Patterns of recent natural selection on genetic loci associated with sexually differentiated human body size and shape phenotypes

Audrey M. Arner[a], Kathleen E. Grogan[a], Mark Grabowski[b,c], Hugo Reyes-Centeno[b,d], George H. Perry[a,b,e,*]

[a]Departments of Anthropology and Biology, Pennsylvania State University, [b]DFG Center for Advanced Studies, University of Tübingen, [c]Research Centre in Evolutionary Anthropology and Palaeoecology, Liverpool John Moores University, [d]Department of Anthropology & William S. Webb Museum of Anthropology, University of Kentucky, [e]Huck Institutes of the Life Sciences, Pennsylvania State University

Audrey M. Arner, ORCID iD: 0000-0001-8882-399X
Kathleen E. Grogan, ORCID iD: 0000-0002-0112-1766
Mark Grabowski, ORCID iD: 0000-0001-7045-9472
Hugo Reyes-Centeno, ORCID iD: 0000-0002-0179-3941
George H. Perry, ORCID iD: 0000-0003-4527-3806

*Corresponding author: G.H.P (ghp3@psu.edu)

Keywords: GWAS, sexual dimorphism, UK Biobank, natural selection, height

Author Contributions: GHP and AMA conceived of the project with input from KEG, MG, and HR-C. AMA performed all of the analyses with input from all other authors. All authors discussed the results. AMA, KEG, and GHP wrote the manuscript with input from MG and HR-C.

Abstract

Levels of sex differences for human body size and shape phenotypes are hypothesized to have adaptively reduced following the adoption of agriculture. In this study, we tested this hypothesis by first identifying thousands of single nucleotide polymorphisms (SNPs) that differentially impact trait variation between British females and males for five phenotypes: height, body mass, hip circumference, body fat percentage, and waist circumference. After confirming the biological plausibility of these SNPs, we then used a population genomic approach to study the recent (within the last 3,000 years; post-agriculture adoption in Britain) evolutionary histories of these loci. We observed significant increases in the frequencies of alleles associated with greater body fat percentage in females (P=0.0038; FDR=0.038), directionally opposite to that predicted by the sex differentiation reduction hypothesis. Otherwise, we found no evidence of positive selection for sex difference-associated alleles for any other trait. Overall, our results challenge the longstanding hypothesis that sex differences adaptively decreased following subsistence transitions from hunting and gathering to agriculture.

Main Text

Introduction

In many vertebrate species, it is not uncommon for morphological phenotypes to have average size and shape differences between females and males (Fairbairn, 1997). Such traits are often generally referred to as 'sexually dimorphic.' However, that term is technically specific to non-overlapping traits; for example, exaggerated ornamentation in male guppies, peacocks, and mandrills (Khramtsova, Davis, & Stranger, 2019; Loyau, Jalme, & Sorci, 2005; Setchell & Jean Wickings, 2005; Sharma et al., 2014). Traits with average phenotype values that differ by sex but with overlapping trait distributions – such as human height and body fat percentage – are better described as sexually differentiated traits.

Some sexually differentiated traits are believed by many researchers to be the result of sexual selection. In species with high inter-male competition for mates, larger males may have a competitive advantage that results in increased fitness (Plavcan, 2001). Perhaps as a result, the magnitude of sexually differentiated phenotypes are often greater in polygynous species with high competition (e.g., gorillas) and lower in monogamous species (e.g., gibbons) (Lindenfors & Tullberg, 1998; Morris, Cunningham, & Carrier, 2019; Plavcan, 2001). Large ranges in the degree of body size and shape sexually differentiated traits are repeatedly observed among closely related groups of species, suggesting the potential for relatively rapid evolution (Plavcan, 2001).

In humans, females and males exhibit significant but relatively subtle differences in many anthropometric phenotypes (Sidorenko et al., 2019). For example, European and African males are an average of approximately 9% taller in height and 15% heavier in body mass than females from the same populations (Ruff, 2002; Stulp & Barrett, 2016). Humans also exhibit sexually differentiated biometric and disease phenotypes, especially those related to immune function (McCombe, Greer, & Mackay, 2009; Natri, Garcia, Buetow, Trumble, & Wilson, 2019).

There is little consensus regarding the evolutionary history of sexually differentiated traits in our species (Dunsworth, 2020; Stulp & Barrett, 2016; Touraille, 2013). Current levels of human sex differences may partially reflect an evolutionary history of complex interactions between our biology and culture, but the timing and direction of any adaptive changes in these patterns are debated. For example, it has been hypothesized that the degree of sexual differentiation for body size phenotypes in Europe likely decreased following recent (within the past ~10,000 years) shifts from a foraging-based subsistence strategy to one relying primarily on food production, in response to less pronounced divisions of labor and mobility (Frayer, 1980). However, other scholars suggest that sexually differentiated phenotypes would respond to selection at too slow an evolutionary rate to respond to recent environmental and cultural changes; thus, any such recent changes are more likely to reflect genetic drift and/or non-genetic responses to environmental changes rather than natural selection (Rogers & Mukherjee, 1992).

In this study, we combined genomic and evolutionary analyses to quantify how recent changes in lifestyle and culture might be affecting the underlying genetic basis of human sexually

differentiated traits. First, we applied a genome wide association study (GWAS)-based approach to identify genetic variants associated with sexually differentiated phenotypes based on data from the UK Biobank study (Sudlow et al., 2015). With the recent availability of a greatly increased participant sample size, this analysis represents a powerful extension of the several previous GWAS-based approaches that have studied the genetic architecture of human sexually differentiated anthropometric traits (Randall et al., 2013). We then used the Singleton Density Score (SDS), a statistic that identifies signatures of polygenic adaptation that acted within the last ~3,000 years (Field et al., 2016), a period following the transition to agriculture in present-day United Kingdom (Woodbridge et al., 2014).

Results

We analyzed genome-wide association study (GWAS) summary statistics for the following five sexually differentiated anthropometric phenotypes that were produced by the Neale Lab ("http://www.nealelab.is/uk-biobank/," n.d.) using data from the UK Biobank (Sudlow et al., 2015): height, body mass, hip circumference, body fat percentage, and waist circumference. We chose these traits given their relevance to the motivating evolutionary hypothesis and because they have been extensively studied from anthropological and/or genomics perspectives (Gray & Wolfe, 1980; Rawlik, Canela-Xandri, & Tenesa, 2016; Sidorenko et al., 2019; Wood et al., 2014). We did not include body mass index following concerns about this metric (Krakauer & Krakauer, 2012).

Identification of phenotype-associated SNPs

We analyzed summary statistics from GWAS that were performed separately in ~194,000 females and ~167,000 males of white British genetic ancestry on ~13.8 million autosomal SNPs ("http://www.nealelab.is/uk-biobank/," n.d.). SNPs with minor allele frequencies < 0.05 and low imputation quality were filtered out. We further restricted our analysis to SNPs that 1) passed these filters in both females and males and 2) had SDS values available (Field et al., 2016), given our motivation to conduct downstream evolutionary analyses. These filtering steps resulted in ~4.4 million genome-wide SNPs for each sex-stratified GWAS. For each phenotype, we identified significantly phenotype-associated SNPs present in females, males, or both using the genome-wide significance threshold of $P = 5x10^{-8}$ commonly applied in UK Biobank studies (de Kovel & Francks, 2019; Loh et al., 2018; Rask-Andersen, Karlsson, Ek, & Johansson, 2019). We identified the following total (not yet pruned for linkage disequilibrium) numbers of phenotype-associated SNPs significant in either females, males, or both: 67,738 for height, 15,669 for body mass, 12,580 for hip circumference, 10,538 for body fat percent, and 7,765 for waist circumference (**Figure 1A-B**; Supplementary Table 1).

SNPs disproportionately associated with female or male trait variation

From the sets of phenotype-associated SNPs that were significant in females, males, or both for each phenotype (SexDiff-associated SNPs), we identified those SNPs with significant differences in the statistical strengths of association for the female vs. male-specific GWAS results (t-SexDiff) (Randall et al., 2013). To account for multiple testing, we performed this analysis with four different false-discovery rate (FDR) cutoffs for each phenotype: 0.05, 0.01, 0.005, and 0.001.

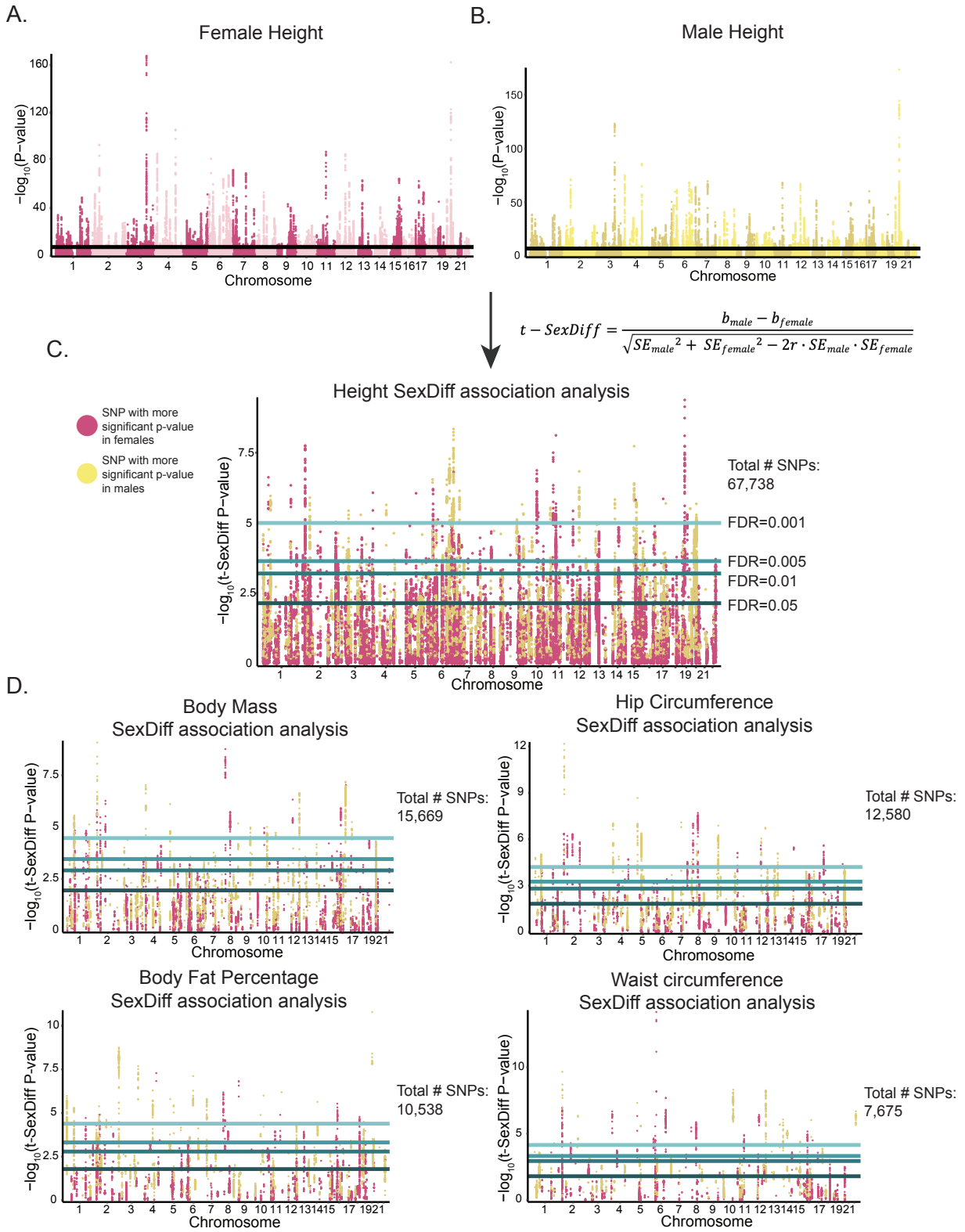

Figure 1. SexDiff-associated SNPs for five anthropometric phenotypes. (A) Manhattan plot depicting -$\log_{10}$ P-values for the association of each genome-wide SNP with female height. The black line corresponds to the genome-wide significance threshold of $P=5x10^{-8}$. (B) Manhattan plot for SNP associations with male height. (C) For each of the 67,738 SNPs significantly associated female and/or male height, we used the equation shown to test whether the SNP was disproportionately associated with height between the sexes (height SexDiff-associated SNPs). The plot depicts -$\log_{10}$ P-values for the t-SexDiff statistic. Green bars correspond to four different FDR cutoffs. (D) SexDiff association analyses for significant phenotype-associated SNPs (number of SNPs included in each analysis is shown to the right of each plot) for four additional anthropometric traits: body mass, hip circumference, body fat percentage, and waist circumference.

At the most stringent FDR cutoff of 0.001, we identified the following number of SexDiff-associated SNPs: 677 for height, 541 for body mass, 808 for hip circumference, 439 for body fat percentage, and 551 for waist circumference (**Figure 1C-D**; Supplementary Table 1). In addition to identifying SexDiff-associated SNPs, for each phenotype we calculated the ratio of SexDiff-associated SNPs at the FDR threshold of 0.001 to the number of phenotype-associated SNPs, which ranged from 0.0010 for height to 0.0718 for waist circumference.

Genes with known roles in sexual differentiation are enriched for SexDiff-associated SNPs

Prior to conducting evolutionary analyses on the anthropometric trait SexDiff-associated SNPs, we assessed their biological plausibility by testing whether these SNPs are significantly more likely to be located within or nearby (+/- 10,000 base pairs) genes previously known to be involved in sexual differentiation (Gene Ontology term GO:0007548) compared to SNPs that are significantly associated with the phenotype but not significantly associated with sex differences (Ashburner et al., 2000; The Gene Ontology Consortium, 2019). Although not all SexDiff-associated SNPs are expected to be located nearby genes already known to be involved in sexual differentiation, we would expect at least some enrichment if we are identifying true SexDiff associations with our approach.

For each t-SexDiff FDR threshold, we determined the number of unique sexual differentiation (GO:0007548) genes with one or more co-localized SexDiff-associated SNPs. We similarly identified the number of all genes (of those in the GO database) that were co-localized with at least one SexDiff-associated SNP and calculated the proportion of the number of GO:0007548 genes to the total number of genes. We conducted the same analysis for the set of SNPs associated with the phenotypes in general but *not* associated with sexual differentiation. Finally, we estimated the GO:0007548 enrichment ratio for SexDiff-associated to non SexDiff-associated SNPs by dividing the two proportions.

For example, at our most stringent FDR threshold (0.001), the 3,016 total SexDiff-associated SNPs were located within or nearby 9 unique GO:0007548 genes and 162 total genes (to be included in Dryad Digital Repository deposition). Thus, the proportion of GO:0007548 genes = 0.0555 (9/162). In contrast, the 100,646 non-SexDiff-associated but still phenotype-associated SNPs for this same FDR threshold were located within or nearby 52 unique GO:0007548 genes and 2,544 total genes, for a proportion of 0.0204, resulting in an enrichment ratio = 2.72 (0.0555/0.0204). In other words, GO:0007548 genes with known roles in sexual differentiation are >2.7 times more likely to be co-localized with one or more SexDiff-associated SNPs than with one or more non-SexDiff-associated but still phenotype-associated SNPs at the FDR<0.001 analysis level.

We repeated this analysis for each FDR threshold (**Figure 2A**; Supplementary Table 2). The enrichment ratio steadily increased with increasingly stringent FDR significance thresholds. This pattern is consistent with expectations if our analyses are identifying true SexDiff-associated loci.

To test whether the proportion of GO:0007548 genes co-localized with at least one SexDiff-associated SNP was significantly greater than expected by chance, we used a simple permutation scheme. From the set of 2,191 total GO classified genes that were co-localized with at least one

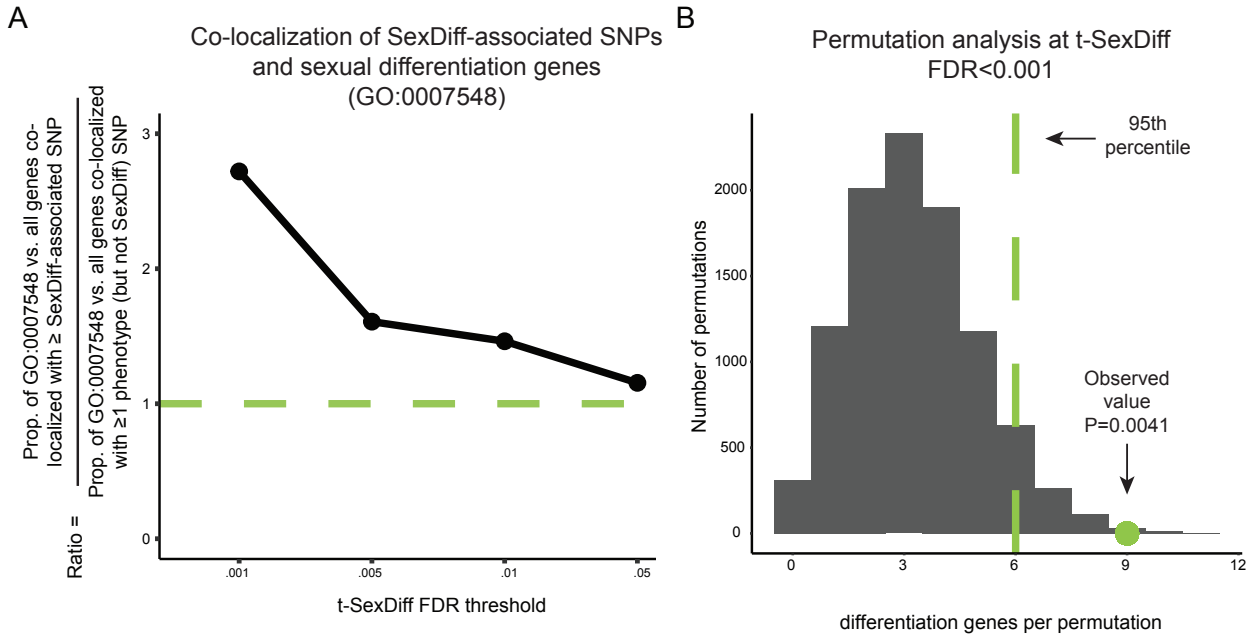

Figure 2. Genes involved in sexual differentiation (GO:0007548) are enriched for SexDiff-associated SNPs. (A) For each t-SexDiff FDR threshold, we computed the proportion of the number of genes in the "Sexual Differentiation" Gene Ontology category (GO:0007548) to the number of all Gene Ontology genes with at least one co-localized SexDiff-associated SNP (+/- 10,000 base pairs). We also computed the same proportion for the set of SNPs significant associated with our studied phenotypes in general but *not* with sexual differentiation. Values shown are ratios of these two proportions at each t-SexDiff FDR threshold. The green line indicates the 1:1 ratio that would be expected in the absence of any disproportionate co-localization between SexDiff-associated SNPs and GO:00007548 genes. (B) Permutation analysis of the number of genes involved in sexual differentiation co-localized with at least one SexDiff-associated SNPs at our most stringent FDR threshold (q<0.001). From the set of 2,191 total GO-classified genes that were co-localized with at least one phenotype-associated SNP, we randomly selected 162 genes, the number of total genes that were co-localized with one or more SexDiff-associated SNPs at the FDR<0.001 cutoff. Of these 162 genes, we counted and recorded the number of GO:0007548 genes represented. We repeated this process 10,000 times and computed an empirical p-value (P=0.0041) as the proportion of permutations with a greater than or equal number of GO:0007548 genes as the observed value for FDR<0.001 SexDiff-associated SNPs (9 genes). Results for similar analyses based on SexDiff FDR thresholds 0.005, 0.01, and 0.05 are shown in Supplementary Figure 1.

phenotype-associated SNP, we randomly selected the same number of total genes that were co-localized with one or more SexDiff-associated SNPs at the FDR cutoff being considered (e.g. 135 genes at FDR<0.001) and counted the number of GO:0007548 genes represented. We repeated this process 10,000 times for each FDR cutoff and compared the resulting distributions to the actual number of GO:0007548 genes co-localized with at least one SexDiff-associated SNP to calculate empirical P-values. The observed proportion of GO:0007548 genes is very unlikely to be explained by chance at the most stringent FDR<0.001 threshold (P=0.0041; **Figure 2B**), with increasing probability for the FDR thresholds of 0.005, 0.01, and 0.05 (P=0.0513, P=0.0829, and P=0.2111, respectively; Supplementary Figure 1).

Altogether, these results suggest that the anthropometric trait SexDiff-associated SNPs we identified – especially those discovered with the most stringent FDR<0.001 threshold – are at least enriched for SNPs that do underlie (or are linked to those that do) sexual differentiation variation. All subsequent analyses were limited to SNPs classified based on the FDR<0.001 analysis.

SexDiff-associated SNPs are associated with greater trait effect sizes in the expected sex

Our statistical approach technically identifies genetic variants with significantly different strengths of association for a given trait between males and females. Before proceeding, we sought to confirm that these identified variants are also disproportionately associated with greater phenotype effect sizes in the sex with the stronger association value.

To do so, for each phenotype we first divided our SexDiff-associated SNPs into those with disproportionately stronger associations in females vs. males. To avoid linkage disequilibrium (LD)-based result inflation, we pruned each of these sets of SNPs to include only the one SNP (if any) with the lowest t-SexDiff FDR value located within each of 1,703 approximately LD-independent blocks of the human genome (Pickrell et al., 2016). For comparison to the disproportionate female and male SexDiff-associated SNPs, we prepared similar LD-pruned sets of all SNPs significantly associated with the phenotype for each of the five traits (to be included in Dryad Digital Repository deposition). Then, for every SNP in each set of pruned SNPs, we calculated the $\log_2$ ratio of the female trait effect size to the male trait effect size.

Each of the five sets of female SexDiff-associated SNPs had strongly positive average ratios, meaning that the female effect sizes for these SNPs were larger than the male effect sizes, and *vice versa* for the male SexDiff-associated SNPs (**Figure 3A**). Mean $\log_2$ ratios were significantly different from 0 for each of the 10 SexDiff-associated SNP subset (one-sided t-tests; $P=3.3\times10^{-5}$ and lower; $FDR=3.3\times10^{-5}$ and lower; Supplementary Tables 3 and 4). Conversely, the mean $\log_2$ ratios for the five sets of general trait-associated SNPs were near 0 (Figure 3A). Mean $\log_2$ ratios for each of the 10 SexDiff-associated SNPs were also significantly different from the corresponding mean $\log_2$ ratios of the phenotype-associated SNPs (Permutation analyses [see methods]; P<0.001, FDR=0.001). These results confirm that our statistical process successfully and reliably identifies SNPs with sex disproportionate effects on phenotypic trait variation. We thus proceeded to study the evolutionary histories of these SNPs.

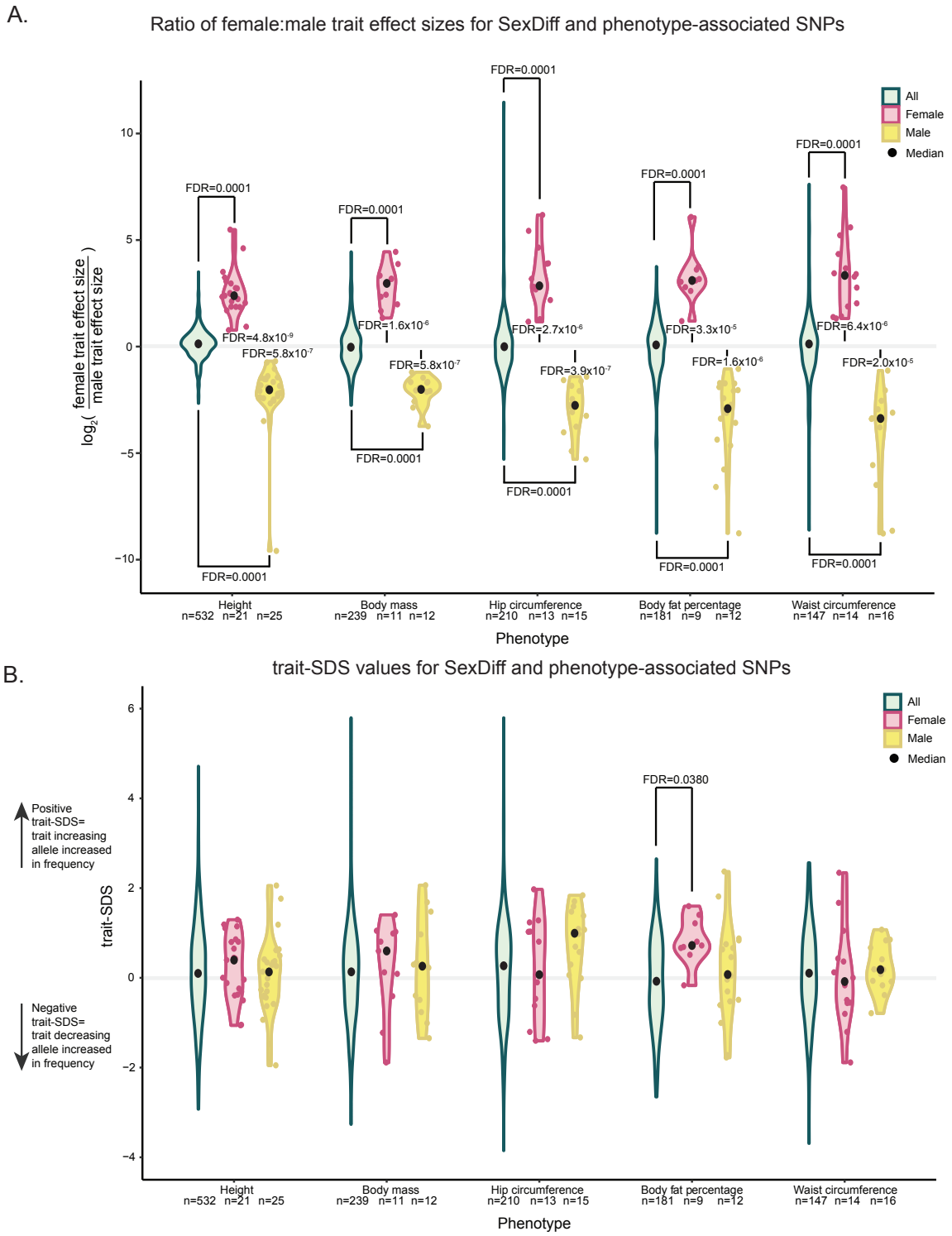

Figure 3. Sex-specific effect size ratios and trait-SDS scores for anthropometric trait-associated SNPs. For each anthropometric trait, a maximum of one SNP per each of 1,703 approximately LD-independent blocks of the human genome was included (within each block, the SNP with the strongest statistical significance) from the sets of i) all SNPs associated with trait variation, and the subsets of those SNPs disproportionately associated with variation in ii) females and iii) males (SexDiff-associated SNPs at FDR<0.001). (A) Violin plots of $\log_2$ ratios of female trait effect size to male effect size (calculated separately for each SNP). Mean $\log_2$ ratios for each set of female and male SexDiff pruned SNPs were compared to those for the corresponding phenotype-association set using a permutation scheme, and to 0 using a one-sided t-test. (B) Violin plots of trait-SDS values. Positive trait-SDS values reflects recent increases in the frequencies of phenotype-increasing alleles, while negative trait-SDS values reflect increases in the frequencies of phenotype-decreasing alleles. The trait-SDS distributions for each set of female and male SexDiff pruned SNPs were compared to those for the corresponding phenotype-association set using a permutation analysis.

Signatures of positive selection on SexDiff-associated alleles

We next tested whether the SexDiff-associated SNPs identified in our analyses are significantly enriched for signatures of recent (~3,000 years) positive selection, using the Singleton Density Score statistic (SDS) (Field et al., 2016). Briefly, alleles affected by recent positive selection are predicted to be found on haplotypes with relatively fewer singleton mutations; the SDS quantifies this pattern.

In turn, the trait-SDS statistic reflects directionality with respect to an associated phenotype. A positive trait-SDS value reflects a recent increase in the frequency of a phenotype-increasing allele, while a negative trait-SDS value reflects an increase in the frequency of a phenotype-decreasing allele (Field et al., 2016). We performed our analyses with previously-published SDS data computed from the whole genome sequences of 3,195 British individuals from the UK10K project (Field et al., 2016; Walter et al., 2015). This dataset is ideal for integration with the UK BioBank genotype-phenotype association data, given study population similarity.

To examine the evolutionary histories of alleles associated with anthropometric trait sex differences, we again considered the sets of LD-pruned SexDiff-associated SNPs divided into those that are disproportionately associated with trait variation in females vs. those disproportionately associated with trait variation in males for each phenotype. We tested whether the average trait-SDS values for the pruned female or male SexDiff-associated SNPs were significantly different than the value for the corresponding pruned set of all phenotype-associated SNPs (**Figure 3B;** Supplementary Table 6), using a permutation analysis. For example, there were n=21 pruned SexDiff-associated SNPs disproportionately associated with female height and n=532 pruned SNPs associated with height generally (but not necessarily with sex differences). We randomly selected 21 of the 532 pruned height-associated SNPs and calculated the average trait-SDS score. We repeated this process 10,000 times and calculated an empirical P-value based on the proportion (multiplied by two; see Methods) of permuted observations with equal or more extreme average trait-SDS values than the observed average trait-SDS for the actual female height SexDiff-associated SNPs (average trait-SDS=0.240; P=0.66). This test was then repeated for the male SexDiff-associated SNPs for height, and the female and male SexDiff-associated SNPs for each of the remaining four anthropometric traits.

For nine of the comparisons, trait-SDS distributions for the sets of pruned SexDiff-associated SNPs were not significantly different (following FDR adjustment for multiple tests) from those for SNPs associated with the corresponding general phenotype (**Figure 3B**; Supplementary Table 5). However, compared to all body fat percentage-associated SNPs (n=181 pruned SNPs; average trait-SDS=-0.124), the average trait-SDS value for the set of female SexDiff-associated SNPs for this phenotype was significantly elevated (n=9 pruned SNPs; average trait-SDS=0.827; P=0.0038; FDR=0.038). In other words, the average frequencies of alleles associated with greater body fat percentage in females have increased in frequency over the past ~3,000 years at a faster rate than expected based on the pattern for SNPs associated with body fat percentage generally, a potential signature of polygenic selection on this sub-trait.

Discussion

Using a sex-stratified GWAS framework for five sexually differentiated anthropometric phenotypes, we identified 3,016 SNPs that were disproportionately associated with either female or male trait variation at a low false discovery rate (FDR<0.001). We confirmed the biological plausibility of these results by showing that genes with known roles in sexual differentiation are significantly enriched for SexDiff-associated SNPs. Together, these results confirm the importance of considering sex differences when investigating the genetic structure of human polygenic traits (Clayton, 2018). We then used a statistic that quantifies changes in the frequencies of alleles underlying polygenic traits over the past ~3,000 years to identify a signature of recent positive selection on SNPs associated with increased female body fat percentage in the British study population.

We must emphasize that inferring selection signals from GWAS data should be approached with great care, as even subtle uncorrected population structure can impact GWAS and downstream results (Berg et al., 2019). For example, data from the GIANT consortium were previously used to identify strong signatures of polygenic selection for height across the genome (Field et al., 2016). However, subtle population structure in the GIANT sample led to effect-size estimate biases, in turn resulting in false signals of polygenic selection for SNPs not crossing the genome-wide significance threshold and impacting results for significant SNPs as well (Berg et al., 2019). In contrast, these issues were much less prevalent using GWAS summary statistics from the UK Biobank, in which population structure is minimized (Barton, Hermisson, & Nordborg, 2019; Berg et al., 2019; Sohail et al., 2019). In light of these considerations, in our study we have i) used UK Biobank GWAS summary statistics only, ii) focused solely on phenotype-associated SNPs below the genome-wide significance threshold, and iii) restricted our evolutionary analyses to direct comparisons between SNPs significantly associated with individual phenotypes and a sub-phenotype (i.e., sex differences).

Our study further demonstrates the value of GWAS-based approaches for testing anthropological hypotheses (Vukelic, Cohen, Sullivan, & Perry, 2017). Concerning the evolution of human body size and shape phenotypes, our results fail to provide support for the prevailing notion of recent (i.e. subsequent to agriculture) adaptive reductions in levels of sex differences for such traits. Specifically, using large samples of genomes from British individuals we did not observe significant differences in the recent evolutionary trajectories of SNPs disproportionately associated with female or male variation in height, body mass, hip circumference, and waist circumference relative to the trajectories of SNPs associated with these traits generally.

Moreover, we found evidence that the average frequencies of alleles disproportionately associated with greater female body fat percentage significantly increased over the past 3,000 years, a pattern consistent with polygenic adaptation. Since SNPs can be pleiotropically associated with multiple phenotypes (Pickrell et al., 2016), we cannot definitively conclude that positive selection acted directly on female body fat percentage. Regardless, at the very least there is still no support for the prevailing hypothesis concerning the decrease of sex differences in recent human evolution. That is, given that females have higher average body fat percentages

than men in historic and contemporary populations, the direction of polygenic adaptation in the population we studied would actually be *opposite* to expectations under hypotheses of recent adaptive reductions in anthropometric trait sex differences in agricultural societies.

Methods

Subjects and dataset generation

We used genome-wide association study (GWAS) summary statistics generated from analyses of UK Biobank data ("http://www.nealelab.is/uk-biobank/," n.d.; Sudlow et al., 2015). The original GWAS analyses were restricted to 361,194 unrelated individuals (194,174 females and 167,020 males) of white British ancestry (based on the combination of self-report and genetic ancestry analysis) who did not have sex chromosome aneuploidies. GWAS summary statistics for each phenotype were computed separately for females and males. Because every phenotype was not defined for all individuals, some of the analyses contributing to our study included fewer than 361,194 individuals (Supplementary Table 7). Approximately 40 million SNPs were originally available for analysis, of which only those with minor allele frequencies > 0.001, an INFO score (imputation quality) > 0.8, and a Hardy-Weinberg Equilibrium p-value > $1x10^{-10}$ were retained, resulting in datasets of 13,791,468 SNPs.

We then subsequently filtered these SNPs with minor allele frequencies < 0.05 and those that were not associated with a singleton density score (SDS; for description, see below) (Field et al., 2016). These filters resulted in a total genome-wide dataset of ~4.4 million SNPs for each phenotype.

We chose five sexually differentiated anthropometric phenotypes from the GWAS summary statistics for analysis (Supplementary Table 7). To identify phenotype-associated SNPs in female and/or male individuals, we applied the commonly used genome wide significance threshold (i.e., to account for the large number of tested SNPs) of $P=5x10^{-8}$ to the male-specific and female-specific GWAS summary statistics (to be included in Dryad Digital Repository deposition).

Scan for SNPs that are disproportionately associated with male or female trait variation

For each genome-wide SNP significantly associated with a phenotype in females, males, or both sexes, we evaluated whether there was a significant difference in the statistical strengths of association for the female vs. male-specific GWAS results using the following t-statistic (t-SexDiff) (Randall et al., 2013).

$$t - SexDiff = \frac{b_{male} - b_{female}}{\sqrt{SE_{male}^2 + SE_{female}^2 - 2r * SE_{male} * SE_{female}}}$$

B_{male} refers to the male-specific beta value and B_{female} refers to the female-specific beta value for each genome-wide SNP. SE_{male} and SE_{female} refer to the male-specific and female-specific standard error for each genome-wide SNP, respectively. The correlation *r* (rho) for each phenotype was calculated as the Spearman rank correlation coefficient between b_{male} and b_{female} using the

cor.test() function in R (version 3.5.1). t-SexDiff was converted to a two-sided P-value using the R function pt(). The effects of multiple testing were considered by computing the False Discovery Rate (FDR) for each t-SexDiff P-value using the R function p.adjust (Benjamini & Hochberg, 1995).

Of the phenotype-associated SNPs that were significantly associated with a given phenotype for females, males, or both sexes, we identified SexDiff-associated SNPs at four different FDR thresholds: <0.05, <0.01, <0.005, and <0.001 (to be included in Dryad Digital Repository deposition).

Assessing the biological plausibility of the SexDiff-associated SNPs

In order to confirm the biological plausibility of our SexDiff-associated SNPs, we tested whether regions of the genome functionally linked to sexual differentiation are more likely than expected by chance to contain one or more of our SexDiff-associated SNPs. Separately for each tSexDiff FDR threshold, we counted the number of unique genes in the Gene Ontology database (Ashburner et al., 2000; The Gene Ontology Consortium, 2019) that contained at least one SexDiff-associated SNP, including within a +/-10,000 base-pair (bp) window around the gene to encompass potential regulatory regions. We then counted the number of these genes with known links to processes of sexual differentiation corresponding to the Gene Ontology (GO) term GO:0007548 and computed the proportion of the numbers of GO:0007548 to all genes co-localized with >= 1 SexDiff-associated SNP. We repeated this analysis for significant phenotype- but not SexDiff-associated SNPs. In the absence of enrichment for SexDiff-associated SNPs, the ratio of these two proportions is expected to equal one.

We then used the following permutation scheme for each t-SexDiff FDR cutoff. There was a total of 2,191 GO-classified genes overlapping one or more phenotype-associated SNP (whether SexDiff-associated or not; each t-SexDiff FDR cutoff started with the same genome-wide significant set of phenotype-associated SNPs so the total number of 2,191 co-localized genes applies to each FDR cutoff). Given the number of these genes overlapping >=1 SexDiff-associated SNP for a given FDR cutoff, we randomly selected that number of unique genes from the pool of 2,191 genes and counted the number of GO:0007548 genes. We then repeated this procedure 10,000 times and computed an empirical P-value as the proportion of permuted data sets with an equal or greater to number of sexual differentiation genes when compared to our observation for the associated FDR threshold.

Assessing the sex-specific effects of SexDiff-associated SNPs

For each phenotype we split our SexDiff-associated SNPs into those that had lower P-values (as identified in the original sex specific GWAS for phenotype-associated SNPs) in females than males (female SexDiff-associated SNPs) and those that had lower p-values in males than females (male SexDiff-associated SNPs). We separately pruned each set of female SexDiff-associated SNPs and male SexDiff-associated SNPs to account for linkage disequilibrium. Specifically, if there was more than one female SexDiff-associated SNP in one of 1,703 approximately LD-independent segments of the genome (Pickrell et al., 2016), we only kept the female SexDiff-associated SNP for that segment with the lowest p-value for association with the phenotype in females. We did the same

for male SexDiff-associated SNPs in males. For each set of pruned SexDiff-associated SNPs, we then calculated the $\log_2$ ratio of the female effect size to male effect size.

The set of phenotype-associated SNPs was pruned in a similar fashion to the above, with a maximum of one SNP per each of the 1,703 approximately LD-independent genome segments, chosen as the SNP with the most significant p-value in the segment regardless of whether it was most significant in females or males (to be included in Dryad Digital Repository deposition) from among those below the genome-wide significance threshold.

We then used a permutation to estimate the probability that the $\log_2$ ratio of female effect size to male effect size was significantly for each of the ten sets of SexDiff-associated SNPs. From the pruned set of phenotype-associated SNP, we randomly selected a number of SNPs equal to the number of observed SexDiff-associated SNPs for that phenotype. We repeated this procedure 10,000 times and computed the proportion of permuted data sets with equal or more extreme average $\log_2$ ratio. We computed FDR values to account for multiple testing.

Finally, we used a one-sided t-test to determine whether our $\log_2$ ratios for each set of SexDiff-associated SNPs was significantly different from zero. We again computed FDR values to account for multiple testing

Identification of signatures of positive selection

To test the hypothesis that SexDiff-associated SNPs have been affected by recent (past ~3,000 years) positive selection in recent human evolution, we used the Singleton Density Score (SDS) statistic (Field et al., 2016). We used a database of genome-wide SNP SDS scores that were computed (Field et al., 2016) using 3,195 whole genome sequences from British individuals the UK10K project (Walter et al., 2015). For the pruned sets of female and male SexDiff-associated SNPs, we fixed the sign of SDS scores so that positive values indicate an increased frequency of the trait-increasing allele.

We used permutations to estimate the probability that the average trait-SDS value for each trait and sex could be observed by chance given the distribution trait-SDS scores for SNPs significantly associated with the corresponding phenotype (regardless of SexDiff-association). From the pruned set of phenotype-associated SNPs, we then randomly selected a number of SNPs equal to the number of observed SexDiff-associated SNPs for that phenotype and the sex and FDR threshold being considered, and we calculated the average trait-SDS score for that set of SNPs. We repeated this procedure 10,000 times and computed the proportion of permuted data sets with equal or more extreme average trait-SDS scores compared to actual result for the observed SexDiff-associated SNPs. This proportion was then multiplied by two to account for the two-tailed nature of this test (i.e. the average trait-SDS values for SexDiff-associated SNPs could have been significantly greater than or less than that for the phenotype-associated SNPs). We computed FDR values to account for multiple testing.

Computational resources

All analyses were conducted in R (version 3.5.1) with Advanced CyberInfrastructure computational resources provided by The Institute for CyberScience at Pennsylvania State University. All scripts are available at https://github.com/audreyarner/dimorphism-evolution.

Data availability

Data from all steps of our analyses will be made available on dryad prior to publication of this manuscript.

Acknowledgments

We thank the DFG Center for Advanced Studies "Words, Bones, Genes, Tools" for their support and S. Marciniak, C. Bergey, M. Hernandez, J. Tung, and D. Puts and the Sex Differences Interest Group for their helpful discussions. This work was funded by NIH grant R01-GM115656 (to G.H.P.); NIH grant F32-GM123634 (to K.E.G.); and DFG grant FOR-22337. A.M.A.'s work on this project was supported by Erickson Discovery, Presidential Leadership Academy Enrichment, and Liberal Arts Enrichment grants from Penn State University. This research was conducted with Advanced CyberInfrastructure computational resources provided by The Institute for CyberScience at Penn State.

References

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Barton, N., Hermisson, J., & Nordborg, M. (2019). Why structure matters. *ELife*, *8*, e45380. https://doi.org/10.7554/eLife.45380

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., … Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *ELife*, *8*, e39725. https://doi.org/10.7554/eLife.39725

Clayton, J. A. (2018). Applying the new SABV (sex as a biological variable) policy to research and clinical care. *Physiology & Behavior*, *187*, 2–5. https://doi.org/10.1016/j.physbeh.2017.08.012

de Kovel, C. G. F., & Francks, C. (2019). The molecular genetics of hand preference revisited. *Scientific Reports*, *9*(1), 5986. https://doi.org/10.1038/s41598-019-42515-0

Dunsworth, H. M. (2020). Expanding the evolutionary explanations for sex differences in the human skeleton. *Evolutionary Anthropology: Issues, News, and Reviews*, 1–9. https://doi.org/10.1002/evan.21834

Fairbairn, D. J. (1997). Allometry for sexual size dimorphism: pattern and process in the coevolution of body size in males and females. *Annual Review of Ecology and Systematics*, *28*(1), 659–687. https://doi.org/10.1146/annurev.ecolsys.28.1.659

Field, Y., Boyle, E. A., Telis, N., Gao, Z., Gaulton, K. J., Golan, D., … Pritchard, J. K. (2016). Detection of human adaptation during the past 2000 years. *Science*, *354*(6313), 760–764. https://doi.org/10.1126/science.aag0776

Frayer, D. W. (1980). Sexual dimorphism and cultural evolution in the Late Pleistocene and Holocene of Europe. *Journal of Human Evolution*, *9*(5), 399–415. https://doi.org/10.1016/0047-2484(80)90050-0

Gray, J. P., & Wolfe, L. D. (1980). Height and sexual dimorphism of stature among human societies. *American Journal of Physical Anthropology*, *53*(3), 441–456. https://doi.org/10.1002/ajpa.1330530314

http://www.nealelab.is/uk-biobank/. (n.d.). Retrieved April 5, 2020, from http://www.nealelab.is/uk-biobank/

Khramtsova, E. A., Davis, L. K., & Stranger, B. E. (2019). The role of sex in the genomics of human complex traits. *Nature Reviews Genetics*, *20*(3), 173–190.

https://doi.org/10.1038/s41576-018-0083-1

Krakauer, N. Y., & Krakauer, J. C. (2012). A New Body Shape Index Predicts Mortality Hazard Independently of Body Mass Index. *PLoS ONE*, *7*(7), e39504. https://doi.org/10.1371/journal.pone.0039504

Lindenfors, P., & Tullberg, B. S. (1998). Phylogenetic analyses of primate size evolution: the consequences of sexual selection. *Biological Journal of the Linnean Society*, *64*(4), 413–447. https://doi.org/10.1006/bijl.1998.0237

Loh, P. R., Genovese, G., Handsaker, R. E., Finucane, H. K., Reshef, Y. A., Palamara, P. F., … Price, A. L. (2018). Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature*, *559*(7714), 350–355. https://doi.org/10.1038/s41586-018-0321-x

Loyau, A., Jalme, M. Saint, & Sorci, G. (2005). Intra- and intersexual selection for multiple traits in the peacock *(Pavo cristatus)*. *Ethology*, *111*(9), 810–820. https://doi.org/10.1111/j.1439-0310.2005.01091.x

McCombe, P. A., Greer, J. M., & Mackay, I. R. (2009). Sexual dimorphism in autoimmune disease. *Current Molecular Medicine*, *9*(9), 1058–1079. https://doi.org/10.2174/156652409789839116

Morris, J. S., Cunningham, C. B., & Carrier, D. R. (2019). Sexual dimorphism in postcranial skeletal shape suggests male-biased specialization for physical competition in anthropoid primates. *Journal of Morphology*, *280*(5), 731–738. https://doi.org/10.1002/jmor.20980

Natri, H., Garcia, A. R., Buetow, K. H., Trumble, B. C., & Wilson, M. A. (2019). The pregnancy pickle: Evolved immune compensation due to pregnancy underlies sex differences in human diseases. *Trends in Genetics*, *35*(7), 478–488. https://doi.org/10.1016/j.tig.2019.04.008

Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y., & Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics*, *48*(7), 709–717. https://doi.org/10.1038/ng.3570

Plavcan, J. M. (2001). Sexual dimorphism in primate evolution. *American Journal of Physical Anthropology*, *116*(S33), 25–53. https://doi.org/10.1002/ajpa.10011

Randall, J. C., Winkler, T. W., Kutalik, Z., Berndt, S. I., Jackson, A. U., Monda, K. L., … Heid, I. M. (2013). Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genetics*, *9*(6), e1003500. https://doi.org/10.1371/journal.pgen.1003500

Rask-Andersen, M., Karlsson, T., Ek, W. E., & Johansson, Å. (2019). Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nature Communications*, *10*(1), 339. https://doi.org/10.1038/s41467-018-08000-4

Rawlik, K., Canela-Xandri, O., & Tenesa, A. (2016). Evidence for sex-specific genetic

architectures across a spectrum of human complex traits. *Genome Biology*, *17*(1), 166. https://doi.org/10.1186/s13059-016-1025-x

Rogers, A. R., & Mukherjee, A. (1992). Quantitative genetics of sexual dimorphism in human body size. *Evolution*, *46*(1), 226–234. https://doi.org/10.1111/j.1558-5646.1992.tb01997.x

Ruff, C. (2002). Variation in human body size and shape. *Annual Review of Anthropology*, *31*(1), 211–232. https://doi.org/10.1146/annurev.anthro.31.040402.085407

Setchell, J. M., & Jean Wickings, E. (2005). Dominance, status signals and coloration in male mandrills *(Mandrillus sphinx)*. *Ethology*, *111*(1), 25–50. https://doi.org/10.1111/j.1439-0310.2004.01054.x

Sharma, E., Künstner, A., Fraser, B. A., Zipprich, G., Kottler, V. A., Henz, S. R., … Dreyer, C. (2014). Transcriptome assemblies for studying sex-biased gene expression in the guppy, *Poecilia reticulata*. *BMC Genomics*, *15*(1), 400. https://doi.org/10.1186/1471-2164-15-400

Sidorenko, J., Kassam, I., Kemper, K. E., Zeng, J., Lloyd-Jones, L. R., Montgomery, G. W., … Visscher, P. M. (2019). The effect of X-linked dosage compensation on complex trait variation. *Nature Communications*, *10*(1), 3009. https://doi.org/10.1038/s41467-019-10598-y

Sohail, M., Maier, R. M., Ganna, A., Bloemendal, A., Martin, A. R., Turchin, M. C., … Sunyaev, S. R. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *ELife*, *8*, e39702. https://doi.org/10.7554/eLife.39702

Stulp, G., & Barrett, L. (2016). Evolutionary perspectives on human height variation. *Biological Reviews*, *91*(1), 206–234. https://doi.org/10.1111/brv.12165

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., … Collins, R. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, *47*(D1), D330–D338. https://doi.org/10.1093/nar/gky1055

Touraille, P. (2013). Human Sex Differences in Height: Evolution due to Gender Hierarchy? In M. Ah-King (Ed.), *Challenging Popular Myths of Sex, Gender and Biology* (pp. 65–75). Springer, Cham. https://doi.org/10.1007/978-3-319-01979-6_7

Vukelic, A., Cohen, J. A., Sullivan, A. P., & Perry, G. H. (2017). Extending Genome-wide Association Study Results to Test Classic Anthropological Hypotheses: Human Third Molar Agenesis and the "Probable Mutation Effect." *Human Biology*, *89*(2), 157. https://doi.org/10.13110/humanbiology.89.2.03

Walter, K., Min, J. L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., … Zhang, W. (2015). The

UK10K project identifies rare variants in health and disease. *Nature*, *526*(7571), 82–90. https://doi.org/10.1038/nature14962

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., … Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*(11), 1173–1186. https://doi.org/10.1038/ng.3097

Woodbridge, J., Fyfe, R. M., Roberts, N., Downey, S., Edinborough, K., & Shennan, S. (2014). The impact of the Neolithic agricultural transition in Britain: a comparison of pollen-based land-cover and archaeological 14C date-inferred population change. *Journal of Archaeological Science*, *51*, 216–224. https://doi.org/10.1016/j.jas.2012.10.025

Figure Captions

Figure 1. SexDiff-associated SNPs for five anthropometric phenotypes. (A) Manhattan plot depicting -$\log_{10}$ P-values for the association of each genome-wide SNP with female height. The black line corresponds to the genome-wide significance threshold of $P=5x10^{-8}$. (B) Manhattan plot for SNP associations with male height. (C) For each of the 67,738 SNPs significantly associated female and/or male height, we used the equation shown to test whether the SNP was disproportionately associated with height between the sexes (height SexDiff-associated SNPs). The plot depicts -$\log_{10}$ P-values for the t-SexDiff statistic. Green bars correspond to four different FDR cutoffs. (D) SexDiff association analyses for significant phenotype-associated SNPs (number of SNPs included in each analysis is shown to the right of each plot) for four additional anthropometric traits: body mass, hip circumference, body fat percentage, and waist circumference.

Figure 2. Genes involved in sexual differentiation (GO:0007548) are enriched for SexDiff-associated SNPs. (A) For each t-SexDiff FDR threshold, we computed the proportion of the number of genes in the "Sexual Differentiation" Gene Ontology category (GO:0007548) to the number of all Gene Ontology genes with at least one co-localized SexDiff-associated SNP (+/- 10,000 base pairs). We also computed the same proportion for the set of SNPs significant associated with our studied phenotypes in general but *not* with sexual differentiation. Values shown are ratios of these two proportions at each t-SexDiff FDR threshold. The green line indicates the 1:1 ratio that would be expected in the absence of any disproportionate co-localization between SexDiff-associated SNPs and GO:00007548 genes. (B) Permutation analysis of the number of genes involved in sexual differentiation co-localized with at least one SexDiff-associated SNPs at our most stringent FDR threshold (q<0.001). From the set of 2,191 total GO-classified genes that were co-localized with at least one phenotype-associated SNP, we randomly selected 162 genes, the number of total genes that were co-localized with one or more SexDiff-associated SNPs at the FDR<0.001 cutoff. Of these 162 genes, we counted and recorded the number of GO:0007548 genes represented. We repeated this process 10,000 times and computed an empirical p-value (P=0.0041) as the proportion of permutations with a greater than or equal number of GO:0007548 genes as the observed value for FDR<0.001 SexDiff-associated SNPs (9 genes). Results for similar analyses based on SexDiff FDR thresholds 0.005, 0.01, and 0.05 are shown in Supplementary Figure 1.

Figure 3. Sex-specific effect size ratios and trait-SDS scores for anthropometric trait-associated SNPs. For each anthropometric trait, a maximum of one SNP per each of 1,703 approximately LD-independent blocks of the human genome was included (within each block, the SNP with the strongest statistical significance) from the sets of i) all SNPs associated with trait variation, and the subsets of those SNPs disproportionately associated with variation in ii) females and iii) males (SexDiff-associated SNPs at FDR<0.001). (A) Violin plots of $\log_2$ ratios of female trait effect size to male effect size (calculated separately for each SNP). Mean $\log_2$ ratios for each set of female and male SexDiff pruned SNPs were compared to those for the corresponding phenotype-association set using a permutation scheme, and to 0 using a one-sided t-test. (B) Violin plots of trait-SDS values. Positive trait-SDS values reflects recent increases in the frequencies of phenotype-

increasing alleles, while negative trait-SDS values reflect increases in the frequencies of phenotype-decreasing alleles. The trait-SDS distributions for each set of female and male SexDiff pruned SNPs were compared to those for the corresponding phenotype-association set using a permutation analysis

Supplementary Figure 1. Permutation enrichment distribution at each FDR threshold. Permutation analysis of the number of genes involved in sexual differentiation for all anthropometric SNPs at every FDR threshold. Data are the frequency of distribution of our results for 10,000 permuted data sets. The empirical P-value represents the probability that the observed value of sexual differentiation genes from our SexDiff-associated SNP pool is equal to or greater than those from a randomly selected set.

Supplementary Table 1: Observed number of SexDiff-associated SNPs at each FDR threshold for every phenotype.

Supplementary Table 2: Observed unique sexual differentiation genes (SDG) and total number of genes for SexDiff-associated SNPs and Non SexDiff-associated SNPs

Supplementary Table 3: Observed $\log_2$ ratio of female to male beta values and p-values for each set of Female SexDiff-associated SNPs

Supplementary Table 4: Observed $\log_2$ ratio of female to male beta values and p-values for each set of Male SexDiff-associated SNPs

Supplementary Table 5: Observed trait-SDS and permutation P-values for each set of Female SexDiff-associated SNPs and Male SexDiff-associated SNPs permutated against phenotype-associated SNPs.

Supplementary Table 6. Observed trait-SDS for each set of pruned phenotype-associated SNP groups.

Supplementary Table 7: Phenotype information