

RESEARCH ARTICLE

# Sex-Specific Selection and Sex-Biased Gene Expression in Humans and Flies

Changde Cheng, Mark Kirkpatrick\*

Department of Integrative Biology, University of Texas, Austin, Texas, United States of America

\* [kirkp@mail.utexas.edu](mailto:kirkp@mail.utexas.edu)



CrossMark  
click for updates

 OPEN ACCESS

**Citation:** Cheng C, Kirkpatrick M (2016) Sex-Specific Selection and Sex-Biased Gene Expression in Humans and Flies. *PLoS Genet* 12(9): e1006170. doi:10.1371/journal.pgen.1006170

**Editor:** Michael W. Nachman, University of California, Berkeley, UNITED STATES

**Received:** March 15, 2016

**Accepted:** June 14, 2016

**Published:** September 22, 2016

**Copyright:** © 2016 Cheng, Kirkpatrick. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data used in this study were downloaded from public repositories, specifically: human SNP data: <http://www.1000genomes.org/data#download>, human expression data: <http://www.gtexportal.org/home/>, and fly expression data: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17013>.

**Funding:** This work was supported by the Swiss National Science Foundation (<http://www.snf.ch/en/funding/programmes/sinergia/Pages/default.aspx>) (Grant CRSII3-147625 to MK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Sexual dimorphism results from sex-biased gene expression, which evolves when selection acts differently on males and females. While there is an intimate connection between sex-biased gene expression and sex-specific selection, few empirical studies have studied this relationship directly. Here we compare the two on a genome-wide scale in humans and flies. We find a distinctive “Twin Peaks” pattern in humans that relates the strength of sex-specific selection, quantified by genetic divergence between male and female adults at autosomal loci, to the degree of sex-biased expression. Genes with intermediate degrees of sex-biased expression show evidence of ongoing sex-specific selection, while genes with either little or completely sex-biased expression do not. This pattern apparently results from differential viability selection in males and females acting in the current generation. The Twin Peaks pattern is also found in *Drosophila* using a different measure of sex-specific selection acting on fertility. We develop a simple model that successfully recapitulates the Twin Peaks. Our results suggest that many genes with intermediate sex-biased expression experience ongoing sex-specific selection in humans and flies.

## Author Summary

Sexual dimorphism, which is evident in virtually all phenotypic traits, results from sex-biased gene expression. The evolution of sex-biased expression, in turn, results from selection that acts differently on males and females. We use genomic data from humans and flies to quantify the relation between ongoing selection pressures and gene expression. A distinctive pattern in both species reveals that many genes are experiencing selection that is currently acting differentially on males and females. These results suggest that the “war between the sexes” is more than a metaphor: it contributes to observable patterns of mortality in humans.

## Introduction

Females and males differ for virtually all phenotypic traits [1]. Sexual dimorphism results from sex-biased gene expression, which evolves in response to selection that acts differently on males and females [2–4]. Thus sex-biased gene expression is intimately linked to sex-specific

**Competing Interests:** The authors have declared that no competing interests exist.

selection [5,6]. What is less clear, however, is the extent to which sex-specific selection is ongoing, and how ongoing selection relates to the strength of sex-biased expression.

To date, the most direct link between sex-specific selection and sex-biased expression comes from a laboratory study of *Drosophila melanogaster*. By comparing gene expression and reproductive fitness in a quantitative genetics design, Innocenti and Morrow [7] identified genes experiencing sexually-antagonistic selection, which is the extreme case of sex-specific selection in which an allele that increases fitness in one sex decreases it in the other. They concluded that 8.5% of loci that show sex-biased expression are experiencing ongoing sexually-antagonistic selection. While this work is a milestone, it leaves key questions unanswered. These include: What is the relation between the strength of ongoing sex-specific selection and sex-biased expression, and how common and how strong is sex-specific selection in natural populations? Answers to these questions are important to our understanding of how sexual dimorphism evolves. Further, the answers will inform us about general issues including constraints to adaptation [8–10] and how genomes evolve [11–13].

We tackle this problem here with a new method that directly quantifies contemporary sex-specific selection on a genomic scale. We reason that Mendelian inheritance ensures that allele frequencies at autosomal loci are equal in males and females at conception, and that sex-specific viability selection will generate genetic divergence between the sexes within a generation. Divergence will occur when selection is sexually-antagonistic, meaning that different alleles are favored in males and females. Divergence also results when selection of different strengths favors the same allele in males and females, a situation sometimes called “sexually-concordant selection” [14]. We use the term “sex-specific selection” here to include both cases.

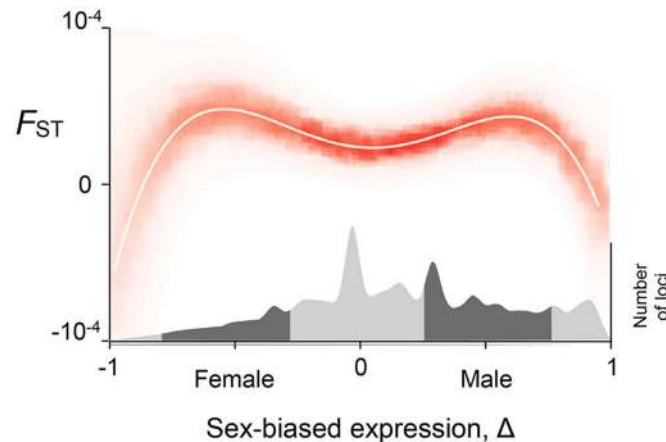
We find that genes with intermediate degrees of sex-biased expression experience the strongest sex-specific selection in humans. We assess the generality of this pattern by reanalyzing the data on flies from Innocenti and Morrow [7]. We again find the The Twin Peaks pattern, which in this case results from sexually-antagonistic (not sexually-concordant) selection, and from selection on fertility (rather than viability). A simple population-genetic model successfully recapitulates the key pattern. Our results suggest that ongoing sex-specific selection is a common feature of the genome in humans and flies.

## Results

### Sex-specific selection and sex-biased expression in humans

We quantify sex differences in viability selection using  $F_{ST}$  between adult males and females [15]. We used allele frequencies at over 6 million autosomal single-nucleotide polymorphisms (SNPs) that appear in the 1000 Genomes Project [16], a database that includes more than 2000 individuals from 26 populations worldwide. For each of 17,839 protein-coding loci, we calculated the average  $F_{ST}$  for all SNPs within transcribed regions within each population. We expect this to give a conservative picture of sex differences in selection since most SNPs are not themselves targets of selection. Rather, they change frequency through hitchhiking, and the  $F_{ST}$  that results at these SNPs is expected to be less than that at a selected SNP in the same gene. Patterns that are qualitatively the same as what we report below also emerge if we use only the single SNP from each locus with the largest value of  $F_{ST}$ . No SNP showed significantly different allele frequencies in males and females at a false discovery rate of 5%. That is not surprising, however. A power analysis presented in the [S1 File](#) shows it is highly unlikely to observe significant divergence at any individual SNP, even with very strong sexually-antagonistic selection, simply because the correction for multiple comparisons is so severe.

We quantify sex-biased expression as the normalized difference between expression in males and females:  $\Delta = (m-f) / (m+f)$ , where  $m$  and  $f$  are the transcript abundances in males



**Fig 1. The strength of sex-specific selection is strongest on human autosomal genes with intermediate sex-biased expression.** The white curve is the best-fit 4<sup>th</sup> degree polynomial and the intensity of red indicates the likelihood that the regression passes through a given value. The average  $F_{ST}$  for the SNPs in a gene is small at  $\Delta = 0$  (unbiased expression), increases to a peak as sex-bias grows, then decreases as  $\Delta$  approaches  $-1$  and  $1$ . The numbers of genes with a given bias are visualized in the density plot in the lower part of the figure; dark gray denotes intermediate sex-biased expression.

doi:10.1371/journal.pgen.1006170.g001

and females. (This measure is highly correlated with the familiar  $\log_2$  of the expression ratio.) A value of  $\Delta = -1$  means a gene is expressed only in females,  $\Delta = 0$  means expression is unbiased, and  $\Delta = 1$  means it is expressed only in males. We use RNAseq data from the Genotype-Tissue Expression Project [17]. Sex-biased expression in adults mainly results from differences between reproductive tissues [18,19], so we used expression levels in testis and ovaries. The data are from 14 males and 6 females [20].

We find that genetic divergence between males and females is related to the degree of sex-biased expression by a distinctive pattern that we call “Twin Peaks” (Fig 1). Average  $F_{ST}$  reaches a minimum when expression is unbiased ( $\Delta = 0$ ). Genetic divergence between the sexes is greatest for genes whose expression is moderately female-biased ( $\Delta \approx -\frac{1}{2}$ ) or moderately male-biased ( $\Delta \approx \frac{1}{2}$ ). After reaching these peaks, genetic divergence then declines again as expression tends towards complete female-bias ( $\Delta = -1$ ) and complete male-bias ( $\Delta = 1$ ).

The Twin Peaks pattern is statistically well-supported. The best fit regression of  $F_{ST}$  on  $\Delta$  is a fourth-degree polynomial with two peaks (S1 Table). We permuted the data  $10^5$  times and found the pattern occurs 1.6% of the time (that is, the pattern is significant at  $p = 0.016$ , details in Materials and Methods). The Twin Peaks pattern is also recovered by fitting a cubic spline using generalized additive models. Twin Peaks are seen within each of the 26 populations when they are analyzed separately. Genetic divergence between the sexes is quite repeatable across populations: taking the SNP with highest  $F_{ST}$  at each locus, the intra-class correlation coefficient is 0.34. Finally, the Twin Peaks also emerge when we measure genetic divergence between the sexes using the statistic  $D_a$  instead of  $F_{ST}$  [21,22] (S2 Table, S3 Fig).

It is plausible that gene expression in adult gonads (on which our analyses are based) is not itself the target of the viability selection that causes the Twin Peaks pattern. Instead, expression in adult gonads could be correlated with expression in other tissues and other life stages that are the actual targets. Unfortunately, we are not able to repeat our analyses with most somatic tissues because the Gene Expression Atlas only provides data on expression averaged across both sexes. We were however able to analyze expression in five sex-specific somatic tissues: ectocervix, fallopian tubes, vagina, and uterus in females, and prostate in males. The Twin Peaks pattern appears again when we use the average expression of the four female-specific

somatic tissues and expression in testes, but not when we use prostate instead of testes. Expression in prostate may not provide a good proxy for gene expression in males, however, as its expression profile is very similar to that of vagina and other female-specific tissues [20].

Sex-specific selection could be acting on regulatory elements as well as coding regions. We therefore repeated the analyses including all noncoding transcripts whose expression is profiled in the Genotype-Tissue Expression Project (34,060 in total) and the 1kb sequence upstream of all coding regions used in the previous analysis. Once again, a significant Twin Peaks pattern appears.

We considered the possibility that the pattern is driven by heterozygosity. Imagine that the strength sex-specific selection does not vary systematically with sex-biased expression, but that heterozygosity follows the Twin Peaks pattern. Then genetic divergence between males and females, measured as  $F_{ST}$ , will also show that pattern. (This prediction follows because changes in allele frequencies caused by selection are proportional to heterozygosity.) We therefore regressed the quantity ( $F_{ST}/pq$ ) onto  $\Delta$ , where  $p$  and  $q$  are the frequency of the two alleles at a given SNP. If the Twin Peaks results from variation of heterozygosity with sex-biased expression, the regression is expected to be flat. This alternative hypothesis is falsified: Twin Peaks are again seen using the polynomial and spline regressions described above (S3 Table). We also fit a regression model that included both heterozygosity and  $\Delta$  as factors. The Twin Peaks pattern remains significant ( $p < 0.01$ ) even after the effect of heterozygosity on  $F_{ST}$  is controlled for.

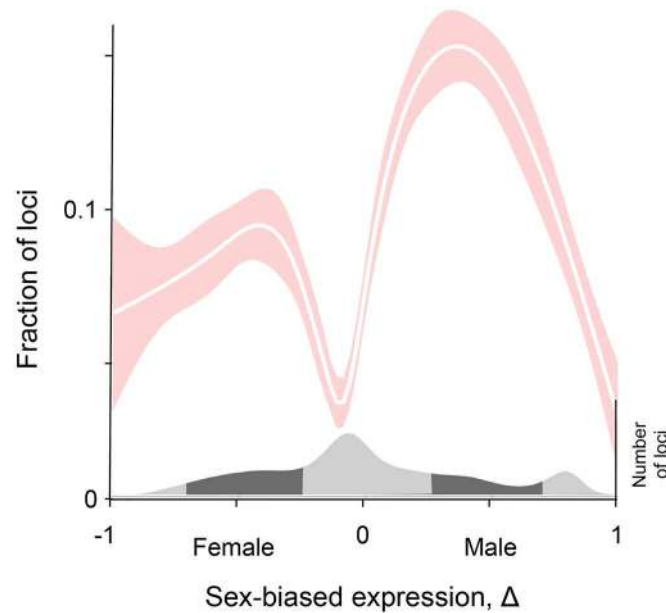
While no individual SNP showed significant  $F_{ST}$  between the sexes, the identities of several highly diverged genes do hold hints of possible connections to sex-specific selection. The gene *RNF212* segregates for alleles that have opposing effects on male and female recombination rates [23,24]. The mean  $F_{ST}$  for the SNPs in this gene is in the top 0.2% of all the loci we analyzed. Other genes in the top 5% of  $F_{ST}$  values are two loci (*LYPLAL1* and *ADAMTS9*) associated with the waist-hip ratio [25]. This is a sexually-dimorphic phenotype that has long been associated with sexual attractiveness [26]. A recent review identified 33 loci that have been linked by multiple genome-wide association studies to sex-specific risk of disease [27]. The most diverged SNPs at these genes have  $F_{ST}$  values that are significantly higher than average ( $p < 4 \times 10^{-5}$  by a permutation test). A final observation is that a set of genes that show evidence of balancing selection [28] also have  $F_{ST}$  values that are significantly higher than the genome-wide average ( $p < 1 \times 10^{-6}$  by a permutation test). This correlation could result if sexually-antagonistic selection itself is maintaining polymorphism at these loci [29], or if sex-specific selection acts on polymorphisms that are maintained by other mechanisms (e.g. overdominance).

These results lead us to conclude that the Twin Peaks pattern in humans results from sex-specific viability selection whose strength varies with the degree of sex-biased gene expression.

## Twin Peaks also occur in flies

To test the generality of the pattern we discovered in humans, we reanalyzed the fly data from Innocenti and Morrow [7]. Here our measure of the strength of sex-specific selection acting on a locus is binary: it takes a value of 1 if they identified it as a target of sexually-antagonistic selection, and is 0 otherwise. Their data are based on fertility selection, rather than viability selection. We again quantified sex-biased gene expression using  $\Delta$  based on their transcriptome data. We then asked if fraction of genes under sexually-antagonistic selection varies systematically with  $\Delta$ .

The Twin Peaks pattern appears once again (Fig 2). The greatest fraction of loci under sexually-antagonistic selection have intermediate sex-biased expression. Only a small fraction of genes experience sexually-antagonistic selection when expression is not sex-biased ( $\Delta = 0$ ) or



**Fig 2. The Twin Peaks pattern in flies.** The relation between the fraction of genes under sexually-antagonistic selection and  $\Delta$  is shown by the best-fit cubic spline (white line), and the 95% confidence interval is shown by the shaded areas. The distribution of gene numbers is shown at the bottom of the figure.

doi:10.1371/journal.pgen.1006170.g002

when it is completely sex-biased ( $\Delta = -1$  or  $1$ ). The pattern is highly statistically significant ( $p < 10^{-5}$ ) by bootstrapping using regressions based both on polynomials and splines (see [Materials and Methods](#)).

The pattern in flies is consistent with that in humans. A difference between the two data sets is that the pattern in humans results from genetic effects on viability, while the pattern in flies reflects effects on fertility and fecundity. Together, these results suggest that a tendency for stronger sex differences in selection to act on genes with intermediate sex-biased expression may be general to diverse forms of selection.

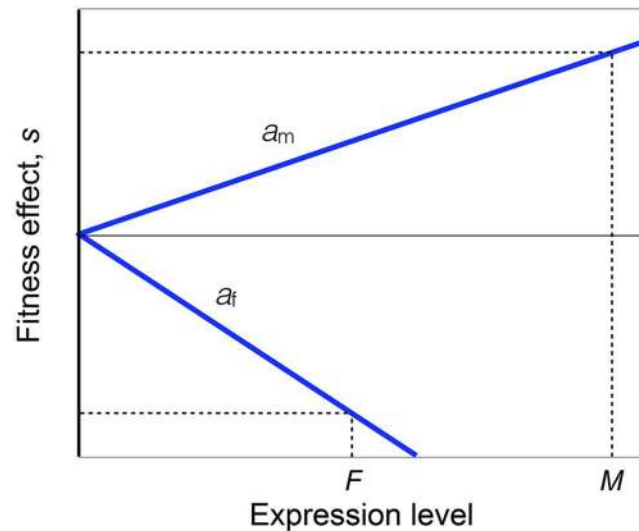
### A simple model explains Twin Peaks

To understand the Twin Peaks pattern, we built a simple model that relates gene expression to fitness. Three key assumptions are invoked: (1) Allele frequencies are at equilibrium under sexually-antagonistic viability selection, which is the special case of sex-specific selection in which selection favors one allele in one sex and the other allele in the other sex; (2) A gene that is not expressed experiences no selection; and (3) At low expression levels, the effects of alleles on viability increase approximately linearly with the amount of expression. These assumptions are illustrated graphically in [Fig 3](#). The model allows selection to be frequency-dependent, and for arbitrary dominance. Further details and the analysis of the model are given in the Materials and Methods section.

This model leads to two predictions regarding how genetic divergence between the sexes varies with the degree sex-biased expression. First, we find that when  $\Delta$  is near to 0,

$$F_{ST} \approx 4 p q \left( \frac{a_m a_f}{a_m + a_f} \right)^2 \Delta^2, \quad (1)$$

where  $a_m$  and  $a_f$  measure how the strength of selection varies with the gene's expression in



**Fig 3. Schematic of the genetic model.** The X-axis shows the expression of a locus, measured as the log of the number of transcripts, in males ( $M$ ) and females ( $F$ ). The Y-axis shows the additive fitness effect of an allele in males ( $s_m$ ) and females ( $s_f$ ). No selection ( $s_m, s_f = 0$ ) occurs when there is no expression ( $M, F = 0$ ). Fitness effects increase with expression at a rate  $a_m$  in males and  $a_f$  in females. At an evolutionary equilibrium,  $s_m = -s_f$ . Our measure  $\Delta$  of sex-bias is approximately equal to twice the difference between  $M$  and  $F$ , normalized by the total of expression in both sexes,  $M + F$ , when  $\Delta$  is small. Further details are given in the Materials and Methods section.

doi:10.1371/journal.pgen.1006170.g003

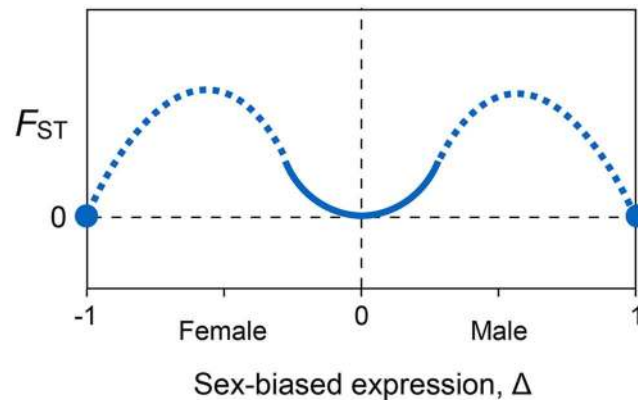
males and females (see Fig 3). We see that allele frequency divergence between the sexes depends on three kinds of quantities: heterozygosity (proportional to  $4pq$ ), the potential for sexually-antagonistic selection (represented by the term in parentheses), and the strength of sex-biased expression (represented by  $\Delta$ ).

All else equal, this result predicts that  $F_{ST}$  will be 0 when expression is unbiased, and it will increase quadratically as expression becomes slightly female- or male-biased. Second, the model predicts that  $F_{ST}$  will also be zero when expression is completely female-biased and completely male-biased. The model does not apply to genes whose expression is strongly but not completely sex-biased. We can, however, make qualitative predictions for these cases by interpolating between the predictions from the first two cases. The pattern that results is Twin Peaks (Fig 4).

To check the qualitative fit of this model to data, we focused on genes in the human dataset whose expression is weakly sex-biased. We took all genes expression lies between the two peaks at  $\Delta = -0.52$  and  $\Delta = 0.72$ , which includes 80% of the genes in the dataset. The best-fit polynomial regression is a quadratic with a significantly positive coefficient ( $p \ll 0.001$ ), as predicted by our model (S4 Table). In principle, we might use the curvature of this regression to estimate the quantity that appears within parentheses in Eq (1). That would be of interest because it is a measure of the potential for sexually-antagonistic selection. We do not take that step here, however, because as we now explain these data may not be adequate to give accurate estimates of selection.

### The strength and prevalence of sex-specific selection in humans

Under the assumptions of our model, the value of  $F_{ST}$  between males and females is proportional to the strength of sexually-antagonistic viability selection (see Materials and Methods). This suggests that one might estimate the strength of sexually-antagonistic selection from these



**Fig 4. Qualitative conclusions from the genetic model.** Genetic divergence between the sexes, measured as  $F_{ST}$ , is minimized when gene expression is unbiased ( $\Delta = 0$ ), then increases quadratically for small degrees of female-biased and male-biased expression (solid curve).  $F_{ST}$  is also 0 when expression is completely female-biased ( $\Delta = -1$ ) or male-biased ( $\Delta = 1$ ) (solid circles). Values of  $F_{ST}$  expected for intermediate sex-dependent expression are interpolated by eye (dashed curves).

doi:10.1371/journal.pgen.1006170.g004

data. Using the average  $F_{ST}$  for a gene biases the estimate downward because the large majority of SNPs are not expected to be direct targets of selection. Using the SNP in a gene with the largest  $F_{ST}$ , on the other hand, biases estimates upward because this extreme value is expected to result partly from sampling variance.

With these caveats in mind, we attempted to make a rough estimate for the strength of sexually-antagonistic selection acting on genes with intermediate sex-biased expression (see [Materials and Methods](#)). Focusing on the YRI population (Yoruba in Ibadan, Nigeria) because it is the most genetically variable, we sampled loci whose values of  $\Delta$  fall close to maxima of the two Twin Peaks and whose values of  $F_{ST}$  are close to the average for genes with that degree of sex-biased expression. Using the average  $F_{ST}$  for all the SNPs in each gene, we estimate that the additive fitness effects of alleles at these loci are  $s = 0.02$  on average. If correct, this result would suggest that sexually-antagonistic selection at these loci is strong. The estimate is based on small numbers of individuals, however, and so it is imprecise. The estimate is also likely to be biased upwards, as we will see shortly.

What fraction of genes experiences sex-specific selection in humans? While it is difficult to quantify that number with precision, we can get a crude idea by summing the number of loci that lie beneath the Twin Peaks, for example those whose sex-biased expression lies between  $0.25 < |\Delta| < 0.75$ . In both humans and flies, those are about 40% of coding genes. These results suggest that sex-specific selection could be common in both species.

### The selection load

Sex differences in autosomal allele frequencies between mature females and males result from differential survival after conception. Unlike other kinds of genetic load [30], this engenders a true mortality load on the population. We can calculate this load under the assumption of our model that the loci are at equilibrium under sexually-antagonistic selection in which an allele that increases viability in one sex decreases it in the other (see [Materials and Methods](#)). The additional mortality incurred by a population due to sexually-antagonistic selection on a single locus with additive effects is simply  $L_{SA} = s$ , where  $s$  is the fitness effect of an allele (positive in one sex, negative in the other). This result suggests that not many loci can be under strong sexually-antagonistic selection or the population would not be viable. For example, sexually-

antagonistic acting on 100 loci with independent fitness effects of  $s = 0.01$  reduces the average viability in the population by 63%.

This consideration argues that the estimates for the strength and prevalence of selection offered in the previous section may be substantial overestimates. Factors that may contribute are discussed below.

## Discussion

This study focuses on the relation between the strength of ongoing sex-specific selection to the degree of sex-biased expression. In humans, we used allele frequencies at 6 million coding SNPs sampled from 26 populations. We quantify the strength of sex differences in selection acting within a generation using  $F_{ST}$  between males and females adults. In flies, we quantified sex-specific selection as the fraction of loci that had been previously identified as targets of sexually-antagonistic selection [7]. In both species, we see a distinctive Twin Peaks pattern in which sex-specific selection is strongest for genes with intermediate sex-biased expression, and is weak when sex-bias expression is either complete or absent. These results suggest that sex-specific selection is ongoing at many genes. We develop a simple genetic model that recapitulates the Twin Peaks pattern.

In a recent study, Lucotte *et al.* [31] used the  $F_{ST}$ -based strategy that we employed to scan the HapMap dataset [32] for SNPs with divergent allele frequencies in males and females. Under their most stringent statistical criteria, they found two X-linked loci that are significantly diverged (at a FDR of 5%). (Those loci do not appear in our analyses because we excluded the sex chromosomes.) Further, they reported that extreme  $F_{ST}$  values are enriched on the X chromosome relative to the autosomes, consistent with what is expected from sexually-antagonistic selection. Taken together with our results from the 1000 Genomes dataset, the consistent picture is that there is evidence of ongoing sex-specific selection in humans.

Other processes besides sex-specific selection might contribute to the sex differences in allele frequencies we observed in humans. Population structure can generate  $F_{ST}$  between the sexes. If populations have different allele frequencies, then differential migration of one sex will generate  $F_{ST}$  between the sexes within a population.  $F_{ST}$  between the sexes will also appear in the data if viability selection acts equally on males and females but the sexes are sampled at different ages.

It seems unlikely, however, that these alternative hypotheses explain the Twin Peaks pattern. While both population structure and age differences can inflate  $F_{ST}$  between the sexes in the human dataset, it is difficult to see how either of those factors would generate a pattern that varies systematically with sex-biased expression. Further, neither population structure nor age differences are factors in the fly dataset. Sex-specific selection therefore appears to be strongly supported.

As we noted in the Results, it is plausible that the Twin Peaks pattern does not result from viability selection on gene expression in adult gonads (which is what our analysis is based on), but rather from selection on correlated traits such as expression in other tissues and/or at other ages. If so, we expect that our results are conservative: the pattern we observed is weaker than what would be seen if divergence in male and female allele frequency was regressed against the true targets of selection.

There is good reason, however, to think that our estimates for the strength of selection and the selection load in humans are biased upwards. The calculations assume loci are at equilibrium under sexually-antagonistic selection in which selection in one sex is exactly counterbalanced by selection in the other. If instead sex-specific selection acts on polymorphisms that are maintained by other forces (e.g. migration or mutation) or that are transient, then the load



could be substantially lower. Unfortunately, it does not seem possible to estimate the strength of selection from our data under these more general conditions. Further, as we discussed earlier, geographical and age structure in the data may contribute to the genetic differences between the sexes.

The fly data also include biases. One results from sex differences in the allometry of body parts. Innocenti and Morrow [7] measured gene expression in the whole bodies of flies. Differences in transcript abundances between tissues that have different relative sizes in males and females will result in sex differences in relative expression, and so bias our  $\Delta$  statistic. It is difficult, however, to see how this bias would vary systematically with the fraction of genes under sexually-antagonistic selection and so produce the Twin Peaks pattern.

Our model suggests that the Twin Peaks pattern is expected to emerge under certain conditions. Two in particular deserve consideration. The first assumption is that selection is sexually-antagonistic, with different alleles favored in males than in females. Our  $F_{ST}$ -based analysis of the human data is not able to discriminate between sexually-antagonistic selection and sexually-concordant selection (when the same allele is favored with different strengths of selection in males and females). We note, however, that the Twin Peaks pattern in flies does result from sexually-antagonistic (and not sexually-concordant) selection. Second, the model assumes that allele frequencies are at equilibrium. Unfortunately, we are unable to generalize either of these two assumptions. But while the quantitative results from the model will not hold if those assumptions are violated, it seems likely that the Twin Peaks pattern will emerge under a broader range of conditions than our model assumes.

Our results and those of Innocenti and Morrow [7] suggest that a substantial number of genes are targets of ongoing sex-specific selection. Why hasn't evolution adjusted expression levels and resolved this tension? Two general ideas have been offered. One is that pleiotropy is extensive and so there is little genetic variation available to decouple expression levels in males and females [2,33–36]. A second suggestion is that selection pressures may fluctuate over time and space, and evolutionary lags prevent the degree of sex-biased expression from being optimized in the current environment [37–39]. This hypothesis is consistent with the high turnover rate seen of genes with sex-biased expression [19,40–42].

The holy grail of evolutionary genomics has been to identify individual genes that underlie adaptation. Our approach shows that even when that goal is not possible, we can nevertheless make important inferences about selection and adaptation by studying genome-wide patterns. The strategy of combining data across genes with a simple model has offered a new perspective on the evolution of sex-biased gene expression. Similar strategies may also be fruitful in studies of other problems.

## Materials and Methods

### A simple model relating $F_{ST}$ between the sexes to sex-biased gene expression

To understand the relation between sex-biased expression and sex-specific selection, we developed a very simple model. We reasoned that sex-specific selection at a locus must vanish when it is not expressed in one of the sexes. We approximate the strength of viability selection acting on two alleles in males and females as functions that are linear with the relative expression levels and that have intercepts at the origin (Fig 3). We further assume that allele frequencies are at equilibrium under sexually-antagonistic selection, such that selection favoring an allele in one sex is balance by selection against that allele in the other sex. We will now use this visualization to make a quantitative connection between sex-biased expression and genetic divergence between the sexes.

Allele frequencies at autosomal loci are equal in males and females among zygotes. Sexually-antagonistic selection acting on viability will cause divergence in allele frequencies between adult males and females. Consider a locus segregating for alleles  $A$  and  $a$ . To measure the impact of sexually-antagonistic selection, we use the familiar  $F_{ST}$  statistic:

$$F_{ST} = \frac{(p_m - p_f)^2}{4 p q}, \quad (2)$$

where  $p_m$  and  $p_f$  are the frequencies of  $A$  in mature males and females,  $p$  is its frequency in zygotes, and  $q = 1 - p$ .

Write the relative viabilities of genotypes  $aa$ ,  $Aa$ , and  $AA$  in males as  $1::1 + h_m S_m::S_m$ . The selection coefficient  $S_m$  and dominance coefficient  $h_m$  can be frequency-dependent, in which case these coefficients take their values at equilibrium (see below). Assuming that zygotes are at Hardy-Weinberg equilibrium, the allele frequency in adult males is

$$p_m \approx p + pq s_m, \quad (3)$$

where  $s_m$  is the additive fitness effect of allele  $A$  in males, defined as

$$s_m = [p + (q - p)h_m]S_m. \quad (4)$$

The approximation of Eq (3) neglects terms of order  $s_m^2$ . An analogous expression gives the additive fitness effect of allele  $A$  in females,  $s_f$ . The divergence between males and females is therefore

$$F_{ST} \approx \frac{1}{4}pq(s_m - s_f)^2. \quad (5)$$

We link selection to expression by assuming that when expression is low, the fitness effects can be approximated by the functions:

$$\begin{aligned} s_m &= a_m M \\ s_f &= a_f F, \end{aligned} \quad (6)$$

where

$$\begin{aligned} M &= \ln(m + 1) \\ F &= \ln(f + 1), \end{aligned} \quad (7)$$

and  $m$  and  $f$  are the absolute expression levels (e.g. number of transcripts per cell or in a tissue). The coefficients  $a_m$  and  $a_f$  reflect how sensitive selection is to the expression level. At a locus under sexually-antagonistic selection, these coefficients have opposite signs. The situation is sketched in Fig 3.

Eqs (6 and 7) have three desirable features. First, the fitness effect of an allele in a given sex is 0 if the gene is not expressed in that sex. Second, for values of  $m$  and  $f$  much larger than 1, changes in transcript abundance have proportional effects on the strength of selection. For example, a 10% increase in expression has the same effect on fitness regardless of whether there are 100 or 1,000 transcripts in the cell. Last, we will see below that the logarithmic scaling for expression used in Eq (7) leads to predictions from the model that can be directly compared to data.

We now assume that allele frequencies are at equilibrium under sexually-antagonistic viability selection, which implies that  $s_m = -s_f$ . This assumption will be violated if fertility selection acts in addition to viability selection, but we do not anticipate that will systematically bias the predictions. (We note that an equilibrium is possible with any degree of dominance if selection

is frequency-dependent.) The equilibrium condition and Eq (6) imply that

$$T = \left( \frac{a_f - a_m}{a_f + a_m} \right) D, \tag{8}$$

where  $T$  is the total expression and  $D$  is the difference in expression (sex-bias) between males and females:

$$\begin{aligned} T &= M + F \\ D &= M - F. \end{aligned} \tag{9}$$

Substituting Eqs (6–9) into Eq (5) gives

$$F_{ST} \approx p q A D^2, \tag{10}$$

where

$$A = \left( \frac{a_m a_f}{a_m + a_f} \right)^2. \tag{11}$$

Eq (10) gives a pleasingly simple relationship between the divergence of allele frequencies between the males and females resulting from sexually-antagonistic selection, on the one hand, and the degree of sex-biased expression, on the other. Three quantities appear on the right side. First is the term  $pq$ , which is proportional to heterozygosity. The second is  $A$ , which is a measure of the potential for sexually-antagonistic selection. Last is  $D^2$ , which is the square of the sex-bias in expression.

We want to express these results in terms of quantities that have been estimated. A natural measure of sex-biased is

$$\Delta = \frac{m - f}{m + f}, \tag{12}$$

where  $m$  and  $f$  are the transcript abundances in males and females. The measure  $\Delta$  is 0 when expression is unbiased, reaches a minimum value of  $\Delta = -1$  when a gene is expressed only in females, and a maximum value of  $\Delta = 1$  when it is expressed only in males. When bias is small and the absolute expression levels are much greater than 1, a Taylor expansion shows that

$$\Delta \approx \frac{1}{2} \ln \left( \frac{m}{f} \right) \approx \frac{1}{2} D. \tag{13}$$

Substituting Eq (13) into Eq (10), we finally have

$$F_{ST} \approx 4 p q A \Delta^2, \tag{14}$$

which appears as Eq (1) in the main text.

This result has several implications. Divergence in allele frequencies between males and females (measured by  $F_{ST}$ ) is predicted to be at a minimum when there is no sex bias in expression ( $\Delta = 0$ ). For small amounts of bias, either positive or negative, divergence should increase quadratically with  $|\Delta|$ . We expect no divergence in male and female allele frequencies at the extremes of completely female-biased and completely male-biased expression ( $\Delta = -1$  and  $\Delta = 1$ , respectively). That prediction follows because either male or female expression is then 0, and under our equilibrium assumption there will be no positive selection acting on the other sex that could generate a difference in allele frequencies in males and females.

To summarize, our model makes three predictions:  $F_{ST}$  will be 0 for a gene with unbiased expression ( $\Delta = 0$ ), it will increase quadratically with  $\Delta$  for small amounts of male- and female-biased expression, and finally it will return to 0 for a gene that are expressed only in females or in males ( $\Delta = -1$  and  $\Delta = 1$ ). The results are shown graphically in Fig 4.

### The strength of sexually-biased selection

We estimated the strength of selection from the frequencies of alleles in females and males. Following the notations and assumptions described in previous sections, we assume that loci under sexually-antagonistic selection and are at equilibrium. That implies that the fitness effect of an allele in females is equal in magnitude and opposite in sign to its effect in males. The total numbers of alleles sampled from females and males are denoted  $N_f$  and  $N_m$ , and the numbers of the minor allele observed are  $n_f$  and  $n_m$ . We assume the samples are independent.

Denote the (unknown) frequency of the minor allele in zygotes at a given locus as  $p$ . Then likelihood of  $s_m$ , the additive fitness effect in males as defined by Eq (4), is

$$L(s_m, p) = B(n_f; N_f, p_f^*)B(n_m; N_m, p_m^*), \tag{15}$$

where  $B()$  is the binomial distribution, and the allele frequencies in the adults from which the alleles are sampled are

$$\begin{aligned} p_m^* &= p + p q s_m \\ p_f^* &= p - p q s_m. \end{aligned} \tag{16}$$

(Eq (16) are approximations that assume  $p q s_m \ll 1$ .) The posterior probability of the allele's fitness effect is

$$P_\beta(s_m) = \int_0^{1/2} L(s_m, p) \Pr(p) dp, \tag{17}$$

where  $\Pr(p)$  is the prior probability density of  $p$ .

For  $\Pr(p)$ , we fit an *ad hoc* function to the distribution of minor allele frequencies observed in the YRI population. That function is:

$$\Pr(p) = 48.4 \operatorname{Re}[\operatorname{Exp}\{-35(x - 0.003)^{0.7}\}] + 3.59(0.5 - x), \tag{18}$$

where  $\operatorname{Re}[x]$  is the real part of  $x$ . The fit of this function to the data is shown in S1 Fig. In practice, we found that using a uniform prior distribution changed the estimates of  $s_m$  very little.

We evaluated posterior probability by numerically integrating Eq (17). We found the maximum *a posteriori* probability (MAP) estimate of  $s_m$  by numerically maximizing that function. The 95% credible interval was found by searching for the values of  $s_m$  that returned a posterior probability equal to 1/20 of the maximum probability.

S2 Fig shows the results for a sample of 15 genes. These were taken from loci with values of  $\Delta$  that fall close to the maxima of the two Twin Peaks and whose values of  $F_{ST}$  averaged across all SNPs in the gene are typical for genes with that degree of sex-biased expression. For each of the genes in the sample, we chose the SNP with the largest  $F_{ST}$ , reasoning that this SNP was more likely to be the actual target of sex-specific selection. All of the MAP estimates are very large ( $|\hat{s}_m| \approx 1/2$ ) but no estimate for  $s_m$  is significantly different from 0.

### The selection load

The selection load caused by sexually-antagonistic viability selection can be calculated by the following argument. The load is the average of the mortality incurred from sexually-

antagonistic selection by males ( $\mu_m$ ) and in females ( $\mu_f$ ):

$$L_{SA} = \frac{1}{2}(\mu_m + \mu_f). \quad (19)$$

Following the notation used earlier, the relative viabilities of genotypes  $aa$ ,  $Aa$ , and  $AA$  in males are  $1 :: 1 + h_m S_m :: S_m$ , and analogous expressions give the relative viabilities in females. Assume Hardy-Weinberg equilibrium, and let  $p$  be the frequency of the allele that is beneficial to males. The mortalities are then:

$$\begin{aligned} \mu_m &\approx 2 p q (1 - h_m) S_m + q^2 S_m \\ \mu_f &\approx 2 p q h_f S_f + p^2 S_f. \end{aligned} \quad (20)$$

We now assume the population is at equilibrium. Then Eqs (3 and 4) imply that

$$[p + (q - p)h_m]S_m = [p + (q - p)h_f]S_f, \quad (21)$$

which on rearranging gives

$$S_f = \left( \frac{p + (q - p)h_m}{p + (q - p)h_f} \right) S_m. \quad (22)$$

Substituting these results back into Eq (19) tells us that the sexually-antagonistic load is

$$L_{SA} = \left( \frac{p + q^2 h_f - p^2 h_m}{2[p + (q - p)h_f]} \right) S_m. \quad (23)$$

This result simplifies substantially when heterozygotes have intermediate fitness ( $h_m = h_f = 1/2$ ):

$$L_{SA} = s, \quad (24)$$

where  $s$  is the additive fitness effect of allele  $A$ .

## Quantifying sex-specific selection using $F_{ST}$

We downloaded SNP data on humans from the 1000 Genomes Project [16]. The sample includes more than 2000 individuals from 26 populations worldwide. The gene annotation file (GRCh38) was downloaded from Ensembl [43]. We filtered the data to include only SNPs in transcribed regions of autosomal protein coding genes. A SNP was excluded from a population if its minor allele appeared as a singleton (28% of all SNPs), and SNPs were removed from all populations if they were monomorphic in more than 5 populations (0.24% of all genes). The resulting dataset includes over 6 million SNPs in 17,839 autosomal genes. In a second series of analyses, we also included potential regulatory elements by analyzing all 34,060 transcripts profiled by the Genotype-Tissue Expression Project [17] and the 1kb upstream regions of protein-coding genes.

$F_{ST}$  between the sexes and the genetic diversity in each population were estimated using the R package *PopGenome* [44]. The  $F_{ST}$  for each gene was summarized using the mean value for all SNPs inside that gene. The results were qualitatively unchanged when we used several alternative methods: the median or maximum  $F_{ST}$  (rather than the mean) of SNPs within a gene, including singleton SNPs, and alternative estimators for  $F_{ST}$ . Likewise, results were not changed when we excluded autosomal genes with paralogs on the X or Y chromosome.

To test for significant  $F_{ST}$  between the sexes at individual SNPs, we applied the False Discovery Rate (FDR) approach using the *qvalue* package from R. Known SNPs inside genes with X

or Y paralogs were removed. No SNP is significant at 5% FDR level. We further used Fisher's method to combine  $p$ -values across multiple populations, and then calculated corresponding  $q$  value. Again, we were not able to find any significant SNP at the 5% FDR level. We conducted a power analysis to determine how likely that result is to occur under different strengths of selection ([S1 File](#)). We used parameter values corresponding to the Yoruban population (which is the most polymorphic and so most likely to show signals of sex-specific selection). We found that even if every one of the more than 3 million SNPs in this population experienced sexually-antagonistic selection with a selection coefficient as large as  $s = 0.25$ , there is about a 96% probability that none of them will show a significant  $F_{ST}$ . It is therefore unsurprising that no SNP showed significant divergence.

To learn if sex-specific selection is more prevalent among genes with certain functions, we identified the SNP with the highest  $F_{ST}$  at each locus, since this SNP is more likely to be the target of selection. Gene set enrichment analysis was performed using online tools from the Database for Annotation, Visualization and Integrated Discovery [45]. Cell adhesion, synapses, neuron development, and several other annotation clusters are highly enriched among the loci that fall in the top 5% of  $F_{ST}$  values (DAVID scores greater than 7; S9 Table).

## Quantifying sex-biased gene expression

RNAseq data for gene expression in humans provided by the Genotype-Tissue Expression Project [17] was queried from the Gene Expression Atlas [46]. We used expression levels in ovaries measured in 6 females and levels in testes measured in 14 males. Relative sex bias in expression is measured as:  $\Delta = (m-f) / (m+f)$ , where  $m$  and  $f$  are the normalized transcript abundances in males and females. Our index of sex bias  $\Delta$  is highly correlated ( $r = 0.95$ ) with the familiar  $\log_2$  ratio of male and female expression.

To determine if the results hold for tissues other than testes and ovaries, we added the data for all of the sex-specific somatic tissues available in the Genotype-Tissue Expression Project. For female-specific tissues, we averaged expression in the ovary, ectocervix, fallopian tubes, vagina, and uterus. For male-specific tissues, we averaged testes and prostate. The Twin Peaks pattern again appears, but with the left-hand peak shifted far to the left. This shift appears to be driven by the prostate: the original pattern reappears using the five female-specific tissues but only testes from males, but no pattern is seen using the five female-specific tissues and prostate. We believe that prostate does not provide a good proxy for gene expression in males: its expression profile is most similar to that of vagina, and is quite similar to other female-specific tissues [20]. We were unable to repeat our analyses with other somatic tissues because the Gene Expression Atlas only provides data on the average expression across both sexes.

## Reanalysis of fruit fly data

Gene expression data collected by Innocenti and Morrow [7] was downloaded from Gene Expression Omnibus (accession number GSE17013) and the probe annotation file from Ensemble. Following Innocenti and Morrow, the  $\log_2$  ratio of male and female gene expression was estimated using the R *Bioconductor* packages [47]. Sex-biased expression was then transformed into our  $\Delta$  statistic (described above). The list of candidate genes under sexually-antagonistic selection was taken from [7]. Results shown in the main text pertain to autosomal loci, but the Twin Peaks pattern also appears when sex-linked loci are included.

## Fitting curves

The relation between sex-specific selection and sex-biased expression was determined by first regressing  $F_{ST}$  onto  $\Delta$  using polynomials. Each population was treated as a replicate. The

optimal polynomial degree was determined using the Akaike information criterion [48] and likelihood ratio tests. Regressions were also fit with cubic splines using generalized additive models (GAM) implemented *mgcv* package [49] in R.

Following [50], we fit fourth degree polynomials to  $10^5$  bootstrap samples of the original data. The Twin Peaks pattern were identified in a particular sample if it met three criteria: the fit was significant ( $p < 0.05$  by ANOVA), the leading coefficient of the polynomial was negative, and the derivative of the polynomial had three real roots in the interval  $[-1, 1]$ . The last two criteria imply that there were two peaks (local maxima with a local minimum between them). A heat map was used to visualize the probability density for each value of  $F_{ST}$  as a function of  $\Delta$ . The result is shown in Fig 1.

A permutation test was used to test the significance of the pattern we observed in the data. We permuted the values of  $\Delta$ , fit a quartic polynomial by regression, and again used the three criteria to determine if the Twin Peaks pattern appeared. This procedure was repeated  $10^5$  times.

The locations of the local minimum and the two local maxima for the polynomial regressions were found by using the *polyroot* function in R [51]. The 95% confidence intervals for these locations were estimated by bootstrapping the data  $10^5$  times.

## Supporting Information

**S1 Fig. Fit of Eq (18) to the minor allele frequency distribution in the YRI population.**  
(PDF)

**S2 Fig. Estimates of the sexually-antagonistic fitness effects at 15 loci.** The loci were chosen at random from those with sex-biased expression levels corresponding to the Twin Peaks, and typical  $F_{ST}$  values for genes with that degree of expression bias. The maximum *a posteriori* probability (MAP) estimates are shown by the large horizontal line, and the 95% credible interval by the whiskers.  
(PDF)

**S3 Fig. The relationship of  $D_a$  vs.  $\Delta$  is shown by the best-fit cubic spline (blue line), and the 95% confidence interval is shown by the shaded area.**  
(PDF)

**S4 Fig. The relationship of  $\pi$  vs  $\Delta$  for is shown by the best-fit cubic spline (blue line), and the 95% confidence interval is shown by the shaded area.**  
(PDF)

**S5 Fig. The probability of detecting significantly different allele frequencies in males and females at any SNP in the YRI population as a function of the strength of sexually-antagonistic selection.**  
(PDF)

**S1 Table. A polynomial regression of  $F_{ST}$  on  $\Delta$  is best fit by a 4<sup>th</sup> degree polynomial.**  
(PDF)

**S2 Table. A polynomial regression of  $D_a$  on  $\Delta$  is best fit by a 4<sup>th</sup> degree polynomial.**  
(PDF)

**S3 Table. A polynomial regression of  $F_{ST} / (2pq)$  on  $\Delta$  is best fit by a 4<sup>th</sup> degree polynomial.**  
(PDF)

**S4 Table.** A polynomial regression of  $F_{ST}$  on  $\Delta$  is best fit by a quadratic when  $\Delta$  is small ( $-0.52 < \Delta < 0.72$ ).

(PDF)

**S5 Table.** A polynomial regression of  $F_{ST}$  on  $\Delta$  is best fit by a 4<sup>th</sup> degree polynomial when we including noncoding genes and regions 1kb upstream of coding regions.

(PDF)

**S6 Table.** A polynomial regression of  $D_a$  on  $\Delta$  is best fit by a 4<sup>th</sup> degree polynomial when we including noncoding genes and regions 1kb upstream of coding regions.

(PDF)

**S7 Table.** A polynomial regression of  $F_{ST} / (2pq)$  on  $\Delta$  is best fit by a 4<sup>th</sup> degree polynomial when we including noncoding genes and regions 1kb upstream of coding regions.

(PDF)

**S8 Table.** A polynomial regression of  $F_{ST}$  on  $\Delta$  is best fit by a quadratic when  $\Delta$  is small ( $-0.52 < \Delta < 0.72$ ) when we including noncoding genes and regions 1kb upstream of coding regions.

(PDF)

**S1 File.** The details of a power analysis that calculates the probability of observing a statistically significant difference between allele frequencies in males and females that results from sexually-antagonistic selection.

(PDF)

## Acknowledgments

We thank D. Houle and A.J. Dagilis for stimulating discussions. We are grateful for useful suggestions made by two reviewers.

## Author Contributions

**Conceived and designed the experiments:** CC MK.

**Performed the experiments:** CC MK.

**Analyzed the data:** CC MK.

**Contributed reagents/materials/analysis tools:** CC MK.

**Wrote the paper:** CC MK.

## References

1. Darwin C. The descent of man, and selection in relation to sex. London: Murray; 1871.
2. Lande R. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution*. 1980; 34: 292–305. doi: [10.2307/2407393](https://doi.org/10.2307/2407393)
3. Rice WR. Sex chromosomes and the evolution of sexual dimorphism. *Evolution*. 1984; 38: 735–742. doi: [10.2307/2408385](https://doi.org/10.2307/2408385)
4. Arnqvist G, Rowe L. Sexual conflict. Princeton University Press; 2005.
5. Vicoso B, Charlesworth B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet*. 2006; 7: 645–653. doi: [10.1038/nrg1914](https://doi.org/10.1038/nrg1914) PMID: [16847464](https://pubmed.ncbi.nlm.nih.gov/16847464/)
6. Ellegren H, Parsch J. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet*. 2007; 8: 689–698. doi: [10.1038/nrg2167](https://doi.org/10.1038/nrg2167) PMID: [17680007](https://pubmed.ncbi.nlm.nih.gov/17680007/)



7. Innocenti P, Morrow EH. The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biol*. 2010; 8: e1000335. doi: [10.1371/journal.pbio.1000335](https://doi.org/10.1371/journal.pbio.1000335) PMID: [20305719](https://pubmed.ncbi.nlm.nih.gov/20305719/)
8. van Doorn GS. Intralocus sexual conflict. *Ann N Y Acad Sci*. 2009; 1168: 52–71. doi: [10.1111/j.1749-6632.2009.04573.x](https://doi.org/10.1111/j.1749-6632.2009.04573.x) PMID: [19566703](https://pubmed.ncbi.nlm.nih.gov/19566703/)
9. Bonduriansky R, Chenoweth SF. Intralocus sexual conflict. *Trends Ecol Evol*. 2009; 24: 280–288. doi: [10.1016/j.tree.2008.12.005](https://doi.org/10.1016/j.tree.2008.12.005) PMID: [19307043](https://pubmed.ncbi.nlm.nih.gov/19307043/)
10. Rice WR. Nothing in genetics makes sense except in light of genomic conflict. *Annu Rev Ecol Evol Syst*. 2013; 44: 217–237. doi: [10.1146/annurev-ecolsys-110411-160242](https://doi.org/10.1146/annurev-ecolsys-110411-160242)
11. Charlesworth D, Charlesworth B. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet Res*. 1980; 35: 205–214. doi: [10.1017/S0016672300014051](https://doi.org/10.1017/S0016672300014051) PMID: [6930353](https://pubmed.ncbi.nlm.nih.gov/6930353/)
12. van Doorn GS, Kirkpatrick M. Turnover of sex chromosomes induced by sexual conflict. *Nature*. 2007; 449: 909–912. doi: [10.1038/nature06178](https://doi.org/10.1038/nature06178) PMID: [17943130](https://pubmed.ncbi.nlm.nih.gov/17943130/)
13. van Doorn GS, Kirkpatrick M. Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics*. 2010; 186: 629–645. doi: [10.1534/genetics.110.118596](https://doi.org/10.1534/genetics.110.118596) PMID: [20628036](https://pubmed.ncbi.nlm.nih.gov/20628036/)
14. Connallon T, Cox RM, Calsbeek R. Fitness consequences of sex-specific selection. *Evolution*. 2010; 64: 1671–1682. doi: [10.1111/j.1558-5646.2009.00934.x](https://doi.org/10.1111/j.1558-5646.2009.00934.x) PMID: [20050912](https://pubmed.ncbi.nlm.nih.gov/20050912/)
15. Kirkpatrick M, Guerrero RF. Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics*. 2014; 197: 531–541. doi: [10.1534/genetics.113.156026](https://doi.org/10.1534/genetics.113.156026) PMID: [24578352](https://pubmed.ncbi.nlm.nih.gov/24578352/)
16. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015; 526: 68–74.
17. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013; 45: 580–585. doi: [10.1038/ng.2653](https://doi.org/10.1038/ng.2653) PMID: [23715323](https://pubmed.ncbi.nlm.nih.gov/23715323/)
18. Wright AE, Mank JE. The scope and strength of sex-specific selection in genome evolution. *J Evol Biol*. 2013; 26: 1841–1853. doi: [10.1111/jeb.12201](https://doi.org/10.1111/jeb.12201) PMID: [23848139](https://pubmed.ncbi.nlm.nih.gov/23848139/)
19. Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, et al. Sexual selection drives evolution and rapid turnover of male gene expression. *Proc Natl Acad Sci*. 2015; 112: 4393–4398. doi: [10.1073/pnas.1501339112](https://doi.org/10.1073/pnas.1501339112) PMID: [25831521](https://pubmed.ncbi.nlm.nih.gov/25831521/)
20. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. The human transcriptome across tissues and individuals. *Science*. 2015; 348: 660–665. doi: [10.1126/science.aaa0355](https://doi.org/10.1126/science.aaa0355) PMID: [25954002](https://pubmed.ncbi.nlm.nih.gov/25954002/)
21. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci*. 1979; 76: 5269–5273. PMID: [291943](https://pubmed.ncbi.nlm.nih.gov/291943/)
22. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol*. 2014; 23: 3133–3157. doi: [10.1111/mec.12796](https://doi.org/10.1111/mec.12796) PMID: [24845075](https://pubmed.ncbi.nlm.nih.gov/24845075/)
23. Kong A, Thorleifsson G, Stefansson H, Masson G, Helgason A, Gudbjartsson DF, et al. Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science*. 2008; 319: 1398–1401. doi: [10.1126/science.1152422](https://doi.org/10.1126/science.1152422) PMID: [18239089](https://pubmed.ncbi.nlm.nih.gov/18239089/)
24. Sandor C, Li W, Coppieters W, Druet T, Charlier C, Georges M. Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle. *PLoS Genet*. 2012; 8: e1002854. doi: [10.1371/journal.pgen.1002854](https://doi.org/10.1371/journal.pgen.1002854) PMID: [22844258](https://pubmed.ncbi.nlm.nih.gov/22844258/)
25. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet*. 2010; 42: 949–960. doi: [10.1038/ng.685](https://doi.org/10.1038/ng.685) PMID: [20935629](https://pubmed.ncbi.nlm.nih.gov/20935629/)
26. Singh D. Adaptive significance of female physical attractiveness: role of waist-to-hip ratio. *J Pers Soc Psychol*. 1993; 65: 293. PMID: [8366421](https://pubmed.ncbi.nlm.nih.gov/8366421/)
27. Gilks WP, Abbott JK, Morrow EH. Sex differences in disease genetics: evidence, evolution, and detection. *Trends Genet*. 2014; 30: 453–463. doi: [10.1016/j.tig.2014.08.006](https://doi.org/10.1016/j.tig.2014.08.006) PMID: [25239223](https://pubmed.ncbi.nlm.nih.gov/25239223/)
28. DeGiorgio M, Lohmueller KE, Nielsen R. A model-based approach for identifying signatures of ancient balancing selection in genetic data. *PLoS Genet*. 2014; 10: e1004561. doi: [10.1371/journal.pgen.1004561](https://doi.org/10.1371/journal.pgen.1004561) PMID: [25144706](https://pubmed.ncbi.nlm.nih.gov/25144706/)
29. Connallon T, Clark AG. Balancing selection in species with separate sexes: insights from Fisher's geometric model. *Genetics*. 2014; genetics.114.165605. doi: [10.1534/genetics.114.165605](https://doi.org/10.1534/genetics.114.165605)
30. Keightley PD. Rates and fitness consequences of new mutations in humans. *Genetics*. 2012; 190: 295–304. doi: [10.1534/genetics.111.134668](https://doi.org/10.1534/genetics.111.134668) PMID: [22345605](https://pubmed.ncbi.nlm.nih.gov/22345605/)

31. Lucotte EA, Laurent R, Heyer E, Ségurel L, Toupance B. Detection of allelic frequency differences between the sexes in humans: a signature of sexually antagonistic selection. *Genome Biol Evol.* 2016; evw090. doi: [10.1093/gbe/evw090](https://doi.org/10.1093/gbe/evw090)
32. Consortium TIH 3. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010; 467: 52–58. doi: [10.1038/nature09298](https://doi.org/10.1038/nature09298) PMID: [20811451](https://pubmed.ncbi.nlm.nih.gov/20811451/)
33. Chippindale AK, Gibson JR, Rice WR. Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proc Natl Acad Sci.* 2001; 98: 1671–1675. doi: [10.1073/pnas.98.4.1671](https://doi.org/10.1073/pnas.98.4.1671) PMID: [11172009](https://pubmed.ncbi.nlm.nih.gov/11172009/)
34. Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H. Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *Am Nat.* 2008; 171: 35–43. doi: [10.1086/523954](https://doi.org/10.1086/523954) PMID: [18171149](https://pubmed.ncbi.nlm.nih.gov/18171149/)
35. Poissant J, Wilson AJ, Coltman DW. Sex-specific genetic variance and the evolution of sexual dimorphism: a systematic review of cross-sex genetic correlations. *Evolution.* 2010; 64: 97–107. doi: [10.1111/j.1558-5646.2009.00793.x](https://doi.org/10.1111/j.1558-5646.2009.00793.x) PMID: [19659596](https://pubmed.ncbi.nlm.nih.gov/19659596/)
36. Griffin RM, Dean R, Grace JL, Rydén P, Friberg U. The shared genome is a pervasive constraint on the evolution of sex-biased gene expression. *Mol Biol Evol.* 2013; 30: 2168–2176. doi: [10.1093/molbev/mst121](https://doi.org/10.1093/molbev/mst121) PMID: [23813981](https://pubmed.ncbi.nlm.nih.gov/23813981/)
37. Delcourt M, Blows MW, Rundle HD. Sexually antagonistic genetic variance for fitness in an ancestral and a novel environment. *Proc R Soc Lond B Biol Sci.* 2009; 276: 2009–2014. doi: [10.1098/rspb.2008.1459](https://doi.org/10.1098/rspb.2008.1459)
38. Rice WR, Chippindale AK. Intersexual ontogenetic conflict. *J Evol Biol.* 2001; 14: 685–693. doi: [10.1046/j.1420-9101.2001.00319.x](https://doi.org/10.1046/j.1420-9101.2001.00319.x)
39. Rice WR, Holland B. The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behav Ecol Sociobiol.* 1997; 41: 1–10. doi: [10.1007/s002650050357](https://doi.org/10.1007/s002650050357)
40. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science.* 2003; 300: 1742–1745. doi: [10.1126/science.1085881](https://doi.org/10.1126/science.1085881) PMID: [12805547](https://pubmed.ncbi.nlm.nih.gov/12805547/)
41. Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature.* 2007; 450: 233–237. doi: [10.1038/nature06323](https://doi.org/10.1038/nature06323) PMID: [17994089](https://pubmed.ncbi.nlm.nih.gov/17994089/)
42. Jiang Z-F, Machado CA. Evolution of sex-dependent gene expression in three recently diverged species of *Drosophila*. *Genetics.* 2009; 183: 1175–1185. doi: [10.1534/genetics.109.105775](https://doi.org/10.1534/genetics.109.105775) PMID: [19720861](https://pubmed.ncbi.nlm.nih.gov/19720861/)
43. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. *Nucleic Acids Res.* 2015; 43: D662–D669. doi: [10.1093/nar/gku1010](https://doi.org/10.1093/nar/gku1010) PMID: [25352552](https://pubmed.ncbi.nlm.nih.gov/25352552/)
44. Pfeifer B, Wittelsbürger U, Onsins SER, Lercher MJ. PopGenome: An efficient swiss army knife for population genomic analyses in R. *Mol Biol Evol.* 2014; msu136. doi: [10.1093/molbev/msu136](https://doi.org/10.1093/molbev/msu136)
45. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2008; 4: 44–57. doi: [10.1038/nprot.2008.211](https://doi.org/10.1038/nprot.2008.211)
46. Kapushesky M, Emam I, Holloway E, Kurnosov P, Zorin A, Malone J, et al. Gene Expression Atlas at the European Bioinformatics Institute. *Nucleic Acids Res.* 2010; 38: D690–D698. doi: [10.1093/nar/gkp936](https://doi.org/10.1093/nar/gkp936) PMID: [19906730](https://pubmed.ncbi.nlm.nih.gov/19906730/)
47. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004; 5: R80. doi: [10.1186/gb-2004-5-10-r80](https://doi.org/10.1186/gb-2004-5-10-r80) PMID: [15461798](https://pubmed.ncbi.nlm.nih.gov/15461798/)
48. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media; 2003.
49. Wood SN. mgcv: GAMs and generalized ridge regression for R. *R News.* 2001; 1: 20–25.
50. Hsiang SM. Visually-weighted regression [Internet]. Rochester, NY: Social Science Research Network; 2013 May. Report No.: ID 2265501. Available: <http://papers.ssrn.com/abstract=2265501>
51. Team RC. R: A language and environment for statistical computing. Vienna, Austria; 2014. URL <http://www.R-Project.org>. 2015;