

The X-Effect: Regulatory Variation Between the Sexes



Kimberly R. Kukurba^{1,2}, Princy Parsana⁴, Kevin S. Smith^{1,2}, Zachary Zappala^{1,2}, Anshul Kundaje¹, Douglas F. Levinson³, Alexis Battle^{4*}, and Stephen B. Montgomery^{1,2*}

Departments of Genetics¹, Pathology², and Psychiatry³, Stanford University School of Medicine, Stanford, California 94305, USA
⁴Department of Computer Science, Johns Hopkins University, Baltimore, Maryland 21218, USA; *Co-senior authors

Introduction

The role of gender in health and disease has slowly been uncovered at the clinical level. For instance, the prevalence of certain diseases (e.g. multiple sclerosis, COPD) is dramatically different and specific medications (e.g. zolpidem) require different dosing in men and women. In order to fully understand the role of gender on health and disease, it is paramount to understand the effect of sex differences on molecular phenotypes. With a large set of gene expression data from over nine hundred individuals, we have the opportunity to uncover sex-biased genetic effects on transcriptional regulation. **Our objectives are to (1) understand sex differences at the molecular level and (2) characterize the differential effects of sex chromosome eQTLs compared to autosomal eQTLs.**

Data Overview and Methods

Data overview. To study the effect of sex on gene regulation, we used the Depression Genes and Network (DGN) cohort [1]. This dataset consists of 51bp single-end RNA-Seq samples from the whole blood of 922 individuals (648 females and 274 males), all of European ancestry. Each individual was genotyped on the Illumina Human Omni1-Quad and missing genotypes were imputed by IMPUTE2 using the 1000 Genomes Phase 1 haplotypes [2].

Expression normalization and eQTL mapping. To correct for technical and biological factors, probabilistic estimation of expression residuals (PEER) was used to infer factors while retaining the sex effects [3]. Association testing was conducted using MatrixeQTL for common variants (MAF > 0.05) [4]. We used Bonferroni correction to account for the number of SNPs tested per gene and subsequently identified eQTLs using gene level significance at FDR 5%.

Expression variation between the sexes

First, we investigated sex-specific patterns of gene expression variation across the genome using the F-test statistic. We observe that females have higher variance on autosomal genes; genes exhibiting higher variance are enriched in apoptosis processes (FDR < 0.05) (Figure 1A). On the X chromosome, males have higher variance than expected due to the presence of a single X chromosome, which is confirmed in our analysis of the proportion of variance explained by genotype (Figure 1B).

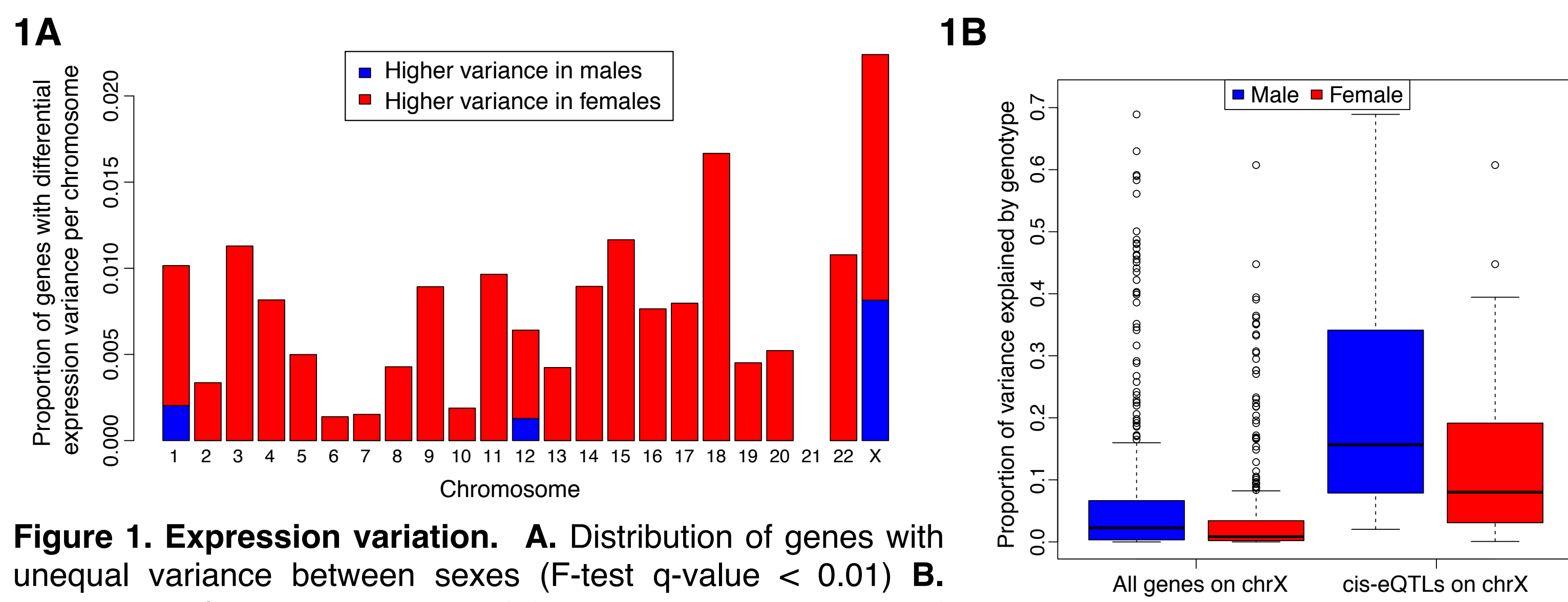


Figure 1. Expression variation. A. Distribution of genes with differential expression variance between sexes (F-test q-value < 0.01). B. Proportion of variance explained by genotype in males and females.

Discovery and characterization of X-chromosome eQTLs

We observe a depletion of *cis*-eQTLs on the X chromosome relative to the autosomes ($P < 10^{-15}$, chi-square test) (Figure 2A). When we evaluate the relationship of effect size and selective constraint, we observe that eQTLs for strongly constrained genes have a lower effect size on the X chromosome compared to autosomes ($P = 3.6 \times 10^{-4}$, Wilcoxon rank-sum test) (Figure 2B), suggesting more effective purifying selection on the X chromosome.

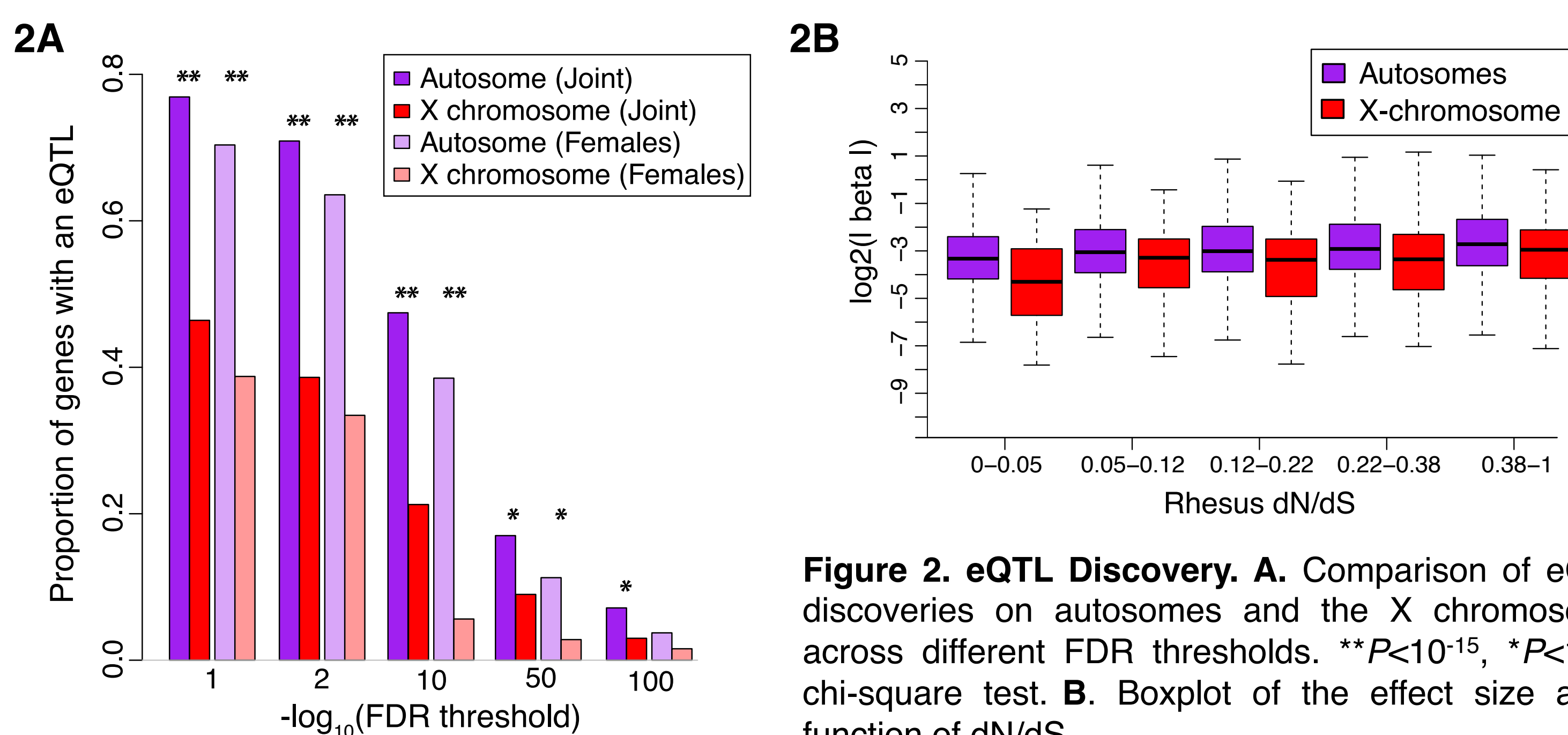


Figure 2. eQTL Discovery. A. Comparison of eQTL discoveries on autosomes and the X chromosome across different FDR thresholds. B. Boxplot of the effect size as a function of dN/dS.

Discovery sex-interacting eQTLs

To understand the effect of genotype-sex interactions on expression, we used a linear model with a genotype-sex interaction term to identify sex-interacting eQTLs. We detect more sex-interacting eQTLs on the X chromosome compared to autosomes (Figure 3A). Sex-interacting eQTLs are not enriched in differentially expression genes, although they are enriched in genes involved in sex differentiation (GO:0007548; OR = 6.6, $P = 4.3 \times 10^{-3}$, Fisher's exact test). We also observe differing eQTL effect sizes between males and females for variants associated with autoimmune diseases, which have sexually dimorphic incidence rates and severity (Figure 3B).

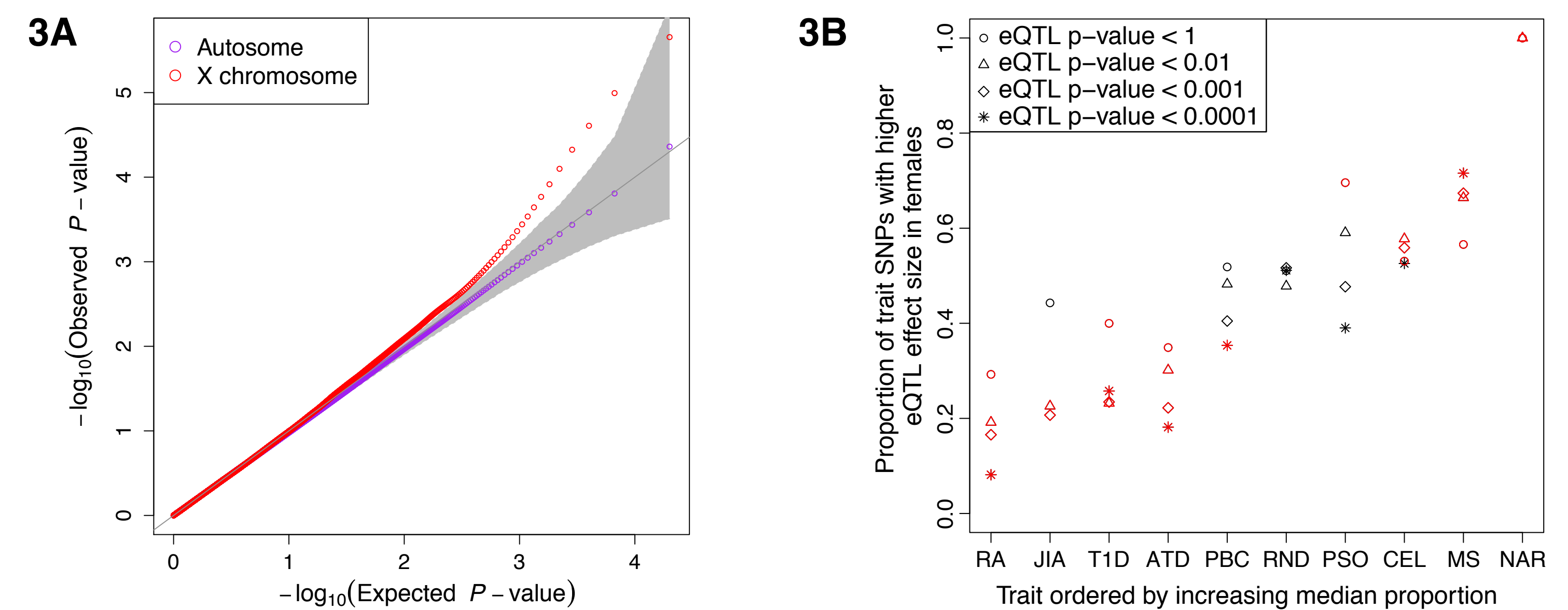


Figure 3. Sex-interacting eQTLs. A. Q-Q plot for genotype-sex interaction tests with 95% genome-wide confidence interval. B. Differing eQTL effect size in males and females for variants associated with Immunobase diseases. Red points indicate Bonferroni-adjusted binomial test P -value < 0.05.

Sex-specific chromatin accessibility

To understand the biological mechanisms of sex-specific gene regulation, we investigated differences in chromatin accessibility between males and females using ATAC-seq [5]. We observe an enrichment of sex-specific expression on the X chromosome (Figure 4A) and enrichment of genes with sex-specific expression in sex-specific chromatin regions (OR = 1.45, $P = 4.7 \times 10^{-6}$, Fisher's exact test) (Figure 4B).

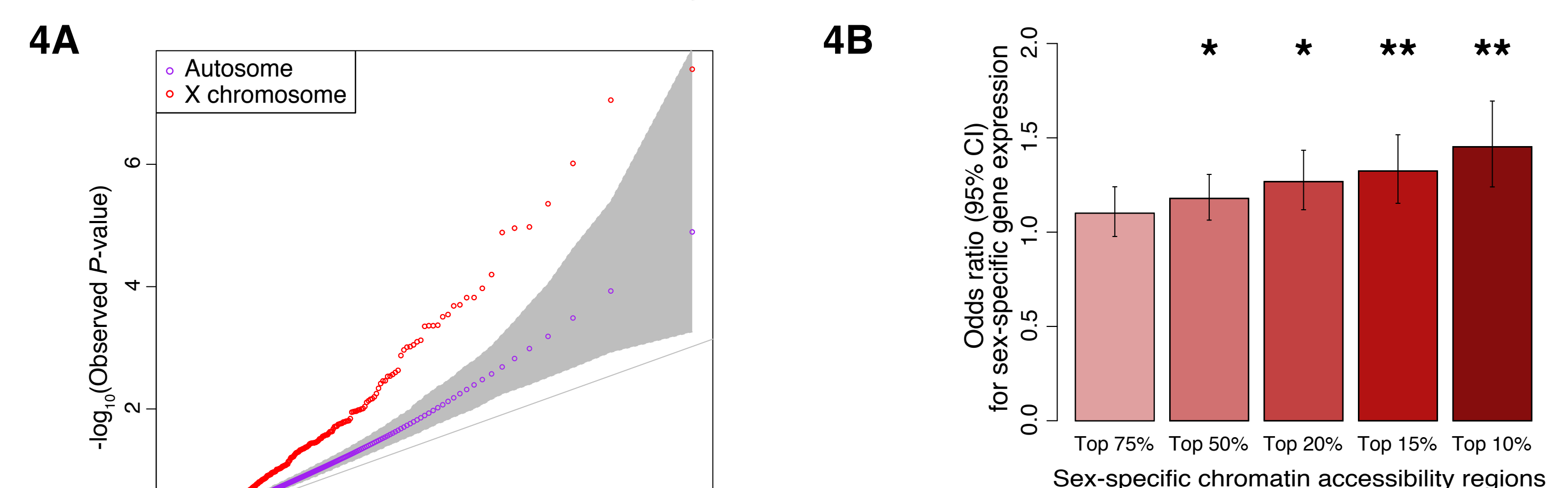


Figure 4. Sex-specific peak discovery. A. Q-Q plot for sex-specific chromatin accessibility tests with 95% genome-wide confidence interval. B. Enrichment of genes with sex-specific expression (FDR 5%) in sex-specific peaks.

We also observe that sex-specific chromatin accessibility regions are enriched for genes with sex-interacting eQTLs, highlighting a potential mechanism for genotype-sex interactions (Figure 5).

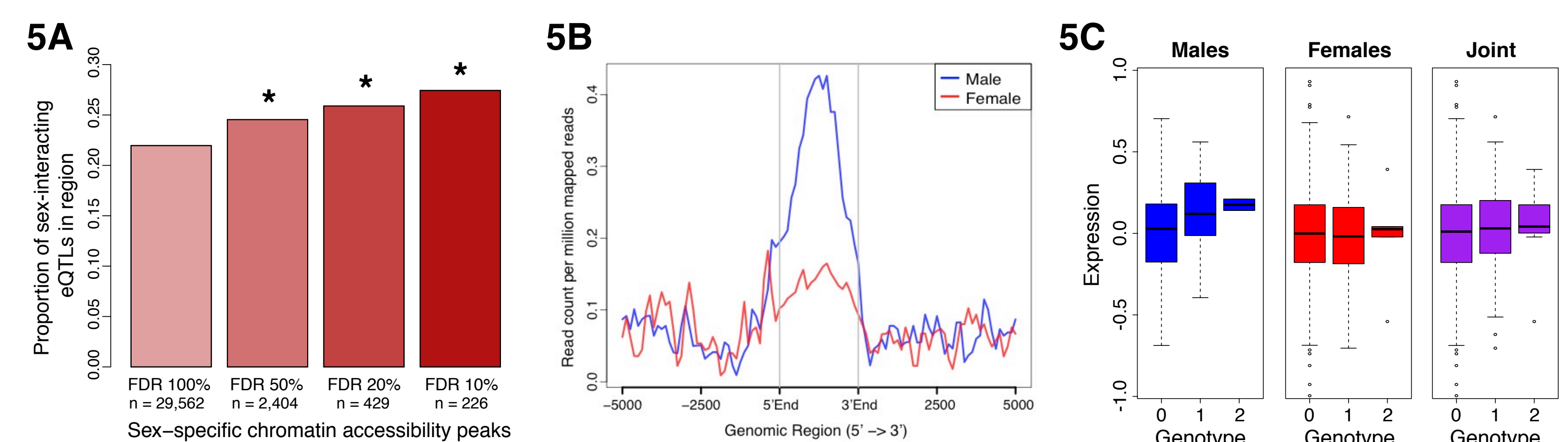


Figure 5. Integration of sex-interacting eQTLs with sex-specific peaks. A. Enrichment of genes with sex-specific expression (FDR 5%) in sex-specific peaks. B-C. Example of a sex-specific chromatin accessibility peak eQTL with a sex-interacting eQTL.

References

- [1] Battle, A., et al. (2014). *Genome Research*, 24(1), 14-24. [2] Howie, B. N., et al. (2009). *PLoS Genetics*, 5(6), e1000529. [3] Stegle, O., et al. (2012). *Nature Protocols*, 7(3), 500-507. [4] Shabalin, A. A. (2012). *Bioinformatics*, 28(10), 1353-1358. [5] Buenrostro, J.D., et al. (2013). *Nature Methods* 10, 1213-8.

Acknowledgments

This research was conducted with government support under and awarded by DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a.