The functional and evolutionary genomics of sexual dimorphism in *Homo sapiens*

Audrey Arner[1,2], Kathleen Grogan[1], George Perry[1,2]
Departments of [1]Anthropology and [2]Biology, Pennsylvania State University, PA 16802

Introduction

Genome wide association studies (GWAS) can be used to identify genetic loci associated with phenotypic variation within a population. These loci can then be studied with population genetic analyses to indirectly test hypotheses about the evolutionary history of the trait. We are applying this two-part approach to study the recent evolutionary history of human body size and shape sexual dimorphism. In humans and other great apes, female and male body size phenotypes significantly diverge during development, despite the mostly common genome shared by both sexes. Women and men exhibit sexual dimorphism, or substantial disparities in many traits, most notably height, weight, and body mass index. While some anthropologists have hypothesized that there has been selection for reduced dimorphism following the transition to agriculture, any recent changes in the degree of sexual dimorphism may simply reflect genetic drift.[1] Here we are identifying genetic loci that differentially influence female and male body size and determining the biological plausibility of these loci through enrichment tests of identified genes in the hypothalamic-pituitary-gonadal, or HPG, axis, which is known to influence sexual differentiation (Figure 3).

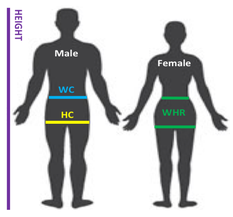


Figure 1: Phenotypic variation between females and males

Figure 2: Example schematic of the effect of genotype on stature

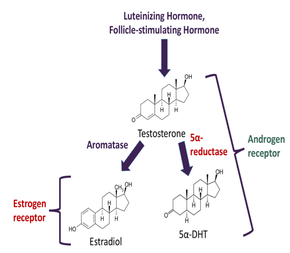
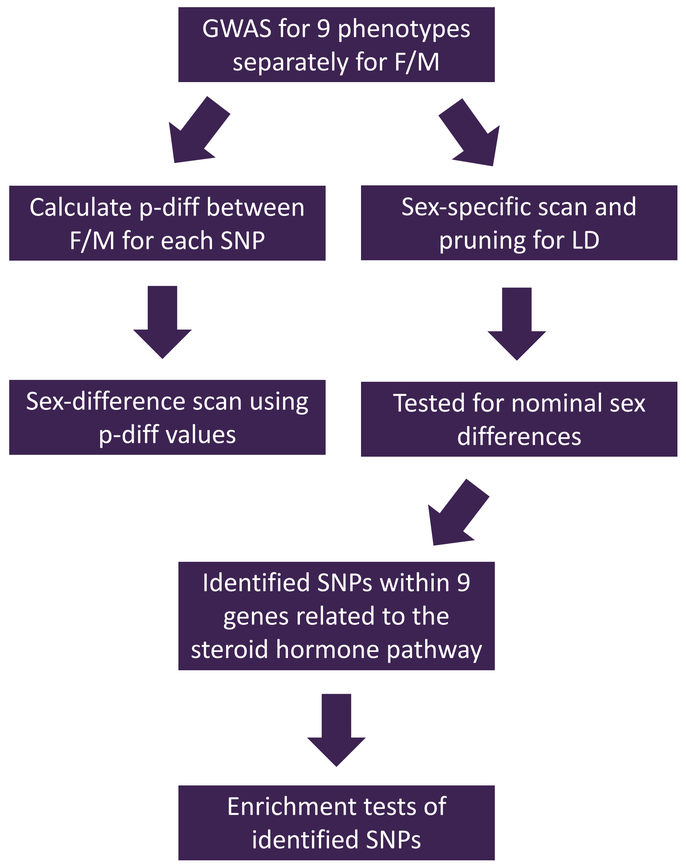


Figure 3: Steroid hormone synthesis

Methods



GWAS were conducted separately in women and men on 2.8 million SNPs for 9 phenotypes (height, weight, BMI, waist circumference (WC), WC adjusted for BMI, hip circumference (HC), HC adjusted for BMI, waist-to-hip ratio (WHR), and WHR adjusted for BMI).[2] False discovery rates (FDR) for this sex-specific scan ranged from 0.3 to 0.001. The data were pruned for linkage disequilibrium (LD) with a 200 kb window. P-values (p-diff) were calculated to test for nominal differences between the women and men-specific beta-estimates with this following t statistic:

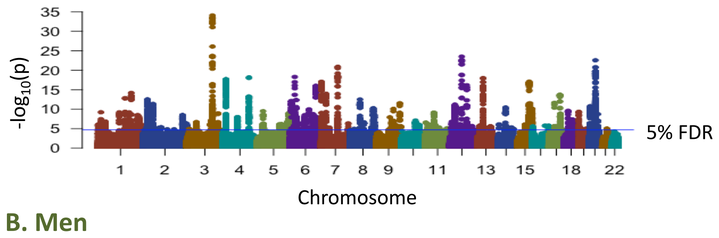
$$t = \frac{b_{men} - bw_{omen}}{\sqrt{SE_{men}{}^2 + SEw_{omen}{}^2 - 2r \cdot SEmen \cdot SEw_{omen}}}$$

A sex-difference scan was also used to test for p-diff significance in the concatenated data set. To assess the biological plausibility of less conservative FDR cutoffs, we identified SNPs located in or nearby (+/- 50 kb) 9 genes related to the HPG axis: AR, ESR1, ESR2, CYP19A1, SRD5A1, SRD5A2, SRD5A3, CGA, and LHCGR (Figure 3). Enrichment tests were used to determine whether SNPs with nominally significant p-diff values were more likely to be located in genes related to the HPG axis compared to SNPs that are also associated with our phenotypes, but equally in males and females.

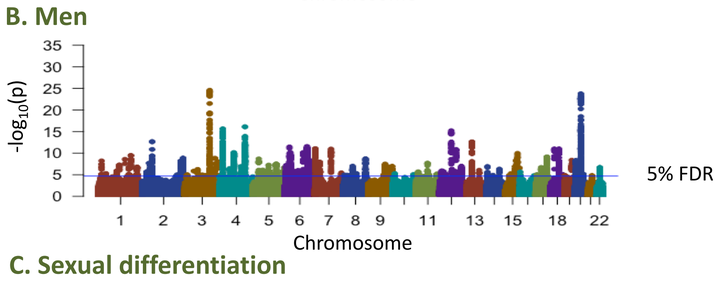
Results

Sample Manhattan Plot for one phenotype (Height)

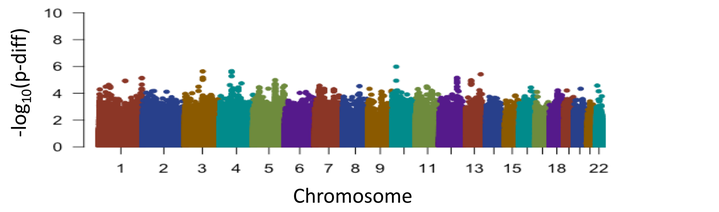
A. Women



B. Men

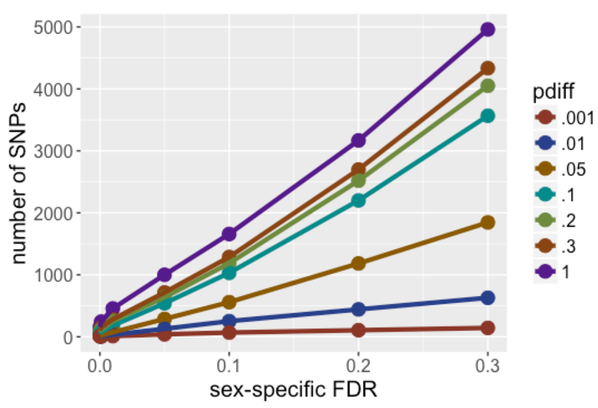


C. Sexual differentiation



Manhattan plots depicting the trait association p-values for each SNP plotted against the chromosomal position. Plots A and B show the results of the sex-specific scan for women and men respectively. The mid-line denotes a 5% false discovery rate. Plot C shows the results of the sex-difference scan.

Number of SNPs at various p-diff values (difference between women and men) and sex-specific FDR cutoffs.

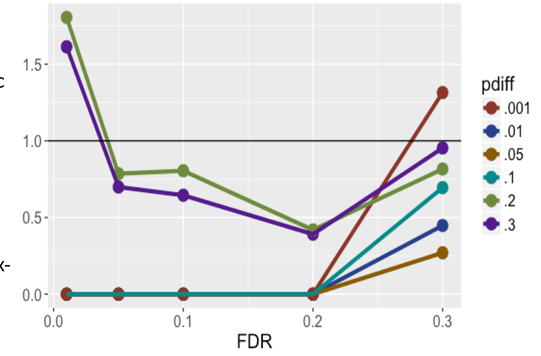


SNPs Identified within 9 genes (AR, ESR1, ESR2, CYP19A1, SRD5A1, SRD5A2, SRD5A3, CGA, and LHCGR) related to the HPG axis

GENE	SNP	FDR (Sex-specific)	P-DIFF	PHENOTYPE
CYP19A1	rs1902586	.11	.21	Height
ESR1	rs2881766	.23	.06	Height
ESR1	rs9383593	.04	.84	Height
ESR1	rs2982710	.01	.20	Height
SRD5A3	rs7663650	.26	.06	Height
LHCGR	rs3884615	.3	.00037	WHRadj

Enrichment analysis of sexually dimorphic SNPs in the steroid hormone pathway genes

The ratio of the observed proportion of sexually dimorphic SNPs in steroid hormone gene regions to the expected proportion of SNPs in steroid hormone genes at various p-diff and sex-specific FDR cutoffs.



For none of the above comparisons was the proportion of sexually dimorphic SNPs in genes related to the HPG axis significantly greater than the expected proportion of genes related to the HPG axis by chance (Fisher's Exact Tests).

Total number of SNPs in the 9 genes at FDR cutoffs

FDR (Sex-specific)	# SNPs overlapping with genes
0.01	1
0.05	2
0.1	2
0.2	3
0.3	6

Future Directions

Repeat methods on a more recent GWAS data set
↓
Broaden list of candidate genes for enrichment test
↓
Transcription factor binding site analysis
↓
Test for positive selection using the Singleton Density Score

Our next step will be to use this framework to identify a much larger number of sexually dimorphic SNPs using the data from a more recently conducted, larger, and more powerful GWAS.[3] In the biological plausibility assessment phase of this analysis, in addition to testing whether the identified SNPs are enriched within the 9 genes related to the HPG axis, we will also test whether these SNPs are enriched for sites in or nearby genome-wide transcription factor binding sites for regulatory proteins known to be involved in sexual differentiation. Finally, we will then use the Singleton Density Score (SDS) statistic[4] to test whether the sexual dimorphism-associated genetic loci have been affected by positive selection in recent human evolution. In brief, favored alleles are predicted be found on haplotypes with relatively fewer point mutations; SDS identifies this pattern in aggregate across the genome for polygenic traits.

Conclusions

We identified between 455 and 4,954 SNPs with sex-specific genetic effects that may be related to dimorphism depending on the statistical cutoffs applied. In order to assess the biological plausibility of our expanded sets of sexually differentiated SNPs at various FDR cutoffs, we tested whether the sexually differentiated SNPs were significantly more likely to be located in or around regions of the genome associated with genes related to steroid hormones and the HPG axis. While we did not observe significant enrichments in this analysis, our ongoing work for this project will include repeating our methods on a larger data set with higher expected power and then using the SDS statistic to test whether the identified sexual dimorphism-associated genetic loci have been affected by positive selection in recent human evolution and if so, in what aggregate direction for each trait.

References & Acknowledgements

1. Wolfe LD and Gray JP. (1982) Subsistence practices and human sexual dimorphism of stature. *JHE,* 11, 575-580.
2. Randall JC, *et al.* (2013) Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in loci for anthropometric traits. *PLOS Gent.,* 9, e1003500.
3. Wood, *et al.* (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics,* 46, 1173-1186.
4. Field Y, *et al*. (2016) Detection of human adaptation during the past 2000 years. *Science,* 354, 760-764.

Acknowledgements: We thank Christina Bergey for computational assistance. We thank the authors of the Randall *et al.* paper for their assistance and clarification. We also thank the members of the Perry lab for their discussions and advice.