

Research

Impact of the X Chromosome and sex on regulatory variation

Kimberly R. Kukurba,^{1,2} Princy Parsana,³ Brunilda Balliu,¹ Kevin S. Smith,¹ Zachary Zappala,^{1,2} David A. Knowles,³ Marie-Julie Favé,⁴ Joe R. Davis,^{1,2} Xin Li,¹ Xiaowei Zhu,⁵ James B. Potash,⁶ Myrna M. Weissman,⁷ Jianxin Shi,⁸ Anshul Kundaje,^{2,9} Douglas F. Levinson,⁵ Philip Awadalla,⁴ Sara Mostafavi,¹⁰ Alexis Battle,^{3,11} and Stephen B. Montgomery^{1,2,9,11}

¹Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA; ²Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA; ³Department of Computer Science, Johns Hopkins University, Baltimore, Maryland 21218, USA; ⁴Sainte-Justine University Hospital Research Centre, Department of Pediatrics, University of Montreal, Montreal, Québec H3T 1J4, Canada; ⁵Department of Psychiatry, Stanford University School of Medicine, Stanford, California 94305, USA; ⁶Department of Psychiatry, University of Iowa Hospitals & Clinics, Iowa City, Iowa 52242, USA; ⁷Department of Psychiatry, Columbia University and New York State Psychiatric Institute, New York, New York 10032, USA; ⁸Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA; ⁹Department of Computer Science, Stanford University, Stanford, California 94305, USA; ¹⁰Department of Statistics, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

The X Chromosome, with its unique mode of inheritance, contributes to differences between the sexes at a molecular level, including sex-specific gene expression and sex-specific impact of genetic variation. Improving our understanding of these differences offers to elucidate the molecular mechanisms underlying sex-specific traits and diseases. However, to date, most studies have either ignored the X Chromosome or had insufficient power to test for the sex-specific impact of genetic variation. By analyzing whole blood transcriptomes of 922 individuals, we have conducted the first large-scale, genome-wide analysis of the impact of both sex and genetic variation on patterns of gene expression, including comparison between the X Chromosome and autosomes. We identified a depletion of expression quantitative trait loci (eQTL) on the X Chromosome, especially among genes under high selective constraint. In contrast, we discovered an enrichment of sex-specific regulatory variants on the X Chromosome. To resolve the molecular mechanisms underlying such effects, we generated chromatin accessibility data through ATAC-seq to connect sex-specific chromatin accessibility to sex-specific patterns of expression and regulatory variation. As sex-specific regulatory variants discovered in our study can inform sex differences in heritable disease prevalence, we integrated our data with genome-wide association study data for multiple immune traits identifying several traits with significant sex biases in genetic susceptibilities. Together, our study provides genome-wide insight into how genetic variation, the X Chromosome, and sex shape human gene regulation and disease.

[Supplemental material is available for this article.]

Many human phenotypes are sexually dimorphic. In addition to males and females having recognizable anatomic and morphological differences, accumulating evidence suggests that they exhibit differences in the prevalence, severity, and age of complex diseases. Classic examples of sex-biased diseases include autoimmune disorders (Whitacre et al. 1999; Whitacre 2001), cardiovascular disease (Lerner and Kannel 1986; Mendelsohn and Karas 2005), cancer susceptibility (Cohn et al. 1996; Naugler et al. 2007), and psychiatric disorders (Breslau et al. 1997; Pigott 1999; Hankin and Abramson 2001). While genetic factors may underlie observed differences, determining the genetic contribution to sexual dimorphism has generally lagged behind the hormonal contribution due to challenges in both study design and statistical power

(Luan et al. 2001; Patsopoulos et al. 2007; Ober et al. 2008). Despite these limitations, several studies have discovered genotype-by-sex interaction effects in human phenotypes, such as anthropometric traits (Heid et al. 2010; Randall et al. 2013), bone mineral density (Liu et al. 2012a), complex diseases (Liu et al. 2012b; Myers et al. 2014), and intermediate cellular phenotypes such as gene expression (Dimas et al. 2012; Yao et al. 2014). To explain the etiology of these sexually dimorphic traits, several mechanisms have been proposed, including those arising due to the X Chromosome (Dobyns et al. 2004; Ober et al. 2008).

Although genome-wide association studies (GWAS) have uncovered numerous loci associated with complex phenotypes on the autosomes, the X Chromosome is significantly underrepresented in such work. Indeed, only one-third of GWAS include the X Chromosome, largely due to specialized analytical methods

¹¹These authors contributed equally to this work.

Corresponding authors: ajbattle@cs.jhu.edu, smontgom@stanford.edu

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.197897.115>. Freely available online through the *Genome Research* Open Access option.

© 2016 Kukurba et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

required for processing and interpreting genetic data on this chromosome (Wise et al. 2013). Furthermore, many large-scale functional genomic studies investigating the effect of genetic variants also exclude the X Chromosome (Dimas et al. 2009; Montgomery et al. 2010; Pickrell et al. 2010; Lappalainen et al. 2013; Battle et al. 2014; The GTEx Consortium 2015). Motivated by the underutilization of the X Chromosome, recent studies have characterized the role of the X Chromosome in the heritability of human phenotypes (Chang et al. 2014; Tukiainen et al. 2014). However, no studies to date have systematically investigated the contribution of the X Chromosome in the context of both regulatory variation and its interaction with sex.

By leveraging a recent, large genetic study of gene expression (Battle et al. 2014), we comprehensively survey the impact of sex and genetic variation on the X Chromosome on human gene expression to improve our understanding of the genetic and molecular basis of sex-biased disease risk. Our study overcomes several limitations of previous eQTL and sex-specific eQTL studies which have either ignored the X Chromosome, conducted analyses in cell lines which may inaccurately reflect in vivo sex differences (Dimas et al. 2012), had insufficient power to detect sex-specific eQTLs (Trabzuni et al. 2013), or focused on only specific variants for sex-specific eQTL analysis (Castagne et al. 2011; Yao et al. 2014). We extend these studies to describe the characteristics of eQTL on the X Chromosome versus the autosomes, address the relationship between sex-specific gene expression and chromatin accessibility, and identify the contribution of multiple eQTLs to informing sex-biased disease risks. Together, our study provides new insight into the genome-wide regulatory mechanisms of sexual dimorphism and the importance of including the X Chromosome and sex in the design, analysis, and interpretation of genetic studies.

Results

To study sex-specific genetic variation in humans, we obtained gene expression data for the Depression Genes and Networks

(DGN) cohort comprised of 922 individuals of European ancestry across the United States (Battle et al. 2014; Mostafavi et al. 2014). Gene and isoform expression were quantified from whole blood 51-bp single-end RNA sequencing data (Methods). Each individual was also genotyped for 737,187 single nucleotide polymorphisms (SNPs) located on the autosomes and X Chromosome on the Illumina HumanOmni1-Quad BeadChip and then imputed using the 1000 Genomes Phase 1 reference panel (Methods).

Sex-specific expression variation across the genome

Sex-specific patterns of gene expression have been well studied in the literature across many organisms (Yang et al. 2006; Ober et al. 2008; Reinius et al. 2008; Trabzuni et al. 2013). However, beyond differences in mean gene expression levels, differences in expression variation may be attributed to interaction effects of genomic variants and both environmental and biological variables, such as sex (Idaghdour and Awadalla 2012). We investigated both sex-specific patterns of gene expression levels and variance across the genome. After correcting for known technical covariates in the expression data (see Methods), we tested if gene expression variance in females is equivalent to that in males. We limited our analysis to genes expressed in both sexes and matched the number of males and females tested via random subsampling to ensure that sample size did not influence measured variance (Methods). We identified 924 genes with sex-specific expression variance (5% FDR). Across the genome, we observed that a higher proportion of genes with sex-specific variance are on the X Chromosome compared to autosomes (9.8% versus 6.4%; P -value = 1.9×10^{-3} , Fisher's exact test) (Fig. 1A). This pattern remains the same for different normalization methods of the gene expression data and variance tests (P -value of Fisher's exact test equals 4.04×10^{-3} and 1.57×10^{-4} when the difference in variance is assessed using the F-test or Brown-Forsythe test, respectively) (Supplemental Figs. S1, S2), after correcting for gender effects on the mean (P -value = 1.57×10^{-4}) and replicates in the ImmVar cohort providing cell-type-specific data (Ye et al. 2014) (P -value $< 5.4 \times 10^{-3}$) (Supplemental Fig. S3), indicating that this is unlikely to be an

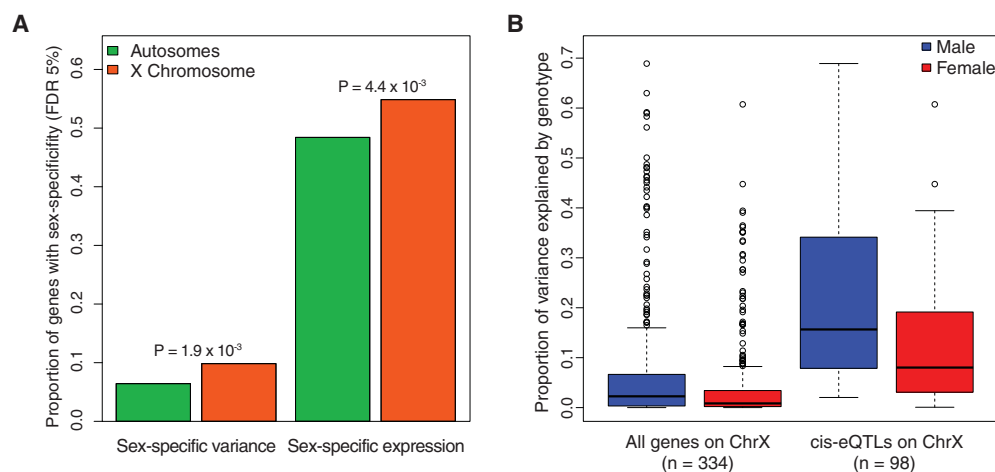


Figure 1. Differential expression variance within the sexes. (A) Comparison of genes with significant sex-specific expression variance (FDR 5%) and sex-specific expression (FDR 5%) on the autosomes and the X Chromosome in the DGN. To test for differences in mean expression and variance, the number of males and females were matched ($n = 274$). One-sided Fisher's exact test P -values above bars indicate significance of higher expression or variance on the X Chromosome relative to autosomes. (B) Proportion of variance explained (PVE) by genotype on the X Chromosome in males and females. To test for the PVE, the number of males and females were matched ($n = 274$). We tested all genes on the X Chromosome and genes with a *cis*-eQTL (Bonferroni adjusted P -value < 0.05).

artifact of differences in cell-type proportions between sexes. As with variance, we observe more genes on the X Chromosome exhibiting sex-specific gene expression compared to autosomes (54.8% versus 48.4%; P -value = 4.4×10^{-3} , Fisher's exact test) (Fig. 1A), an observation predicted by theory (Rice 1984) and seen in other species (Ranz et al. 2003; Yang et al. 2006). To help account for any potential bias in the relationship of variance and mean, we tested whether genes with sex-specific expression variance are genes with a significant difference in their mean expression between the sexes (FDR < 0.05, Welch's two-sample t -test) and detected no significant enrichment (P -value = 0.39, χ^2 test).

To understand why the X Chromosome exhibits unique patterns of expression variance relative to the autosomes, we tested if the hemizyosity of the X Chromosome contributes to greater expression variance in males. To test this hypothesis, we estimated the proportion of expression variance explained by common genetic variants (MAF > 0.05) in males and females for each gene on the X Chromosome. We found that the mean estimated percentage of variance explained by genotype is 1.7-fold higher for males than females (2.3% versus 0.9%; P -value = 6.14×10^{-5} , Wilcoxon rank-sum test) (Fig. 1B). This difference remained when restricting only to genes with a *cis*-eQTL (Bonferroni-adjusted P -value < 0.01; 21.0% versus 12.5%; P -value = 6.4×10^{-5} , Wilcoxon rank-sum test) (Fig. 1B). In contrast, if we evaluate the proportion of variance explained by genotype on an autosomal chromosome, we observe no significant difference between males and females (P -value = 0.557 and 0.686, Wilcoxon rank-sum test) (Supplemental Fig. S4). We then asked if the autosomal genes with sex-specific expression variance are enriched in specific biological processes and found that these genes are enriched in cell death and regulation of apoptosis (Supplemental Table S1). These observations suggest that divergent regulation of cell death occurs between the sexes, supporting previous studies that have found sex-specific alterations in apoptosis within immune and neural cell populations (Molloy et al. 2003; Lang and McCullough 2008).

Identification and characterization of X-Chromosome *cis*-eQTLs

We tested for *cis*-eQTL on the X Chromosome and autosomes using linear regression. At a genome-wide level FDR of 5%, we detect eQTLs on 74.8% of autosomal genes and 43.7% of X Chromosome genes. For a range of FDR thresholds, we find a depletion of eQTLs on the X Chromosome relative to autosomes (two-sided χ^2 test, P -value = 9.4×10^{-37} at FDR 1%) (Fig. 2A). To determine if either male hemizyosity of the X Chromosome or female X-inactivation

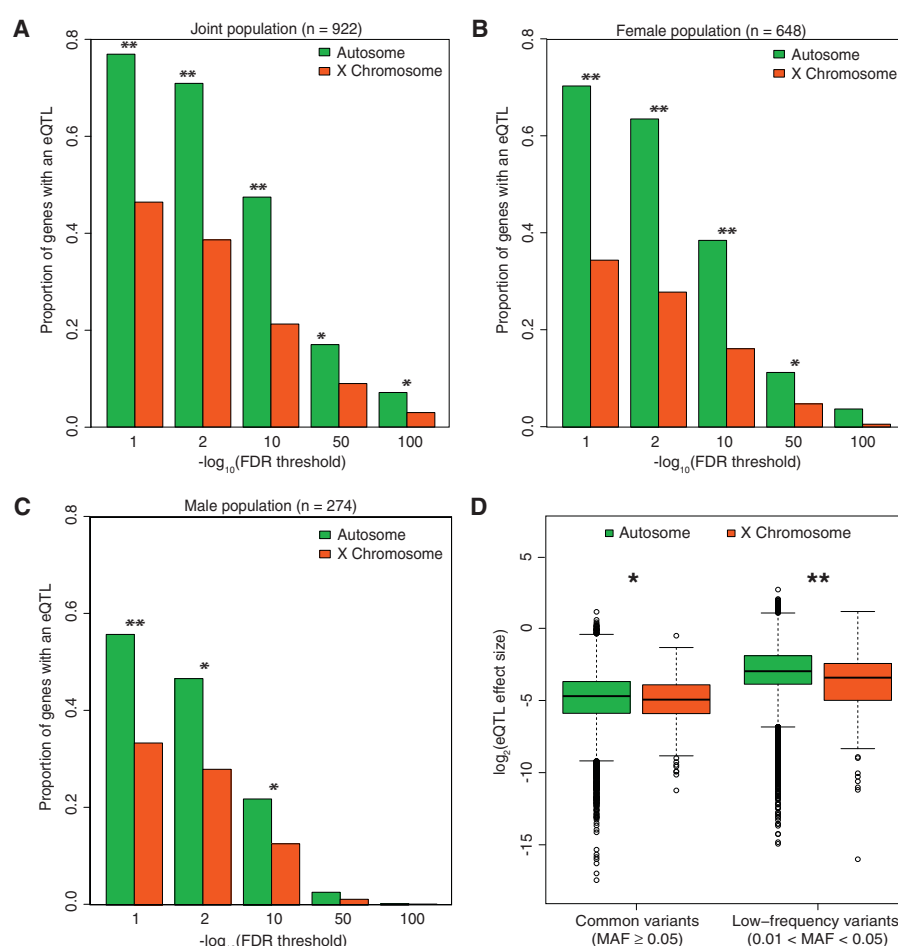


Figure 2. Characterization of eQTLs on the X Chromosome. (A) Proportion of genes with an eQTL at different FDR thresholds discovered in the joint ($n = 922$), (B) female ($n = 648$), and (C) male ($n = 274$) populations. (***) P -value < 1×10^{-15} , (*) P -value < 0.05, Bonferroni adjusted χ^2 test. (D) Comparison of eQTL effect size between autosomes and the X Chromosome for common variants (MAF ≥ 0.05) and low-frequency variants ($0.01 < \text{MAF} < 0.05$). The difference in effect size is statistically significant for both common ([*] P -value = 0.0382, two-sided Wilcoxon rank-sum test) and low-frequency variants ([**] P -value = 1.21×10^{-7}).

influenced this observed depletion, we also detected eQTLs in just the male and female populations alone. In both male and female populations, we observed a depletion of eQTLs on the X Chromosome across multiple FDR thresholds (P -value < 10^{-15} , χ^2 test) (Fig. 2B,C), compared to autosomal eQTL rates. The depletion of X Chromosome eQTLs further replicated in multiple individual cell types, including monocytes from the ImmVar cohort and lymphoblastoid cells from the Geuvadis cohort (Supplemental Table S2). To ensure that this depletion of eQTLs on the X Chromosome was not due to lower power in the X Chromosome compared to the autosomes, e.g., lower minor allele frequency (MAF) or mean gene expression, we compare the distributions of MAF and mean expression for all SNPs and genes that were considered in the eQTL analyses. We observe only a slightly smaller proportion of SNPs with $0.05 < \text{MAF} < 0.15$ and genes with lower mean expression in the X Chromosome compared to the autosomes, suggesting that we are at least as powered to detect eQTLs in the X Chromosome as we are in the autosomes. We also compared the MAF, F_{ST} , and expression levels of detected eQTLs and observed that X Chromosome eQTLs have lower MAFs (0.19

versus 0.23; t -test P -value = 3.8×10^{-7}), higher F_{ST} (0.0847 versus 0.0287; t -test P -value = 0.0433), and slightly lower mean expression levels (t -test P -value = 2.0×10^{-3}) relative to autosomal eQTLs (Supplemental Fig. S5A–C). Based on these observations, we can speculate that the X Chromosome's unique evolution and selective pressures, which are known to contribute to reduced genetic diversity compared to autosomes (Keinan et al. 2009; Gottipati et al. 2011) and affect dosage effects on expression levels (Gupta et al. 2006; Kharchenko et al. 2011), may contribute to the depletion of eQTLs.

eQTL effect sizes on the X Chromosome

To understand the properties of X Chromosome eQTLs, we compared effect sizes to those of autosomal eQTLs within females. For eQTLs detected from common variants ($MAF \geq 0.05$), we observe that the effect size on the X Chromosome is 1.17-fold lower relative to autosomes (P -value = 3.8×10^{-2} , Wilcoxon rank-sum test) (Fig. 2D). For eQTLs detected from low-frequency variants ($0.01 < MAF < 0.05$), we observe that effect size on the X Chromosome is 1.34-fold lower relative to autosomes (P -value = 1.2×10^{-7} , Wilcoxon rank-sum test), suggesting increased purifying selection for large eQTL effects on the X Chromosome versus the autosomes.

Previous studies have found that hemizyosity of the X Chromosome in males may cause unusual patterns of evolution, including lower nucleotide diversity and more efficient removal of deleterious alleles on the X Chromosome compared to the autosomes (Andolfatto 2001; Schaffner 2004; Lu and Wu 2005; Vicoso and Charlesworth 2006). To test the relationship of eQTL effect size and purifying selection on the X Chromosome, we obtained the ratios of nonsynonymous to synonymous substitutions (d_N/d_S) of human-rhesus genes as an indicator of selective constraint, quantile binned the d_N/d_S ratios, and compared the eQTL effect sizes on the X Chromosome to autosomes within each bin. For genes under the greatest degree of constraint (lower 20% quantile bin, d_N/d_S 0–0.05), we observe significantly lower effect sizes on the X Chromosome (P -value = 3.6×10^{-4} , Wilcoxon rank-sum test) (Supplemental Fig. S5D). For genes under less constraint (higher d_N/d_S ratios), we observe similar but less significant patterns (P -value < 5.0×10^{-2}). This evidence suggests that X Chromosome genes experience increased purifying selection on regulatory variation relative to the autosomes.

Identification and characterization of sex-interacting *cis*-eQTL

To identify sex-interacting *cis*-eQTLs, we used a linear model with a genotype-sex interaction term. Genotype-by-sex interactions occur when the effect of genotype on expression differs between males and females. For example, the genetic effects on expression may be present in only one sex, may have different magnitudes of effect, or may even have opposing directions of effect in the two sexes (Supplemental Fig. S6). We observe an enrichment of sex-specific eQTLs on the X Chromosome compared to the autosomes (Wilcoxon rank-sum test, P -value = 8.2×10^{-4} and 1.9×10^{-5} for the top 50 and 500 associations, respectively) (Fig. 3A). After adjusting for the number of SNPs tested per gene using Bonferroni and subsequently identifying sex-interacting eQTLs using gene-level significance at 5% FDR, we discover six eQTLs (four autosomal and two X Chromosome eQTLs) with significant sex interactions (Table 1). If we restrict our tests to variants previously detected as *cis*-eQTLs (Bonferroni adjusted P -value < 0.1), we improve power but the total number of discoveries is unaffected

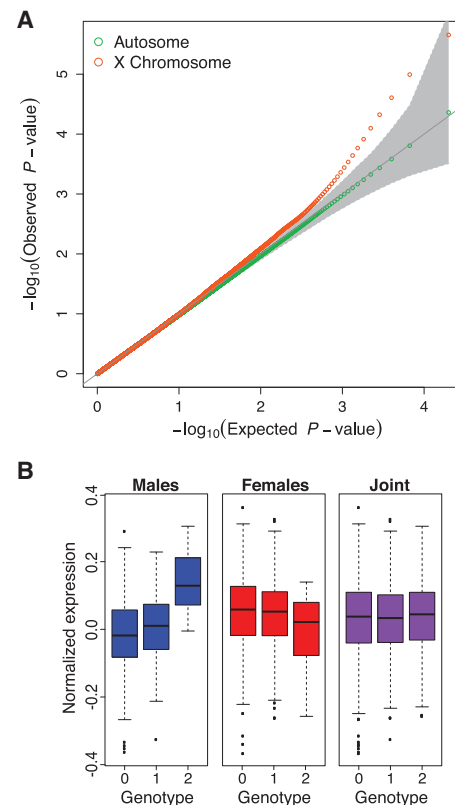


Figure 3. Discovery of sex-interacting eQTLs. (A) Quantile-quantile (QQ) plot describing the sex-interacting eQTL association P -values for SNPs tested within 1 Mb of genes on the X Chromosome (orange) and the autosomes (green) with 95% confidence interval (gray). (B) Sex-interacting eQTL (q -value = 0.0198) for *DNAAF1* (dynein, axonemal, heavy chain 1), a protein-coding gene involved in microtubule motor activity, ATPase activity, and sperm motility.

(Supplemental Fig. S7). Notably, two of the six sex-interacting eQTLs discovered in our DGN cohort were also identified as sex-interacting eQTLs in other cohorts (Supplemental Table S3). Specifically, a genotype-sex interaction was previously identified for *NOD2* at rs9302752 (nominal P -value = 9.03×10^{-13}) in the Framingham Heart Study (Yao et al. 2014); rs9302752 is in high linkage disequilibrium (LD) ($R^2 = 0.93$) with the top sex-interacting eQTL variant we identified for *NOD2* in the DGN cohort (nominal P -value = 5.1×10^{-10}). We also reanalyzed the genotype and expression data from the ImmVar (Ye et al. 2014) and CARTaGENE cohorts (Awadalla et al. 2013; Hussin et al. 2015) to test if our sex-interacting eQTLs replicated in other cohorts (Supplemental Table S3); we observed that the sex-interacting eQTL at Chr 11: 62735958:D-*BSCL2* replicated in the CARTaGENE cohort (P -value = 2.89×10^{-3}), and the remaining variants were not testable due to expression levels or lack of replication.

We additionally used a binomial generalized linear mixed model (Knowles et al. 2015) to test for a sex-interacting allele-specific expression QTL (aseQTL). Specifically, we tested if alleles at heterozygous loci show allele-specific expression (ASE) and if the magnitude of ASE effects differs in males and females. Two of our six significant sex-interacting eQTL genes had heterozygous loci with sufficient read depth to measure allelic imbalance. Of these two genes, we observed that *NOD2* has a significant sex-interacting aseQTL (P -value = 2.5×10^{-3}) (Supplemental Fig. S8A)

Table 1. Discovery of sex-interacting eQTLs (FDR 5%)

Chr	Variant	Gene	eQTL discovery in populations								
			Genotype-sex interaction term			Males		Females		Joint	
			β	<i>P</i> -value	FDR	β	<i>P</i> -value	β	<i>P</i> -value	β	<i>P</i> -value
5	rs66668391	<i>WDR36</i>	-0.17	7.60×10^{-11}	1.80×10^{-4}	0.21	4.90×10^{-15}	0.07	1.20×10^{-4}	0.08	4.80×10^{-13}
16	rs4785448	<i>NOD2</i>	-0.15	5.10×10^{-10}	1.20×10^{-3}	-0.62	$<10^{-16}$	-0.48	$<10^{-16}$	-0.59	$<10^{-16}$
11	Chr 11: 62735958:D	<i>BSCL2</i>	-0.24	5.50×10^{-10}	1.70×10^{-3}	0.41	$<10^{-16}$	0.28	2.90×10^{-14}	0.24	$<10^{-16}$
X	rs147067421	<i>MAP7D3</i>	0.19	7.70×10^{-9}	2.50×10^{-3}	-0.18	2.50×10^{-7}	-0.01	7.00×10^{-1}	-0.08	2.00×10^{-5}
X	Chr X: 119134201:D	<i>RHOXF1</i>	0.63	1.50×10^{-8}	6.10×10^{-3}	-0.95	8.10×10^{-16}	-0.64	3.90×10^{-9}	-0.6	$<10^{-16}$
3	rs9846315	<i>DNAH1</i>	-0.08	1.10×10^{-8}	2.00×10^{-2}	0.06	1.20×10^{-6}	0	8.20×10^{-1}	0	6.50×10^{-1}

and *DNAH1* has a possible sex-interacting eQTL with small effect but falls slightly below the level of significance (P -value = 1.0×10^{-1}) (Supplemental Fig. S8B), likely due in part to higher variance of allelic ratio observed at this locus. All variants tested for sex interactions in the DGN cohort can be browsed at <http://montgomerylab.stanford.edu/resources.html>.

Of the six genes with sex-interacting eQTLs, *DNAH1* on Chromosome 3 was distinct because this SNP-gene pair was not a significant eQTL in the female or joint population analysis (Fig. 3B). Although this gene exhibits similar expression between the sexes in whole blood (nominal negative binomial P -value = 0.81), it shows the highest expression in the testis compared to the remaining 46 tissues tested in the GTEx Project. We investigated this gene further and found that it is a force-producing protein with ATPase activity involved in sperm motility and flagellar assembly (UniProt ID Q9P2D7). The sex-specific function of this gene prompted us to test if the sex-interacting eQTLs were enriched in sex-specific biological processes. To evaluate this hypothesis, we tested the top 100 sex-interacting eQTLs for enrichment in genes involved in sex differentiation (GO:0007548) and discovered a modest enrichment (odds ratio = 6.6 [1.17–18.2], P -value = 4.3×10^{-3} , Fisher's exact test). When testing all biological processes, no specific process passed multiple testing corrections; however, the top GO terms were linked to sex-specific biological processes (e.g., reproductive structure development and response to estrogen stimulus) (Supplemental Table S4).

To determine if sex-interacting eQTL discovery is driven by differences in gene expression between males and females, we tested for enrichment sex-interacting eQTLs within differentially expressed genes (FDR 10%). We observed no significant enrichment of genes with sex-specific expression for the top 100 sex-interacting eQTLs (odds ratio = 0.4 [0.05–1.58], P -value = 3.3×10^{-1} , Fisher's exact test) compared to the background of genes expressed in whole blood. Concordant with a previous study (Dimas et al. 2012), this demonstrates that sex-interacting eQTLs likely do not arise as a consequence of expression differences between the sexes and may result from other factors that differ in a sex-specific matter, such as transcription factor activity, hormone receptors, and chromatin accessibility.

Sex-specific chromatin accessibility

To identify the molecular mechanisms of sex-specific gene regulation, we generated and investigated differences in chromatin accessibility between males and females. We measured chromatin accessibility of peripheral blood mononuclear cells (PBMCs) from 20 individuals matched for age, ethnicity, and sex using the assay for transposase-accessible chromatin followed by sequencing

(ATAC-seq) (Buenrostro et al. 2013). Using a negative binomial model, we identified 577 (0.69%) sex-specific chromatin accessibility regions at FDR 10% (see Methods).

Considering the unique heterochromatic state of the X Chromosome in females, we asked if the distribution of sex-specific open chromatin regions differed between the autosomes and the X Chromosome (Fig. 4A). We observed an enrichment of sex-specific open chromatin regions on the X Chromosome (Wilcoxon rank-sum test, P -value = 3.5×10^{-10} and 1.4×10^{-4} for the top 50 and 500 associations, respectively). Following this observation, we hypothesized that regions on the X Chromosome are more likely to have greater chromatin accessibility in females due to genes escaping X-inactivation. Indeed, we found that X Chromosome regions are more likely to have higher chromatin accessibility in females than males compared to autosomal regions (odds ratio = 2.44 [1.99–3.01], P -value $< 2.2 \times 10^{-16}$, Fisher's exact test) (Supplemental Fig. S9). To further interpret this, we then tested if the top sex-specific peaks on the X Chromosome are more likely to be genes escaping X-inactivation (Carrel and Willard 2005; Park et al. 2010). We observed that the top 10% of sex-specific peaks on the X Chromosome are enriched for genes escaping X-inactivation compared to genes in sex-shared peaks (44.6% versus 6.5%; odds ratio = 9.60 [2.70–34.49], P -value = 1.6×10^{-4} , Fisher's exact test), indicating that a large number of the X Chromosome sex-specific peaks inform regions of escape from X-inactivation.

Integration of sex-specific chromatin accessibility with expression

One intuitive mechanism for sex-specific gene expression and sex-interacting eQTLs is sex-specific chromatin accessibility. Therefore, we sought to identify if differential open chromatin between the sexes was associated with sex-specific expression and genotype-sex interactions. First, we tested if genes with sex-specific expression (FDR 5%) are enriched in regions with sex-specific chromatin accessibility. We observed that genes with sex-specific chromatin accessibility 40 kb upstream of the gene TSS were associated with sex-specific expression (odds ratio = 1.45 [1.24–1.69], P -value = 4.7×10^{-6} , Fisher's exact test) (Fig. 4B). We then investigated the directionality of enrichment and observed opposing directions of effect (i.e., genes with elevated expression in a particular sex were more likely to have depressed chromatin signal in the same sex and vice versa) (odds ratio = 2.38 [1.13– ∞], Fisher's exact test P -value = 2.2×10^{-2}). We speculate that genes with elevated expression may exhibit a depressed chromatin signal due to transcription factor occupancy or other chromatin modifications, as observed by others (Gilbert et al. 2004; Schneider and Grosschedl 2007; Natarajan et al. 2012).

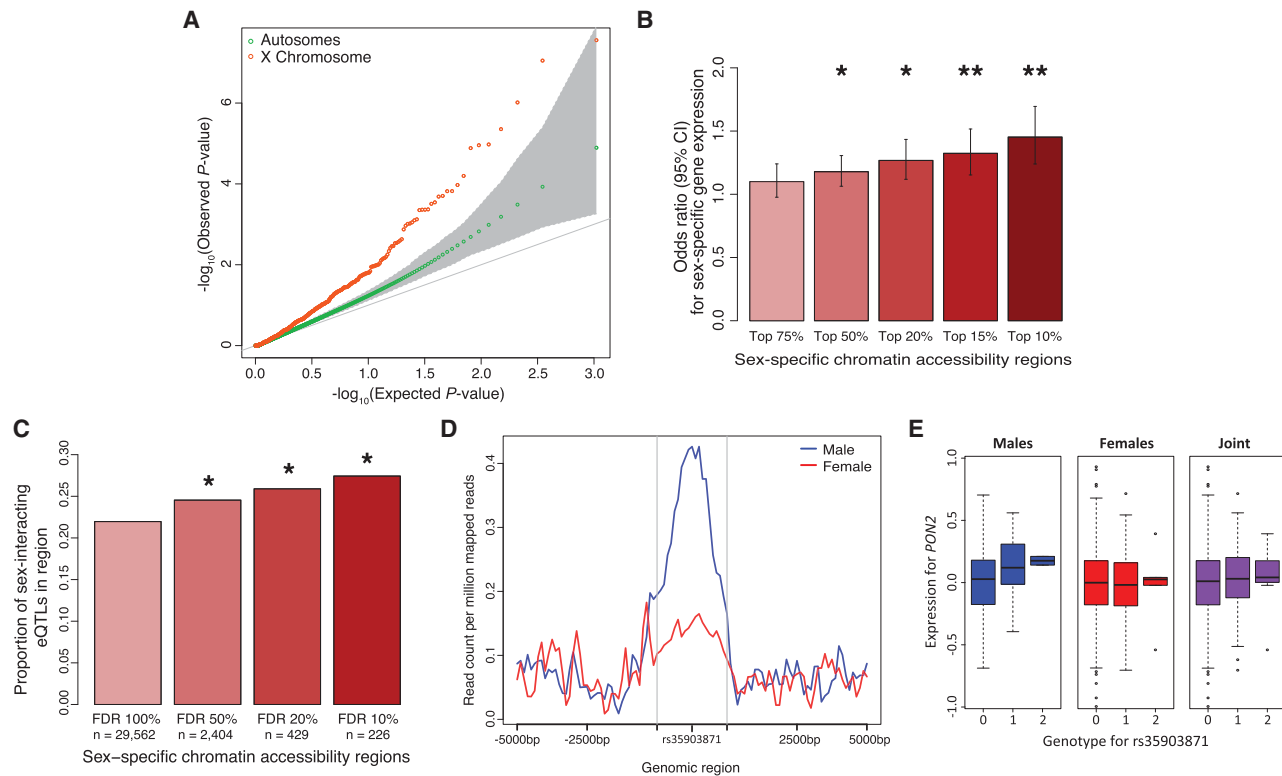


Figure 4. Discovery of sex-specific chromatin accessibility regions. (A) QQ plot for tests of differential chromatin accessibility between the sexes. 95% genome-wide confidence interval in gray. (B) Enrichment of genes with differential expression between the sexes (FDR 5%) with differential chromatin accessibility (varying thresholds) 40 kb upstream. (***) P -value $< 10^{-5}$, (*) P -value $< 10^{-3}$, Fisher's exact test. (C) Proportion of sex-interacting eQTL genes (P -value < 0.05) in differential chromatin accessibility regions (varying thresholds). (*) P -value $< 5.0 \times 10^{-2}$, Fisher's exact test. (D) Chromatin accessibility peak located at Chr 7: 95,063,722–95,064,222 and 5000 bp upstream and downstream. This region has differential chromatin accessibility between males and females (nominal P -value = 4.1×10^{-4} , Q -value = 7.5×10^{-2}). (E) Sex-interacting eQTL for *PON2* and rs35903871 located at Chr 7: 95,063,972 (nominal P -value = 8.0×10^{-3}).

Next, we investigated if sex-specific chromatin accessibility regions are enriched in sex-interacting eQTL genes by testing for genotype-sex interactions for variants within 1 Mb of the gene. We observe that genes with a sex-interacting eQTL (nominal P -value < 0.1) are likely to have differential chromatin accessibility between the sexes (odds ratio = 1.36 [0.99–1.81], P -value = 5.1×10^{-2} , Fisher's exact test) (Fig. 4C). To ensure that we did not observe this enrichment by chance, we permuted the sexes for eQTL testing and tested sex-interacting eQTLs in sex-specific chromatin regions and observed no enrichment (Supplemental Fig. S10). We also investigated the directionality of enrichment and observed opposing directions of effect (i.e., eQTLs with elevated effect sizes in a particular sex were more slightly likely to have depressed chromatin signal in the same sex and vice versa) (odds ratio = 1.75 [1.19– ∞], Fisher's exact test P -value = 5.8×10^{-2}). As with sex-specific expression, eQTLs with higher effect in a particular sex may have decreased chromatin signal due to the occupancy of transcription factors at these regions (Gilbert et al. 2004; Schneider and Grosschedl 2007; Natarajan et al. 2012). In addition, we asked if sex-specific chromatin accessibility regions were enriched in sex-interacting eQTL variants. We tested the genotype-sex interaction term for the variant closest to the region mid-point and the closest five genes and observed weak but not statistically significant enrichment (Supplemental Fig. S11). One example of a sex-interacting eQTL in a sex-specific chromatin accessibility region is illustrated in Figure 4D, in which a SNP

(rs35903871) is located in a region with greater chromatin accessibility in males (P -value = 4.1×10^{-4} , FDR 7.5%) and exhibits genotype-sex interaction effects on *PON2* expression ($\beta = 0.13$; P -value = 8.2×10^{-3}). If tested separately in each sex, this gene-SNP pair is an eQTL in males (P -value = 5.5×10^{-3}) but not females (P -value = 5.0×10^{-1}). Interestingly, *PON2* does not exhibit sex-specific gene expression (P -value = 6.3×10^{-1} , FDR 89.2%) but has been associated with sex-specific effects in oxidative stress responses (Cheng and Klaassen 2012; Giordano et al. 2013; Polonikov et al. 2014).

Integration of sex-specific effects with human disease

In humans, sexual dimorphism is observed in the incidence rates and severity of many common diseases, including cardiovascular, immune, and neurological diseases. For example, a recent study (Myers et al. 2014) identified six sex-specific asthma risk loci (P -value $< 1 \times 10^{-6}$), three of which exhibit genotype-sex interactions on expression in our study (nominal P -value < 0.05). Unfortunately, previous limitations on statistical power and study design have provided challenges for identifying significant genotype-sex effects in many disease association studies (Brookes et al. 2004; Wang et al. 2008). Given the role of regulatory variation in determining disease risk (Nica et al. 2010; Nicolae et al. 2010), we asked if genetic variants identified through genome-wide association studies have cumulatively different eQTL effects in males and females. To test this, we obtained trait-associated

variants from ImmunoBase and examined the effect size of each GWAS variant on expression in males and females separately. We found that for the majority of autoimmune diseases, the disease-associated loci exhibit a bias in effect size in one sex (Fig. 5; see Methods). For example, a significant proportion of genetic variants associated with multiple sclerosis, a disease with well-documented sex-specific disparity in prevalence and severity (Whitacre et al. 1999; Whitacre 2001), exhibit a bias toward females (P -value = 5.6×10^{-7} , binomial exact test). We performed an identical analysis using data available from the NHGRI-EBI GWAS catalog (Welter et al. 2014) and observe similar patterns (Supplemental Fig. S12). These observations indicate that disease-associated variants have different cumulative effects on gene expression in males and females underlying potential sex-biases in disease prevalence and severity.

Discussion

In this study, we comprehensively evaluated the effect of the X Chromosome and sex on regulatory variation. First, we demonstrated that genes on the X Chromosome are more likely to have sex-specific expression compared to genes on the autosomes reflecting a study in mice (Yang et al. 2006) and a recent report in the human brain (Trabzuni et al. 2013). We observed similar differences for sex-specific gene expression variance between the X and autosomes, highlighting further interactions with sex. As expected, we identified that a portion of this difference was due to the hemizyosity of the X Chromosome in males where the exposure of individual alleles is not balanced by random X-inactivation. However, as with gene expression, sex-specific gene expression variance was exhibited genome-wide and we observed genes with higher sex-specific expression variance were more likely to be involved in apoptosis and regulation of cell death, consistent with previous reports highlighting sex differences in regulation of cell death (Du et al. 2004; Li et al. 2005; Jog and Caricchio 2013).

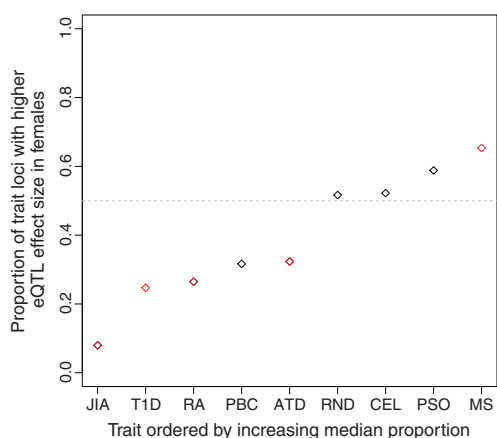


Figure 5. Disease associated variants with sex-biased eQTLs. Proportion of independent (LD-pruned) variants with higher eQTL effect sizes in females for GWAS variants of traits in ImmunoBase. Each trait variant tested had to be a significant eQTL variant (nominal P -value < 0.001). Red data points indicate traits with eQTL effect sizes that are significantly different between sexes (Bonferroni adjusted P -value < 0.05). (ATD) Autoimmune thyroid disease, (CEL) celiac disease, (JIA) juvenile idiopathic arthritis, (MS) multiple sclerosis, (PBC) primary biliary cirrhosis, (PSO) psoriasis, (RA) rheumatoid arthritis, (RND) random eQTL variants, (T1D) type 1 diabetes.

When considering the impact of genetic variation on gene expression for the X Chromosome versus autosomes, we observed a depletion of eQTLs and large eQTL effect sizes on the X Chromosome, supporting previous observations of more efficient purifying selection on the X Chromosome relative to the autosomes (Vicoso and Charlesworth 2006; Singh and Petrov 2007; Gottipati et al. 2011). In contrast, we tested the effect of genotype-by-sex interactions across the transcriptome and observed an enrichment of sex-interacting eQTLs on the X Chromosome. This increase reflects the increased proportion of sex-specific expression effects on the X Chromosome but also suggests the possibility that sex-specific eQTLs on the X Chromosome may not experience the same selective pressure as other eQTLs on the X Chromosome. To hone in on the mechanisms underlying sex-interacting expression and eQTLs, we generated chromatin accessibility data using ATAC-seq in 20 additional individuals (10 male and 10 female). Differences in chromatin accessibility are well known to influence gene expression and the exposure of genetic variants (Thurman et al. 2012). We identified that sex-specific chromatin accessibility was enriched for genes with sex-specific expression and for variants with genotype-by-sex interactions. However, we have so far not been able to identify any particular mechanism driving differential chromatin accessibility, despite preliminary evaluation of possibilities such as hormone receptor TF binding. This highlights a continued challenge and opportunity in mapping sex-dependent regulatory effects.

We discover that complex traits may have variants with cumulatively different effects between sexes, leading to potential sex-biases in their disease prevalence and severity. Previous analyses within the Framingham Heart Study specifically focused on identifying which GWAS variants exhibit genotype-by-sex interactions and discovered 14 of 11,672 tested had significant effects (Yao et al. 2014). This suggests that strong sex-specific genetic effects on gene expression may not underlie many complex trait-associated variants. Here, by considering the cumulative effect sizes of eQTL between sexes, we were able to identify well-known sex-biased diseases, suggesting that accumulation of sex-biased genetic disease risk could be polygenic across a broad range of variants and genes.

Many challenges and opportunities remain in the study of the combined effects of genetics and sex on gene expression. Indeed, despite correction of hidden covariates, blood is a heterogeneous tissue where differences in cell type and diverse environmental factors may still confound discovery of differences that can be causally attributed to sex. These challenges must be weighed against in vitro studies of sex differences that may not reflect hormone biology or other in vivo factors that expose meaningful differences in the activity of regulatory variants between sexes. As studies of diverse tissues increase their sample sizes (The GTEx Consortium 2015), we expect future studies will have comparable power to broadly expose the characteristics of X Chromosome and sex-specific eQTLs across a range of tissues. Moreover, in our study, the detection and interpretation of differences in expression variances between the two genders is complicated by the fact that the means of the two gender distributions are not equal (Sun et al. 2013). We attempt to overcome this issue by regressing out the mean effect of gender and testing for variance effects in the gender-corrected residuals. However, despite this correction, residual gender effects might still be present, if, e.g., gender has a nonlinear effect on the mean, and thus the biological and/or clinical significance of such differences in variances will always need to be determined on nonstatistical grounds. Further, our study identifies

differences in purifying selection for eQTL on the X Chromosome versus the autosomes; however, as many common variants are likely neutral, future studies will benefit from considering low frequency and rare regulatory variants, as it is increasingly evident that personal genome interpretation will require careful integration of X-Chromosome biology between sexes.

Methods

Study cohort for expression analysis

In the discovery of associations between genotype, expression, and sex, we used the Depression Genes and Networks cohort (Battle et al. 2014; Mostafavi et al. 2014). The cohort is comprised of 922 individuals (648 females and 274 males) of European ancestry between the ages of 21 and 60 yr within the United States. A detailed description of the recruitment and phenotype data for this cohort is provided elsewhere (Battle et al. 2014; Mostafavi et al. 2014). See [Supplemental Methods](#) for details on processing of genetic and transcriptome data.

Differential gene expression and variance analysis

Differential expression analysis was conducted using a negative binomial model in the R package DESeq (Anders and Huber 2010) to identify genes with sex-specific expression (FDR 5%) in a matched number of males and females ($n = 274$). Using DESeq, the samples were corrected for a number of technical factors previously identified (Battle et al. 2014; see [Supplemental Methods](#)).

To detect genes with sex-specific expression variance, we used technical factor-corrected and sample-matched data ($n = 274$). This was further compared to regression with PEER residuals. We applied three different testing strategies. First, we tested for differences in variance between males and females using an *F*-test. Second, we tested for normality of gene expression traits using the Shapiro-Wilk normality test. If normality was not rejected, we tested for differences in expression variance between males and females using an *F*-test; otherwise, we tested for differences in expression variance using the Brown-Forsythe Levene-type test. In the last approach, we quantile-normalized the log-transformed PEER residuals to a $N(0, 1)$ and ran a linear regression with quantile-normalized log-transformed PEER residuals as response and gender as a covariate. Again, if the normality assumption for the regression residuals was not rejected, we tested for differences between males and females in regression residual expression variance using either an *F*-test or a Brown-Forsythe Levene-type test. We used here the regression residuals, instead of the original response, to distinguish between genes that show genuine DV (since mean effects are removed in residuals) and genes that show DV due to DE.

cis-QTL mapping

We performed association testing for expression QTL using the R package Matrix eQTL (Shabalin 2012), after correction for hidden covariates ([Supplemental Methods](#); [Supplemental Fig. S13](#)). Variants within 1 Mb upstream of the transcription start site (TSS) and 1 Mb downstream from the transcription end site (TES) of a gene were tested using a linear regression model and accounting for sex as a covariate. We model the association between a candidate variant and gene expression by a linear regression:

$$y = \mu + \beta_1 \times \text{genotype} + \beta_2 \times \text{sex} + \varepsilon,$$

where y denotes the observed expression level of the gene, μ the mean expression level across the population, β_1 the regression co-

efficient of genotype, β_2 the regression coefficient of sex, and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. In the primary QTL analysis, we only tested variants with $\text{MAF} \geq 0.05$. In the low-frequency QTL analysis, we tested variants with MAF in the range 0.01–0.05. To control for multiple testing, we used Bonferroni correction to account for the number of variants tested per gene, retained the best association per gene, and controlled for FDR at the gene-level significance by the Benjamini-Hochberg method (Benjamini and Hochberg 1995).

Identification of sex-interacting QTLs

We tested for genotype-by-sex (G×S) interactions in the full sample ($n = 922$) using the interaction model in the R package Matrix eQTL (Shabalin 2012). This model was used to test for equality of effect sizes between males and females. We model the interaction between genotype and sex by adding an interaction term to the linear regression:

$$y = \mu + \beta_1 \times \text{genotype} + \beta_2 \times \text{sex} + \beta_3 \times (\text{genotype} \times \text{sex}) + \varepsilon,$$

where y , μ , β_1 , and ε are as above, and β_2 and β_3 the regression coefficient of sex and the interaction between genotype and sex. For each gene, variants within 1 Mb upstream of the transcription start site (TSS) and 1 Mb downstream from the transcription end site (TES) of a gene were tested. We required variants to have a $\text{MAF} \geq 0.05$ in the male and female samples. To remove cases where genotype is collinear with sex, we excluded variants with a rank less than three between genotype and sex. All sex-interacting eQTLs tested in the DGN cohort can be visualized at <http://montgomerylab.stanford.edu/resources.html>.

Sample collection for open chromatin profiling

We obtained buffy coat samples from 20 healthy donors from the Stanford Blood Center (Stanford, CA). Male and female donors were equally represented for downstream differential analyses. To control for age and ethnicity, the samples selected were restricted to Caucasians between 18 and 45 yr of age. Age, ethnicity, and healthy status were self-reported by donors. ATAC-seq was applied to isolated PBMCs, and differentially accessible peaks were quantified (see [Supplemental Methods](#)).

Differential open chromatin accessibility analysis

To identify regions with differential chromatin accessibility between males and females, we used the R package DESeq (Anders and Huber 2010). The data were normalized by the effective library size, and variance was estimated for each group. We tested for differential chromatin accessibility using a model based on the negative binomial distribution. To account for multiple testing, we adjusted the *P*-values by the Benjamini-Hochberg method; regions with a $\text{FDR} < 0.05$ were considered to be regions with differential open chromatin. We also evaluated the effect of removing hidden factors using SVA on the differential chromatin accessibility analysis and observed no differences in the number of significant peaks identified at multiple thresholds ($P\text{-value} = 0.2, 0.1, 0.01, 0.001$). To ensure that %GC bias and sequencing depth did not influence the differential chromatin accessibility analysis, we confirmed that there were no significant differences in these metrics between the sexes ([Supplemental Fig. S14](#)).

Trait analysis

We obtained trait-associated SNPs from ImmunoBase (available at www.immunobase.org; accessed on April 2, 2015) and the NHGRI-EBI GWAS catalog (Welter et al. 2014) (available at www.ebi.ac.uk/

gwas; accessed on April 2, 2015). We only considered independent variants ($R^2 < 0.5$) with a genome-wide association threshold of P -value $< 1 \times 10^{-8}$. Independent variants were selected using an LD pruning strategy such that the most strongly associated GWAS SNP for each trait for each LD block ($R^2 < 0.5$) was selected. In the DGN cohort, we calculated the absolute effect size of each independent GWAS variant on expression for genes within 1 Mb in the male ($n = 274$) and randomly subsampled female ($n = 274$) samples. We considered a GWAS variant to have a sex-specific effect if the effect size in one sex was 1.2-fold greater than the opposite sex. For variants with a 1.2-fold greater effect in either sex, we ran a two-sided binomial exact test to determine if one sex had greater effects than the opposite sex. Only traits with at least 20 significant GWAS variants were considered in the trait analysis (Fig. 5; Supplemental Fig. S15).

Data access

Genotype, raw RNA-seq, quantified expression, and covariate data for the DGN cohort are available by application through the National Institute of Mental Health (NIMH) Center for Collaborative Genomic Studies on Mental Disorders. Instructions for requesting access to data can be found at the NIMH Repository and Genomics Resource (RGR; https://www.nimhgenetics.org/access_data_biomaterial.php), and inquiries should reference the “Depression Genes and Networks study (D. Levinson, PI).” Open chromatin accessibility data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE69749.

Acknowledgments

We thank Stephen J. Galli for critical review of the manuscript. K. R.K. is supported by DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. S.B.M. is supported by the Edward Mallinckrodt Jr. Foundation. A.B. and S.B.M. are supported by the NIH (Office of the Director, R01MH101814; and National Human Genome Research Institute, R01HG008150).

References

- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.
- Andolfatto P. 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol Biol Evol* **18**: 279–290.
- Awadalla P, Boileau C, Payette Y, Idaghdour Y, Goulet JP, Knoppers B, Hamet P, Laberge C; CARTaGENE Project. 2013. Cohort profile of the CARTaGENE study: Quebec’s population-based biobank for public health and personalized genomics. *Int J Epidemiol* **42**: 1285–1299.
- Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi J, Mei R, et al. 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res* **24**: 14–24.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol* **57**: 289–300.
- Breslau N, Davis GC, Andreski P, Peterson EL, Schultz LR. 1997. Sex differences in posttraumatic stress disorder. *Arch Gen Psychiatry* **54**: 1044–1048.
- Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. 2004. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* **57**: 229–236.
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–1218.
- Carrel L, Willard HF. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**: 400–404.
- Castagne R, Zeller T, Rotival M, Szymczak S, Truong V, Schillert A, Tregouet DA, Munzel T, Ziegler A, Cambien F, et al. 2011. Influence of sex and genetic variability on expression of X-linked genes in human monocytes. *Genomics* **98**: 320–326.
- Chang D, Gao F, Slavney A, Ma L, Waldman YY, Sams AJ, Billing-Ross P, Madar A, Spritz R, Keinan A. 2014. Accounting for eXcentricities: Analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PLoS One* **9**: e113684.
- Cheng X, Klaassen CD. 2012. Hormonal and chemical regulation of paraoxonases in mice. *J Pharmacol Exp Ther* **342**: 688–695.
- Cohn BA, Wingard DL, Cirillo PM, Cohen RD, Reynolds P, Kaplan GA. 1996. Differences in lung cancer risk between men and women: examination of the evidence. *J Natl Cancer Inst* **88**: 1867.
- Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez-Arcelus M, Sekowska M, et al. 2009. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**: 1246–1250.
- Dimas AS, Nica AC, Montgomery SB, Stranger BE, Raj T, Buil A, Giger T, Lappalainen T, Gutierrez-Arcelus M, McCarthy MI, et al. 2012. Sex-biased genetic effects on gene regulation in humans. *Genome Res* **22**: 2368–2375.
- Dobyns WB, Filauro A, Tomson BN, Chan AS, Ho AW, Ting NT, Oosterwijk JC, Ober C. 2004. Inheritance of most X-linked traits is not dominant or recessive, just X-linked. *Am J Med Genet A* **129A**: 136–143.
- Du L, Bayir H, Lai Y, Zhang X, Kochanek PM, Watkins SC, Graham SH, Clark RS. 2004. Innate gender-based proclivity in response to cytotoxicity and programmed cell death pathway. *J Biol Chem* **279**: 38563–38570.
- Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA. 2004. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118**: 555–566.
- Giordano G, Tait L, Furlong CE, Cole TB, Kavanagh TJ, Costa LG. 2013. Gender differences in brain susceptibility to oxidative stress are mediated by levels of paraoxonase-2 expression. *Free Radic Biol Med* **58**: 98–108.
- Gottipati S, Arbiza L, Siepel A, Clark AG, Keinan A. 2011. Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. *Nat Genet* **43**: 741–743.
- The GTEx Consortium. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**: 648–660.
- Gupta V, Parisi M, Sturgill D, Nuttall R, Doctolero M, Dudko OK, Malley JD, Eastman PS, Oliver B. 2006. Global analysis of X-chromosome dosage compensation. *J Biol* **5**: 3.
- Hankin BL, Abramson LY. 2001. Development of gender differences in depression: an elaborated cognitive vulnerability–transactional stress theory. *Psychol Bull* **127**: 773–796.
- Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, Thorleifsson G, Zillikens MC, Speliotes EK, Magi R, 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.
- Hussin JG, Hodgkinson A, Idaghdour Y, Grenier JC, Goulet JP, Gbeha E, Hip-Ki E, Awadalla P. 2015. Recombination affects accumulation of damaging and disease-associated mutations in human populations. *Nat Genet* **47**: 400–404.
- Idaghdour Y, Awadalla P. 2012. Exploiting gene expression variation to capture gene-environment interactions for disease. *Front Genet* **3**: 228.
- Jog NR, Caricchio R. 2013. Differential regulation of cell death programs in males and females by Poly (ADP-Ribose) Polymerase-1 and 17 β estradiol. *Cell Death Dis* **4**: e758.
- Keinan A, Mullikin JC, Patterson N, Reich D. 2009. Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nat Genet* **41**: 66–70.
- Kharchenko PV, Xi R, Park PJ. 2011. Evidence for dosage compensation between the X chromosome and autosomes in mammals. *Nat Genet* **43**: 1167–1169.
- Knowles DA, Davis JR, Raj A, Zhu X, Potash JB, Weissman MM, Shi J, Levinson DF, Mostafavi S, Montgomery SB, et al. 2015. Allele-specific expression reveals interactions between genetic variation and environment. *bioRxiv* doi: <http://dx.doi.org/10.1101/025874>.
- Lang JT, McCullough LD. 2008. Pathways to ischemic neuronal cell death: Are sex differences relevant? *J Transl Med* **6**: 33.
- Lappalainen T, Sammeth M, Friedlander MR, ‘t Hoen PAC, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**: 506–511.
- Lerner DJ, Kannel WB. 1986. Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population. *Am Heart J* **111**: 383–390.
- Li H, Pin S, Zeng Z, Wang MM, Andreasson KA, McCullough LD. 2005. Sex differences in cell death. *Ann Neurol* **58**: 317–321.

- Liu CT, Estrada K, Yerges-Armstrong LM, Amin N, Evangelou E, Li G, Minster RL, Carless MA, Kammerer CM, Oei L, et al. 2012a. Assessment of gene-by-sex interaction effect on bone mineral density. *J Bone Miner Res* **27**: 2051–2064.
- Liu LY, Schaub MA, Sirota M, Butte AJ. 2012b. Sex differences in disease risk from reported genome-wide association study findings. *Hum Genet* **131**: 353–364.
- Lu J, Wu CI. 2005. Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. *Proc Natl Acad Sci* **102**: 4063–4067.
- Luan JA, Wong MY, Day NE, Wareham NJ. 2001. Sample size determination for studies of gene-environment interaction. *Int J Epidemiol* **30**: 1035–1040.
- Mendelsohn ME, Karas RH. 2005. Molecular and cellular basis of cardiovascular gender differences. *Science* **308**: 1583–1587.
- Molloy EJ, O'Neill AJ, Grantham JJ, Sheridan-Pereira M, Fitzpatrick JM, Webb DW, Watson RW. 2003. Sex-specific alterations in neutrophil apoptosis: the role of estradiol and progesterone. *Blood* **102**: 2653–2659.
- Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. 2010. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**: 773–777.
- Mostafavi S, Battle A, Zhu X, Potash JB, Weissman MM, Shi J, Beckman K, Haudenschield C, McCormick C, Mei R, et al. 2014. Type I interferon signaling genes in recurrent major depression: increased expression detected by whole-blood RNA sequencing. *Mol Psychiatry* **19**: 1267–1274.
- Myers RA, Scott NM, Gauderman WJ, Qiu W, Mathias RA, Romieu I, Levin AM, Pino-Yanes M, Graves PE, Villarreal AB, et al. 2014. Genome-wide interaction studies reveal sex-specific asthma risk alleles. *Hum Mol Genet* **23**: 5251–5259.
- Natarajan A, Yardimci GG, Sheffield NC, Crawford GE, Ohler U. 2012. Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res* **22**: 1711–1722.
- Naugler WE, Sakurai T, Kim S, Maeda S, Kim K, Elsharkawy AM, Karin M. 2007. Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production. *Science* **317**: 121–124.
- Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET. 2010. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet* **6**: e1000895.
- Nicolae DL, Gamazon E, Zhang W, Duan SW, Dolan ME, Cox NJ. 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* **6**: e1000888.
- Ober C, Loisel DA, Gilad Y. 2008. Sex-specific genetic architecture of human disease. *Nat Rev Genet* **9**: 911–922.
- Park C, Carrel L, Makova KD. 2010. Strong purifying selection at genes escaping X chromosome inactivation. *Mol Biol Evol* **27**: 2446–2450.
- Patsopoulos NA, Tatsioni A, Ioannidis JP. 2007. Claims of sex differences: an empirical assessment in genetic associations. *JAMA* **298**: 880–893.
- Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**: 768–772.
- Pigott TA. 1999. Gender differences in the epidemiology and treatment of anxiety disorders. *J Clin Psychiatry* **60**(Suppl 18): 4–15.
- Polonikov AV, Ivanov VP, Bogomazov AD, Freidin MB, Illig T, Solodilova MA. 2014. Antioxidant defense enzyme genes and asthma susceptibility: gender-specific effects and heterogeneity in gene-gene interactions between pathogenetic variants of the disease. *Biomed Res Int* **2014**: 708903.
- Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, Kilpelainen TO, Esko T, Magi R, Li S, et al. 2013. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* **9**: e1003500.
- Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**: 1742–1745.
- Reinius B, Saetre P, Leonard JA, Blekhan R, Merino-Martinez R, Gilad Y, Jazin E. 2008. An evolutionarily conserved sexual signature in the primate brain. *PLoS Genet* **4**: e1000100.
- Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**: 735–742.
- Schaffner SF. 2004. The X chromosome in population genetics. *Nat Rev Genet* **5**: 43–51.
- Schneider R, Grosschedl R. 2007. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev* **21**: 3027–3043.
- Shabalin AA. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**: 1353–1358.
- Singh ND, Petrov DA. 2007. Evolution of gene function on the X chromosome versus the autosomes. *Genome Dyn* **3**: 101–118.
- Sun X, Elston R, Morris N, Zhu X. 2013. What is the significance of difference in phenotypic variability across SNP genotypes? *Am J Hum Genet* **93**: 390–397.
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82.
- Trabzuni D, Ramasamy A, Imran S, Walker R, Smith C, Weale ME, Hardy J, Ryten M, North American Brain Expression Consortium. 2013. Widespread sex differences in gene expression and splicing in the adult human brain. *Nat Commun* **4**: 2771.
- Tukiainen T, Pirinen M, Sarin AP, Ladvall C, Kettunen J, Lehtimäki T, Lokki ML, Perola M, Sinisalo J, Vlachopoulou E, et al. 2014. Chromosome X-wide association study identifies loci for fasting insulin and height and evidence for incomplete dosage compensation. *PLoS Genet* **10**: e1004127.
- Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet* **7**: 645–653.
- Wang C, Cheng Y, Liu T, Li Q, Fillingim RB, Wallace MR, Staud R, Kaplan L, Wu R. 2008. A computational model for sex-specific genetic architecture of complex traits in humans: implications for mapping pain sensitivity. *Mol Pain* **4**: 13.
- Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, et al. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**: D1001–D1006.
- Whitacre CC. 2001. Sex differences in autoimmune disease. *Nat Immunol* **2**: 777–780.
- Whitacre CC, Reingold SC, O'Looney PA. 1999. A gender gap in autoimmunity. *Science* **283**: 1277–1278.
- Wise AL, Gyi L, Manolio TA. 2013. eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am J Hum Genet* **92**: 643–647.
- Yang X, Schadt EE, Wang S, Wang H, Arnold AP, Ingram-Drake L, Drake TA, Lusk AJ. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res* **16**: 995–1004.
- Yao C, Joeanes R, Johnson AD, Huan T, Esko T, Ying S, Freedman JE, Murabito J, Lunetta KL, Metspalu A, et al. 2014. Sex- and age-interacting eQTLs in human complex diseases. *Hum Mol Genet* **23**: 1947–1956.
- Ye CJ, Feng T, Kwon HK, Raj T, Wilson MT, Asinowski N, McCabe C, Lee MH, Frohlich I, Paik HI, et al. 2014. Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* **345**: 1254665.

Received August 6, 2015; accepted in revised form April 18, 2016.



Impact of the X Chromosome and sex on regulatory variation

Kimberly R. Kukurba, Princy Parsana, Brunilda Balliu, et al.

Genome Res. 2016 26: 768-777 originally published online April 21, 2016

Access the most recent version at doi:[10.1101/gr.197897.115](https://doi.org/10.1101/gr.197897.115)

Supplemental Material <http://genome.cshlp.org/content/suppl/2016/05/10/gr.197897.115.DC1>

References This article cites 73 articles, 16 of which can be accessed free at:
<http://genome.cshlp.org/content/26/6/768.full.html#ref-list-1>

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Affordable, Accurate
Sequencing.



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>
