Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues

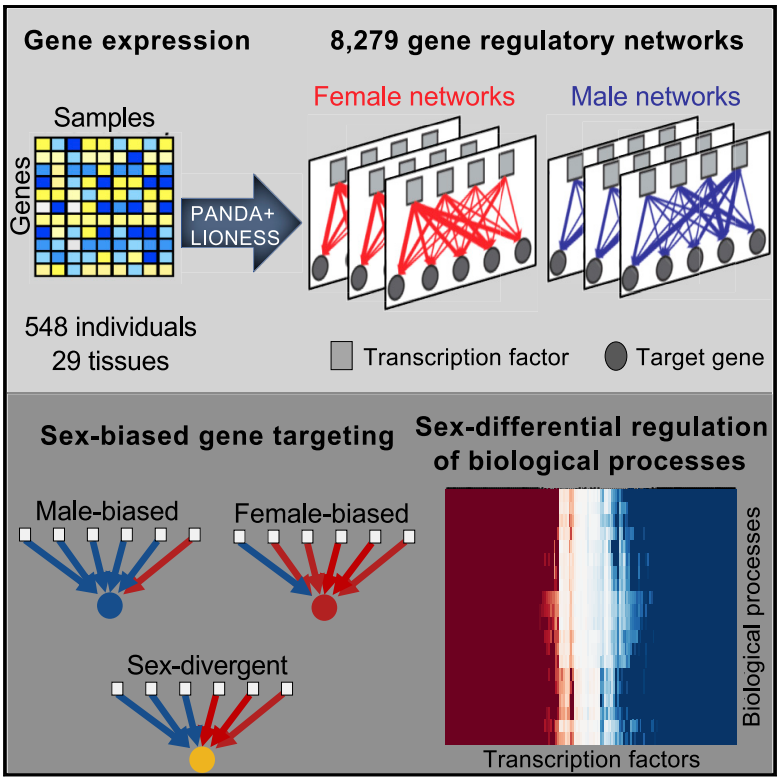
Camila Lopes-Ramos, Cho-Yi Chen, Marieke Kuijjer, Joseph Paulson, Abhijeet Sonawane, Maud Fagny, John Platig, Kimberly Glass, John Quackenbush, Dawn Demeo

Cell Reports

Article

Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues

Graphical Abstract



Highlights

- Sex differences are evident in sample-specific gene regulatory networks
- TF sex-biased targeting of genes is independent of their differential expression
- Sex-biased target genes are enriched for tissue-related functions and diseases
- Rich public resource that includes 8,279 gene regulatory networks of 29 tissues

Authors

Camila M. Lopes-Ramos, Cho-Yi Chen, Marieke L. Kuijjer, ..., Kimberly Glass, John Quackenbush, Dawn L. DeMeo

Correspondence

johnq@hsph.harvard.edu (J.Q.), dawn.demeo@channing.harvard.edu (D.L.D.)

In Brief

In analyzing 8,279 sample-specific gene regulatory networks, Lopes-Ramos et al. find significant sex differences in gene regulation across all 29 tissues analyzed. Transcription factors have sex-biased targeting patterns, and sex-biased target genes are associated with each tissue's function and with diseases specific to that tissue.

Lopes-Ramos et al., 2020, Cell Reports *31*, 107795
June 23, 2020
https://doi.org/10.1016/j.celrep.2020.107795

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues

Camila M. Lopes-Ramos,[1] Cho-Yi Chen,[2] Marieke L. Kuijjer,[3] Joseph N. Paulson,[4] Abhijeet R. Sonawane,[5] Maud Fagny,[6] John Platig,[5] Kimberly Glass,[1,5] John Quackenbush,[1,5,7,*] and Dawn L. DeMeo[5,8,9,*]

[1]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA
[2]Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan
[3]Centre for Molecular Medicine Norway (NCMM), Nordic EMBL Partnership, University of Oslo, Oslo, Norway
[4]Department of Biostatistics, Product Development, Genentech Inc., San Francisco, CA, USA
[5]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[6]Genetique Quantitative et Evolution-Le Moulon, Universite Paris-Saclay, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, Centre National de la Recherche Scientifique, AgroParisTech, Gif-sur-Yvette, France
[7]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA
[8]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA
[9]Lead Contact
*Correspondence: johnq@hsph.harvard.edu (J.Q.), dawn.demeo@channing.harvard.edu (D.L.D.)
https://doi.org/10.1016/j.celrep.2020.107795

SUMMARY

Sex differences manifest in many diseases and may drive sex-specific therapeutic responses. To understand the molecular basis of sex differences, we evaluated sex-biased gene regulation by constructing sample-specific gene regulatory networks in 29 human healthy tissues using 8,279 whole-genome expression profiles from the Genotype-Tissue Expression (GTEx) project. We find sex-biased regulatory network structures in each tissue. Even though most transcription factors (TFs) are not differentially expressed between males and females, many have sex-biased regulatory targeting patterns. In each tissue, genes that are differentially targeted by TFs between the sexes are enriched for tissue-related functions and diseases. In brain tissue, for example, genes associated with Parkinson's disease and Alzheimer's disease are targeted by different sets of TFs in each sex. Our systems-based analysis identifies a repertoire of TFs that play important roles in sex-specific architecture of gene regulatory networks, and it underlines sex-specific regulatory processes in both health and disease.

INTRODUCTION

Human development, physiology, and disease manifest differently in males and females. Sex differences are prevalent in many human diseases, including cancers, diabetes, neurological disorders, cardiovascular, and autoimmune diseases (Clocchiatti et al., 2016; Morrow, 2015; Ober et al., 2008). Differences between the sexes in incidence, prevalence, severity, and response to treatment can complicate our understanding of and hinder our ability to prevent, treat, and cure diseases. Biological factors and mechanisms that drive sex differences are understudied and poorly understood.

Sex divergence in health and disease is partially driven by the inherent inequality of the sex chromosomes, such as the effects of the expression of Y chromosome genes, differences in dosage of X chromosome genes, and epigenetic effects (Arnold, 2017). The influence of the sex hormones is also a contributor. Limited sex differences are revealed when focusing on individual gene expression levels. Previous studies have found sex-biased differences in gene expression in sex chromosome genes and in autosomal genes, yet the fold changes are generally small for autosomes (Gershoni and Pietrokovski, 2017; InanlooRahatloo et al., 2017; Mayne et al., 2016). However, small expression changes may be associated with large phenotypic effects, and other genome-wide analyses have demonstrated significant sex differences in chromatin accessibility (Kukurba et al., 2016; Sugathan and Waxman, 2013), indicating that a combination of sex-biased factors may be reflected in variation in the regulatory networks that control gene expression and gene regulation in each sex.

Gene regulatory network modeling synthesizes the complex interactions between transcription factors (TFs) and their target genes into a unified framework. A phenotype can be defined by a characteristic network, whereas changes in network structure between groups can provide insight into phenotypic drivers. Our previous work in analyzing gene regulatory networks in chronic obstructive pulmonary disease (Glass et al., 2014) and in colon cancer (Lopes-Ramos et al., 2018) found significant sex differences in regulatory features, including those involving genes not encoded on the sex chromosomes, and provided insight into the underlying biological processes active in males and females, but not found by standard differential expression analyses. Therefore, we used a systems-based approach to integrate multi-omics data with the goal of gaining insight into the molecular basis of sex differences across human tissues.

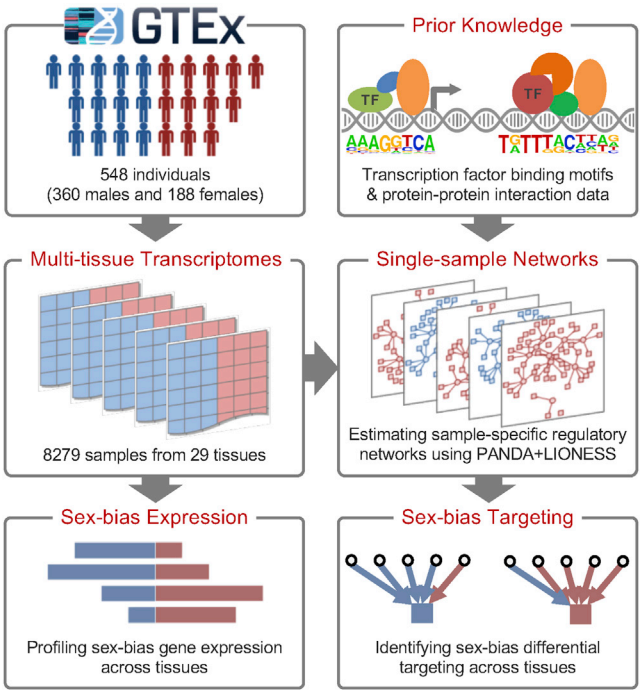


Figure 1. Study Overview
Schematic overview of our study. See also Table S1.

Figure 1 provides an overview of our study. To study sex bias in regulatory processes, we performed differential expression analysis followed by gene regulatory network analysis using a large dataset collected for the Genotype-Tissue Expression (GTEx) project (Battle et al., 2017). We used two network modeling methods, PANDA (Glass et al., 2013) and LIONESS (Kuijjer et al., 2019), to infer sample-specific gene regulatory networks in each of the 8,279 samples collected from 29 healthy human tissue types (28 solid tissues and whole blood). Generating sample-specific gene regulatory networks through LIONESS is a substantial advance, enabling the use of statistical methods to compare network properties between phenotypic groups and correcting for covariates.

We compared the networks between males and females in each tissue and found significant sex differences in gene regulatory networks across all tissues. Most importantly, many TFs have sex-biased and tissue-biased targeting patterns of genes associated with the tissue's function and diseases. Our results provide a repertoire of sex-biased regulatory processes and their regulatory drivers. Furthermore, they underline the importance of considering systems-level differences in gene regulation to understand sex differences contributing to health and disease.

RESULTS

Gene Expression Data Pre-processing

We preprocessed and normalized gene expression data from the GTEx project (version 6.0) in a tissue-aware manner using YARN, a software pipeline designed to perform quality control, gene filtering, and normalization, allowing gene expression levels to be comparable between tissues while preserving information regarding tissue-specific expression (Paulson et al., 2017). The data included expression information for 30,243 genes, including 1,074 encoded on the sex chromosomes. For the analyses shown here, we only considered tissues with samples collected from both men and women with more than 30 samples in either of the sexes. This included 8,279 samples representing 29 tissue types from 548 research subjects (360 males and 188 females). Table S1 shows the demographic information, including a comparison of covariates by sex (age, race, body mass index [BMI], RNA integrity number [RIN], and RNA isolation kit).

Sex Differences in Gene Expression

We used voom (Law et al., 2014) to identify genes that were differentially expressed between males and females for each tissue. We defined differential expression using a cutoff of false discovery rate (FDR) < 0.05 and absolute fold change ≥ 1.5. We also evaluated a transcriptomic signal-to-noise ratio (tSNR) as a measure of the overall distance between male and female transcriptomes. For the tSNR, the signal was defined as the Euclidean distance of average gene expression profiles between males and females, and the noise was defined as the overall variation among individuals. Although the voom differential expression analysis allowed us to identify specific genes with sex differences in their expression, the tSNR measures the overall divergence of transcriptomes between males and females and has the advantage of not depending on a particular threshold.

We included autosomal and sex chromosome genes in all analyses, unless noted otherwise. An overview of the differential gene expression by sex is shown by tissue in Figure 2A, and the complete list of differentially expressed genes is shown in Table S2. Breast, adipose (subcutaneous), thyroid, skeletal muscle, and skin had the largest number of differentially expressed genes by sex, whereas the gastrointestinal tract had the least, except for esophagus (mucosa). In breast, 4,181 genes were differentially expressed (4,009 were autosomal) by sex, followed by adipose (subcutaneous) with 482 genes (431 autosomal). In contrast, all other tissues had a much smaller number of differentially expressed genes. The median number of differentially expressed genes by sex across all tissues was 64 (28 autosomal), which represents 0.2% of analyzed genes (0.1% of autosomal genes).

We found no correlation between sample sizes and number of differentially expressed genes detected across tissues (Pearson's $r = -0.038$, $p = 0.85$). As expected, the number of differentially expressed genes depended on the threshold applied (Table S2). We found that the number of differentially expressed genes and the tSNR values shown in Figure 2A are highly correlated across tissues (Pearson's $r = 0.94$, $p = 4 \times 10^{-14}$). Most tissues had tSNRs significantly higher than expected by chance ($p < 0.05$, permutation test); exceptions included the intestine terminal ileum, stomach, whole blood, and colon (transverse) (Table S3). However, the tSNR mean differences were small, and when we repeated the analysis excluding sex chromosome genes, the tSNRs in most tissues were not statistically significant, suggesting minimal gene expression differences between the sexes.

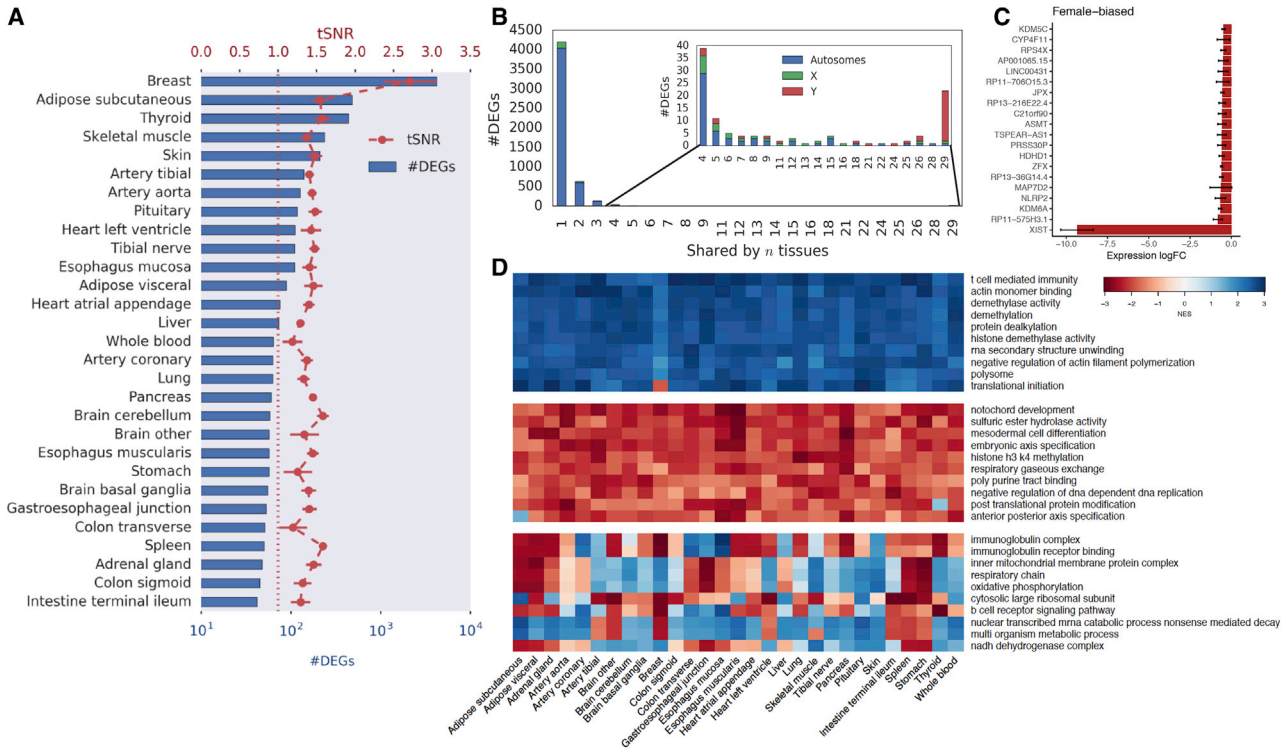


Figure 2. Sex Bias in Gene Expression

(A) Number of differentially expressed genes (DEGs) (absolute fold change $\geq$ 1.5 and FDR < 0.05), and male versus female tSNR across 29 tissues. The red line represents the tSNR value expected in the case of no male versus female differences. Error bars represent the standard deviation (SD) of tSNR values across 10,000 random samplings of 30 males and 30 females.

(B) Number of DEGs and number of tissues that share DEGs. Figure inset: same as the main figure, but the minimum number of shared tissues shown is 4.

(C) Top 20 female-biased DEGs based on the average log-fold-change expression values across all tissues. The bar plot shows the average ± SD.

(D) Enriched GO terms in males (blue) and females (red). The three heatmaps show the 10 GO terms with the highest average normalized enrichment score (NES) across all tissues (consistent male enrichment), lowest average NES across all tissues (consistent female enrichment), and highest SD across all tissues (most variable sex enrichment across tissues).

See also Tables S2 and S3.

Most differentially expressed autosomal genes were found in only one or two tissues (Figure 2B). In contrast, differentially expressed genes shared by many tissues were enriched for sex chromosome genes. Figure 2C shows the top 20 genes overexpressed in females compared with males based on the average log fold change across all tissues. As expected, most of these genes are encoded on the X chromosome. Similarly, the top genes overexpressed in males compared with females are on the Y chromosome. We performed a pre-ranked gene set enrichment analysis (GSEA) (Subramanian et al., 2005) based on voom's weighted t statistic to identify sex-biased enrichment of Gene Ontology (GO) terms in each tissue. We found male- and female-biased genes are annotated to different biological processes, for example, methylation and the immune system (Figure 2D).

To assess whether the differences in gene expression might be reflecting differences in the mixtures of cell types in the bulk sample used for RNA sequencing (RNA-seq), we compared the expression of 10 cell-type marker genes between males and females, as reported previously (Kim-Hellmuth et al., 2019): *FASN* (adipocytes); *CDH1* and *CLDN7* (epithelial cells); *AFP* (hepatocytes); *KRT10* (keratinocytes); *MYH7*, *TNNI1*, and *MYOG* (myocytes); *GAD1* (neurons); and *STX3* (neutrophils). The cell-type marker genes were not differentially expressed by sex in the tissues, except for breast. In breast, the epithelial cells gene markers (*CDH1* and *CLDN7*) demonstrated higher expression in tissue from males, whereas the adipocytes gene marker (*FASN*) demonstrated higher expression in tissue from females (FDR < 0.05 and fold change $\geq$ 1.5).

Sex Differences in Gene Regulatory Networks

One mechanism for sex differences in gene expression is through gene regulation by sex hormone receptors. Thus, we evaluated whether differentially expressed genes by sex were enriched for motifs of estrogen receptor 1 (ESR1), estrogen receptor 2 (ESR2), or androgen receptor (AR). We found that differentially expressed genes were not significantly enriched for these sex hormone receptor motifs, suggesting that these receptors are not the sole regulators of the sex differences we observed (Table S4). Therefore, we used gene regulatory network analysis to gain additional insight into sex-specific gene regulatory processes.

We used PANDA and LIONESS (Glass et al., 2013; Kuijjer et al., 2019) to infer gene regulatory network models for each sample in each tissue and compared those network models in samples from males and females (Figure 1). We first used PANDA to estimate population-based networks by integrating TF binding motifs (CIS-BP) (Weirauch et al., 2014), gene expression profiles (GTEx), and protein-protein interactions (StringDB) (Szklarczyk et al., 2015). The algorithm uses message passing to integrate a prior network (obtained by mapping TF motifs to the genome) with protein-protein interaction and gene expression data and to optimize the structure of the network given the data. PANDA produced a directed network of TFs to their target genes, comprising 644 TFs and 30,243 target genes.

LIONESS estimates networks for each sample in a population by serially leaving each sample out, calculating a network with and without that sample, and using linear interpolation to estimate the network for the left-out sample. We used LIONESS to estimate individual gene regulatory networks for each sample in the population. This approach produced 8,279 networks, one for each RNA-seq transcriptome from the 548 GTEx research subjects.

Each of the 8,279 estimated network models is structured as a bipartite graph, with regulatory edges connecting TF and gene nodes and edge weights representing the likelihood of the regulatory interaction between a TF and its target genes. We can use these models to test whether regulatory edges differ significantly between males and females, to identify the differentially regulated genes, and to determine the TFs driving these regulatory differences.

We performed a broad comparison of the male and female networks. We used linear regression models (limma; Ritchie et al., 2015) to compare the weight of each edge between males and females in each tissue and identified the significantly different edge weights (FDR < 0.05). A higher edge weight indicates a higher likelihood of a regulatory interaction between the TF and its target gene. We use the term higher targeting interchangeably with higher edge weight. For all analyzed tissues, we were able to identify gene regulatory network edges whose weights differ significantly between males and females (Table S5). Breast, thyroid, and brain (basal ganglia) had the largest number of sex-biased edges; approximately 7 million, 2 million, and 1 million edges (corresponding to 36%, 10%, and 5% of total edges) were significantly different between the sexes (FDR < 0.05), respectively. Across all tissues, the median number of significant sex-biased edges was 68,293 (0.35% of total edges in the network). There was no significant correlation between sample sizes and number of significant sex-biased edges found across tissues (Pearson's $r = 0.05$, $p = 0.83$). Most significant sex-biased edges were found in only one tissue, whereas the sex-biased edges shared across most tissues targeted sex chromosome genes (Figure S1). The complete list of sex-biased edges and the tissues in which they were found to be statistically significant is shown in Data S1.

Next, we evaluated the distribution of sex-biased edges around genes (Figure 3A). Each gene is connected to 644 TFs, because the generated networks are complete graphs. Genes with more than 5% of their edges with significantly different weights between males and females (FDR < 0.05) were defined as differentially targeted genes (Table S6). We recognize three classes of differentially targeted genes: (1) male-biased genes (the proportion of sex-biased edges in the male direction is greater than 0.6), (2) female-biased genes (the proportion of sex-biased edges in the female direction is greater than 0.6), and (3) sex-divergent genes (the proportion of sex-biased edges in the male- and female-biased directions is between 0.4 and 0.6). Quantifying the direction of sex-biased edges allowed us to identify genes that have a similar number of male- and female-biased edges, which represent genes that are differentially targeted in both males and females but by a different set of TFs (named sex-divergent genes). As expected, breast had the largest number of differentially targeted genes (25,994). However, all other tissues showed a large number of differentially targeted genes, varying from 87 to 14,361 genes, with a median of 169 genes (0.6% of analyzed genes). Overall, breast, thyroid, and brain (basal ganglia) had the largest number of differentially targeted genes, and the gastrointestinal tract had the least, except for esophagus (Table S6). There were no significant correlation between sample sizes and number of differentially targeted genes found across tissues (Pearson's $r = 0.12$, $p = 0.56$).

Figure 3B shows that the number and patterns of differentially targeted genes are tissue specific. Although some tissues have a balanced number of differentially targeted genes across the three classes, some tissues are enriched for specific classes. Brain (other) and whole blood are the tissues with the largest proportion of differentially targeted genes in the sex-divergent class (Table S6). The sex-divergent class, represented by genes targeted in both males and females but by a different set of TFs, is particularly interesting for understanding sex differences in health and disease and for therapeutics development.

On average, across the 29 tissues, 87% of the differentially expressed genes were also identified as differentially targeted ($p < 0.05$, Fisher's exact test). However, we found that on average, 70% of the differentially targeted genes were not differentially expressed (Table S6). These results suggest that limited sex differences can be found when considering only the mRNA expression levels, whereas more pronounced sex differences can be found when analyzing transcriptional regulation.

Sex-Biased Targeting of X Chromosome Genes

Across all tissues, the median number of differentially targeted genes on the X chromosome was 36 (minimum of 21 genes and maximum of 913 genes) of the 1,018 analyzed genes. X chromosome inactivation (XCI) silences one of the two X chromosomes in females, compensating for transcriptional dosages between the sexes. However, XCI is incomplete and some genes are expressed from both alleles. We evaluated whether differentially targeted genes are enriched in specific XCI categories, previously defined as escape (n = 82), variable escape (n = 89), inactive (n = 388), or unknown (n = 459) (Tukiainen et al., 2017). Differentially targeted genes are enriched in escape genes compared with variable escape, inactive, and unknown genes and with autosomes. Across all tissues, a median of 33% of the X chromosome escape genes are differentially targeted compared with 0.3% of autosomal genes (two-sided paired Wilcoxon rank-sum test, $p = 3.73 \times 10^{-9}$, Figure S2A). The direction of sex bias depends on the region in which the escape genes are located. A median of 17% of the escape genes that are not in the pseudoautosomal regions (PARs) are

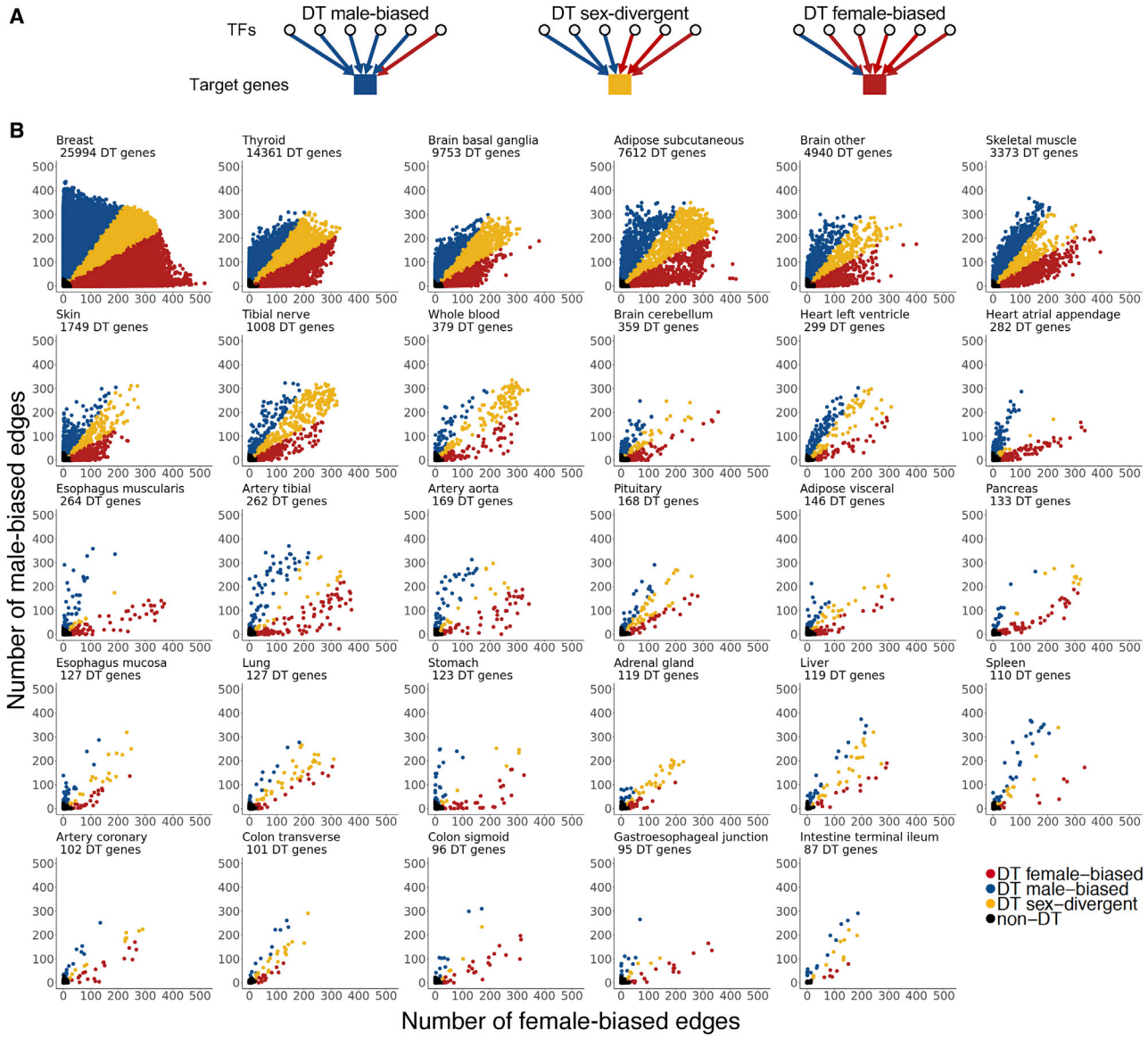


Figure 3. Sex-Biased Targeting in Gene Regulatory Networks

(A) Schematic representation of the three classes of differentially targeted (DT) genes: male biased, sex divergent, and female biased.

(B) Scatterplots of all genes (n = 30,243), indicating the number of female-biased edges (x axis) and the number of male-biased edges (y axis) at FDR < 0.05. Genes with more than 5% of their edges significantly different between males and females (FDR < 0.05) were defined as DT; the number of DT genes in each tissue is noted underneath the tissue name. Blue points represent DT male-biased genes (the proportion of sex-biased edges in the male direction is greater than 0.6), yellow points represent DT sex-divergent genes (the proportion of sex-biased edges in the male- and female-biased directions is between 0.4 and 0.6), red points represent DT female-biased genes (the proportion of sex-biased edges in the female direction is greater than 0.6), and black points represent non-DT genes (less than 5% of the edges are sex biased).

See also Tables S5 and S6 and Data S1.

female biased, and 7% are male biased (p = 0.006, Figure S2B). In contrast, a median of 13% of the escape genes in PARs are male biased, and 7% are female biased (p = 0.048, Figure S2C). Consistent with previous findings of expression differences (Tukiainen et al., 2017), our results show that male-female targeting differences reflect incomplete XCI, with enrichment of differentially targeted genes in escape genes compared with other X chromosome genes and autosomes (Figure S2D).

An Example of Differential Targeting by the TF MAZ

We found sex-biased regulatory network structures across many tissues, as illustrated by a high number of genes that are differentially targeted between males and females. To characterize how TFs might drive these regulatory sex differences, we evaluated how TF targeting patterns correlate with the expression of their target genes and associate with biological pathways.

We began by selecting a single TF to investigate. To choose that example TF, we performed a TF motif enrichment analysis (Fisher's exact test) on the genes differentially targeted between males and females in all tissues. Motifs for sex hormone receptors were not the most significantly enriched. Instead, the motif for MAZ, MYC-associated zinc finger protein, was significantly enriched across many tissues (Table S7). MAZ, which encodes a zinc finger TF, is ubiquitously expressed and has 13,270 target genes (genes with the TF binding motif in the promoter region). Thus, MAZ is involved in many cellular processes, including urogenital (Haller et al., 2018), neuronal (Goldie et al., 2017), and tumor development (Cogoi et al., 2013; Luo et al., 2016; Yu et al., 2017).

We investigated the regulatory role of MAZ in brain (basal ganglia), because this tissue had the third-largest number of sex-biased network edges, even though it had relatively few differentially expressed genes. To evaluate the regulatory potential of MAZ on its target genes, we started by choosing the autosomal target gene with the highest differential expression in female (*PRSS30P*) and the one in male (*FRG1B*). *PRSS30P* has higher average expression and higher average targeting (i.e., higher average edge weight) by MAZ in females (Figure 4A). In contrast, *FRG1B* has higher average expression and higher average targeting by MAZ in males (Figure 4B). We investigated the correlation between targeting and expression of *PRSS30P* and *FRG1B* across all samples. For both genes, we observed that the weight of the edge from *MAZ* was linearly correlated with the gene's expression (R^2 = 0.77 for *PRSS30P* and R^2 = 0.52 for *FRG1B*, Figure 4C). *MAZ* was not differentially expressed between males and females, and its expression was not correlated with the expression of either *PRSS30P* or *FRG1B* (R^2 = 0, Figure 4D).

We extended this analysis to include all 13,270 target genes of MAZ. We found that differential targeting by MAZ is positively correlated with the differential expression of its target genes (R^2 = 0.36, Figure 4E). This relationship is also true when excluding sex chromosome genes (R^2 = 0.36, Figure 4F). Genes with no sex-biased differential expression can be differentially targeted by MAZ. For example, *CASP2*, involved with synaptic dysfunction and neuronal death in Huntington's disease and Alzheimer's disease (Hermel et al., 2004; Pozueta et al., 2013; Troy et al., 2000), is not differentially expressed but does have significantly higher targeting by MAZ in females (FDR = 0.025, Figure 4F). Similarly, *CUEDC2* is not differentially expressed but has significantly higher targeting in males (FDR = 0.038). *CUEDC2* has been shown to modulate ESR1 protein levels regardless of the presence of 17β-estradiol (Pan et al., 2011).

To determine whether these observations were specific to brain (basal ganglia), we repeated this analysis in each of the other tissues. We found that the correlation between MAZ's targeting patterns and its targets' expression levels varies across tissues (Figure 5A). In pituitary and colon (transverse), for example, differential targeting by MAZ largely explains the differential expression of its target genes (R^2 = 0.88 and 0.53, respectively) (Figures 5B and 5C). However, in adipose (subcutaneous), there was no relationship between targeting and expression (R^2 = 0) (Figure 5D). Although highly expressed across the tissues, MAZ is not differentially expressed by sex in the tissues. Thus, these network analyses reveal a sex-biased regulatory potential of TFs that is independent of both their mRNA expression level differences and the differential expression of their target genes.

TF Differential Targeting Patterns across 29 Tissues

We repeated our differential targeting analysis for all TFs and across all tissues, providing a holistic view of the landscape of TFs potentially driving sex differences. To visualize the results, we ranked all 644 TFs in our network models based on the correlation between the TF differential targeting and the differential expression of its target genes (R^2) (Figure 6; Table S8). We found many TFs with targeting patterns that are well correlated with the differential expression of their target genes and many without correlation between differential targeting and expression. These patterns are ubiquitous across the tissues, but the specific TFs and the number of TFs involved vary by tissue.

Sex hormone receptors play an important role in sex differential regulation of genes (van Nas et al., 2009). We found that although sex hormone receptors were often not the highest-ranked differentially targeting TFs, ESR1, ESR2, and AR exhibit sex-biased regulatory patterns. Similar to MAZ, these sex hormone receptors were not differentially expressed between males and females, but they exhibit strong differential targeting patterns in several tissues, including breast (Figures S3–S5), heart (left ventricle), and whole blood (Figure S6). Overall, these results suggest that various TFs, including many that are not sex hormone receptors, drive the sex-biased regulatory programs active in different tissues.

Differential Regulation of Biological Processes

Finally, we sought to characterize how sex bias in targeting patterns may influence the regulation of biological processes. By doing this, we hoped to gain insights into how gene regulation, manifested as differential targeting of genes between the sexes, might drive sex differences and influence human health and disease.

To systematically characterize the biological functions associated with differential targeting by TFs in each tissue, we ran preranked GSEA to identify the GO terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched for genes differentially targeted between males and females. We performed one enrichment analysis for each TF with expression-targeting correlation greater than 0.3 in a given tissue; genes were ranked based on their level of differential targeting (t statistics from limma) by the corresponding TF in the tissue.

Again, we used brain (basal ganglia) as an example. Figure 7A shows the 20 GO terms and 20 KEGG pathways with the highest variability in their GSEA normalized enrichment score across differentially targeting TFs. We found sex-biased enrichment patterns for several brain-related biological processes, including the GO terms regulation of synapse structure or activity and neurotransmitter transport. In addition, several brain-related diseases showed sex-biased enrichment, such as Alzheimer's disease and Parkinson's disease. These diseases are known to have sex differences in incidence and severity that are higher in women for Alzheimer's disease and higher in men for Parkinson's disease (Mielke et al., 2014; Pringsheim et al., 2014).

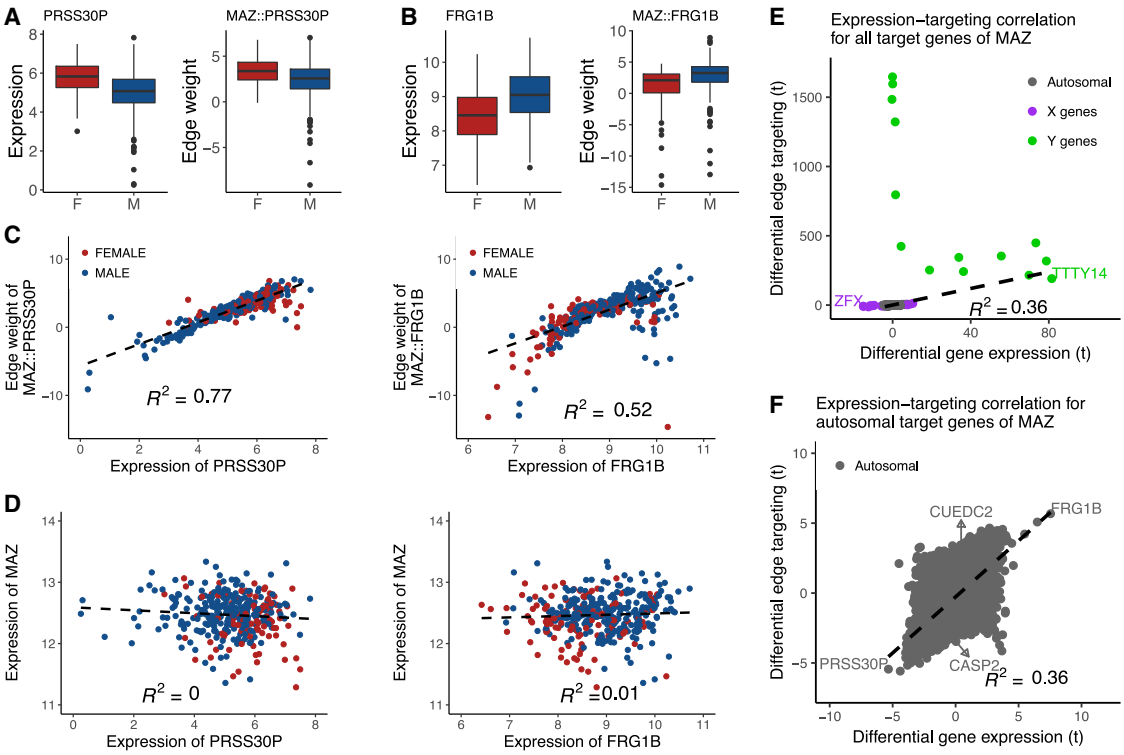


Figure 4. Differential Targeting by the TF MAZ in Brain (Basal Ganglia)

(A) Expression levels of *PRSS30P* (left) and MAZ to *PRSS30P* edge weights (right) across male and female brain (basal ganglia) samples.

(B) Expression levels of *FRG1B* (left) and MAZ to *FRG1B* edge weights (right) across male and female brain (basal ganglia) samples.

(C) Plots showing the expression of the target gene versus the MAZ to target gene edge weight.

(D) Plots showing the expression of the target gene versus *MAZ* gene expression. In (C) and (D), each data point represents a single sample in brain (basal ganglia).

(E) Scatterplot of MAZ target genes showing the genes' differential expression levels (t statistics) by the genes' differential edge weight targeting levels by MAZ (t statistics). Mismapping of sequencing reads to the sex chromosomes may result in spurious expression of Y chromosome genes in females.

(F) Same as (E), but without sex chromosome genes.

M, male; F, female. See also Figures S3–S5.

In Figure 7B, we present the 20 most significant TFs (based on the GSEA FDR value) that differentially target genes annotated to Alzheimer's disease and Parkinson's disease. Although the genes annotated to these diseases only partially overlap (Jaccard index of 0.46), most TFs found to be significantly different in their targeting patterns (FDR < 0.05) were the same in both diseases (100 of the 114 unique significant TFs, Jaccard index of 0.88). This could indicate that TFs play important roles in sex-biased regulation in the normal and diseased brain. For example, consistent with the β-amyloid plaques being the major pathological lesions in Alzheimer's disease brain, MAZ (male biased) has been shown to colocalize with these plaque structures (Jordan-Sciutto et al., 2000), and STAT2 (female biased) is known to participate in the SGK1-STAT1/STAT2 pathway to protect against β-amyloid-induced apoptosis (Hsu et al., 2009).

When we repeated this analysis for other tissues, we found similar patterns of sex-biased regulation of biological processes (Figure S7). Many GO terms and KEGG pathways were enriched for genes targeted by different sets of TFs in each sex. These included cellular defense response, regulation of leukocyte-mediated immunity, and primary immunodeficiency in whole blood. In heart (atrial appendage), not only were biological processes important for the tissue function enriched (such as heart process and cardiac cell development), but we also found sex-biased targeting of dilated cardiomyopathy. We further found that differentially targeted genes were enriched for autoimmune diseases known to have pronounced sex differences: maturity-onset diabetes of the young in pancreas and autoimmune thyroid disease in thyroid. Overall, despite limited differential expression, TFs have sex-biased targeting patterns that are consistent with known differences in disease and tissue function.

DISCUSSION

Sex differences have been recognized among the most significantly understudied aspects of human disease. Historically, sex has not been properly taken into account, many experimental studies have been done only on males, the sex chromosomes have been excluded from analyses, and sequence mapping protocols have not account for sex chromosome biases (Khramtsova et al., 2019). The 2015 guidelines from the National Institutes of Health (NOT-OD-15-102) stated that sex should be

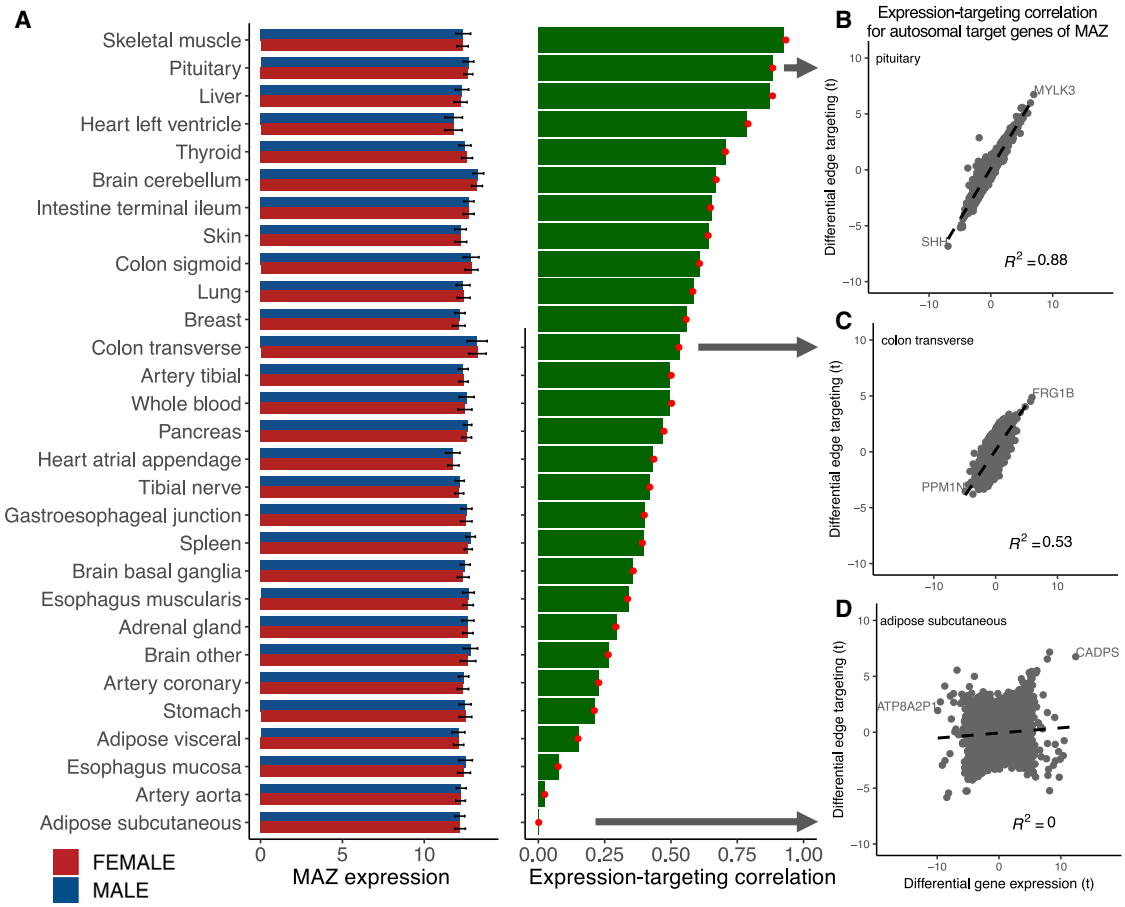


Figure 5. Expression Levels and Expression-Targeting Correlation of MAZ across 29 Tissues

(A) Expression levels (average ± SD, left panel) and expression-targeting correlation values (R^2, right panel) across 29 tissues for MAZ. Green bars represent the correlation considering all target genes, and red dots represent the correlation considering only autosomal target genes.

(B–D) Three example tissues that have different levels of MAZ expression-targeting correlation: (B) pituitary, (C) colon (transverse), and (D) adipose (subcutaneous). MAZ, a highly expressed TF with no differential expression by sex, exhibits tissue-dependent expression-targeting correlation.

See also Figure S6.

considered in research designs, analyses, and reporting in vertebrate animal and human studies. However, there are many challenges in studying sex as a biological variable. This includes ensuring adequate sample sizes for well-powered analyses of each sex and issues with sequencing data mapping, including poor mapping quality and bias from both highly similar regions between the sex chromosomes and the lack on Y chromosome in females.

Several studies have found sex differences in the expression levels of autosomal genes, although the expression differences are generally small. The differentially expressed genes between males and females are enriched for biological processes with known sex differences, such as reproduction, fat biogenesis, muscle contraction, and cardiomyopathy (Gershoni and Pietrokovski, 2017; InanlooRahatloo et al., 2017). In an analysis of a small GTEx pilot dataset (1,641 tissue samples from 175 individuals), Melé et al. (2015) identified 135 genes with sex-biased expression globally across the tissues (mostly on sex chromosomes). Using Weighted Correlation Network Analysis (WGCNA) (Langfelder and Horvath, 2008), they identified groups of genes with correlated expression enriched for GO terms that include spermatid, ectoderm, and epidermis development. In interpreting these results, WGCNA is a correlation-based method that captures general patterns of coexpression but does not distinguish between TFs and their targets and thus does not identify sex-biased regulatory processes. In contrast, our analysis is based on PANDA (Glass et al., 2013) and LIONESS (Kuijjer et al., 2019), which explicitly model gene regulation by TFs and allow relevant regulatory differences to be identified.

We analyzed sex differences in gene expression and gene regulatory networks in 8,279 samples from 29 healthy tissues in the GTEx version 6.0 cohort. Consistent with the other analyses cited earlier, we found that the number of genes that are differentially expressed between males and females is small (median of 64 genes) in most tissues, with the exception of breast. We found that in general, differentially expressed genes between the sexes are enriched for methylation processes (based on pathway enrichment analysis). These results suggest that gene regulatory

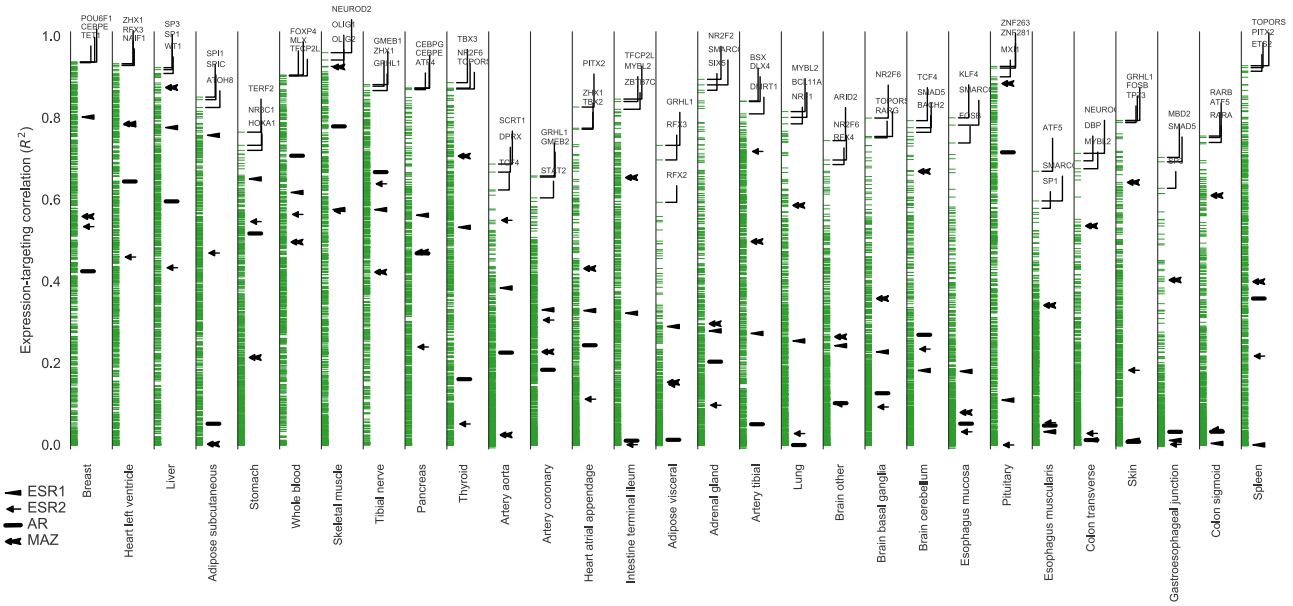


Figure 6. TF Differential Targeting Patterns across 29 Tissues

Expression-targeting correlation value (including both autosomal and sex chromosome genes) for each TF (green ticks) in each tissue. Tissues are ordered based on the ESR1 expression-targeting correlation value. For each tissue, the names of the top 3 TFs with the highest expression-targeting correlation values are annotated, and the locations of ESR1, ESR2, AR, and MAZ are marked by arrows. See also Table S8.

processes, such as epigenetic variability, may be drivers of sex differences in human health and, by extension, human diseases.

Using gene regulatory network analysis, we observed that male and female networks have significant differences in their regulatory structure, meaning that genes in males and females are often regulated by different TFs, or combinations of TFs, whether those target genes are differentially expressed or not. Finding sex-differential targeting of genes in the absence of strong differential expression (considering a 1.5-fold-change cutoff) can reveal latent sex differences in gene regulation, which might become important under diverse conditions of age, stress, disease development, and therapeutic response. For example, our recent study evaluating the regulatory networks of colon tumor samples before chemotherapy revealed that male and female patients are primed to respond differently to therapy even though they had not yet been challenged (Lopes-Ramos et al., 2018). We found strong sex-biased regulation of genes associated with drug metabolism, despite no differential expression, and determined that higher targeting of the drug metabolism pathway was associated with a higher survival outcome for women treated with chemotherapy. Genes that are not differentially expressed can be targeted in both sexes but through different sets of TFs. Thus, identifying sex-biased molecular regulation is crucial, and modulating sex-divergent genes may require sex-aware drug development and treatment protocols.

Finding regulatory drivers in the absence of strong differential expression is another important advantage of network analysis. Previous studies comparing phenotypic groups have found that commonly expressed regulators can target different genes and play specific regulatory roles despite little or no difference in expression level (Glass et al., 2014, 2015; Lopes-Ramos et al., 2017; Neph et al., 2012; Sonawane et al., 2017). Several mechanisms can mediate differential targeting by TFs, such as TF protein abundance and conformation, establishment of TF interactions and cooperativity, and epigenetic modifications that interfere with TF binding (Lambert et al., 2018; Yin et al., 2017).

Gonadal hormones play a strong role in sex differences in gene expression (van Nas et al., 2009). Our network models captured the differential targeting pattern of the sex hormone receptors (ESR1, ESR2, and AR) even though the TFs were not differentially expressed. Differential edge weights between sex hormone receptors and their target genes across individuals could be attributed to the differential levels of the ligands, such as estrogens and progesterone, influencing the receptor activity. We found that the sex hormone receptors were not the sole TFs mediating the sex-biased regulatory processes. Instead, various TFs mediate the sex-biased regulatory program in each tissue, with the number and the specific TFs varying among tissues. In each tissue, differences in gene regulation between the sexes were often associated with the tissue's function and the diseases that affect the tissue. For example, we found that many TFs with sex-biased differential targeting regulate genes associated with diseases with recognized sex-biased manifestations, such as Alzheimer's disease, Parkinson's disease, diabetes, autoimmune thyroid disease, and cardiomyopathy. These findings reveal latent sex differences in gene regulation, which might be important drivers of sex-biased manifestations during disease development and progression.

In our study, we note several limitations, including the mismapping of sequencing reads to the sex chromosomes in the GTEx data and our inclusion of the sex chromosome genes to estimate our networks. For biological coherence and to account

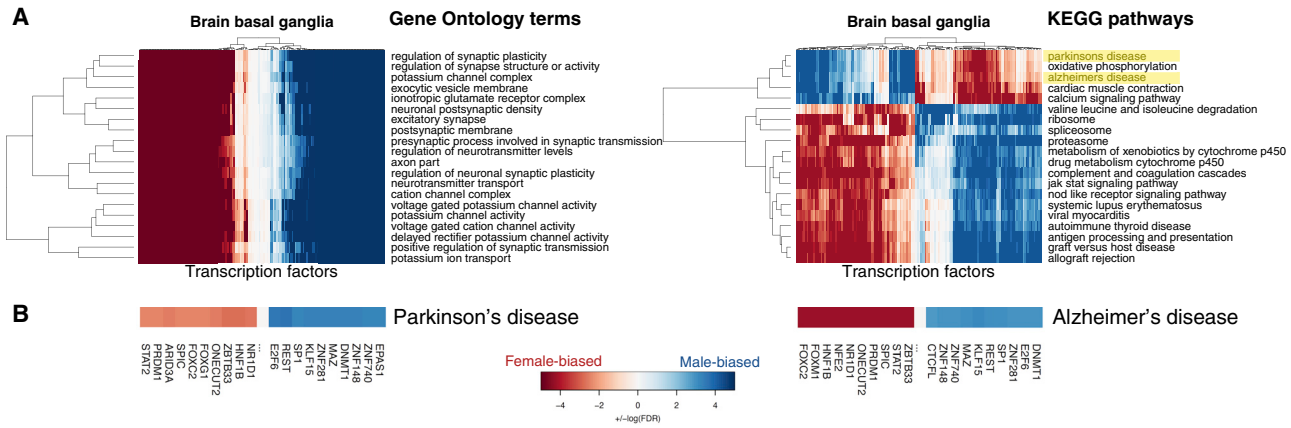


Figure 7. Sex-Biased Targeting of Biological Processes in Brain (Basal Ganglia)

(A) Top 20 GO terms and 20 KEGG pathways enriched for genes differentially targeted by TFs with high expression-targeting correlation values ($R^2 > 0.3$). GO terms and KEGG pathways selected are those with the highest NES standard deviation across the TFs. Enrichment for males is shown in blue and for females is shown in red.

(B) Top 10 male-biased (blue) and female-biased (red) TFs differentially targeting genes annotated to Parkinson's disease, and Alzheimer's disease. TFs were selected based on their FDR significance.

See also Figure S7.

for spurious correlation in females, we set the edge weights to Y chromosome genes in females to the minimum value found, which can cause significant inflation of male-biased Y chromosome edges. As network methods evolve, we need to develop better approaches that can balance these types of technical and biological artifacts to improve and extend comparisons of male and female networks and to understand how these regulatory differences contribute to sex differences in health and disease. Our analyses were done on the GTEx dataset, and need to be repeated in additional datasets to confirm that our findings are generalizable.

Our analysis identified statistically significant sex differences in regulatory processes and the regulatory relationships between TFs and their target genes. We found that gene regulatory processes vary by sex across tissues and that many TFs have sex-biased targeting patterns. Genes that are differentially targeted between males and females are enriched for disease and tissue-related functions. These findings are a significant step forward in understanding how sex differences manifest in gene regulatory networks and underscore the importance of looking beyond gene expression.

Most importantly, we found regulatory differences between the sexes in every tissue, indicating that sex-biased regulation of cellular processes is systemic, rather than isolated to a small number of tissues in which one might argue that differences in biological function are driving differences in regulation. Males and females often do not manifest disease in the same way, or respond identically to treatment, and the exploration of the causes of these differences should remain an area of active study. More fully exploring sex-biased patterns of gene regulation is crucial not only for understanding how sex-specific biological processes drive health and disease but also for the development of precision therapeutics that will best treat disease in an individual, accounting for sex.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- METHOD DETAILS
  - GTEx dataset
  - Network inference using PANDA and LIONESS
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Differential expression analysis
  - Transcriptome-based signal-to-noise ratio
  - Enrichment analysis for sex hormone receptor motifs
  - Analysis of differential network targeting
  - Transcription factors' differential targeting

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.107795.

ACKNOWLEDGMENTS

This work was supported by grants from the U.S. National Institutes of Health, including grants from the National Heart, Lung, and Blood Institute (5P01HL105339, 5R01HL111759, 5P01HL114501, K25HL133599, and K25HL140186), the National Cancer Institute (1R35CA220523, 5P50CA127003, 1R35CA197449, 1U01CA190234, and 5P30CA006516), and the National Institute of Allergy and Infectious Diseases (5R01AI099204). Additional funding was provided through a grant from the NVIDIA Foundation and grants from the Norwegian Research Council, Helse Sør-Øst, and University of Oslo through the Centre for Molecular Medicine Norway (NCMM). This work

was conducted under dbGaP-approved protocol 9112 (dbGaP: phs000424.v6.p1).

AUTHOR CONTRIBUTIONS

Conceptualization, C.M.L.-R., C.-Y.C., M.L.K., J.N.P., A.R.S., M.F., J.P., K.G., J.Q., and D.L.D.; Methodology, C.M.L.-R., C.-Y.C., M.L.K., K.G., J.Q., and D.L.D.; Formal Analysis, C.M.L.-R. and C.-Y.C.; Investigation, C.M.L.-R. and C.-Y.C.; Resources, J.Q. and D.L.D.; Data Curation, J.N.P. and J.Q.; Writing – Original Draft, C.M.L.-R. and C.-Y.C.; Writing – Review & Editing, C.M.L.-R., C.-Y.C., M.L.K., J.N.P., A.R.S., M.F., J.P., K.G., J.Q., and D.L.D.; Visualization, C.M.L.-R. and C.-Y.C.; Supervision, J.Q. and D.L.D.; Funding Acquisition, M.L.K., K.G., J.Q., and D.L.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 14, 2016
Revised: April 1, 2020
Accepted: May 29, 2020
Published: June 23, 2020

REFERENCES

Arnold, A.P. (2017). A general theory of sexual differentiation. J. Neurosci. Res. *95*, 291–300.

Battle, A., Brown, C.D., Engelhardt, B.E., and Montgomery, S.B.GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)-Analysis Working Group; Statistical Methods groups-Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site-NDRI; Biospecimen Collection Source Site-RPCI; Biospecimen Core Resource-VARI; Brain Bank Repository-University of Miami Brain Endowment Bank; Leidos Biomedical-Project Management; ELSI Study; Genome Browser Data Integration &Visualization-EBI; Genome Browser Data Integration &Visualization-UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

Clocchiatti, A., Cora, E., Zhang, Y., and Dotto, G.P. (2016). Sexual dimorphism in cancer. Nat. Rev. Cancer *16*, 330–339.

Cogoi, S., Zorzet, S., Rapozzi, V., Géci, I., Pedersen, E.B., and Xodo, L.E. (2013). MAZ-binding G4-decoy with locked nucleic acid and twisted intercalating nucleic acid modifications suppresses KRAS in pancreatic cancer cells and delays tumor growth in mice. Nucleic Acids Res. *41*, 4049–4064.

Dewey, M. (2019). metap: meta-analysis of significance values. R Packag. *version 1*.

Gershoni, M., and Pietrokovski, S. (2017). The landscape of sex-differential transcriptome and its consequent selection in human adults. BMC Biol. *15*, 7.

Glass, K., Huttenhower, C., Quackenbush, J., and Yuan, G.-C. (2013). Passing messages between biological networks to refine predicted interactions. PLoS ONE *8*, e64832.

Glass, K., Quackenbush, J., Silverman, E.K., Celli, B., Rennard, S.I., Yuan, G.-C., and DeMeo, D.L. (2014). Sexually-dimorphic targeting of functionally-related genes in COPD. BMC Syst. Biol. *8*, 118.

Glass, K., Quackenbush, J., Spentzos, D., Haibe-Kains, B., and Yuan, G.-C. (2015). A network model for angiogenesis in ovarian cancer. BMC Bioinformatics *16*, 115.

Goldie, B.J., Fitzsimmons, C., Weidenhofer, J., Atkins, J.R., Wang, D.O., and Cairns, M.J. (2017). miRNA Enriched in Human Neuroblast Nuclei Bind the MAZ Transcription Factor and Their Precursors Contain the MAZ Consensus Motif. Front. Mol. Neurosci. *10*, 259.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Haller, M., Au, J., O'Neill, M., and Lamb, D.J. (2018). 16p11.2 transcription factor *MAZ* is a dosage-sensitive regulator of genitourinary development. Proc. Natl. Acad. Sci. USA *115*, E1849–E1858.

Hermel, E., Gafni, J., Propp, S.S., Leavitt, B.R., Wellington, C.L., Young, J.E., Hackam, A.S., Logvinova, A.V., Peel, A.L., Chen, S.F., et al. (2004). Specific caspase interactions and amplification are involved in selective neuronal vulnerability in Huntington's disease. Cell Death Differ. *11*, 424–438.

Hicks, S.C., Okrah, K., Paulson, J.N., Quackenbush, J., Irizarry, R.A., and Bravo, H.C. (2018). Smooth quantile normalization. Biostatistics *19*, 185–198.

Hsu, W.L., Chiu, T.H., Tai, D.J.C., Ma, Y.L., and Lee, E.H.Y. (2009). A novel defense mechanism that is activated on amyloid-β insult to mediate cell survival: role of SGK1-STAT1/STAT2 signaling. Cell Death Differ. *16*, 1515–1529.

InanlooRahatloo, K., Liang, G., Vo, D., Ebert, A., Nguyen, I., and Nguyen, P.K. (2017). Sex-based differences in myocardial gene expression in recently deceased organ donors with no prior cardiovascular disease. PLoS ONE *12*, e0183874.

Jordan-Sciutto, K.L., Dragich, J.M., Caltagarone, J., Hall, D.J., and Bowser, R. (2000). Fetal Alz-50 clone 1 (FAC1) protein interacts with the Myc-associated zinc finger protein (ZF87/MAZ) and alters its transcriptional activity. Biochemistry *39*, 3206–3215.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. *44* (*D1*), D457–D462.

Khramtsova, E.A., Davis, L.K., and Stranger, B.E. (2019). The role of sex in the genomics of human complex traits. Nat. Rev. Genet. *20*, 173–190.

Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Wucher, V., Kasela, S., Castel, S.E., Hamel, A.R., Viñuela, A., Roberts, A.L., et al. (2019). Cell type specific genetic regulation of gene expression across human tissues. bioRxiv. https://doi.org/10.1101/806117.

Kuijjer, M.L., Tung, M.G., Yuan, G., Quackenbush, J., and Glass, K. (2019). Estimating Sample-Specific Regulatory Networks. iScience *14*, 226–240.

Kukurba, K.R., Parsana, P., Balliu, B., Smith, K.S., Zappala, Z., Knowles, D.A., Favé, M.-J., Davis, J.R., Li, X., Zhu, X., et al. (2016). Impact of the X Chromosome and sex on regulatory variation. Genome Res. *26*, 768–777.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Lopes-Ramos, C.M., Paulson, J.N.J.N., Chen, C.-Y.C.-Y., Kuijjer, M.L.M.L., Fagny, M., Platig, J., Sonawane, A.R.A.R., DeMeo, D.L.D.L., Quackenbush, J., and Glass, K. (2017). Regulatory network changes between cell lines and their tissues of origin. BMC Genomics *18*, 723.

Lopes-Ramos, C.M., Kuijjer, M.L., Ogino, S., Fuchs, C.S., DeMeo, D.L., Glass, K., and Quackenbush, J. (2018). Gene Regulatory Network Analysis Identifies Sex-Linked Differences in Colon Cancer Drug Metabolism. Cancer Res. *78*, 5538–5547.

Luo, W., Zhu, X., Liu, W., Ren, Y., Bei, C., Qin, L., Miao, X., Tang, F., Tang, G., and Tan, S. (2016). MYC associated zinc finger protein promotes the invasion and metastasis of hepatocellular carcinoma by inducing epithelial mesenchymal transition. Oncotarget *7*, 86420–86432.

Mayne, B.T., Bianco-Miotto, T., Buckberry, S., Breen, J., Clifton, V., Shoubridge, C., and Roberts, C.T. (2016). Large Scale Gene Expression Meta-Analysis Reveals Tissue-Specific, Sex-Biased Gene Expression in Humans. Front. Genet. *7*, 183.

Melé, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., et al.;

GTEx Consortium (2015). Human genomics. The human transcriptome across tissues and individuals. Science *348*, 660–665.

Mielke, M.M., Vemuri, P., and Rocca, W.A. (2014). Clinical epidemiology of Alzheimer's disease: assessing sex and gender differences. Clin. Epidemiol. *6*, 37–48.

Morrow, E.H. (2015). The evolution of sex differences in disease. Biol. Sex Differ. *6*, 5.

Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E., and Stamatoyannopoulos, J.A. (2012). Circuitry and dynamics of human transcription factor regulatory networks. Cell *150*, 1274–1286.

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

Pan, X., Zhou, T., Tai, Y.-H., Wang, C., Zhao, J., Cao, Y., Chen, Y., Zhang, P.-J., Yu, M., Zhen, C., et al. (2011). Elevated expression of CUEDC2 protein confers endocrine resistance in breast cancer. Nat. Med. *17*, 708–714.

Paulson, J.N.J.N., Chen, C.-Y.C.-Y., Lopes-Ramos, C.M., Kuijjer, M.L.M.L., Platig, J., Sonawane, A.R.A.R., Fagny, M., Glass, K., and Quackenbush, J. (2017). Tissue-aware RNA-Seq processing and normalization for heterogeneous and sparse data. BMC Bioinformatics *18*, 437.

Pozueta, J., Lefort, R., Ribe, E.M., Troy, C.M., Arancio, O., and Shelanski, M. (2013). Caspase-2 is required for dendritic spine and behavioural alterations in J20 APP transgenic mice. Nat. Commun. *4*, 1939.

Pringsheim, T., Jette, N., Frolkis, A., and Steeves, T.D.L. (2014). The prevalence of Parkinson's disease: a systematic review and meta-analysis. Mov. Disord. *29*, 1583–1590.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Sonawane, A.R., Platig, J., Fagny, M., Chen, C.-Y., Paulson, J.N., Lopes-Ramos, C.M., DeMeo, D.L., Quackenbush, J., Glass, K., and Kuijjer, M.L. (2017). Understanding Tissue-Specific Gene Regulation. Cell Rep. *21*, 1077–1088.

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

Sugathan, A., and Waxman, D.J. (2013). Genome-wide analysis of chromatin states reveals distinct mechanisms of sex-dependent gene regulation in male and female mouse liver. Mol. Cell. Biol. *33*, 3594–3610.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Troy, C.M., Rabacchi, S.A., Friedman, W.J., Frappier, T.F., Brown, K., and Shelanski, M.L. (2000). Caspase-2 mediates neuronal cell death induced by beta-amyloid. J. Neurosci. *20*, 1386–1392.

Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M.A., Marshall, J.L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., et al.; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)-Analysis Working Group; Statistical Methods groups-Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site-NDRI; Biospecimen Collection Source Site-RPCI; Biospecimen Core Resource-VARI; Brain Bank Repository-University of Miami Brain Endowment Bank; Leidos Biomedical-Project Management; ELSI Study; Genome Browser Data Integration &Visualization-EBI; Genome Browser Data Integration &Visualization-UCSC Genomics Institute, University of California Santa Cruz (2017). Landscape of X chromosome inactivation across human tissues. Nature *550*, 244–248.

van Nas, A., Guhathakurta, D., Wang, S.S., Yehya, N., Horvath, S., Zhang, B., Ingram-Drake, L., Chaudhuri, G., Schadt, E.E., Drake, T.A., et al. (2009). Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks. Endocrinology *150*, 1235–1249.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, 1431–1443.

Yin, P., and Fan, X. (2001). Estimating R^2 Shrinkage in Multiple Regression: A Comparison of Different Analytical Methods. J. Exp. Educ. *69*, 203–224.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*, eaaj2239.

Yu, Z.-H., Lun, S.-M., He, R., Tian, H.-P., Huang, H.-J., Wang, Q.-S., Li, X.-Q., and Feng, Y.-M. (2017). Dual function of MAZ mediated by FOXF2 in basal-like breast cancer: Promotion of proliferation and suppression of progression. Cancer Lett. *402*, 142–152.

STAR★METHODS

KEY RESOURCES TABLE

Deposited Data		
Sample-specific gene regulatory networks	This paper	https://grand.networkmedicine.org/tissues
GTEx RNA-Seq data (version 6)	Battle et al., 2017	https://www.ncbi.nlm.nih.gov/gap dbGaP Study Accession: phs000424v6.p1
MSigDB (c5.all.v6.0.symbols.gmt; c2.cp.kegg.v6.2.symbols.gmt)	Subramanian et al., 2005	https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp#C5
Software and Algorithms		
R	version 3.6.0 (2019-04-26)	https://www.r-project.org
limma (3.40.2)	Ritchie et al., 2015	https://bioconductor.org/packages/release/bioc/html/limma.html
PANDA (netZooM 0.1)	Glass et al., 2013	https://github.com/netZoo/netZooM
LIONESS (netZooM 0.1)	Kuijjer et al., 2019	https://github.com/netZoo/netZooM
GSEAPreranked (Java command line version 2-2.2.4)	Subramanian et al., 2005	https://www.gsea-msigdb.org/gsea/login.jsp

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Dawn L. DeMeo (dawn.demeo@channing.harvard.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The sample-specific gene regulatory networks generated in this study are available at https://grand.networkmedicine.org/tissues. The network algorithms PANDA (Glass et al., 2013) and LIONESS (Kuijjer et al., 2019) were used as implemented in netZooM 0.1 (https://github.com/netZoo/netZooM).

METHOD DETAILS

GTEx dataset

We downloaded the Genotype-Tissue Expression (GTEx) version 6.0 RNA-Seq dataset (phs000424v6.p1, 2015-10-05 released) from dbGaP (approved protocol #9112). Using YARN in Bioconductor (https://bioconductor.org/packages/release/bioc/html/yarn.html), we performed quality control by removing sex-misannotated samples, gene filtering, tissue-aware normalization using qsmooth (Hicks et al., 2018), and grouped related body regions based on gene expression similarity (further details described in Paulson et al., 2017). "Skin" includes samples from the lower leg (sun exposed) and from the suprapubic region (sun unexposed); brain (basal ganglia) category includes caudate, nucleus accumbens, and putamen; brain (cerebellum) category includes cerebellar hemisphere and cerebellum; brain (other) category includes amygdala, anterior cingulate cortex (BA24), cortex, frontal cortex (BA9), hippocampus, hypothalamus, spinal cord (cervical c-1), and substantia nigra.

This YARN-filtered dataset included gene expression profiles from 9,435 samples across 38 tissues. We removed sex-specific tissues (prostate, testis, uterus, vagina, and ovary), cell lines (EBV-transformed lymphocytes and transformed fibroblasts), and those with fewer than 30 samples in either of the sexes (kidney cortex and minor salivary gland). The final dataset contained 8,279 samples from 29 tissues (28 solid organ tissues and whole blood) and 548 research subjects (188 females and 360 males), and included expression information for 30,243 genes (including 1,074 genes encoded on the sex chromosomes).

Network inference using PANDA and LIONESS

Sample-specific networks were reconstructed using PANDA (Glass et al., 2013) and LIONESS (Kuijjer et al., 2019) as implemented in netZooM 0.1 (https://github.com/netZoo/netZooM). PANDA begins with a "prior" network consisting of TFs connected to target

genes, which is generated by mapping TF motifs to the genome. PANDA then uses a message passing algorithm to iteratively refine the network model by integrating gene expression and protein-protein interaction (PPI) data. The PPI and the TF/target gene prior networks were generated as described by Sonawane et al. (2017). Briefly, TF motifs with direct and inferred evidence were downloaded from the Catalog of Inferred Sequence Binding Preferences (Weirauch et al., 2014). For each TF, the motif with highest information content was selected and subsequently mapped to the human genome using FIMO (Grant et al., 2011). An edge in the TF/target gene prior network indicates a significant TF motif match ($p < 10^{-5}$) within the promoter region of Ensembl genes (GRCh37.p13), defined as [-750, +250] around the transcription start site. To generate the PPI network, the interactions between all TFs in the prior network were obtained and weighted based on interaction scores from StringDB v10 (Szklarczyk et al., 2015).

We applied PANDA to estimate one aggregate regulatory network including auto- and allosomes for the complete dataset, which included 9,435 gene expression samples. We then applied LIONESS to infer individual networks for each sample in the population. LIONESS uses an iterative process that leaves out each individual in a population, estimates the network with and without that individual, and then interpolates edge weights to derive an estimate for the network active in that single individual. We applied LIONESS to compare the aggregate PANDA network to PANDA networks that iteratively left out each of the 8,279 samples in the 29 tissues retained for analysis in this study. This produced 8,279 complete, weighted gene regulatory networks that consisted of 644 TFs targeting 30,243 genes (nearly 20 million edges).

In females, a non-zero weight can be found for edges to Y chromosome genes due to spurious correlation values with non-existing Y chromosome genes in females. These correlations result from sequencing reads that mis-align to the reference genome (which contains the Y chromosome), and sequencing reads normalization bias. Therefore, we replaced the edge weights to all Y chromosome genes in females to the minimum edge weight found for the Y chromosome genes across samples (to be consistent with the non-existence of Y chromosome genes in females). This had the effect of down-weighting 36,064 edges that were connected to the 56 Y chromosome genes in the female networks. The entire collection of reconstructed networks used in this study are available on: https://grand.networkmedicine.org/tissues. For a detailed description of the network methods and their interpretation, see Methods S1: Network inference using PANDA and LIONESS, related to STAR Methods.

QUANTIFICATION AND STATISTICAL ANALYSIS

Differential expression analysis

Differential expression analysis was performed using voom to transform RNA-seq read counts to log counts per million (log-cpm) with associated precision weights, followed by linear modeling and empirical Bayes procedure using limma (Law et al., 2014). We used the following linear regression model to detect sex-biased gene expression in each tissue:

$$Y \sim \beta 0 + \beta 1 Batch + \beta 2 Race + \beta 3 Age + \beta 4 BMI + \beta 5 RIN + \beta 6 Sex + \varepsilon$$

where Y is the gene expression, *Batch* denotes the type of RNA isolation batch (PAXgene, QIAGEN, and Trizol), *Race* denotes the race of the subject (Asian, Black or African American, White, American Indian or Alaska Native, and unknown), *Age* denotes the age of the subject, *BMI* denotes the body mass index of the subject, *RIN* denotes the sample RNA integrity number, and *Sex* denotes the reported sex of the subject. In brain and skin, we also accounted for multiple-samples from the same individual collected from different tissue subregions, adding the subregion from where the sample was collected as a covariate in the regression model.

We calculated the variance inflation factor (VIF) to detect multicollinearity between covariates in the model. All covariates have VIFs close to 1, suggesting that these data are generally free of multicollinearity between variables. The p values for the estimated coefficient of *Sex* were adjusted to control for the false discovery rate (FDR) from multiple testing using the Benjamini-Hochberg (BH) method. Positive t-statistics indicate male-biased gene expression; and negative t-statistics indicate female-biased gene expression. Genes with absolute fold-change ≥ 1.5 and FDR < 0.05 were considered differentially expressed.

We performed Gene Ontology (GO) enrichment analysis using pre-ranked Gene Set Enrichment Analysis (GSEA) (Java command line version 2-2.2.4) (Subramanian et al., 2005), and GO term assignments available in the Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp#C5) (“c5.all.v6.0.symbols.gmt”). For each tissue, we performed a GSEA on the list of genes ranked by the voom t-statistics. Figure 2D shows heatmaps indicating GO term enrichment. We selected the GO terms with a positive Normalized Enrichment Score (NES) across all tissues, which indicates GO enrichment in males, then we calculated the average NES, and selected the ten GO terms with the highest average NES value (first panel). We repeated this for negative NES to select the top ten GO terms enriched in females (second panel). Note that by doing this we calculated the NES average considering only the pathways that were found enriched in the same direction in all tissues. Finally, to plot GO terms changing the sex-enrichment direction across tissues, we selected the ten GO terms with the highest standard deviation of NES values across the tissues (third panel).

Transcriptome-based signal-to-noise ratio

We also investigated sex bias in gene expression by calculating a transcriptome-based signal-to-noise ratio (tSNR) to quantify the overall divergence of gene expression profiles between males and females. For the tSNR, the signal was defined as the Euclidean distance of average gene expression profiles between groups (in this case, males and females), and the noise was defined as the overall variation among individuals. The tSNR between females (*X*) and males (*Y*) was calculated as:

$$tSNR(X,Y) = \frac{\overline{X} - \overline{Y}_2}{\sqrt{\frac{\sigma_X^2}{F} + \frac{\sigma_Y^2}{M}}}$$

$$\sigma_X^2 = \frac{\sum_{i=1}^{F} x_i - \overline{X}_2^2}{F - 1}$$

$$\sigma_Y^2 = \frac{\sum_{i=1}^{M} y_i - \overline{Y}_2^2}{M - 1}$$

where F and M represent the number of female and male samples, respectively. $\overline{X}$is a vector representing the average expression of genes across female samples, and $\overline{Y}$ is the average expression of genes across male samples. Here, x_i is the expression of each gene in sample i, y_i is the expression of each gene in sample i.

Instead of calculating a single tSNR per tissue, we performed a boot-strapping analysis that allowed us to estimate a tSNR value that is robust to sample-size and sex-imbalance. For each tissue, we randomly sampled 30 males and 30 females ($F = M = 30$) and computed the corresponding tSNR. We repeated this 10,000 times and averaged across the computed tSNR values. We report these average tSNR values in Figure 2 and Table S3. To estimate the statistical significance of these average tSNR values, we derived an empirical p value. To do this, we performed 10,000 additional tests in which we selected 30 random samples for each group, irrespective of the sex labels. P values were calculated as the percentage of these tests in which the "random" tSNR value was greater than the calculated average tSNR. We performed this tSNR analysis considering the expression of all genes, and then repeated considering the expression of only autosomal genes.

Enrichment analysis for sex hormone receptor motifs

We performed a transcription factor (TF) motif scan by mapping TF motifs from the Catalog of Inferred Sequence Binding Preferences (Weirauch et al., 2014) to promoter regions within the genome (described below). The output of this scan was a set of target genes for each of 644 TFs. To find if sex hormone receptor motifs were enriched in the promoter regions of the differentially expressed genes, we performed a TF motif enrichment analysis for three sex hormone receptor TFs: estrogen receptor 1 (ESR1), estrogen receptor 2 (ESR2), and androgen receptor (AR). For each of these TFs, we used Fisher's Exact Test to evaluate the overlap between a set of differentially expressed genes in each tissue with the target genes of that TF. The p values were adjusted using the BH method. The results are presented in Table S4.

Analysis of differential network targeting

To compare the weight of each edge between male and female networks, we used a linear regression model (limma package in Bioconductor (Ritchie et al., 2015)) similar to the differential gene expression analysis described above, and adjusting for the same covariates. We used the following linear regression model to detect sex-biased edges in each tissue, where Z is the edge weight and the covariates as defined previously:

$$Z \sim \beta 0 + \beta 1 Batch + \beta 2 Race + \beta 3 Age + \beta 4 BMI + \beta 5 RIN + \beta 6 Sex + \varepsilon$$

The p values were adjusted for FDR using the BH method. Edges with FDR < 0.05 were considered sex-biased.

Each gene is connected to 644 TFs as the generated networks are complete graphs. Genes with more than 5% of their edges significantly different between males and females (FDR < 0.05) were defined as differentially targeted genes. To identify the direction of the differential targeting, we calculated the proportion of male-biased edges for each gene, which is the number of male-biased edges divided by the total number of sex-biased edges. We recognized three classes of differentially targeted genes: 1) male-biased genes (the proportion of male-biased edges is greater than 0.6); 2) female-biased genes (the proportion of male-biased edges is less than 0.4); and 3) sex-divergent genes (the proportion of male-biased edges is between 0.4 and 0.6), in this class genes have a similar number of male- and female-biased edges, but are targeted by a different set of TFs. These results are represented as scatterplots of the number of female-biased edges by the number of male-biased edges (Figure 3), and listing the differentially targeted genes by tissue (Table S6).

Transcription factors' differential targeting

To characterize how TFs drive regulatory sex differences, we evaluated: 1) the TF motifs enriched for differentially targeted genes; 2) the correlation between a TF's targeting patterns and the expression of its target genes; and 3) the association of a TF's targeting patterns with biological processes (enrichment analysis).

We performed a TF motif enrichment analysis to find TF motifs enriched in the promoter regions of the differentially targeted genes. For each TF in each tissue, we compared the number of differentially targeted genes with or without the TF motif (based on the motif scan described above) using a Fisher's Exact Test. The p values were adjusted using the BH method. For each TF motif, we

calculated a combined p value across all tissues using Fisher's method as implemented in the *sumlog* function in the metap R package (Dewey, 2019). The results are presented in Table S7.

We calculated the expression-targeting correlation for each TF in each tissue. For this, we selected a TF's target genes (genes with the TF motif in the promoter region), then we calculated the correlation between the differential expression levels of the target genes and the differential targeting levels from the TF. More specifically, we calculated the Spearman correlation between the t-statistics of differential expression of the target genes and the t-statistics of differential targeting of edge weights from the TF to its target genes. We reported the coefficient of determination (R^2), adjusted by the Wherry's formula (Yin and Fan, 2001) for all target genes and for autosomal target genes (Figure 6; Table S8). The adjusted R^2 was calculated as:

$$R^2 = 1 - \left(1 - r^2\right)\frac{n-1}{n-p}$$

where r is the Spearman correlation coefficient, n is the sample size (in this case the number of the TF target genes), p is the number of predictor variables (2 in this case).

To systematically characterize the biological processes associated with each TF's sex-biased targeting patterns, we ran pre-ranked GSEA. In each tissue, we performed one enrichment analysis for each TF with expression-targeting correlation greater than 0.3. Genes were ranked based on their level of differential targeting (t-statistics from the limma analysis) by the corresponding TF in the tissue. Knowing that the differential targeting of the Y chromosome genes is an extreme outlier when comparing male and female networks, and no additional biological processes would be found enriched in males when all the Y chromosome genes are ranked at the top of the differentially targeted genes list, we removed the Y chromosome genes for this enrichment analysis.

We performed two separate analyses, one using GO annotations (as described previously), and the other one using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (Kanehisa et al., 2016) downloaded from MSigDB: "c2.cp.kegg.v6.2.symbols.gmt." The results for selected tissues are shown in a heatmap in Figures 7 and S7. For these figures, we selected the 20 GO terms and 20 KEGG pathways with the highest NES standard deviation across the evaluated TFs. As an example, we also zoom in and show the top 20 TFs most statistically significant (based on FDR) for a biological process relevant to the tissue.