# Sex differences in inter-individual gene expression variability across human tissues

**Authors:**

Samuel Khodursky[1], Caroline S. Jiang[2], Eric B. Zheng[1], Roger Vaughan[2], Daniel R. Schrider[3], Li Zhao[1,*]

**Affiliations:**

[1]Laboratory of Evolutionary Genetics and Genomics, The Rockefeller University, New York, NY 10065, USA

[2]Department of Biostatistics, The Rockefeller University, New York, NY 10065, USA

[3]Department of Genetics, University of North Carolina, Chapel Hill, NC 27514, USA

Correspondence to: lzhao@rockefeller.edu

**Abstract**

Understanding phenotypic sex differences has long been a goal of biology from both a medical and evolutionary perspective. Although much attention has been paid to mean differences in phenotype between the sexes, little is known about sex differences in phenotypic variability. To gain insight into sex differences in inter-individual variability at the molecular level, we analyzed RNA-seq data from 43 tissues from the Genotype-Tissue Expression project (GTEx). Within each tissue, we identified genes that show sex differences in gene expression variability. We found that these sex-differentially variable (SDV) genes are associated with various important biological functions, including sex hormone response, immune response, and other signaling pathways. By analyzing single-cell RNA sequencing data collected from breast epithelial cells, we found that genes with sex differences in gene expression variability in breast tissue tend to be expressed in a cell-type specific manner. We looked for an association between SDV expression and Graves' disease, a well-known heavily female-biased disease, and found a significant enrichment of Graves' associated genes among genes with higher variability in females in thyroid tissue. This suggests a role for SDV expression in the context of sex-biased disease. We then

examined the evolutionary constraints acting on genes with sex differences in variability and found that they exhibit evidence of increased selective constraint. Through analysis of sex-biased eQTL data, we found evidence that SDV expression may have a genetic basis. Finally, we propose a simple evolutionary model for the emergence of sex-differentially variable expression from sex-specific constraint.

## Introduction

Identifying, characterizing, and understanding phenotypic differences between the sexes has long been a major goal of biology. A very wide variety of traits in humans show significant sex differences including morphological, behavioral, metabolic, immune, and disease traits (Chella Krishnan et al., 2018; Choleris et al., 2018; Klein and Flanagan, 2016; Ober et al., 2008). These trait differences are likely in part due to sex differences in gene expression and regulation (Naqvi et al., 2019; Parsch and Ellegren, 2013; Williams and Carroll, 2009). As a phenotype, gene expression is both relatively straightforward to measure and highly quantitative. As a result, many studies have examined differences in mean gene expression levels between the sexes in many species (Catalán et al., 2012; Gershoni and Pietrokovski, 2017; Khodursky et al., 2020; Lopes-Ramos et al., 2020; Naqvi et al., 2019; Oliva et al., 2020a; Parsch and Ellegren, 2013). Sex-biased expression (SBE) is thought to arise from sex-specific genetic architectures which are ultimately established by sex-determination systems initiated during early development (Williams and Carroll, 2009). In addition to underlying sexually dimorphic phenotypes, genes with SBE have been found to possess many interesting evolutionary and genomic properties (Parsch and Ellegren 2013). For example, SBE has been associated with increased rates of evolution at both the coding and expression levels in *Drosophila* and other animals (Khodursky et al., 2020; Meisel et al., 2012; Parsch and Ellegren, 2013; Ranz et al., 2003). In humans, genes with SBE have been associated with reduced selective constraint (Gershoni and Pietrokovski, 2017; Naqvi et al., 2019). Additionally, genes with SBE have been shown to be distributed across the genome in a non-random fashion with respect to the X chromosome – thereby linking the evolution of sex-biased genes to the unique

evolutionary forces experienced by the X chromosome (Khodursky et al., 2020; Oliva et al., 2020a; Parisi et al., 2003).

Relative to sex differences in phenotypic means, sex differences in phenotypic variabilities have been less thoroughly studied. In the biomedical sciences there has been a long-standing assumption that females must generally be more phenotypically variable due to the menstrual or estrous cycles; however, there has been little empirical evidence to support this assumption (Beery and Zucker, 2011). In fact, a meta-analysis of microarray studies in mice and humans revealed that males may have slightly higher expression variability (Itoh and Arnold, 2015) and a meta-analysis of various traits in mice indicated that sex differences in phenotypic variability are trait dependent (Zajitschek et al., 2020). Despite growing interest in understanding sex differences in phenotypic variability, the biological and evolutionary properties of genes with sex differences in gene expression variability have not, to our knowledge, been explored in humans. Here, we define genes with sex-differentially variable (SDV) expression to be genes which show significant differences in inter-individual gene expression variability between the sexes. From a medical standpoint, genes with SDV expression may underlie diseases for which there are sex differences in prevalence or outcome. For example, one could imagine a disease partially caused by the aberrant expression of a certain gene (akin to the liability threshold model; (Falconer, 1967)). If one sex has more variable expression of the gene in a normal healthy population, this may help explain a higher prevalence of that disease in that sex. Additionally, since more variable expression has been associated with reduced effectiveness of drug targets it is possible that SDV expression may account for sex differences in drug effectiveness, which are known to exist (Simonovsky et al., 2019; Soldin and Mattison, 2009).

Understanding the evolutionary properties of these genes may give us novel insight into the evolution of sex differences. For example, genes with sex differences in expression variability may evolve under sex-specific constraints (e.g. purifying or stabilizing selection) and be of particular functional importance for a given sex. Here we examine the biological properties and evolutionary constraints associated with SDV expression across 43 human tissues.

## Results

### Identification and characterization of genes with sex differences in gene expression variability

We obtained RNA-seq count data for 43 tissues from the GTEx project (version 8). Since we sought to identify genes with sex differences in gene expression variability, we first compared several statistical methods using power simulations. Specifically, we compared three tests: the Fligner-Killeen test, Levene's test, and a test based on generalized additive models for location, scale and shape (GAMLSS) ( Rigby and Stasinopoulos, 2005) . Since RNA-seq count data is known to follow an approximately negative binomial distribution (Robinson and Smyth, 2007; Robinson et al., 2010), we performed power simulations by drawing random samples from negative binomial distributions with identical means but different variances by varying the overdispersion term (see methods). At each parameter combination, 1000 pairs of samples were drawn, and tests were performed at an alpha of 0.05. We found that GAMLSS had the highest power across a range of parameters, and this was especially true when testing for differences in variance between samples of unequal size (**Supplementary Figure 1A-D**). Both GAMLSS and Levene's test managed to maintain an adequate false positive rate (FPR) across a range of parameters. Beyond increased power GAMLSS offers other advantages over the other two tests: GAMLSS negative binomial models allow for the independent parametrization of both the mean and the non-Poisson variation (overdispersion) terms as a function of variables and covariates of interest, including library size (de Jong et al., 2019; Rigby and Stasinopoulos, 2005).

When considering the inter-sample variation of gene expression, it is natural to consider the coefficient of variation: the standard deviation divided by the mean. This provides a mean-normalized statistic for variability. Notably, it is thought that the Poisson component of the coefficient of variation in RNA-seq data is due to technical variation while the overdispersion component of the coefficient of variation contains the true biological variation (McCarthy et al., 2012). The overdispersion parameter fit by GAMLSS can therefore be considered to be the square of the biological coefficient of variation (de Jong et al., 2019; McCarthy et al., 2012). Observed overdispersion values and effect sizes in GTEx data

104 were similar to the simulated values, as determined by models fit using GAMLSS (**Supplementary**
105 **Figure 1E**).
106 To determine the effect of sex on gene expression variation we fit three negative binomial
107 generalized additive models using the R library 'gamlss' for each expressed gene in each tissue (Rigby
108 and Stasinopoulos, 2005). To reduce the risk of confounding due to unequal sampling across various
109 populations, we limited our analysis to European-American individuals. Significance for the effect of sex
110 on overdispersion and mean expression level was determined through a pair of likelihood-ratio tests (see
111 methods). All models included covariates for age, RNA integrity number (RIN), ischemic time, and the
112 first five genotype principal components (to control for population structure) for both mean and
113 overdispersion terms. We initially considered genes to be sex-differentially variable (SDV) if the effect of
114 sex on overdispersion was significant at FDR<0.05. Since we observed that outliers dramatically
115 influenced estimates of variation in some genes, as a secondary test we performed randomization tests by
116 randomly permuting the sex labels on genes with initially significant coefficients for the effect of sex on
117 overdispersion and refitting the GAMLSS models (see methods) (**Figure 1A**). Coefficients for the effect
118 of sex on overdispersion were saved after each permutation. We used 1000 rounds of permutation to
119 generate empirical null distributions for the effect of sex on overdispersion. Initially significant genes
120 with FDR-adjusted empirical p-values greater than 0.05 in the permutation test were no longer considered
121 to be SDV (see methods). Results from all tested genes are included in **Supplementary Table 1** in
122 Github webpage.
123 Across 43 tissues, we identified 3891 unique genes with SDV expression in at least one tissue out
124 of a total 22069 tested genes (**Figure 1A, B and Supplementary Table 1**). We were able to identify
125 SDV genes in 42 out of the 43 tested tissues (**Figure 1B**). Examples of two genes with SDV expression in
126 the pituitary gland and breast tissue are shown in **Figure 1C**. A total of 2778 unique genes showed higher
127 variability in males in at least one tissue, while 1394 genes showed higher variability in females in at least
128 one tissue. This imbalance between the sexes was largely driven by breast tissue. Surprisingly, 281 genes
129 had higher variability in males in one tissue and higher variability in females in at least one other tissue.

130 The vast majority (3342) of SDV genes only showed SDV expression in one tissue while 549 genes
131 showed SDV expression in multiple tissues.

132 Breast tissue had the largest number of SDV genes with 2046 genes showing significantly higher
133 variation in males and 127 genes showing significantly higher variation in females (**Figure 1B**). Apart
134 from a few tissues, SDV genes were not consistently expressed at higher or lower levels than non-SDV
135 genes across tissues (in terms of both CPM and TPM), suggesting that our discovery of SDV genes was
136 not significantly biased towards genes of high or low expression (**Supplementary Figure 2**).

137 Given that differences in variation are generally difficult to detect between groups, we performed
138 a replication analysis using data from breast tissue to help validate our statistical approach. First, we
139 randomly split individuals into discovery and replication sets – 70% of males and females were placed in
140 the discovery set while the rest were placed in the replication set. We identified genes with sex
141 differences in overdispersion in the discovery set using the combined likelihood ratio test and permutation
142 approach described above (both at FDR<0.05). These genes were then tested in the replication set: genes
143 with p-values for the overdispersion term less than 0.05 in both the likelihood ratio test and permutation
144 test were considered 'replicated' if the sex with higher variability in the discovery set was also more
145 variable in the replication set. We observed a replication rate of 19.5% (**Supplementary Figure 3**). This
146 replication rate is limited by reduced power due to the relatively small size of the replication set. To
147 determine the significance of this replication rate we generated an empirical null distribution under the
148 null hypothesis that our statistical approach was simply randomly sampling genes. Therefore, we
149 randomly sampled genes from the discovery set and calculated the replication rate for each sample. The
150 median replication rate under the null hypothesis was found to be 8.2% and out of 10,000 random
151 samples none had a replication rate higher than 10.5% (**Supplementary Figure 3**). This null replication
152 rate should not be considered as a type 1 error rate. Since we are randomly sampling genes from the set of
153 all expressed genes, we are also capturing some true SDV genes in each sample. One would expect the
154 replication rate under random sampling to be higher than 5% even if the FDR is properly controlled at
155 5%.

156 Since our GAMLSS framework allowed us to simultaneously test for sex differences in both
157 mean and overdispersion we examined the relationship between SDV genes and genes with sex-biased
158 expression (SBE). In several tissues a majority of SDV genes also displayed sex-biased expression
159 (**Supplementary Figure 4**). Although a minority of SDV genes in breast tissue, 464 genes with
160 significant sex differences in both mean and overdispersion showed discordance, with higher
161 overdispersion but lower mean expression in one sex versus the other (**Supplementary Figure 4**).
162 Specifically, 387 of those genes simultaneously had significantly higher overdispersion in males and
163 significantly higher mean expression in females.
164 Given that genes with sex-biased expression had been shown to be overrepresented on the X
165 chromosome in a variety of tissues and species (Khodursky et al., 2020; Oliva et al., 2020a; Parisi et al.,
166 2003), we wondered if SDV genes were similarly enriched on the X chromosome. However, we found no
167 consistent enrichment of SDV genes on the X chromosome; the enrichment patterns were highly tissue-
168 dependent (**Supplementary Figure 5**).

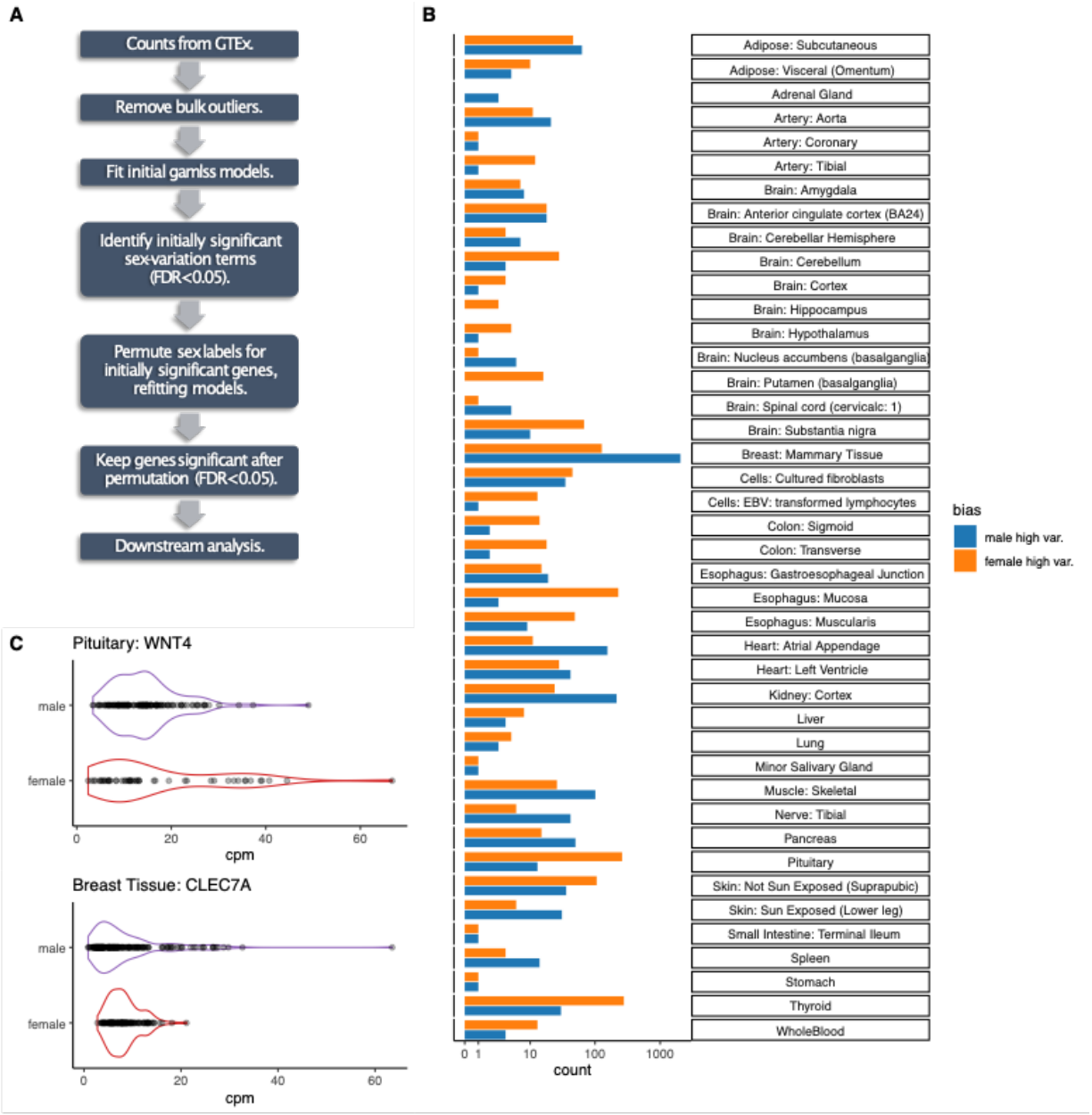


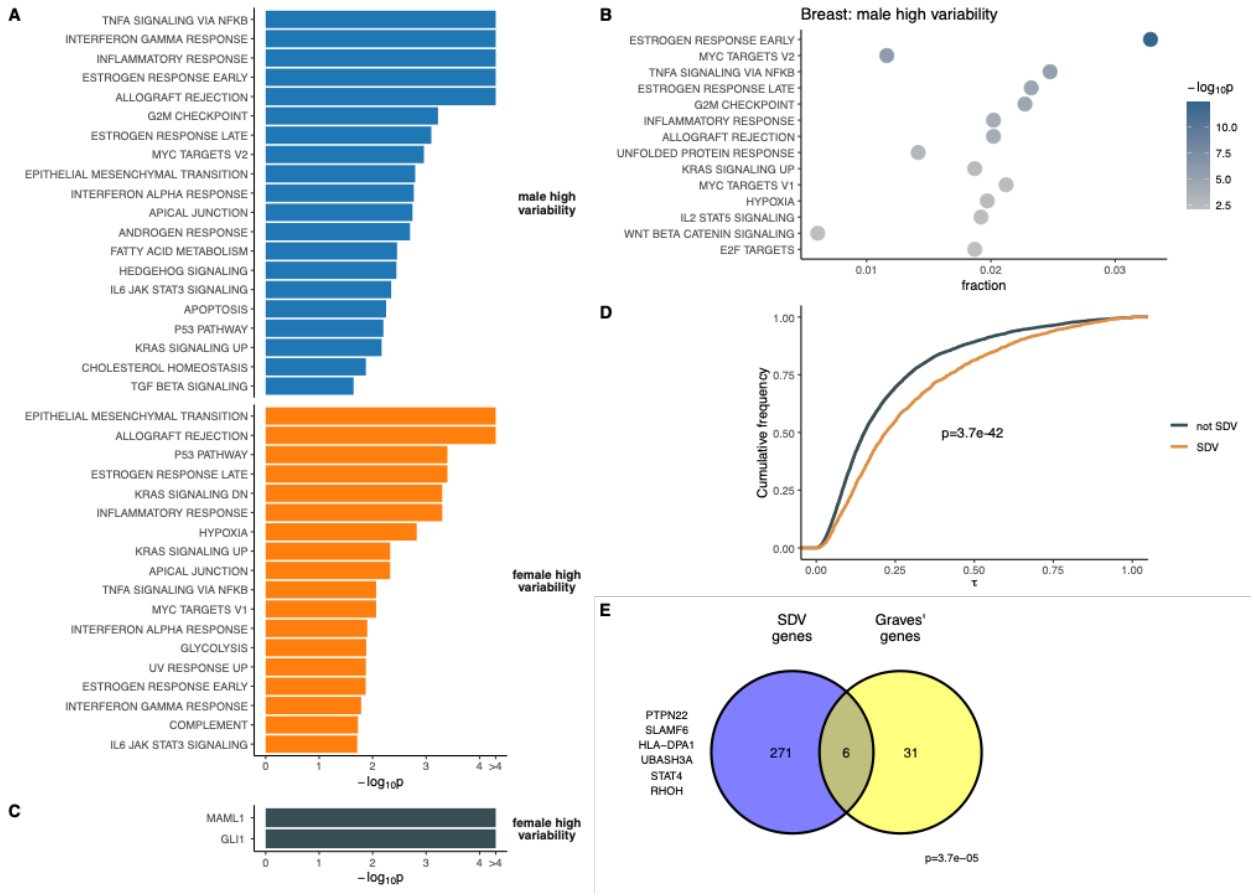

**Figure 1. The identification of genes with sex-differentially-variable gene expression.** A) Outline of pipeline to identify SDV genes. B) The number of genes with SDV expression identified in each tissue. The X-axis has a pseudo-log scale. C) Two examples of SDV genes with higher variation in females (WNT4 in the pituitary gland) and in males (CLEC7A in breast tissue), respectively.

To identify biological processes associated with SDV gene expression, we performed an overrepresentation analysis using gene sets obtained from the hallmark gene set collection of MSigDB (Liberzon et al., 2015; Subramanian et al., 2005). The hallmark gene set collection consists of non-redundant and refined gene sets associated with various biological processes and states (Liberzon et al.,

179 2015). The hallmark gene set collection is ideal for our analysis since we identified few SDV genes in
180 most tissues – a larger, more redundant, and less refined collection would introduce more noise while
181 increasing the multiple testing burden for our overrepresentation analysis. We first performed an
182 enrichment analysis at the tissue level and aggregated results across tissues using an empirical approach
183 based on Fisher's combined probability method (**Figure 2A, B, C**) (see methods). Following the meta-
184 analysis, we found that genes with higher variation in males were overrepresented in sets related to
185 estrogen and androgen response, immune response, and cancer-associated pathways (**Figure 2A**). Genes
186 with increased variation in females were similarly overrepresented in sets related to immune response,
187 estrogen response, and cancer-associated pathways (**Figure 2A)**. Notably, there was overlap in multiple
188 sets between the two sexes even though the lists of SDV genes did not show much overlap. At the single-
189 tissue level, genes with higher variation in males in breast tissue were overrepresented in sets related to
190 estrogen response and immune response (**Figure 2B**). This suggests that genes of particular biological
191 importance to one sex (genes related to estrogen response in female breast tissue) may show higher
192 variation in the opposite sex. To identify transcription factors whose targets are overrepresented in the set
193 of SDV genes we downloaded the GTRD dataset from MSigDB (Liberzon et al., 2015; Subramanian et
194 al., 2005). The GTRD collection contains sets of transcription factors and their predicted binding targets
195 (within the promoter region) (Kolmykov et al., 2021). We first performed an enrichment analysis at the
196 individual tissue level, followed by a meta-analysis across tissues using the same modified Fisher's
197 combined probability approach mentioned above (**see methods**). Following the meta-analysis, we found
198 that genes with higher variability in females were significantly enriched for targets of MAML1 and GLI1
199 (**Figure 2C**). GLI1 is a transcription factor that is part of the Hedgehog signaling pathway, which is
200 involved in differentiation during development. However, in fully differentiated tissues GLI1 can act as
201 an oncogene has been implicated in a wide variety of cancers, including glioblastoma, breast cancer, and
202 colon cancer (Avery et al., 2021; Clement et al., 2007; Sun et al., 2014). MAML1 is a transcription factor
203 involved in Notch signaling and is a transcriptional co-activator of Notch receptors (Wu et al., 2000).

204 Additionally, MAML1 has been shown to interact cooperatively with GLI1 to regulate hedgehog
205 signaling (Quaranta et al., 2017).
206
207 **Single-cell RNA-seq data from breast epithelial cells shows that genes with SDV expression tend to**
208 **be expressed in a cell type-specific manner**
209 To get further biological insight into genes with SDV expression in breast tissue, which had the most
210 SDV genes identified out of any tissue by far, we analyzed a previously published single cell RNA-seq
211 dataset (Nguyen et al., 2018). With count data from individual 4 (female), we applied a standard single
212 cell RNA-seq pipeline using the 'R' library 'Seurat' (Hao et al., 2020)(see methods). Our analysis
213 recovered the three major cell types found in the breast epithelium: type 1 and type 2 luminal cells along
214 with basal cells (Nguyen et al., 2018)(**Supplementary Figure 6**). To determine the cell type specificity of
215 each gene we combined raw count data from each cell to generate a pseudo-bulk count matrix for each
216 cell type and calculated CPM values for each gene. We then applied a standard measure of tissue
217 specificity, $\tau$, to the log-transformed CPM values across the three cell types (see methods).
218 We found that genes exhibiting SDV expression were expressed in a significantly more cell-type-
219 specific manner than non-SDV genes (**Figure 2D**). If there exist significant sex differences in inter-
220 individual cell-type composition variation, such cell-type-specific expression could lead to sex differences
221 in expression variability in bulk tissue samples. Overall, these results suggest that SDV expression may
222 be partly due to sex differences in inter-individual cell-type composition variation.
223
224 **SDV genes in the thyroid with higher variability in females are associated with Graves' disease**
225 Given the preponderance of sex-biased diseases (Mauvais-Jarvis et al., 2020), we wondered if genes with
226 SDV expression could be associated with a classically sex-biased disease. Hence, we decided to examine
227 Graves' disease: a common autoimmune disease of the thyroid which results in hyperthyroidism (McIver
228 and Morris, 1998). The incidence of Graves' disease is almost ten-fold higher in females than in males
229 (McIver and Morris, 1998). We obtained GWAS summary data from the NHGRI-EBI GWAS catalog

compiled from 5 studies on the genetics of Graves' disease in various populations (Buniello et al., 2019; Chu et al., 2011; Cooper et al., 2012; Nakabayashi et al., 2011; Sakaue et al., 2021; Zhao et al., 2013). Of the 37 genes implicated in Graves' disease which were expressed in thyroid tissue, 6 displayed SDV expression with higher variation in females. This corresponds to a significant 9.41-fold enrichment of Graves' associated genes among SDV genes with higher variability in females (p=3.7e-5, hypergeometric test) (Figure 2E). Although many of the genes implicated in Graves' disease are expressed in immune cells, a characteristic of Graves' disease is the infiltration of immune cells into the thyroid (Dayan et al., 1991; McIver and Morris, 1998; TOTTERMAN et al., 1979). Overall, this analysis suggests a role for SDV expression in sex-biased diseases.



242 **Figure 2. Biological properties of SDV genes.** A) Following meta-analysis SDV genes with higher
243 variation in males and females are enriched for terms related to various signaling pathways, immune
244 response, and sex-hormone response. B) Genes with higher variability in males are enriched for terms
245 related to estrogen and immune response in breast tissue. C) Following meta-analysis genes with higher
246 variability in females are enriched for targets of two transcription factors. D) SDV genes identified in
247 breast tissue show highly cell-type specific expression in breast epithelial tissue. E) Genes with higher
248 variability in females in the thyroid are significantly enriched for genes associated with Graves' disease.
249

250 **Genes with sex-differentially variable expression show evidence of increased selective constraint**

251 **acting on gene expression and sequence**

252 Given that genes with sex-biased expression (SBE) have been shown to be associated with
253 relaxed selective constraint at both the sequence and expression levels, we asked whether SDV genes
254 would similarly be associated with reduced constraint (Gershoni and Pietrokovski, 2017; Naqvi et al.,
255 2019). Stabilizing selection acting on gene expression is thought to reduce *cis*-regulatory variation and
256 reduce the minor allele frequency (MAF) of *cis*-regulatory variants (Glassberg et al., 2019; Josephs et al.,
257 2015). Therefore, if SDV expression is associated with reduced selective constraint, one would expect an
258 enrichment of *cis*-regulatory variation in SDV genes. To determine whether SDV genes were enriched or
259 depleted for *cis*-regulatory variation, we obtained expression quantitative trait loci (eQTL) summary data
260 for each tissue from GTEx (v8). For each tissue with at least 20 autosomal SDV genes, we tested for
261 enrichment of genes with significant eQTLs (eGenes, q-value<0.1) within that tissue's set of SDV genes.
262 Only genes that were tested in both studies in a given tissue were considered. Contrary to expectation, we
263 observed a depletion of eGenes within the set of SDV genes in most tissues, and this depletion was
264 significant in three tissues: sun-exposed skin, breast tissue, and pituitary glands (**Figure 3A**). We also
265 tested whether the median fold enrichment of eGenes across the 20 tested tissues (median fold enrichment
266 = 0.92) was lower than expected by chance. To this end, we randomly permuted the SDV labels in each
267 tissue (genes were either SDV or not SDV in each tissue) and calculated a median fold enrichment across
268 all tissues. In total 10,000 permutations were performed and for each round of permutation, the median
269 fold enrichment of eGenes in the set of SDV genes was recorded. We found that the observed median fold

270 enrichment of eGenes among SDV genes was significantly lower than expected by chance suggesting that
271 expression levels of SDV genes are evolving under elevated constraint (p=4e-04) (**Figure 3B).**
272 We examined if the observed depletion of *cis*-regulatory variation in SDV genes in breast tissue
273 could be explained by their elevated cell-type specificity. For example, one could imagine that most
274 expression variation (across bulk samples) in cell-type specific genes is driven by sample-to-sample
275 differences in cell-type composition and not by *cis*-regulatory genetic factors. Therefore, one might
276 expect to observe a depletion of eQTLs in cell-type specific genes. Upon examination, we found no
277 significant relationship between cell-type specificity and the presence or absence of a significant eQTL
278 (p=0.93, Mann-Whitney *U* test)(**Supplementary Figure 7**).
279 Although our analysis of eQTL data provided some evidence that expression levels of SDV genes
280 are evolving under increased constraint, a depletion of *cis*-regulatory variation might also be expected if
281 the gene expression variation observed in SDV genes was due to environmental factors and/or *trans*-
282 regulatory genetic variation to a larger extent than in non-SDV genes. We therefore sought to quantify
283 selective constraint using a complementary—and more direct—approach. To this end, we used 20-species
284 phastCons scores to look for further evidence that the *cis*-regulatory regions of SDV genes were
285 evolutionary highly constrained (Siepel et al., 2005). In particular, we examined the sequence
286 conservation in the proximal promoter regions of genes – 250 bp and 100 bp upstream from the
287 transcription start sites (TSSs). The phastCons score of a nucleotide can be interpreted as the probability
288 that the nucleotide resides in a conserved element (Siepel et al., 2005). We found that genes with SDV
289 expression in at least one tissue showed significantly elevated phastCons scores compared to genes which
290 did not exhibit SDV expression in any tissues (p=2.7e-7 for 250 bp upstream and p=1.6e-08 for 100 bp
291 upstream, Mann-Whitney U test) (**Figure 3C, D**). Only genes expressed in all tissues with at least 20
292 SDV genes were considered to remove bias, i.e., a gene is more likely to be SDV if it's expressed in more
293 tissues and genes expressed in many tissues tend to be more evolutionarily constrained (Karczewski et al.,
294 2020) .

295 Finally, we asked whether the increased constraint acting on the expression levels of SDV genes
296 would be reflected in other metrics of dosage sensitivity. To do this, we compared LOEUF scores in SDV
297 genes and non SDV genes. LOEUF (loss-of-function observed/expected upper bound fraction) is a score
298 that reflects the tolerance of a protein to loss-of-function mutations (Karczewski et al., 2020). The scores
299 range from 0 to 2, with higher scores indicating higher tolerance to loss-of-function mutations and
300 therefore reduced selective constraint. We only considered genes expressed in all tissues with at least 20
301 SDV genes and found that genes with SDV in at least one tissue had significantly reduced LOEUF scores
302 compared to genes without SDV expression in any tissue (p=2.2e-12, Mann-Whitney U test) (**Figure 3E**),
303 corroborating that SDV genes tend to be more constrained at expression level.

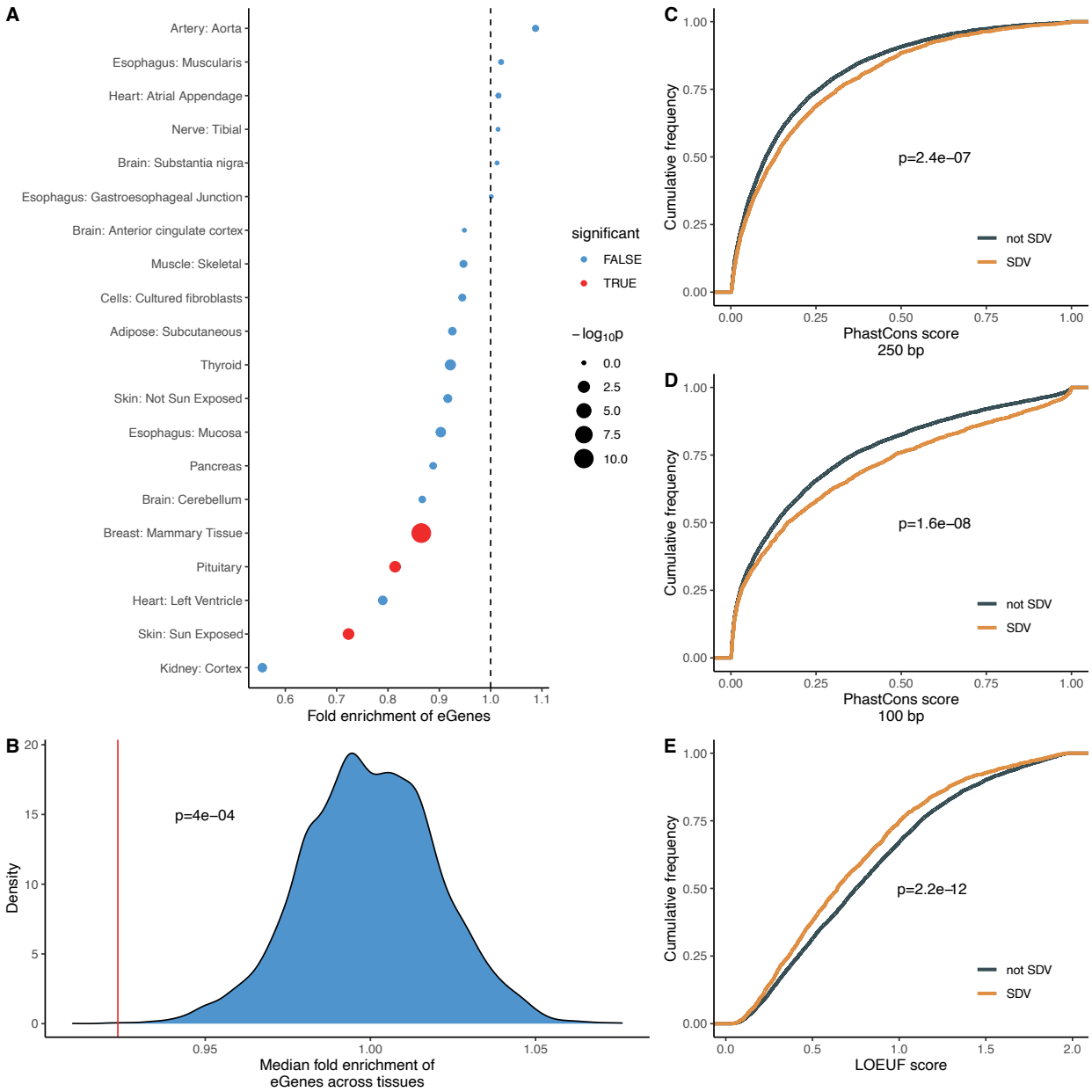



**Figure 3. SDV genes are associated with increased constraint.** A) 14 out of 20 tested tissues with at least 20 SDV genes show a depletion in eGenes among SDV genes B) Across the 20 tested tissues the median fold enrichment of eGenes is significantly lower than expected by chance. The red line shows the observed median fold-enrichment, while the blue distribution shows the empirical null distribution. C) and D) Genes with SDV expression in at least one tissue show significantly elevated mean phastCons scores in proximal promoters (250 bp and 100 bp upstream of TSS) indicating increased conservation. E) Genes with SDV expression in at least one tissue show significantly reduced LOEUF scores, indicating more intolerance to loss of function mutations. For C), D), and E) SDV genes were considered to be SDV if they showed SDV expression in at least one tissue out of the 20 tested tissues. Only genes expressed in all 20 tested tissues were considered to remove bias.

317 **SDV genes in breast tissue show evidence of sex-specific genetic regulation**

318 To determine if SDV genes showed evidence of sex-specific genetic regulation, we downloaded sex-
319 biased eQTL (sbQTL) data from GTEx (Oliva et al., 2020). We restricted our analysis to breast tissue
320 given that it was found to have both the largest number of SDV genes (in our study) and by far the largest
321 number of sbQTLs of any tissue (Oliva et al., 2020). To test if genes with sbQTLs were enriched for
322 SDV genes we performed a gene set enrichment analysis (GSEA). By only considering genes tested in
323 both analyses and ranking genes according to their corrected $-\log_{10}$(p-values) in the sbQTL analysis we
324 found that SDV genes were significantly enriched at lower p-values (p=1.7e-2) (**Supplementary Figure**
325 **8**). Additionally, we found that all 13 SDV genes with higher variation in males and significant sbQTLs
326 (FDR<0.25, the threshold used in Oliva et al., 2020) had eQTLs with larger effect sizes in males than
327 females. This represents a significant enrichment of genes with larger eQTL effect sizes in males relative
328 to the set of all genes with significant sbQTLs (p=8e-4, hypergeometric test). These results indicate that
329 SDV genes are enriched for sex-specific genetic architectures regulating gene expression and suggest that
330 SDV expression may have an underlying genetic component.

331

332 **Discussion**

333 In this work, we examined a largely unappreciated aspect of transcriptomic sex differences: genes with
334 sex differences in inter-individual expression variation. In particular, we focused on genes that showed
335 sex differences in inter-individual expression overdispersion. We termed these genes sex-differentially
336 variable (SDV) genes. Across 43 tested tissues, we found that roughly 18% of expressed genes showed
337 SDV patterns of expression in at least one tissue (**Figure 1B**). Our results were undoubtedly affected by
338 the unequal number of samples in different tissues – we had more power to detect SDV genes in some
339 tissues than others. Nevertheless, we identified more SDV genes in breast tissue than in any other tissue.
340 Issues of sample size aside, from a physiological perspective this seems logical since the breast is the only
341 tissue tested known to have functional significance in only one sex: the production of milk is essentially
342 female-specific. Other studies have found evidence that low inter-individual transcriptional variation in a

343 gene is indicative of increased constraint acting on the gene (Fair et al., 2020). Since the vast majority
344 (94%) of SDV genes identified in breast tissue showed higher expression variability in males than in
345 females, this is consistent with the idea that expression levels should be more functionally constrained in
346 female breast tissue than in male breast tissue.

347 To gain biological insight into the nature of SDV genes we performed a gene set
348 overrepresentation analysis followed by meta-analysis. We found that SDV genes were enriched for genes
349 from sets related to sex hormone response, immune response, and various signaling and cancer-associated
350 pathways (**Figure 2A, B**). Additionally, we found that SDV genes were enriched for the targets of two
351 transcription factors implicated in cancer (**Figure 2C**). The overrepresentation of SDV genes in gene sets
352 related to various biological pathways and transcription factors indicates that SDV expression may be
353 caused by SDV activity at the pathway level. In breast tissue, the overrepresentation of genes with high
354 variability in males in sets related to estrogen response suggests that genes of particular biological
355 importance to one sex may show higher variation in the opposite sex.

356 In order to further examine the biological significance of SDV genes in breast tissue we analyzed
357 previously published single-cell RNA sequencing data of breast epithelial tissue. We found that SDV
358 genes (identified in the GTEx data) were expressed in a more cell-type specific manner (**Figure 2D**).
359 These results indicate that SDV genes may be more likely to be involved in cell-type specific functions.
360 Additionally, this suggests that sex differences in gene expression variability may be partially driven by
361 sex differences in cellular composition variability across individuals.

362 To examine a possible link between SDV expression and disease, we analyzed summary GWAS
363 data from a classically female-biased disease: Graves' disease. We found a significant enrichment of
364 Graves' associated genes in the set of SDV genes with higher variability in females in thyroid tissue
365 (**Figure 2E**). These results indicate that SDV expression may be either an indicator or cause of sex-biased
366 disease. In the future, the role of SDV expression in sex-biased disease should be thoroughly investigated.

367 We found that SDV genes tended to be depleted for *cis*-regulatory variation overall (**Figure 3A.**
368 **B**). Additionally, we showed that SDV genes show significantly increased constraint in both proximal

369 promoter and protein-coding sequences. These results are consistent with increased constraint acting on
370 the expression levels of SDV genes and suggest that SDV genes may be evolving under increased sex-
371 specific constraint acting on gene expression.

372 We propose that SDV expression can arise through sex-specific increases in stabilizing selection
373 acting on gene expression. To test this, we performed a forward Monte Carlo simulation of the evolution
374 of a gene evolving under sex-specific stabilizing selection (**Figure 4**). Our model assumes that the
375 expression depends on the additive contributions of a shared genetic locus (this locus contributes equally
376 to both sexes), two sex-specific loci (each locus only contributes to expression in one sex), and noise (to
377 model environmental effects) (see methods). Therefore, an assumption of the model is some degree of
378 sex-specific genetic architecture for SDV genes, which is supported by empirical evidence
379 (**Supplementary Figure 8**). The three loci were in linkage equilibrium. Increased stabilizing selection in
380 one sex results in a difference in inter-individual phenotypic variances between the two sexes – with the
381 sex experiencing increased selection having relatively reduced variation (**Figure 4A, B**). Additionally,
382 increased selection in one sex results in reduced genetic variation in the set of shared genetic loci –
383 consistent with our observations of reduced *cis*-regulatory variation in SDV genes (**Figure 4C**). These
384 results are qualitatively consistent with predictions from quantitative genetics theory. Within a single
385 generation, stronger stabilizing selection in one sex – for example, females – would lead to reduced
386 additive genetic variance in females compared to males. Assuming that phenotypic variance is the sum of
387 additive genetic and environmental variances, and that environmental variances are the same in both
388 sexes, it implies that phenotypic variance will be lower in females than in males. Additionally, increased
389 stabilizing selection in females will decrease the additive genetic covariance between the sexes to a larger
390 extent than identical stabilizing selection in both sexes. This corresponds to our observation of reduction
391 in the variance of the shared genetic locus under conditions of increased selection in females (**Figure 4C**).
392 A more detailed description of the theoretical predictions is presented in the **Appendix**.

393 Overall, our results suggest that genes with sex differences in inter-individual gene expression
394 variability are possibly involved in sex-specific biological functions and may be evolving under sex-

specific constraint acting on gene expression. These results highlight the possible importance of SDV genes in a functional and evolutionary sense, but future work needs to be done to properly understand the role of SDV genes in human health.

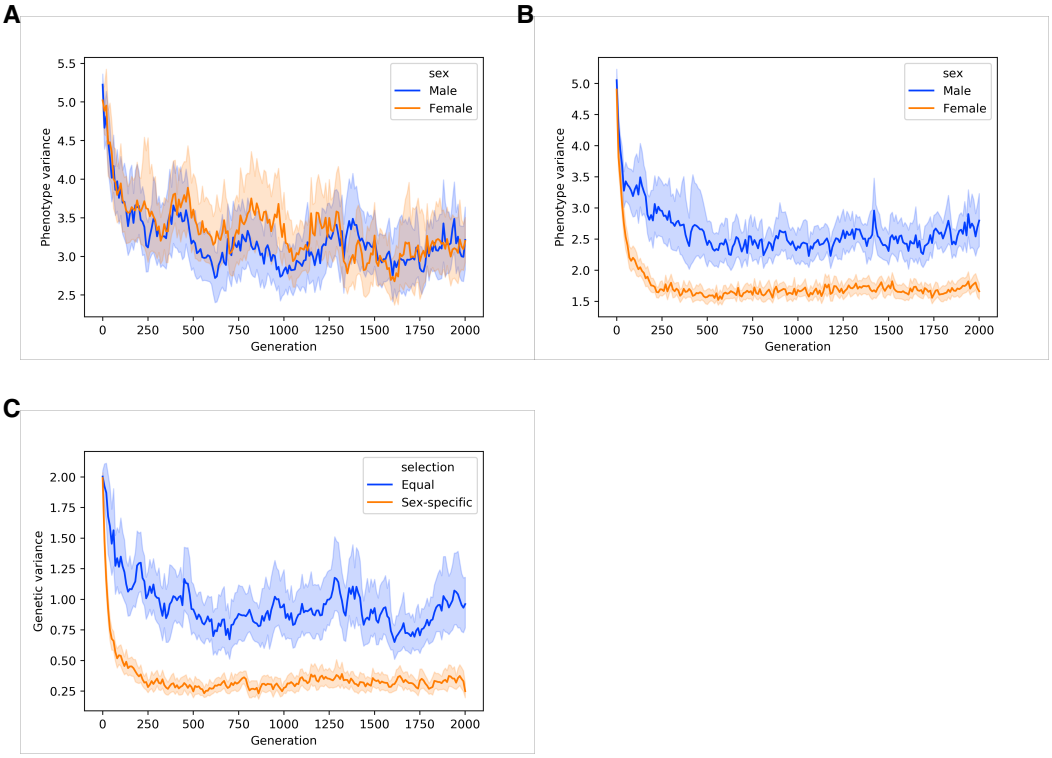



**Figure 4**. **Simulations showing that SDV expression can arise through sex-specific constraint.** Simulation results showing the output from 30 simulations in both the equal selection and sex-specific selection conditions. A) Equal constraint in both sexes results in similar phenotypic variance in both sexes. B) Increased sex-specific constraint acting in females reduces the phenotypic (expression) variation across individuals. C) Increased constraint in females results in reduced genetic variance, consistent with our observations of reduced *cis*-regulatory variation in SDV genes.

## Methods

### Power Simulations

Random samples were generated from negative binomial distributions using the 'R' function 'rnegbin' from the library 'MASS'. To simulate the effect of 'sex' on the overdispersion parameter, samples were drawn from two negative binomial distributions with equal mean ($\mu$) parameters and different overdispersion ($\sigma$, or equivalently $\frac{1}{\theta}$ in the 'MASS' documentation) parameters. Similarly, to determine false positive rates samples were drawn from identical distributions. To restrict the search space, $\mu$ was fixed at 1500 for all simulations and the size of both samples combined was fixed at 400. These parameters were typical of the gene-expression levels and sample sizes found in GTEx data. Moreover, varying the mean parameter only marginally affected power calculations. For the unbalanced group sizes simulations, the size of one group was 320 while the size of the other was 80. Both the baseline overdispersion parameter ($\sigma$) and the difference in overdispersion between groups ($\Delta\sigma$) were varied over the course of the simulation experiments. The variances of the two samples were compared using the Fligner-Killeen test ('fk.test' function from the 'R' library 'npsm'), Levene's test ('levene.test' function from the 'R' library 'lawstat'), and a likelihood-ratio test implemented using negative binomial GAMLSS models. The more complex GAMLSS model parametrized the mean ($\mu$) and the overdispersion ($\sigma$) parameters as functions of sex, $\sigma \sim sex$ and $\mu \sim sex$, while the simplified model parametrized only the mean in terms of sex with the overdispersion being modeled as constant. The p-value for the significance of the sex overdispersion term was saved. At each parameter combination 1000 sample pairs were drawn and p-values from each of the three tests were saved. Power was calculated as the fraction of 1000 simulations where the p-value fell below an alpha of 0.05. False positive rate was calculated as the faction of 1000 simulations where the p-value fell below and alpha of 0.05 when the two underlying distributions being sampled for a sample pair were identical. It was observed that power for Levene's test markedly improved when log-transforming the sampled count data, while the Fligner-Killeen test showed largely unaltered power: hence samples were transformed as ln(counts + 1) before testing with Levene's test and the Fligner-Killeen test. Raw samples were used for the GAMLSS likelihood-ratio test.

435 **Identification of SDV genes**

436 For each tissue, SDV genes were identified separately using RNA-seq count data downloaded from
437 GTEx-v8 (dbGAP accession number # phs000424.v8.p2). To reduce confounding due to population
438 structure, only European Americans were considered. First, counts per million (CPM) values were
439 calculated using edgeR and genes with no expression (CPM=0) in at least one individual were removed.
440 To identify individuals that were global expression outliers, dimensionality reduction was then performed
441 on $\log_2$ transformed CPM values using principal component analysis (PCA) with the 'R' function
442 'prcomp'. Since outliers can greatly affect estimates of variation, we removed individuals whose global
443 expression profiles deviated significantly from other individuals in a given tissue. Individuals whose
444 global expression patterns were more than 1.5 interquartile ranges higher than the third quartile or lower
445 than the first quartile along any principal component explaining more than 5% of the total variance in the
446 dataset were considered outliers and removed from further analysis.

447 Following global outlier removal three negative binomial models were fit to the count data (one
448 gene at a time) using gamlss. The first and most complex model (model 1) parametrized the mean and
449 non-Poisson variation terms of the negative binomial distribution ( $\mu$ and $\sigma$, respectively) as $\mu \sim sex +$
450 $age + ischemic\ time + RIN + PC1 + PC2 + PC3 + PC4 + PC5 + \ offset(offset)$ and $\sigma \sim sex +$
451 $age + ischemic\ time + RIN + PC1 + PC2 + PC3 + PC4 + PC5$ , where $RIN$ represents RNA integrity
452 number and the 5 PCs represent genotype principal components obtained from GTEx . The offset term, to
453 account for library size was calculated as $\log(library\ size\ * normalization\ factors)$ , where the
454 normalization factors were calculated using the 'calcNormFactors' function from edgeR using the
455 'upperquartile' method. Model 2 had the same form as model 1 for $\mu$ , however for the formula for the
456 non-Poisson variation was $\sigma \sim age + ischemic\ time + RIN + 5PCs$. Model 3 had the same formula as
457 model 1 for $\sigma$ however the formula for the mean was $\mu \sim age + ischemic\ time + RIN + 5PCs +$
458 $offset(offset)$ . To test for the significance of the effect of sex on non-Poisson variation ($p_{\sigma,sex}$ ) a
459 likelihood ratio test was performed comparing model 1 and model 2. To test for the significance of the

460 effect of sex on mean expression ($p_{\mu,sex}$ ) a likelihood ratio test was performed comparing model 1 and
461 model 3. The coefficients corresponding to the sex terms for both $\mu$ and $\sigma$ ($\beta_{\mu,sex}$ and $\beta_{\sigma,sex}$ ) from model
462 1 were recorded for each gene. After every gene was tested in a given tissue p-values for $\mu$ and $\sigma$ were
463 corrected independently using the Benjamini-Hochberg (BH) procedure using the 'R' function 'p.adjust'.
464 Unsurprisingly we noticed that individual expression outliers at the gene level dramatically
465 affected our estimates of $p_{\sigma,sex}$ and $\beta_{\sigma,sex}$. To account for this and to generally validate the robustness of
466 our models we performed randomization tests for all genes with a BH corrected $p_{\sigma,sex}$ <0.05. For each
467 gene the sex labels were randomly permuted, model 1 was fit to the randomized data, and $\beta_{\sigma,sex}$ values
468 were recorded. We performed 1000 permutations for every gene to generate an empirical null distribution
469 for $\beta_{\sigma,sex}$. An empirical p-value was calculated as the fraction of random permutations where
470 $abs(\beta_{\sigma,sex,perm}) > abs(\beta_{\sigma,sex,obs})$. These empirical p-values were then corrected using the BH
471 procedure. In the end genes were considered SDV if the original BH corrected p-values ($p_{\sigma,sex}$) were less
472 than 0.05 and the BH corrected empirical p-values were less than 0.05.
473
474 **Hallmark enrichment analysis**
475 Gene set annotations for collection 'H' (hallmark gene sets) were downloaded from MSigDB using the
476 'R' function 'msigdbr' from the package 'msigdbr'. Only tissues with at least 20 male or female high
477 variability genes were considered. Enrichment of gene sets for male high variability and female high
478 variability genes was calculated separately. Enrichment p-values for each gene set in each tissue were
479 calculated using a hypergeometric test (using the 'R' function 'phyper'). For each gene set a meta-
480 analysis was performed across all tested tissues using a variation of Fisher's method. This approach was
481 implemented as follows. For each gene set a test statistic was calculated by taking -2 times the sum of the
482 log p-values in each tested tissue. In an ordinary implementation of Fisher's method this test statistic
483 would have a chi-squared distribution under the null hypothesis. However, given that the underlying p-
484 values are based on a discrete test (the hypergeometric test) a core assumption of Fisher's method is

485 violated: namely that for an individual tissue p-values should be uniformly distributed under the null
486 hypothesis. To overcome this, we generated empirical null distributions for the test statistic described
487 above using a permutation approach: within each tissue we permuted the labels of each gene (not SDV,
488 female high variability, and male high variability) and calculated a p-value for the label of interest
489 (female high variability or male high variability) using a hypergeometric test. This permutation procedure
490 was done for every tissue tested – finally resulting in a null value for the sum of log p-values test statistic
491 described above. For each gene set 10,000 such permutations were performed to generate an empirical
492 null distribution for the test statistic. The empirical p-values for each gene set were calculated by finding
493 the fraction of permutations resulting in a test statistic larger than the one observed. These combined p-
494 values across gene sets were corrected using the BH procedure and gene sets with corrected p-
495 values<0.05 were considered significant. Gene sets that were highly tissue/developmentally specific were
496 not tested in the meta-analysis. The 7 removed hallmark gene sets were pancreas beta cells, adipogenesis,
497 myogenesis, spermatogenesis, bile acid metabolism, heme metabolism, and xenobiotic metabolism.
498
499 **Transcription factor enrichment analysis**
500 The analysis was performed in a fashion identical to the hallmark enrichment analysis except that the
501 gene sets from collection 'C3' subcollection 'GTRD' were used.
502
503 **Single-cell RNA sequencing analysis**
504 The count matrix for individual 4 was downloaded from GSE113197 (Nguyen et al., 2018). All
505 clustering and marker identification analysis was performed using the 'R' library 'Seurat' version 3.2.3.
506 Cells which had fewer than 500 and more than 6000 expressed features were removed. Cells with more
507 than 10% of their transcripts coming from mitochondrial genes were removed. The counts were
508 normalized using the function 'SCTransform'. Both dimensionality reduction (using the function
509 'RunUMAP') and clustering (function: 'FindClusters') were performed using the first 10 principal
510 components of the normalized counts. L1, L2, and basal cells were identified based on their expression of

511 markers “KRT18”, “SPLI”, and “KRT14”. Clusters expressing “KRT18” and “SPLI” were deemed to be
512 L1 cells, clusters only expressing “KRT18” were deemed to be L2 cells, and clusters expressing
513 “KRT14” were taken to be basal cells.

514 To calculate the cell-type specificity of each gene raw counts were aggregated within each of the
515 three cell types, generated 3 pseudo-bulk count vectors (one for each cell type). For each cell-type CPMs
516 were calculated by multiplying the pseudo-bulk count vector by 1 million and dividing by the sum of the
517 vector. The CPM values were then $\log_2$ transformed. Cell-type specificity was calculated using the
518 standard metric for tissue specificity $\tau = \frac{sum(1 - x_{hat})}{length(x_{hat}) - 1}$, where $x_{hat}$ is the vector representing a given
519 gene’s expression across the 3 cell-types divided by the maximum expression of that gene in any of the
520 three cell-types.

521

522 **Graves’ disease analysis**

523 Summary GWAS statistics were obtained from the NHGRI-EBI GWAS catalog for study EFO_0004237,
524 downloaded on 08/04/2022 (Buniello et al., 2019).

525

526 **eQTL analysis**

527 Summary statistics for single-tissue eQTL mapping performed in European-American individuals were
528 downloaded from the GTEx-v8 website. Genes with at least one eQTL with q-value<0.1 were considered
529 to be eGenes. A two-sided Fisher’s exact test, as implemented in the ‘fisher.test’ function, was used to
530 test for the enrichment/depletion of eGenes among SDV genes in each tissue. Only tissues with at least 20
531 autosomal SDV genes were considered. Only autosomal genes with expression high enough to be tested
532 in both the original eQTL study and our study were considered. To test if the median fold enrichment of
533 eGenes across tissues was significant we randomly permuted the SDV labels in each tissue and calculated
534 a median fold enrichment across all tissues. In total 10,000 permutations were performed and for each
535 round of permutation the median fold enrichment of eGenes across the 20 tissues was recorded. The

536 empirical p-value was calculated as the fraction of all permutations where the fold enrichment was at least
537 as low as the observed fold enrichment.
538
539 **PhastCons analysis**
540 Genome-wide phastCons scores (in hg38 coordinates) generated from a 20 species alignment were
541 downloaded from http://hgdownload.cse.ucsc.edu/goldenPath/hg38/phastCons20way/. Transcription start
542 sites (TSS) for all annotated genes were downloaded from BioMart using the 'R' package biomaRt
543 (Durinck et al., 2005, 2009). The most distal TSS was used for each gene. Only genes expressed in all
544 tissues containing at least 20 SDV genes were tested (20 tissues overall). A gene was considered 'SDV' if
545 it displayed 'SDV' expression in at least 1 of the 20 tested tissues, otherwise it was considered to be 'not
546 SDV'. Requiring genes to be expressed in all 20 tissues was done to remove bias, i.e., a gene is more
547 likely to be SDV if it's expressed in more tissues and genes expressed in many tissues tend to be more
548 evolutionarily constrained (Karczewski et al., 2020). Further increasing the number of tissues (by
549 decreasing the minimum number of SDV genes allowed in each tissue) reduces power by decreasing the
550 intersection of genes expressed across all tissues. Mean phastCons scores were obtained for each gene in
551 a region 100 bp and 250 bp upstream from their TSSs.
552
553 **Loss of function analysis**
554 Loss of function metrics were downloaded from https://gnomad.broadinstitute.org/downloads . In
555 particular the v2.1.1 loss of function metrics by gene were used. The analysis was similar to the analysis
556 for PhastCons score: only genes expressed across all tissues with at least 20 SDV genes were considered.
557 LOEUF scores were compared between 'SDV' and 'not SDV' genes.
558
559 **Sex-specific stabilizing selection simulation**
560 The simulation assumed a diploid, diecious, and randomly mating population of 500 individuals. In each
561 individual the phenotype (expression) was modelled as the additive contribution of 3 genetic loci and

562 environmental noise. One of the genetic loci had equal effect in both sexes while each of the two other
563 genetic loci only had effect in one sex or the other. The allelic effects of each genetic locus in each
564 individual were initialized as draws from a standard normal distribution (mean=0 and standard
565 deviation=1). The environmental component was also taken to be a random number drawn from a
566 standard normal distribution. Similarly, mutation resulted in the addition of a number drawn from a
567 standard normal distribution to the allelic effect and the mutation rate was set to 5e-3 per generation at
568 each genetic locus in each individual. The fitness of each individual was taken to be $w = e^{-s(G+E)^2}$,
569 where '$s$' represented the strength of selection, '$G$' was the total sum of genetic allelic effects (from the
570 shared and sex-specific locus), and '$E$' was the environmental effect. For the simulations with equal
571 selection in both sexes '$s$' was 0.005, while in simulation with sex-specific selection '$s$' was 0.005 in
572 males and 0.05 in females. Selection only affected viability and had no impact on fecundity. An
573 individual survived to reproduce if a random draw from a uniform distribution on [0, 1] was less
574 than or equal to the individual's fitness; therefore, the fitness represents the probability of an
575 individual surviving to reproduce. In each generation 100 males and 100 females were chosen and
576 paired at random with replacement. Each pair produced 5 offspring and the sex of each offspring was
577 determined randomly. The simulations were run for a total of 2000 generations. Every 10 generations the
578 variance of the phenotype (across individuals in the population) was recorded along with the variance of
579 the allelic effect of the shared genetic component. For both the equal selection and sex-specific selection
580 models 30 independent simulations were run.
581
582 **GSEA analysis of sbQTL data**
583 Sex-biased QTL (sbQTL) data from breast tissue was downloaded from GTEx v8. Genes were only
584 considered if they were tested in our analysis and in the sbQTL analysis. Genes were ranked according to
585 corrected -$\log_{10}$(p-values) where the (Bonferroni) correction was made to account for multiple SNPs

tested per gene. GSEA analysis was performed using the 'R' function 'fgsea' from the library 'fgsea' (Korotkevich et al., 2021).

## Appendix

### Theory behind sex-specific stabilizing selection

The evolution of a single quantitative trait in males and females can be predicted from the 2-by-2 additive genetic variance-covariance matrix:

$$\boldsymbol{G} = \begin{bmatrix} G_m & B \\ B & G_f \end{bmatrix}. \qquad (1)$$

Where the diagonal terms represent the additive genetic variances in males and females, while the identical off-diagonal terms represent the additive genetic covariance between the sexes (Barker et al., 2010; Lande, 1980). In the absence of directional selection, the change in the genetic variance-covariance matrix due to stabilizing selection over the course of a generation is (Barker et al., 2010; Phillips and Arnold, 1989):

$$\Delta\boldsymbol{G} = \boldsymbol{G\gamma G}. \qquad (2)$$

Where $\boldsymbol{\gamma}$ is a matrix describing the strength of stabilizing selection within each sex:

$$\boldsymbol{\gamma} = \begin{bmatrix} \gamma_m & 0 \\ 0 & \gamma_f \end{bmatrix}. \qquad (3)$$

For stabilizing selection to occur the diagonal terms must be negative with decreasing values indicating stronger stabilizing selection. Substituting (3) and (1) into (2) yields:

$$\Delta\boldsymbol{G} = \begin{bmatrix} \Delta G_m & \Delta B \\ \Delta B & \Delta G_f \end{bmatrix} = \begin{bmatrix} \gamma_m G_m^2 + \gamma_f B^2 & B\left(\gamma_m G_m + \gamma_f G_f\right) \\ B\left(\gamma_m G_m + \gamma_f G_f\right) & \gamma_f G_f^2 + \gamma_m B^2 \end{bmatrix}. \qquad (4)$$

By assuming that initially the trait has identical genetic variance in males and females (which is assumed at the start of our simulation): we can use the simplification $G_m = G_f = G$. This leads to the following equations for $\Delta G_m$ and $\Delta G_f$ :

$$\Delta G_m = \gamma_m G^2 + \gamma_f B^2 \qquad (5a)$$

$$\Delta G_f = \gamma_f G^2 + \gamma_m B^2 \ . \qquad (5b)$$

In the case of stronger stabilizing selection in females: $\gamma_f < \gamma_m < 0$. Additionally, theoretical and empirical evidence suggests that generally the magnitude of intra-sex genetic variance is larger than the magnitude of inter-sex genetic covariance: $G > B$ (Barker et al., 2010). Therefore, $(5a, b)$ yields: $\Delta G_f < \Delta G_m < 0$. Following the assumptions made in our simulation (namely that the phenotypic variance in each sex is just the sum of the environmental variances and the additive genetic variances ($G$), and that the environmental variance is the same in both sexes) leads to the result that phenotypic variance will decrease to a larger extent in females than in males over the course of a generation. This is what we observed (however, over multiple generations) in our simulation (**Figure 4A, B**).

To further relate our simulation results to the theory described above we can formulate the phenotypes and their variances from the simulation as follows:

$$P_m = g_1 + g_2 + \varepsilon \qquad (6a)$$

$$P_f = g_1 + g_3 + \varepsilon \ . \qquad (6b)$$

$P_m$ and $P_f$ are the phenotypes in males and females respectively. $g_1$, $g_2$, and $g_3$ are the genetic contributions of the shared genetic locus, the male specific genetic locus, and the female specific genetic locus. The environmental contribution to the phenotypes is $\varepsilon$ . From $(6a, b)$, since all the phenotypic components are independent (so their covariances are 0), the phenotypic variances are:

$$Var(P_m) = Var(g_1) + Var(g_2) + Var(\varepsilon) \qquad (7a)$$

$$Var(P_f) = Var(g_1) + Var(g_3) + Var(\varepsilon) \qquad (7b)$$

and the phenotypic covariances are:

$$Cov(P_m, P_f) = Var(g_1) + Var(\varepsilon) \ . \qquad (7c)$$

From $(7a, b, c)$ it becomes apparent that $G_m$ corresponds to $var(g_1) + var(g_2)$ , $G_f$ corresponds to $var(g_1) + var(g_3)$ , and $B$ corresponds to $var(g_1)$. From the off-diagonal terms in equation $(4)$ it is evident that increasing the strength of selection in females (decreasing $\gamma_f$) makes $\Delta B$ more negative – which is consistent with the observed relative reduction in the genetic variance in the shared locus $(var(g_1))$ (**Figure 4C**).

**Data access and code availability**

The code generated for this work has been deposited on GitHub: https://github.com/LiZhaoLab/SDV_exp.

**Acknowledgments**

We thank Zhao lab members for discussions about the project, Jason Mezey and Christina Leslie for helpful discussions during early stages of the project, Mike Love for statistical advice, and Andy Clark for the suggestion of power analysis. We thank Matt Hahn and Ziyue Gao for their comments and suggestions for a recent version of the manuscript. We thank Jason Banfelder and Bala Jayaraman from The Rockefeller University High Performance Computing (HPC) Center for their dedicated work in these extraordinary times.

**Funding**

This work was supported by National Institutes of Health (NIH) MIRA R35GM133780, the Robertson Foundation, a Rita Allen Foundation Scholar Program, and a Vallee Scholar Program (VS-2020-35), a Monique Weill-Caulier Career Scientist Award, and an Alfred P. Sloan Research Fellowship (FG-2018-10627) to L.Z. C.S.J. was supported by UL1 TR001866 from the National Center for Advancing Translational Sciences (NCATS), NIH Clinical and Translational Science Award (CTSA) program. E.B.Z. was supported by a Medical Scientist Training Program grant from the National Institute of

657 General Medical Sciences of the NIH under award number: T32GM007739 to the Weill
658 Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program.

659 **Competing interest statement**

660 The authors declare that no competing interests exist.

661

662 **References**

663 Avery, J.T., Zhang, R., and Boohaker, R.J. (2021). GLI1: A Therapeutic Target for Cancer. Front. Oncol.
664 *11*.

665 Barker, B.S., Phillips, P.C., and Arnold, S.J. (2010). A test of the conjecture that G-matrices are more
666 stable than B-matrices. Evolution *64*, 2601–2613.

667 Beery, A.K., and Zucker, I. (2011). Sex bias in neuroscience and biomedical research. Neurosci.
668 Biobehav. Rev. *35*, 565–572.

669 Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A.,
670 Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published
671 genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47*,
672 D1005–D1012.

673 Catalán, A., Hutter, S., and Parsch, J. (2012). Population and sex differences in Drosophila melanogaster
674 brain gene expression. BMC Genomics *13*, 654.

675 Chella Krishnan, K., Mehrabian, M., and Lusis, A.J. (2018). Sex differences in metabolism and
676 cardiometabolic disorders. Curr. Opin. Lipidol. *29*, 404–410.

677 Choleris, E., Galea, L.A.M., Sohrabji, F., and Frick, K.M. (2018). Sex differences in the brain:
678 Implications for behavioral and biomedical research. Neurosci. Biobehav. Rev. *85*, 126–145.

679 Chu, X., Pan, C.-M., Zhao, S.-X., Liang, J., Gao, G.-Q., Zhang, X.-M., Yuan, G.-Y., Li, C.-G., Xue, L.-
680 Q., Shen, M., et al. (2011). A genome-wide association study identifies two new risk loci for Graves'
681 disease. Nat. Genet. *43*, 897–901.

682 Clement, V., Sanchez, P., de Tribolet, N., Radovanovic, I., and Ruiz i Altaba, A. (2007). HEDGEHOG-

683 GLI1 Signaling Regulates Human Glioma Growth, Cancer Stem Cell Self-Renewal, and Tumorigenicity.
684 Curr. Biol. *17*, 165–172.
685 Cooper, J.D., Simmonds, M.J., Walker, N.M., Burren, O., Brand, O.J., Guo, H., Wallace, C., Stevens, H.,
686 Coleman, G., Franklyn, J.A., et al. (2012). Seven newly identified loci for autoimmune thyroid disease.
687 Hum. Mol. Genet. *21*, 5202–5208.
688 Dayan, C.M., Londei, M., Corcoran, A.E., Grubeck-Loebenstein, B., James, R.F., Rapoport, B., and
689 Feldmann, M. (1991). Autoantigen recognition by thyroid-infiltrating T cells in Graves disease. Proc.
690 Natl. Acad. Sci. *88*, 7415–7419.
691 Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005).
692 BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis.
693 Bioinformatics *21*, 3439–3440.
694 Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of
695 genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.
696 Fair, B.J., Blake, L.E., Sarkar, A., Pavlovic, B.J., Cuevas, C., and Gilad, Y. (2020). Gene expression
697 variability in human and chimpanzee populations share common determinants. Elife *9*.
698 Falconer, D.S. (1967). The inheritance of liability to diseases with variable age of onset, with particular
699 reference to diabetes mellitus. Ann. Hum. Genet. *31*, 1–20.
700 Gershoni, M., and Pietrokovski, S. (2017). The landscape of sex-differential transcriptome and its
701 consequent selection in human adults. BMC Biol. *15*, 7.
702 Glassberg, E.C., Gao, Z., Harpak, A., Lan, X., and Pritchard, J.K. (2019). Evidence for Weak Selective
703 Constraint on Human Gene Expression. Genetics *211*, 757 LP – 772.
704 Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby,
705 C., Zagar, M., et al. (2020). Integrated analysis of multimodal single-cell data. BioRxiv
706 2020.10.12.335331.
707 Itoh, Y., and Arnold, A.P. (2015). Are females more variable than males in gene expression? Meta-
708 analysis of microarray datasets. Biol. Sex Differ. *6*, 1–9.

709 de Jong, T. V, Moshkin, Y.M., and Guryev, V. (2019). Gene expression variability: the other dimension
710 in transcriptome analysis. Physiol. Genomics *51*, 145–158.
711 Josephs, E.B., Lee, Y.W., Stinchcombe, J.R., and Wright, S.I. (2015). Association mapping reveals the
712 role of purifying selection in the maintenance of genomic variation in gene expression. Proc. Natl. Acad.
713 Sci. *112*, 15390 LP – 15395.
714 Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L.,
715 Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified
716 from variation in 141,456 humans. Nature *581*, 434–443.
717 Khodursky, S., Svetec, N., Durkin, S.M., and Zhao, L. (2020). The evolution of sex-biased gene
718 expression in the Drosophila brain. Genome Res. *30*, 874–884.
719 Klein, S.L., and Flanagan, K.L. (2016). Sex differences in immune responses. Nat. Rev. Immunol. *16*,
720 626–638.
721 Kolmykov, S., Yevshin, I., Kulyashov, M., Sharipov, R., Kondrakhin, Y., Makeev, V.J., Kulakovskiy, I.
722 V, Kel, A., and Kolpakov, F. (2021). GTRD: an integrated view of transcription regulation. Nucleic Acids
723 Res. *49*, D104–D111.
724 Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast
725 gene set enrichment analysis. BioRxiv 60012.
726 Lande, R. (1980). Sexual Dimorphism, Sexual Selection, and Adaptation in Polygenic Characters.
727 Evolution (N. Y). *34*, 292–305.
728 Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The
729 Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.
730 Lopes-Ramos, C.M., Chen, C.-Y., Kuijjer, M.L., Paulson, J.N., Sonawane, A.R., Fagny, M., Platig, J.,
731 Glass, K., Quackenbush, J., and DeMeo, D.L. (2020). Sex Differences in Gene Expression and
732 Regulatory Networks across 29 Human Tissues. Cell Rep. *31*, 107795.
733 Mauvais-Jarvis, F., Bairey Merz, N., Barnes, P.J., Brinton, R.D., Carrero, J.-J., DeMeo, D.L., De Vries,
734 G.J., Epperson, C.N., Govindan, R., Klein, S.L., et al. (2020). Sex and gender: modifiers of health,

735 disease, and medicine. Lancet *396*, 565–582.

736 McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-
737 Seq experiments with respect to biological variation. Nucleic Acids Res. *40*, 4288–4297.

738 McIver, B., and Morris, J.C. (1998). THE PATHOGENESIS OF GRAVES' DISEASE. Endocrinol.
739 Metab. Clin. North Am. *27*, 73–89.

740 Meisel, R.P., Malone, J.H., and Clark, A.G. (2012). Faster-X Evolution of Gene Expression in
741 Drosophila. PLOS Genet. *8*, e1003013.

742 Nakabayashi, K., Tajima, A., Yamamoto, K., Takahashi, A., Hata, K., Takashima, Y., Koyanagi, M.,
743 Nakaoka, H., Akamizu, T., Ishikawa, N., et al. (2011). Identification of independent risk loci for Graves'
744 disease within the MHC in the Japanese population. J. Hum. Genet. *56*, 772–778.

745 Naqvi, S., Godfrey, A.K., Hughes, J.F., Goodheart, M.L., Mitchell, R.N., and Page, D.C. (2019).
746 Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. Science (80-
747 . ). *365*, eaaw7317.

748 Nguyen, Q.H., Pervolarakis, N., Blake, K., Ma, D., Davis, R.T., James, N., Phung, A.T., Willey, E.,
749 Kumar, R., Jabart, E., et al. (2018). Profiling human breast epithelial cells using single cell RNA
750 sequencing identifies cell diversity. Nat. Commun. *9*, 2028.

751 Ober, C., Loisel, D.A., and Gilad, Y. (2008). Sex-specific genetic architecture of human disease. Nat.
752 Rev. Genet. *9*, 911–922.

753 Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A.D.H., Cotter, D.J., Parsana, P.,
754 Kasela, S., Balliu, B., Viñuela, A., et al. (2020a). The impact of sex on gene expression across human
755 tissues. Science (80-. ). *369*, eaba3066.

756 Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A.D.H., Cotter, D.J., Parsana, P.,
757 Kasela, S., Balliu, B., Viñuela, A., et al. (2020b). The impact of sex on gene expression across human
758 tissues. Science (80-. ). *369*, eaba3066.

759 Parisi, M., Nuttall, R., Naiman, D., Bouffard, G., Malley, J., Andrews, J., Eastman, S., and Oliver, B.
760 (2003). Paucity of Genes on the Drosophila X Chromosome Showing Male-Biased Expression. Science

761 (80-. ). *299*, 697–700.

762 Parsch, J., and Ellegren, H. (2013). The evolutionary causes and consequences of sex-biased gene

763 expression. Nat. Rev. Genet. *14*, 83–87.

764 Phillips, P.C., and Arnold, S.J. (1989). Visualizing Multivariate Selection. Evolution (N. Y). *43*, 1209–

765 1222.

766 Quaranta, R., Pelullo, M., Zema, S., Nardozza, F., Checquolo, S., Lauer, D.M., Bufalieri, F., Palermo, R.,

767 Felli, M.P., Vacca, A., et al. (2017). Maml1 acts cooperatively with Gli proteins to regulate sonic

768 hedgehog signaling pathway. Cell Death Dis. *8*, e2942–e2942.

769 Ranz, J.M., Castillo-Davis, C.I., Meiklejohn, C.D., and Hartl, D.L. (2003). Sex-dependent gene

770 expression and evolution of the Drosophila transcriptome. Science (80-. ). *300*, 1742–1745.

771 Rigby, R.A., and Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape.

772 J. R. Stat. Soc. Ser. C (Applied Stat. *54*, 507–554.

773 Robinson, M.D., and Smyth, G.K. (2007). Moderated statistical tests for assessing differences in tag

774 abundance. Bioinformatics *23*, 2881–2887.

775 Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for

776 differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

777 Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T.,

778 Yamamoto, K., Akiyama, M., et al. (2021). A cross-population atlas of genetic associations for 220

779 human phenotypes. Nat. Genet. *53*, 1415–1424.

780 Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth,

781 J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect,

782 worm, and yeast genomes. Genome Res. *15*, 1034–1050.

783 Simonovsky, E., Schuster, R., and Yeger-Lotem, E. (2019). Large-scale analysis of human gene

784 expression variability associates highly variable drug targets with lower drug effectiveness and safety.

785 Bioinformatics *35*, 3028–3037.

786 Soldin, O.P., and Mattison, D.R. (2009). Sex differences in pharmacokinetics and pharmacodynamics.

787 Clin. Pharmacokinet. *48*, 143–157.

788 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A.,
789 Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-
790 based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–
791 15550.

792 Sun, Y., Wang, Y., Fan, C., Gao, P., Wang, X., Wei, G., and Wei, J. (2014). Estrogen promotes stemness
793 and invasiveness of ER-positive breast cancer cells through Gli1 activation. Mol. Cancer *13*, 137.

794 TOTTERMAN, T.H., ANDERSSON, L.C., and HÄYRY, P. (1979). EVIDENCE FOR THYROID
795 ANTIGEN-REACTIVE T LYMPHOCYTES INFILTRATING THE THYROID GLAND IN
796 GRAVES'DISEASE. Clin. Endocrinol. (Oxf). *11*, 59–68.

797 Williams, T.M., and Carroll, S.B. (2009). Genetic and molecular insights into the development and
798 evolution of sexual dimorphism. Nat. Rev. Genet. *10*, 797–804.

799 Wu, L., Aster, J.C., Blacklow, S.C., Lake, R., Artavanis-Tsakonas, S., and Griffin, J.D. (2000). MAML1,
800 a human homologue of Drosophila mastermind, is a transcriptional co-activator for NOTCH receptors.
801 Nat. Genet. *26*, 484–489.

802 Zajitschek, S.R., Zajitschek, F., Bonduriansky, R., Brooks, R.C., Cornwell, W., Falster, D.S., Lagisz, M.,
803 Mason, J., Senior, A.M., Noble, D.W., et al. (2020). Sexual dimorphism in trait variability and its eco-
804 evolutionary and statistical implications. Elife *9*.

805 Zhao, S.-X., Xue, L.-Q., Liu, W., Gu, Z.-H., Pan, C.-M., Yang, S.-Y., Zhan, M., Wang, H.-N., Liang, J.,
806 Gao, G.-Q., et al. (2013). Robust evidence for five new Graves' disease risk loci from a staged genome-
807 wide association analysis. Hum. Mol. Genet. *22*, 3347–3362.

808

809