Sex-biased genetic effects on gene regulation in humans

Antigone S. Dimas[1,2,3], Alexandra C. Nica[1], Stephen B. Montgomery[1], Barbara E. Stranger[4,5,6], Towfique Raj[4,5,6], Alfonso Buil[1], Thomas Giger[1], Tuuli Lappalainen[1], Maria Gutierrez-Arcelus[1], MuTHER Consortium, Mark I. McCarthy[2,7,8], Emmanouil T. Dermitzakis[1].

[1.] Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1211, Switzerland

[2.] Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

[3.] Biomedical Sciences Research Center Al. Fleming, 16672 Vari, Greece

[4.] Department of Medicine, Division of Genetics, Brigham and Women's Hospital, Boston, MA, 02115, USA

[5.] Harvard Medical School, Boston, MA, 02115, USA

[6.] Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA. 02142, USA

[7.] Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LJ, UK

[8.] Oxford NIHR Biomedical Research Centre, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LJ, UK

Correspondence should be addressed to ETD (emmanouil.dermitzakis@unige.ch)

ABSTRACT

Human regulatory variation, reported as expression quantitative trait loci (eQTLs), contributes to differences between populations and tissues. The contribution of eQTLs to differences between sexes however has not been investigated to date. Here we explore regulatory variation in females and males and demonstrate that 12-15% of autosomal eQTLs function in a sex-biased manner. We show that genes possessing sex-biased eQTLs are expressed at similar levels across the sexes and highlight cases of genes controlling sexually dimorphic and shared traits that are under the control of distinct regulatory elements in females and males. This study illustrates that sex provides important context that can modify the effects of functional genetic variants.

INTRODUCTION

The majority of traits that distinguish the two sexes develop secondarily to the development of the ovaries and testes (WILLIAMS and CARROLL 2009). Most studies of sexual dimorphism have focused on the impact of hormones or on the genetic contribution of sex chromosomes. However there is growing evidence that genetic variation on the autosomes contributes to sexual dimorphism (HEID *et al.* 2010; OBER *et al.* 2008). Sex-specific QTLs for sexually dimorphic traits such as lifespan and HDL-cholesterol have been detected respectively in Drosophila (NUZHDIN *et al.* 1997) and mouse (KORSTANJE *et al.* 2004). Sex-specific eQTLs have also been detected in mice (YANG *et al.* 2006), but whether such effects on expression regulation are also seen in humans has not been explored to date.

To address the above question we used gene expression levels quantified in EBV-transformed B-cells (lymphoblastoid cell lines or LCLs) from four HapMap populations (CEU: 54 females (F), 55 males (M); CHB: 42 F, 38 M; JPT: 40 F, 42 M; YRI: 53 F, 55 M) (STRANGER *et al.* 2012). A recent study examining the impact of EBV transformation on LCL expression and methylation profiles has shown that LCLs recapitulate naturally occurring expression variation in primary B cells (CALISKAN *et al.* 2011). A large fraction of sex-specific epigenetic and gene expression effects are therefore likely to be maintained in LCLs. Sex-specific effects due to hormones however cannot be captured by this study since LCLs are grown in the absence of

hormones. We stratified each population sample by sex and performed association of SNP genotype with mRNA levels using Spearman rank correlation (SRC) in each sex separately (sex-stratified analysis). Differences between LCLs derived from females and males are expected because: a) LCLs retain an important fraction of their heritable epigenetic and methylation profiles (McDaniell *et al.* 2010), and b) cell lines derived from females and males differ genotypically due to the presence of sex chromosomes (XX vs. XY).

RESULTS AND DISCUSSION

We tested all SNPs mapping in a 2 Mb window centred on the transcription start site (TSS) of genes, defined cis eQTLs as SNPs with SRC $p<10^{-5}$, and eQTL-genes as genes with at least one cis eQTL. We subsequently increased our stringency by ensuring equal levels of estimated false discovery rate (FDR) for eQTLs that were detected in both sexes (shared) and eQTLs that were detected in one sex only (threshold-based sex discordant or TBSD). We found that at an estimated FDR of 13-17%, approximately one third of eQTL-genes were TBSD (Table 1, Fig. S1). In CEU for example, from a total of 178 and 151 eQTL-genes detected in females and males respectively, 68 and 41 were TBSD. The remaining two thirds of eQTL-genes were shared across sexes for each population.

We also carried out expression association testing for the whole population sample (whole sample study), not stratifying by sex, to enable comparison of findings. In this analysis we detected almost all shared, and 60-78% of TBSD eQTL-genes identified in the sex-stratified study (Table 1). In CEU for example, we detected 109 of the 110 shared eQTL-genes, as well as 41 out of 68, and 26 out of 41 TBSD eQTL-genes that were identified in females and males respectively. Notably, 22-40% of TBSD eQTL-genes (corresponding to 10-15 % of the sex-stratified study total discoveries) were not discovered in the whole sample analysis. This fraction of eQTLs is likely to harbour the true variants that exert effects in a sex-dependent manner and pooling of both sexes achieves a greater sample size, but dilutes rather than strengthens the statistical signal (Fig. S2).

To evaluate the strength of TBSD signals, we applied a continuous (vs. threshold-based) measure of significance. We considered eQTLs detected in one sex only (discovery sex)

and explored their SRC nominal p-values in the other sex (non-discovery sex). Applying the methodology (q-values) described in (STOREY and TIBSHIRANI 2003) we estimated the fraction of the distribution that is enriched for statistically significant effects (π_1) and found that this corresponds, on average, to 72%. Therefore we estimate that 28% (standard deviation 1%) of TBSD associations had negligible effects in the non-discovery sex and we define this subset of eQTLs as sex-biased. In CEU for example 28% of the 109 TBSD discoveries corresponds to 30.52 eQTL-genes. At an estimated FDR of 17% we expect 25.33 true discoveries. This brings the fraction of truly sex-biased discoveries to ~12% of the total (219). Similarly, truly sex-biased eQTL-genes range from 12-14% in CHB, JPT, YRI.

As an independent means to evaluate sex-biased eQTLs, we tested TBSD SNP-expression probes using ANOVA with a SNP x sex interaction term to capture sex-biased effect size patterns. We observed enrichment of low p-values (<=0.05) for the interaction term in overwhelming excess compared to what is expected by chance (Table S1) and estimate that 44-50% of TBSD associations are sex-biased. Taken together with estimated FDR, ANOVA results put the fraction of truly sex-biased eQTLs at approximately 15%. This is consistent with our previous estimate of 12-14%. Furthermore, we highlight that the variance in gene expression explained by the SNP x sex interaction term is significantly higher for sex-biased genes (Fig. S3). It is worth noting that ANOVA highlights both differences in statistical significance and effect size (fold change or slope). This fold change difference constitutes an additional measure of sex bias (NICA *et al.* 2011).

Sampling effects arising from small sample size and population specificity of eQTLs (STRANGER *et al.* 2012) hamper cross-population replication. Indeed, we observed low levels of replication for both shared and TBSD eQTLs. To overcome this limitation, we endeavoured to replicate CEU eQTLs in a population-based cohort (MuTHER) of female twins from the UK (NICA *et al.* 2011). Our expectation was that CEU female-TBSD eQTLs would replicate at higher levels than CEU male-TBSD eQTLs in MuTHER twins. Indeed we observed higher levels of replication of female-TBSD eQTLs with low p-value enrichment (STOREY and TIBSHIRANI 2003) π_1=0.81 than male-TBSD eQTLs where π_1= 0.53.

To address whether TBSD and (ANOVA-defined) sex-biased eQTL discovery is driven by differences in gene expression levels between females and males, we compared $\log_2$ expression medians across the sexes for: a) all autosomal genes tested (Fig. 1A), b) eQTL-genes (Fig. 1B), c) TBSD eQTL-genes (Fig. 1C), and d) sex-biased eQTL-genes (Fig. 1D). In all cases we detected very high correlation between female and male expression medians (r^2=0.98-0.99, $p<10^{-4}$) and in line with other studies (IDAGHDOUR *et al.* 2008; ZHANG *et al.* 2007; ZHANG *et al.* 2009) we did not detect significant differences between females and males. This demonstrates that TBSD and sex-biased eQTLs do not arise as a consequence of expression level differences between the sexes. Therefore, sex bias in genetic effects of gene regulation is not of the same nature as differential expression of genes between sexes.

Although almost all shared eQTL-genes detected in the sex-stratified study were also detected in the whole sample analysis, we identified five cases of female-male shared eQTL-genes that were not discovered when pooling the two sexes into a single analysis (Fig. 2A-E, Table S2). In these cases, although the eQTL-gene is shared, there are independent regulatory elements in each sex. These eQTL-SNPs have negligible significance in the non-discovery sex (Table S2, Fig. S4-8) explaining why such signals are likely to be diluted when both sexes are analyzed simultaneously. These cases include genes (see below, Figures 2A-E created using the UCSC Genome Browser (KENT *et al.* 2002), http://genome.ucsc.edu) with a role in gamete formation, fertility and sexual dimorphism, but also genes involved in processes that are not linked to perceived sex-related traits. This suggests that there may be a sex-biased dimension for traits that, to date, are considered to have similar biology across sexes. *SPO11* (CEU, Fig. 2A, Fig. S4) is involved in meiotic recombination (BELLANI *et al.* 2010), spermatocyte formation, it is expressed in oocytes, and both female and male knockout mice are infertile (BELLANI *et al.* 2010). *CKLF* (JPT, Fig. 2B, Fig. S5) is a chemokine with a role in muscle development and neuronal migration (WANG *et al.* 2010). Its expression is increased in systemic lupus erythematosus (SLE) and in rheumatoid arthritis (RA), diseases that are nine and three times more common in women respectively. *MRFAP1L1* (JPT, Fig. 2C, Fig. S6) is thought to have a role in spermatogenesis through its interaction with *TSNAX* (RUAL *et al.* 2005), a gene involved in spermatogenesis, neuronal regulation and genome stability (JAENDLING and MCFARLANE 2010). *ODF2L* (YRI, Fig. 2D, Fig. S7) interacts with *PRSS23* (STELZL *et*

al. 2005), a serine protease involved in proteolytic degradation of extracellular matrix components, an essential process for ovulation (WAHLBERG *et al.* 2008). Finally, *PSAP* (YRI, Fig. 2E, Fig. S8) is a conserved glycoprotein involved in the development of the reproductive and nervous systems (HU *et al.* 2010). It has a developmental role in prostate cancer, its inactivation in mice leads atrophy of the male reproductive system and its down-regulation decreases metastatic prostate cancer cell adhesion, migration and invasion (HU *et al.* 2010).

To gain a better understanding of the biology behind eQTLs in females and males, we explored the properties of TBSD and sex-biased eQTLs. We found that the direction of allelic effects was consistent across the sexes (Fig. S9) suggesting that if an eQTL allele increases/decreases expression in one sex, it will have the same direction of effect in the other sex. Similarly to cell-type-specific (DIMAS *et al.* 2009) and population-specific (STRANGER *et al.* 2012) eQTLs, we found that TBSD and sex-biased eQTLs (at similar levels of FDR to shared eQTLs) have lower effects and span a range of distances from the TSS, whereas shared eQTLs have higher effects and cluster around the TSS (Fig. S10). This trend is also revealed when plotting the significance differential across the sexes for eQTLs that are sex-biased and those with ANOVA interaction $p>0.05$ vs. distance to TSS (Fig. S11).

To quantify the impact of regulatory variants on gene expression, we measured the expression fold change between the two homozygote classes (Fig. 3A and B and Fig. S12A and B). As expected, we found that sex-biased eQTLs exert a higher fold change in expression in the discovery sex, whereas shared eQTLs tend to result in similar fold changes in females and males. Notably however, there are cases of shared eQTLs where large differences in fold change across the sexes are observed. This implies that distinct regulatory effects are exerted in each sex, even in cases where the eQTL has been designated as shared using threshold-based significance criteria. In CEU for example, a shared eQTL for *PNMAL1* (Fig. 3A) shows a higher fold change in females compared to males (2.7 vs. 1.3 respectively). Similarly *WBSCR27* (Fig. 3B) shows a higher fold change in males compared to females (1.1 vs. 2.9 respectively). Currently, very little is known about these genes, but the above observation may help elucidate their role. Furthermore, this observation highlights the value of considering parameters of sex bias beyond statistical significance (e.g. effect size).

To obtain an overview of eQTL-gene biological functions we used the DAVID Functional Annotation tool (HUANG DA *et al.* 2009) and interrogated Gene Ontology biological processes for TBSD and sex-biased vs. shared genes. We observed only weak enrichment for GO terms linked to reproduction and meiosis possibly due to that fact that we are underpowered for this analysis. A more interesting hypothesis however is that sex-biased gene regulation affects perceived sexually dimorphic and other traits in a similar way. Given the role of regulatory variation in shaping complex traits and determining disease risk (NICA *et al.* 2010), we also interrogated the overlap between sex-biased eQTLs and GWAS SNPs (HINDORFF *et al.* 2009) and found two cases of overlap, both for diseases with a well-established sex imbalance (Table S3, Fig. S13). In both cases, regulatory trait concordance scores (NICA *et al.* 2010) (RTC), an index integrating eQTL and GWAS data to detect disease-causing regulatory effects, were equal to one (RTC=1) suggesting that the disease and regulatory signal are overlapping. rs167769 is associated with eosinophilic esophagitis (ROTHENBERG *et al.* 2010), a disease four times more common in young males. The eQTL-gene for this SNP is *STAT6*, a component of the JAK-STAT signalling cascade, which is implicated in other inflammatory diseases (BARRETT *et al.* 2008). rs2872507 is associated with RA (STAHL *et al.* 2010) and Crohn disease (BARRETT *et al.* 2008). The GWAS-reported locus on 17q12-q21 is a region under complex gene regulation and harbours multiple disease signals (VERLAAN *et al.* 2009). Most studies have focused on the potential role of *IKZF3* and *ORMDL3* in disease, but we report eQTL-gene *ZPBP2* and bring this gene to the fore as a potential mediator of the disease association.

In this study we searched for eQTLs separately in females and males of four HapMap populations and found that approximately 12-15% of detected eQTLs, and their underlying regulatory variants, are sex-biased. This bias extends to a subset of genes that possess the same eQTL in both sexes, but which exerts a very distinct regulatory effect, measured as resulting fold change in gene expression, in each sex. We highlight that a fraction of eQTLs that have been detected to date in eQTL studies where no distinction by sex is made, is driven by strong effects in one sex. Furthermore, we report that sex-related effects can impact traits where no sexual dimorphism has been observed to date. Given the prominence of sex-biased effects, this study emphasizes the

importance of considering each sex separately in genomic studies to uncover new disease and trait variants.

METHODS

RNA preparation

Total RNA was extracted from lymphoblastoid cell lines (LCLs) grown in a hormone-free environment, in the presence of phenol red•Na (0.0053 g/L, Sigma, Saint Louis, Missouri, United States). LCLs were from 379 individuals of four HapMap populations. The numbers of individuals of each population include: CEU: 109 Caucasians living in Utah USA, of northern and western European ancestry, CHB: 80 Han Chinese from Beijing, China, JPT: 82 Japanese in Tokyo, Japan, YRI: 108 Yoruba in Ibadan, Nigeria (International HapMap Constortium 2005; Coriell, Camden, New Jersey, United States). Two in vitro transcription (IVT) reactions were performed as one-quarter scale Message Amp II reactions (Ambion, Austin, Texas, United States) for each RNA extraction using 200 ng of total RNA as previously described (STRANGER *et al.* 2005). 1.5 µg of the cRNA was hybridized to an array (STRANGER *et al.* 2012; STRANGER *et al.* 2007b).

Gene expression quantification

To assay transcript levels in LCLs, we used Illumina's commercial whole genome expression array, Sentrix Human-6 Expression BeadChip version 2 (Illumina, San Diego, California, United States (KUHN *et al.* 2004)). These arrays utilize a bead pool with ~48,000 unique bead types (one for each of 47,294 transcripts, plus controls), each with several hundred thousand gene- specific 50mer probes attached. On a single BeadChip, six arrays were run in parallel (STRANGER *et al.* 2007b). Each of the two IVT reactions from the 379 samples was hybridized to one array each, so that each cell line had two replicate hybridizations. cRNA was hybridized to arrays, and subsequently labelled with Cy3- streptavidin (Amersham Biosciences, Little Chalfont, United Kingdom) and scanned with a Bead Station (Illumina) as previously described (STRANGER *et al.* 2005). Samples were processed in an order randomized with respect to population of origin and IVT batch.

Raw expression data normalization

With the Illumina bead technology, a single hybridization of RNA from one cell line to an array produces on average approximately 30 intensity values for each of 47,294 bead types. These background-corrected values for a single bead type are subsequently summarized by Illumina software and output to the user as a set of 47,294 intensity values for each individual hybridization. In our experiment, each cell line was hybridized to two arrays, thus resulting in two reported intensity values (as averages of the values from the 30 beads per probe) for each of the 47,294 bead types. Hybridization intensity values were normalized on a $\log_2$ scale quantile normalization method (BOLSTAD *et al.* 2003) across replicates of a single individual followed by a median normalization method across all individuals of the four populations. These normalized expression data were used as input for the expression analysis.

Selection of probes to analyze

Of the 47,294 probes for which we collected expression data, we selected a set of 22,744 probes to analyze (21,800 autosomal and 955 on chromosome X). We included in our analyses each probe that mapped to an Ensembl gene, but not to more than one Ensembl gene (Ensembl 49 NCBI Build 36), and we excluded probes mapping to the Y chromosome. The resulting sets of 21,800 autosomal and 955 chromosome X probes were analyzed in the association analyses and correspond to 17,673 unique autosomal and 741 unique chromosome X Ensembl genes respectively (total 18,414 genes).

Genetic variation

Single nucleotide polymorphisms (SNPs) for the same 379 HapMap individuals of CEU, CHB, JPT, and YRI, were selected (Release version 2) for use in the association analyses. SNPs that were included in association analyses fulfilled three criteria: a) were present in both females and males in each population, b) had a MAF>= 5% in each population, and c) had less than 20% missing data. This corresponds to between 1.1 million and 1.3 million SNPs per population.

Association testing

We performed association of SNP genotype with probe expression levels using Spearman rank correlation (SRC) as previously described (DIMAS *et al.* 2009; MONTGOMERY *et al.* 2010; STRANGER *et al.* 2007a; STRANGER *et al.* 2007b). All SNPs mapping in a 2 Mb window centred on genes' TSS were tested. We carried out two

rounds of association testing. In the first round we stratified each of the four population samples by sex (CEU: 54 F, 55 M; CHB: 42 F, 38 M; JPT: 40 F, 42 M; YRI: 53 F, 55 M) and for each population tested individuals of each sex separately (sex-stratified study). Expression association was carried out for autosomal genes. In the second round of association testing we tested all individuals of the population sample irrespective of sex in the same association test (whole sample study). In both the sex stratified study and the whole sample study we applied an empirical threshold of SRC p-value$<10^{-5}$ to define eQTLs.

Permutations and FDR estimation

We performed 10,000 permutations of expression phenotypes relative to genotypes for females and males of the four populations (i.e. CEU_F, CEU_M, CHB_F, CHB_M, JPT_F, JPT_M, YRI_F, YRI_M). In each case, permutations were carried out for 2,560 probes (2,062 genes) chosen at random and resulted in a matrix of SRC permuted $-\log_{10}$ p-values of 2,560 probes by 10,000 permutations. We calculated the average permuted $-\log_{10}$ p-value across probes and created a ranked distribution of permuted $-\log_{10}$ p-values which was subsequently used to estimate false discovery rate (FDR) (STOREY and TIBSHIRANI 2003). This is sufficient since SRC permutation thresholds tend to be very tight around the mean so we did not require to perform permutations for all genes.

Determination of significant associations and designation of shared, threshold-based sex discordant, and sex-biased eQTL-genes

We used an initial threshold-based significance filter to define cis eQTLs as those SNPs with a nominal SRC p$<10^{-5}$. eQTL-genes were defined as those genes with at least one cis eQTL and for all analyses we kept the most significant SNP association per gene. Empirically, we have found that most SNP-expression probe associations that pass the SRC p$<10^{-5}$ threshold correspond to a permutation threshold of 0.01 - 0.001 (DIMAS *et al.* 2009; STRANGER *et al.* 2012; STRANGER *et al.* 2007b). Applying this significance threshold to each population, we determined eQTL-genes that were: a) detected in both sexes (shared eQTL-genes), and b) detected in one sex only. We subsequently filtered eQTL-genes detected in one sex only to FDR levels corresponding to shared eQTL-genes. These FDR-filtered eQTL-genes detected in one sex only comprise the threshold-based sex discordant (TBSD) eQTL-genes detected in each population.

To evaluate the threshold-based sex results we employed a continuous measure of significance. For each population, we took all TBSD SNP-expression probes and extracted their SRC nominal p-values in the non-discovery sex. For these SNP-expression probes we quantified enrichment for low p-values using q-value criteria (STOREY and TIBSHIRANI 2003).

We estimated the number of truly sex-biased eQTLs genes as follows: from the q-value analysis we expect 28% of discoveries to be true nulls. For CEU, 28% of the 109 TBSD eQTLs (68 in females + 41 in males) corresponds to 30.52 genes. Given that estimated FDR is at 17%, 17% of 30.52 equals 25.33 true discoveries. 25.33 true discoveries corresponds to ~12% of the total number of 219 eQTL-genes (110 shared + 68 in females + 41 in males). Similarly we estimated numbers of sex-biased eQTL-genes for CHB, JPT, YRI.

As a further means of evaluating TBSD eQTLs, we carried out ANOVA with an interaction term for SNP x sex, which captures sex biased patterns of associations. For each population, we took all TBSD SNP-expression probes and registered those with ANOVA interaction term $p<=0.05$. This subset of TBSD SNP-expression probes constitutes the sex-biased eQTL-genes used in further analyses.

To assess levels of replication of TBSD associations we used the MuTHER dataset (NICA *et al*. 2011). MuTHER is a population-based cohort of female twins from the UK with expression levels quantified in LCLs, fat and skin. In this study only unrelated individuals were considered at a time by separating twins from the same pair and by performing two independent eQTL analyses. We sought to replicate CEU TBSD associations in LCLs from each twin population by testing for low p-value enrichment in MuTHER of CEU female-TBSD and CEU male-TBSD SNP-expression probes using q-values. Reported π_1 values are the average of the two twin populations.

Properties of eQTLs and eQTL-genes

For shared and sex-biased SNP-expression probes in each population_sex, we calculated the fold change in median $-\log_2$ expression values between major and minor allele homozygotes. To compare direction of the allelic effects for shared and TBSD eQTLs (and as a consequence of sex-biased eQTLs since these are a subset of TBSD

eQTLs) we explored the direction of the Spearman correlation coefficient (rho) for the union of significant SNP-expression probes across the sexes, for each population. To obtain an overview of eQTL-gene biological functions we used the DAVID Functional Annotation tool (HUANG DA *et al.* 2009) and interrogated GO Biological Processes (FAT) for shared and TBSD eQTL-genes vs. all 18,414 genes tested.

Overlap of sex-biased eQTLs with GWAS disease/trait SNPs

The NHGRI GWAS catalogue (HINDORFF *et al.* 2009) (www.genome.gov/gwastudies accessed 23 November 2010) was queried for GWAS SNPs that overlapped with eQTLs detected in this study.

DATA ACCESS

The expression data reported in this paper have been deposited in the Array Express (http://www.ebi.ac.uk/arrayexpress/) database (Series Accession Number E-MTAB-264 and E-MTAB-198). Furthermore, all TBSD eQTL results are provided as supplementary material.

ACKNOWLEDGMENTS

This study was supported by funds from Louis-Jeantet Foundation, the Swiss National Foundation and "NCCR Frontiers in Genetics" and the Wellcome Trust to E.T.D. Financial support was also from NIH (R01 MH090941) and Wellcome Trust (MuTHER 081917/Z/07/Z) to E.T.D and M.I.M. A.S.D. was funded by the Wellcome Trust (083270/Z/07/Z), ENGAGE (HEALTH-F4-2007-201413) and Marie Curie FP7-PEOPLE-2010-IEF. The authors would like to thank Vital-IT for the computational resources to perform the analysis and Anastasia Komnou for her help in figure preparation.

AUTHOR CONTRIBUTIONS

The study was designed by A.S.D. and E.T.D. Analyses were carried out by A.S.D. Additional analyses were performed by A.C.N., S.B.M., B.E.S., A.B., T.R., T.G., T.L.,

and M.G.A. The manuscript was written by A.S.D and E.T.D. with contributions from M.I.M and from the other authors.

Competing financial interests

The authors declare no competing financial interests.

REFERENCES

BARRETT, J. C., S. HANSOUL, D. L. NICOLAE, J. H. CHO, R. H. DUERR *et al.*, 2008 Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet **40:** 955-962.

BELLANI, M. A., K. A. BOATENG, D. MCLEOD and R. D. CAMERINI-OTERO, 2010 The expression profile of the major mouse SPO11 isoforms indicates that SPO11beta introduces double strand breaks and suggests that SPO11alpha has an additional role in prophase in both spermatocytes and oocytes. Mol Cell Biol **30:** 4391-4403.

BOLSTAD, B. M., R. A. IRIZARRY, M. ASTRAND and T. P. SPEED, 2003 A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics **19:** 185-193.

CALISKAN, M., D. A. CUSANOVICH, C. OBER and Y. GILAD, 2011 The effects of EBV transformation on gene expression levels and methylation profiles. Hum Mol Genet **20:** 1643-1652.

DIMAS, A. S., S. DEUTSCH, B. E. STRANGER, S. B. MONTGOMERY, C. BOREL *et al.*, 2009 Common regulatory variation impacts gene expression in a cell type-dependent manner. Science **325:** 1246-1250.

HEID, I. M., A. U. JACKSON, J. C. RANDALL, T. W. WINKLER, L. QI *et al.*, 2010 Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet **42:** 949-960.

HINDORFF, L. A., P. SETHUPATHY, H. A. JUNKINS, E. M. RAMOS, J. P. MEHTA *et al.*, 2009 Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A **106:** 9362-9367.

HU, S., N. DELORME, Z. LIU, T. LIU, C. VELASCO-GONZALEZ *et al.*, 2010 Prosaposin down-modulation decreases metastatic prostate cancer cell adhesion, migration, and invasion. Mol Cancer **9:** 30.

HUANG DA, W., B. T. SHERMAN and R. A. LEMPICKI, 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc **4:** 44-57.

IDAGHDOUR, Y., J. D. STOREY, S. J. JADALLAH and G. GIBSON, 2008 A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs. PLoS Genet **4:** e1000052.

JAENDLING, A., and R. J. MCFARLANE, 2010 Biological roles of translin and translin-associated factor-X: RNA metabolism comes to the fore. Biochem J **429:** 225-234.

KENT, W. J., C. W. SUGNET, T. S. FUREY, K. M. ROSKIN, T. H. PRINGLE *et al.*, 2002 The human genome browser at UCSC. Genome Res **12:** 996-1006.

KORSTANJE, R., R. LI, T. HOWARD, P. KELMENSON, J. MARSHALL *et al.*, 2004 Influence of sex and diet on quantitative trait loci for HDL cholesterol levels in an SM/J by NZB/BlNJ intercross population. J Lipid Res **45:** 881-888.

KUHN, K., S. C. BAKER, E. CHUDIN, M. H. LIEU, S. OESER *et al.*, 2004 A novel, high-performance random array platform for quantitative gene expression profiling. Genome Res **14:** 2347-2356.

MCDANIELL, R., B. K. LEE, L. SONG, Z. LIU, A. P. BOYLE *et al.*, 2010 Heritable individual-specific and allele-specific chromatin signatures in humans. Science **328:** 235-239.

MONTGOMERY, S. B., M. SAMMETH, M. GUTIERREZ-ARCELUS, R. P. LACH, C. INGLE *et al.*, 2010 Transcriptome genetics using second generation sequencing in a Caucasian population. Nature **464:** 773-777.

NICA, A. C., S. B. MONTGOMERY, A. S. DIMAS, B. E. STRANGER, C. BEAZLEY *et al.*, 2010 Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet **6:** e1000895.

NICA, A. C., L. PARTS, D. GLASS, J. NISBET, A. BARRETT *et al.*, 2011 The Architecture of Gene Regulatory Variation across Multiple Human Tissues: The MuTHER Study. PLoS Genet **7:** e1002003.

NUZHDIN, S. V., E. G. PASYUKOVA, C. L. DILDA, Z. B. ZENG and T. F. MACKAY, 1997 Sex-specific quantitative trait loci affecting longevity in Drosophila melanogaster. Proc Natl Acad Sci U S A **94:** 9734-9739.

OBER, C., D. A. LOISEL and Y. GILAD, 2008 Sex-specific genetic architecture of human disease. Nat Rev Genet **9:** 911-922.

ROTHENBERG, M. E., J. M. SPERGEL, J. D. SHERRILL, K. ANNAIAH, L. J. MARTIN *et al.*, 2010 Common variants at 5q22 associate with pediatric eosinophilic esophagitis. Nat Genet **42:** 289-291.

RUAL, J. F., K. VENKATESAN, T. HAO, T. HIROZANE-KISHIKAWA, A. DRICOT *et al.*, 2005 Towards a proteome-scale map of the human protein-protein interaction network. Nature **437:** 1173-1178.

STAHL, E. A., S. RAYCHAUDHURI, E. F. REMMERS, G. XIE, S. EYRE *et al.*, 2010 Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet **42:** 508-514.

STELZL, U., U. WORM, M. LALOWSKI, C. HAENIG, F. H. BREMBECK *et al.*, 2005 A human protein-protein interaction network: a resource for annotating the proteome. Cell **122:** 957-968.

STOREY, J. D., and R. TIBSHIRANI, 2003 Statistical significance for genomewide studies. Proc Natl Acad Sci U S A **100:** 9440-9445.

STRANGER, B. E., M. S. FORREST, A. G. CLARK, M. J. MINICHIELLO, S. DEUTSCH *et al.*, 2005 Genome-wide associations of gene expression variation in humans. PLoS Genet **1:** e78.

STRANGER, B. E., M. S. FORREST, M. DUNNING, C. E. INGLE, C. BEAZLEY *et al.*, 2007a Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science **315:** 848-853.

STRANGER, B. E., S. B. MONTGOMERY, A. S. DIMAS, L. PARTS, O. STEGLE *et al.*, 2012 Patterns of cis regulatory variation in diverse human populations. PLoS Genet **8:** e1002639.

STRANGER, B. E., A. C. NICA, M. S. FORREST, A. DIMAS, C. P. BIRD *et al.*, 2007b Population genomics of human gene expression. Nat Genet **39:** 1217-1224.

VERLAAN, D. J., S. BERLIVET, G. M. HUNNINGHAKE, A. M. MADORE, M. LARIVIERE *et al.*, 2009 Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. Am J Hum Genet **85:** 377-393.

WAHLBERG, P., A. NYLANDER, N. AHLSKOG, K. LIU and T. NY, 2008 Expression and localization of the serine proteases high-temperature requirement factor A1, serine protease 23, and serine protease 35 in the mouse ovary. Endocrinology **149:** 5070-5077.

WANG, Z. Z., G. LI, X. Y. CHEN, M. ZHAO, Y. H. YUAN *et al.*, 2010 Chemokine-like factor 1, a novel cytokine, induces nerve cell migration through the non-extracellular Ca2+-dependent tyrosine kinases pathway. Brain Res **1308:** 24-34.

WILLIAMS, T. M., and S. B. CARROLL, 2009 Genetic and molecular insights into the development and evolution of sexual dimorphism. Nat Rev Genet **10:** 797-804.

YANG, X., E. E. SCHADT, S. WANG, H. WANG, A. P. ARNOLD *et al.*, 2006 Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res **16:** 995-1004.

ZHANG, W., W. K. BLEIBEL, C. A. ROE, N. J. COX and M. EILEEN DOLAN, 2007 Gender-specific differences in expression in human lymphoblastoid cell lines. Pharmacogenet Genomics **17:** 447-450.

ZHANG, W., R. S. HUANG, S. DUAN and M. E. DOLAN, 2009 Gene set enrichment analyses revealed differences in gene expression patterns between males and females. In Silico Biol **9:** 55-63.

Figure legends:

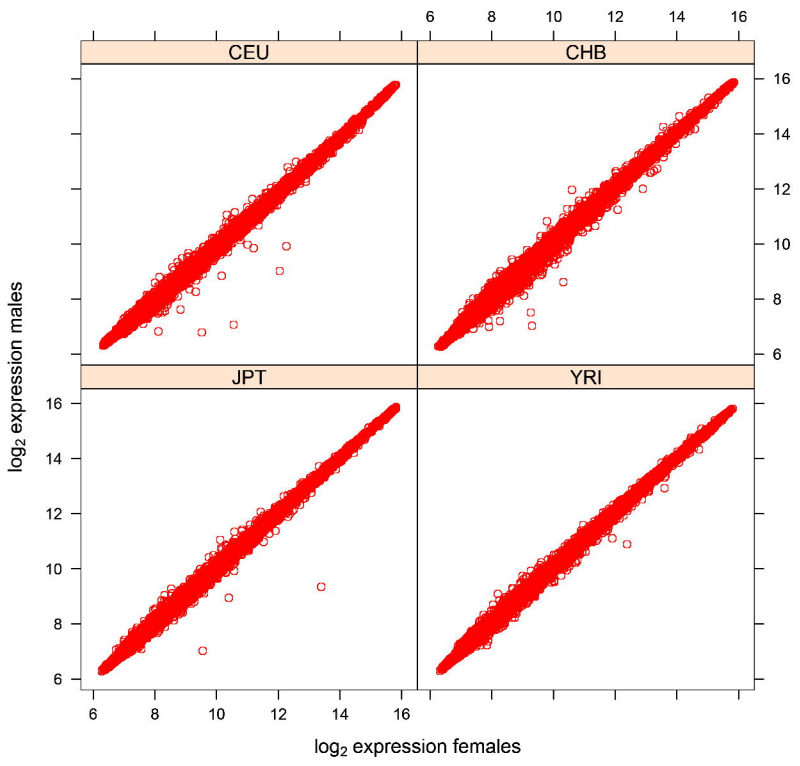
Figure 1. Comparison of female vs. male expression medians. Log_2 expression medians were compared across the sexes for (A) all 21,800 probes, (B) for eQTL-genes, (C) for TBSD eQTL-genes, and (D) for sex-biased eQTL-genes. In all cases the correlation between log_2 expression levels in females and males was found to be very high (r^2=0.99 and $p<10^{-4}$) suggesting that expression level differences between sexes are not the primary driver of eQTL discovery.

Figure 2. Shared eQTL-genes regulated by independent regulatory variants in females (F) and males (M). (A) In CEU rs6025625 (CEU F) and rs3787152 (CEU M), mapping within 760 kb of each other, are associated with *SPO11* expression levels. *SPO11* is essential for meiotic recombination and both F and M knockout mice are infertile. (B) In JPT rs2271025 (JPT F) and rs3826161 (JPT M) (r^2=0.26, D'=0.64) are eQTLs for *CKLF*, a chemokine with a role in muscle development and neuronal migration. (C) In JPT rs11734984 (JPT F) and rs10020189 (JPT M), mapping within 580 kb of each other, are eQTLs for *MRFAP1L1* a gene that is likely to have a role in spermatogenesis through its interaction with *TSNAX*. (D) In YRI rs506733 (YRI F) and rs12097932 (YRI M) (r^2=0.00, D'=0.09) are eQTLs for *ODF2L* a gene thought to have

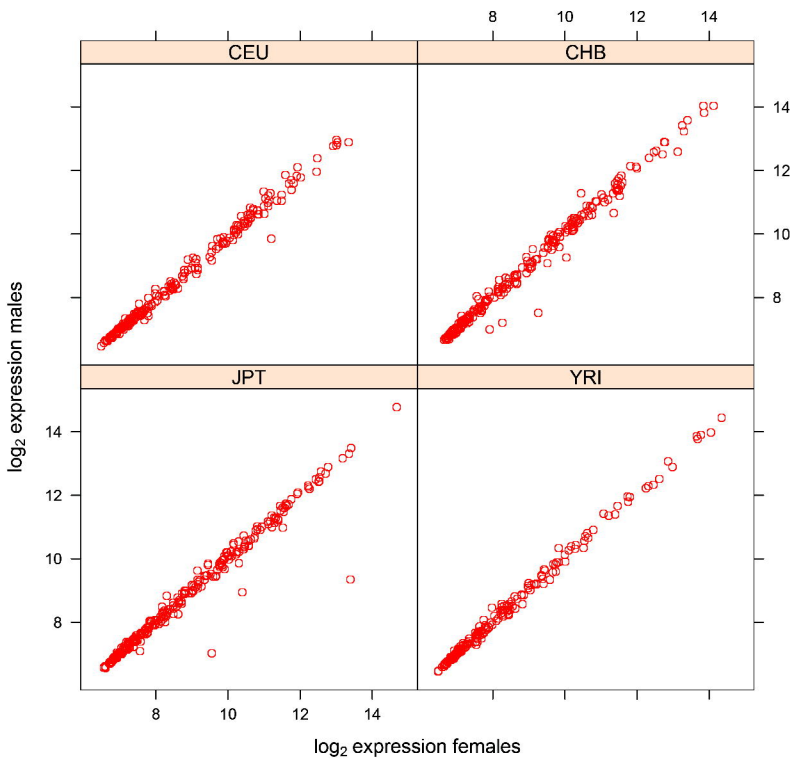
a role in ovulation through its interaction with *PRSS23*. (E) In YRI rs877663 (YRI F) and rs730722 (YRI M) (r^2=0.04, D'=0.43) are eQTLs for *PSAP*, a gene encoding a conserved glycoprotein involved in the development of the reproductive and nervous systems which has been linked to prostate cancer. Red boxes indicate eQTL-genes. r^2 and D' calculated for SNPs within 500 kb of each other. Figures made using the UCSC Genome Browser (http://genome.ucsc.edu).

Figure 3. Expression fold change between major and minor allele homozygotes for shared and sex-biased eQTLs. Sex-biased eQTLs show greater expression fold change in the discovery sex. Expression fold change in females (x axis) and males (y axis) for 65 SNP-expression probes in CEU females (A) and for 59 SNP-expression probes in CEU males (B). Although sex-biased eQTLs show greater fold change differences in the discovery sex, there are cases of shared eQTLs that also display prominent fold change differences between females and males e.g. i) *PNMAL1* (fold change: CEU F 2.7; CEU M 1.3), ii) *WBSCR27* (fold change: CEU F 1.1; CEU M 2.9). Expression fold change in females and males for 62 SNP-expression probes in YRI females (C) and for 62 SNP-expression probes in YRI males (D). Cases of shared eQTLs that display prominent fold change differences between females and males include: i) *S100A13* (fold change: YRI F 1.4; YRI M 0.9), ii) *CPA4* (fold change: YRI F 0.7; YRI M 1.9). Shared eQTLs: black, sex-biased eQTLs: red.

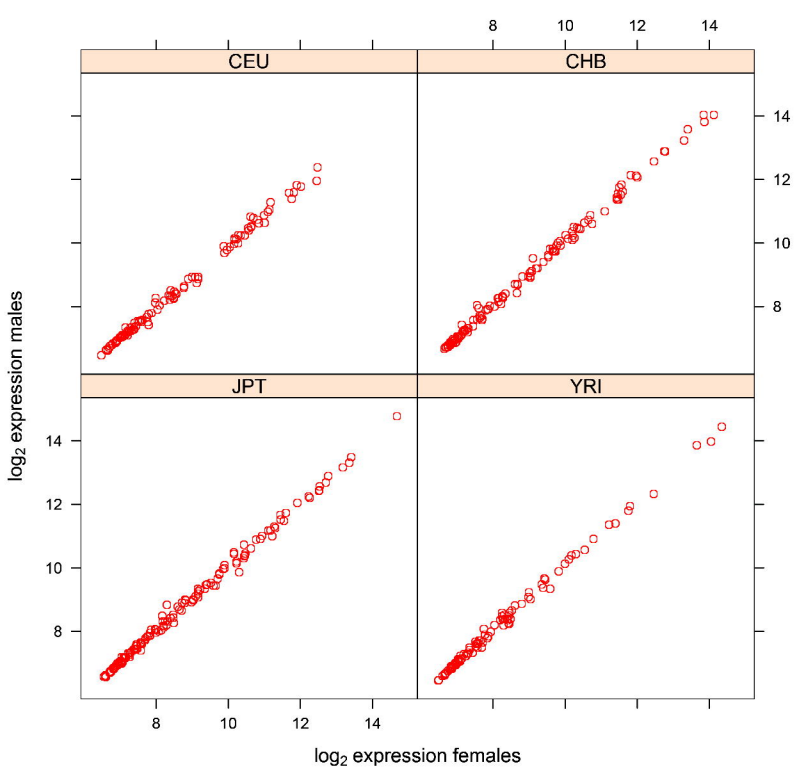
A


Expression median (all genes)
CEU
CHB
JPT
YRI
log2 expression males
log2 expression females

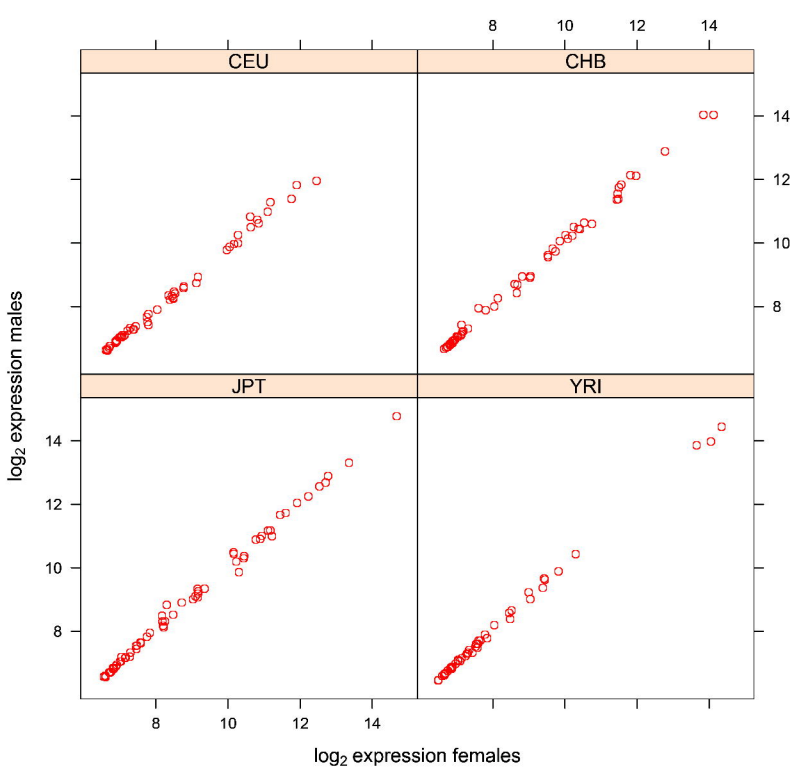
B


Expression median (eQTL-genes)
CEU
CHB
JPT
YRI
log2 expression males
log2 expression females

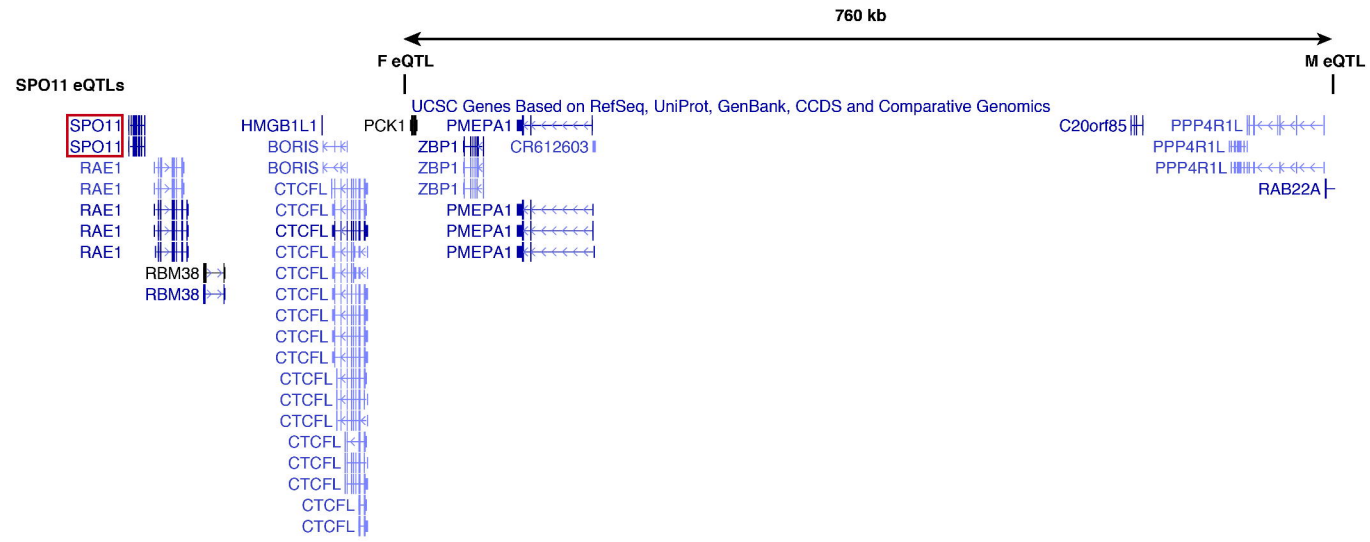
C


Expression median (TBSD eQTL-genes)
CEU
CHB
JPT
YRI
log2 expression males
log2 expression females

D


Expression median (sex-biased eQTL-genes)
CEU
CHB
JPT
YRI
log2 expression males
log2 expression females

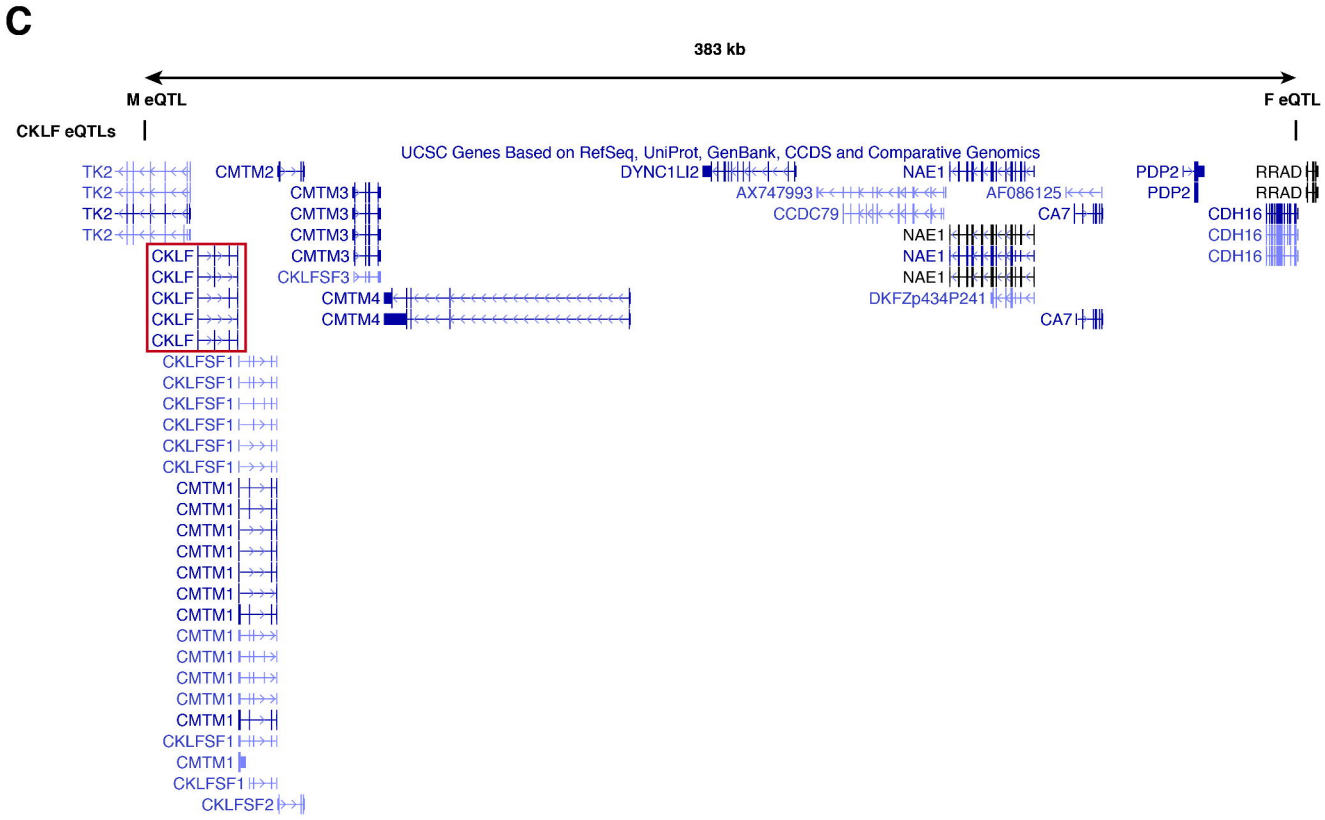
A


760 kb
F eQTL
M eQTL
SPO11 eQTLs
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
SPO11
SPO11
RAE1
RAE1
RAE1
RAE1
RAE1
RBM38
RBM38
HMGB1L1
BORIS
BORIS
CTCFL
CTCFL
CTCFL
CTCFL
CTCFL
CTCFL
CTCFL
CTCFL
CTCFL
CTCFL
CTCFL
CTCFL
CTCFL
CTCFL
PCK1
PMEPA1
ZBP1
ZBP1
ZBP1
CR612603
PMEPA1
PMEPA1
PMEPA1
C20orf85
PPP4R1L
PPP4R1L
PPP4R1L
RAB22A

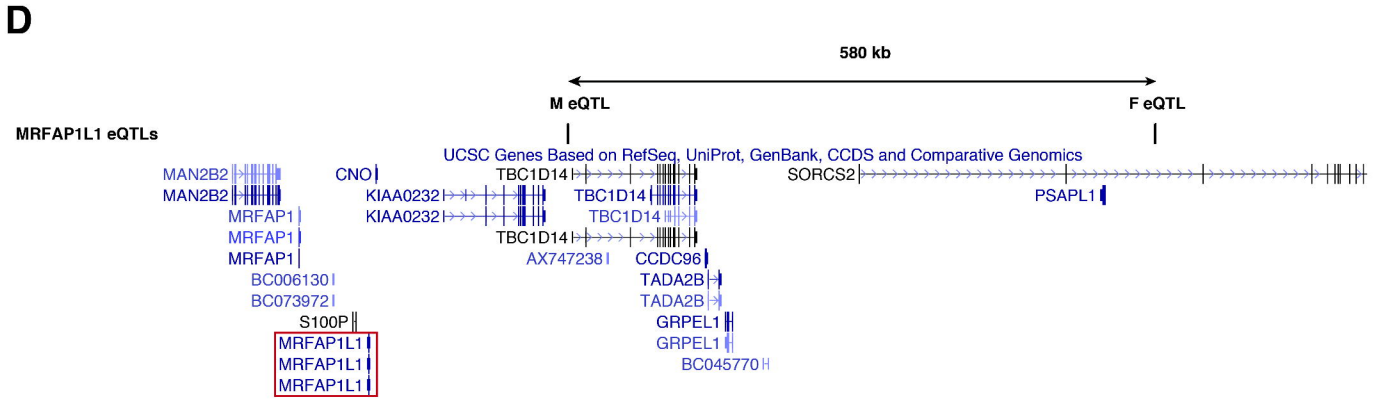
B


59 kb
M eQTL
F eQTL
PSAP eQTLs
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
CDH23
CDH23
CDH23
CDH23
CDH23
KIAA1774
PSAP
PSAP
CHST3
SPOCK2
SPOCK2
BC127952
UNQ269
ASCC1
ASCC1
ASCC1
ASCC1

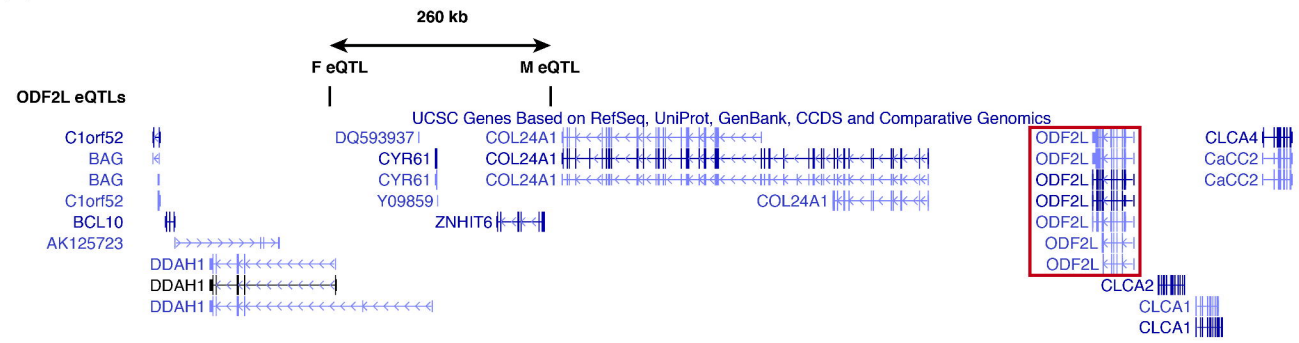
C


383 kb
M eQTL
F eQTL
CKLF eQTLs
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
TK2
TK2
TK2
TK2
CKLF
CKLF
CKLF
CKLF
CKLF
CKLFSF1
CKLFSF1
CKLFSF1
CKLFSF1
CKLFSF1
CKLFSF1
CMTM1
CMTM1
CMTM1
CMTM1
CMTM1
CMTM1
CMTM1
CMTM1
CMTM1
CMTM1
CMTM1
CMTM1
CKLFSF1
CMTM1
CKLFSF1
CKLFSF2
CMTM2
CMTM3
CMTM3
CMTM3
CMTM3
CKLFSF3
CMTM4
CMTM4
DYNC1LI2
AX747993
CCDC79
NAE1
NAE1
NAE1
NAE1
DKFZp434P241
CA7
CA7
AF086125
PDP2
PDP2
RRAD
RRAD
CDH16
CDH16
CDH16

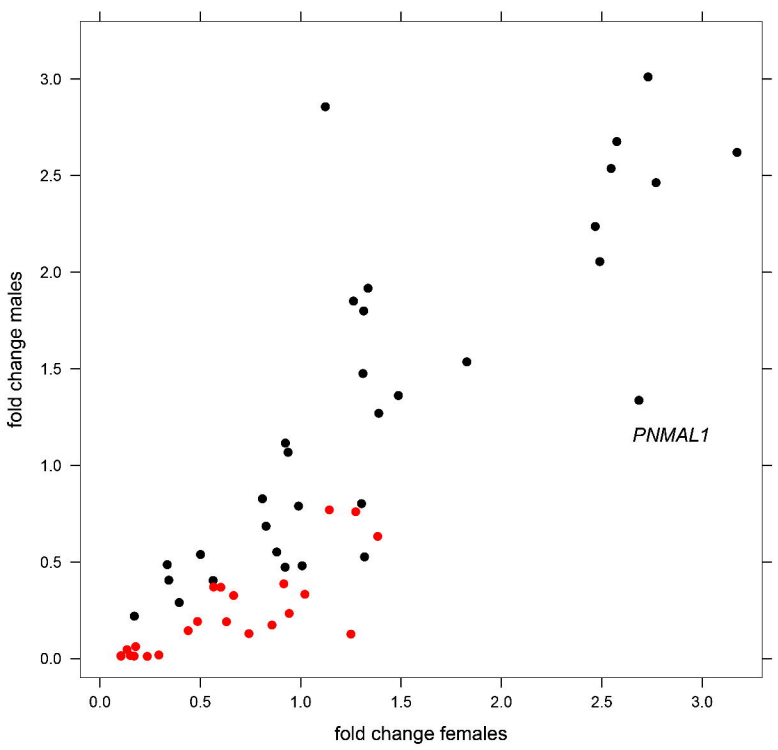
D


580 kb
M eQTL
F eQTL
MRFAP1L1 eQTLs
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
MAN2B2
MAN2B2
MRFAP1
MRFAP1
MRFAP1
BC006130
BC073972
S100P
MRFAP1L1
MRFAP1L1
MRFAP1L1
CNO
KIAA0232
KIAA0232
TBC1D14
TBC1D14
TBC1D14
TBC1D14
AX747238
CCDC96
TADA2B
TADA2B
GRPEL1
GRPEL1
BC045770
SORCS2
PSAPL1

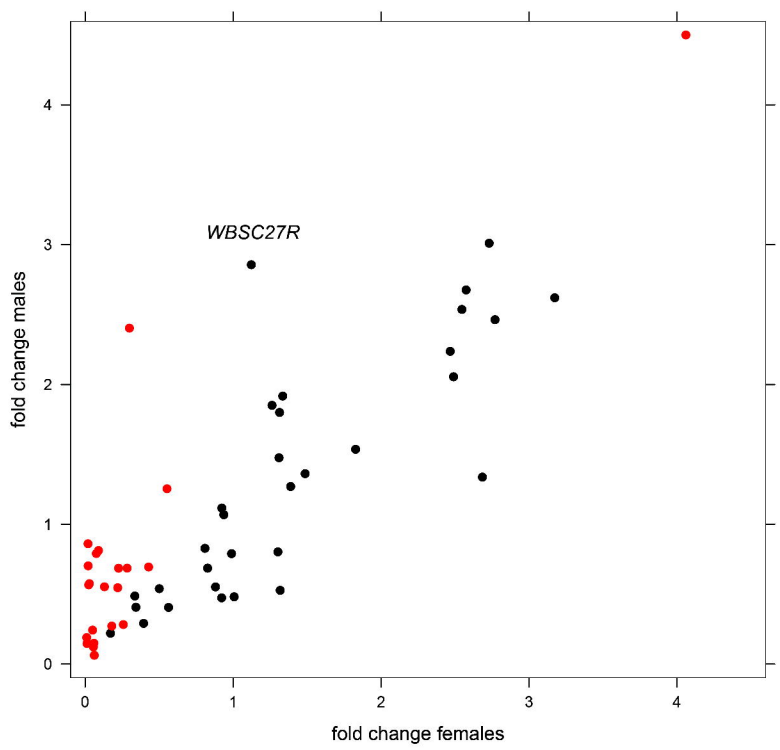
E


260 kb
F eQTL
M eQTL
ODF2L eQTLs
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
C1orf52
BAG
BAG
C1orf52
BCL10
AK125723
DDAH1
DDAH1
DDAH1
DQ593937
CYR61
CYR61
Y09859
ZNHIT6
COL24A1
COL24A1
COL24A1
COL24A1
ODF2L
ODF2L
ODF2L
ODF2L
ODF2L
ODF2L
ODF2L
CLCA2
CLCA1
CLCA1
CLCA4
CaCC2
CaCC2

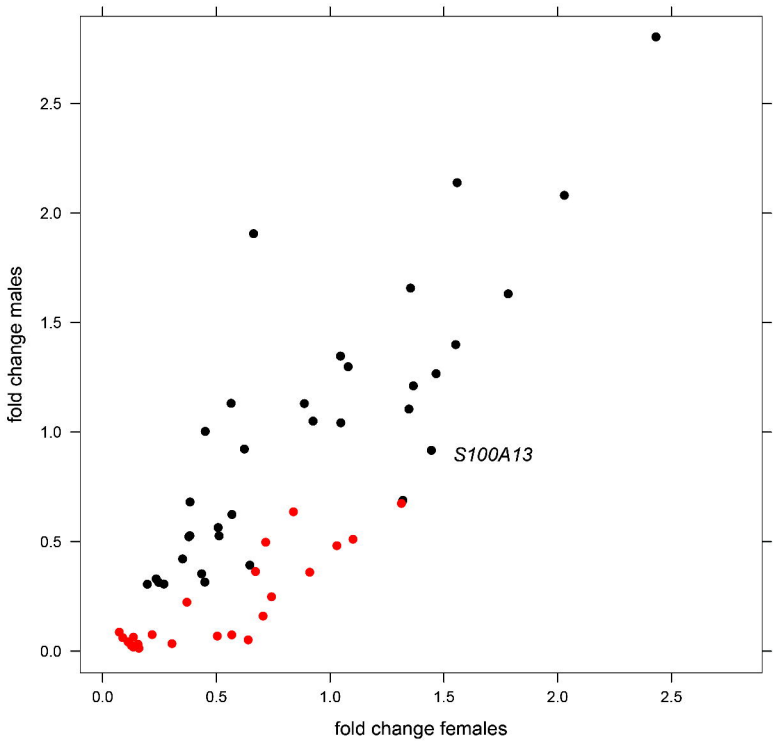
A


Fold change for CEU F eQTLs
fold change males
fold change females
PNMAL1
0.0
0.5
1.0
1.5
2.0
2.5
3.0

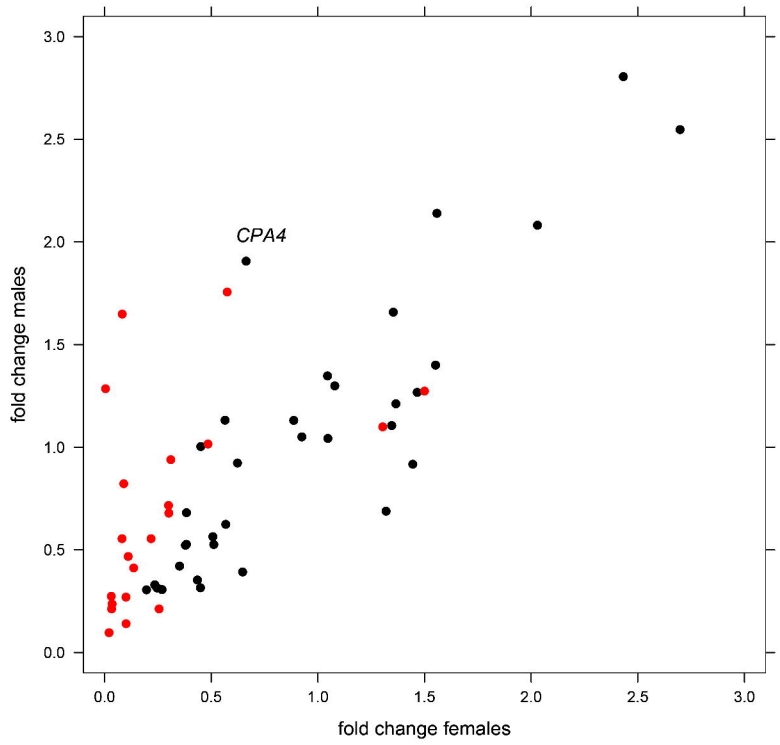
B


Fold change for CEU M eQTLs
fold change males
fold change females
WBSC27R
0
1
2
3
4

C


Fold change for YRI F eQTLs
fold change males
fold change females
S100A13
0.0
0.5
1.0
1.5
2.0
2.5

D


Fold change for YRI M eQTLs
fold change males
fold change females
CPA4
0.0
0.5
1.0
1.5
2.0
2.5
3.0