
32. van der Wijst, M., et al., *The single-cell eQTLGen consortium*. Elife, 2020. **9**.
33. Tigchelaar, E.F., et al., *Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics*. BMJ Open, 2015. **5**(8): p. e006772.
34. Schoenmaker, M., et al., *Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study*. European Journal of Human Genetics, 2006. **14**(1): p. 79-84.
35. Deelen, J., et al., *Employing biomarkers of healthy ageing for leveraging genetic studies into human longevity*. Experimental Gerontology, 2016. **82**: p. 166-174.
36. Lin, B.D., et al., *The Genetic Overlap Between Hair and Eye Color*. Twin Res Hum Genet, 2016. **19**(6): p. 595-599.
37. Hofman, A., et al., *The Rotterdam Study: 2014 objectives and design update*. Eur J Epidemiol, 2013. **28**(11): p. 889-926.
38. Hofman, A., et al., *The Rotterdam Study: 2016 objectives and design update*. Eur J Epidemiol, 2015. **30**(8): p. 661-708.
39. Huisman, M.H., et al., *Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology*. J Neurol Neurosurg Psychiatry, 2011. **82**(10): p. 1165-70.
40. McCarthy, S., et al., *A reference panel of 64,976 haplotypes for genotype imputation*. Nat Genet, 2016. **48**(10): p. 1279-83.
41. Das, S., et al., *Next-generation genotype imputation service and methods*. Nat Genet, 2016. **48**(10): p. 1284-1287.
42. Zhernakova, D.V., et al., *Identification of context-dependent expression quantitative trait loci in whole blood*. Nat Genet, 2017. **49**(1): p. 139-145.
43. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner*. Bioinformatics, 2013. **29**(1): p. 15-21.
44. Boomsma, D.I., et al., *The Genome of the Netherlands: design, and project goals*. Eur J Hum Genet, 2014. **22**(2): p. 221-7.
45. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data*. Bioinformatics, 2015. **31**(2): p. 166-9.
46. Westra, H.J., et al., *Systematic identification of trans eQTLs as putative drivers of known disease associations*. Nat Genet, 2013. **45**(10): p. 1238-1243.