Genome Biology

RESEARCH

Open Access

# Survey of gene, lncRNA and transposon transcription patterns in four mouse organs highlights shared and organ-specific sex-biased regulation

Qinwei Kim-Wee Zhuang[1,2†], Klara Bauermeister[1†], Jose Hector Galvez[3], Najla Alogayil[1], Enkhjin Batdorj[1], Fernando Pardo Manuel de Villena[4,5], Teruko Taketo[6,7,8], Guillaume Bourque[1,2,3*] and Anna K. Naumova[1,6,8*]

[†]Qinwei Kim-Wee Zhuang and Klara Bauermeister contributed equally to the work.

*Correspondence: guil.bourque@mcgill.ca; anna.naoumova@mcgill.ca

[1] Department of Human Genetics, McGill University, Montreal, QC H3A 1C7, Canada
[2] Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto 606-8303, Japan
[3] Canadian Centre for Computational Genomics, Montreal, QC H3A 0G1, Canada
[4] Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA
[5] Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599, USA
[6] The Research Institute of the McGill University Health Centre, Montreal, QC H4A 3J1, Canada
[7] Department of Surgery, McGill University, Montreal, QC H4A 3J1, Canada
[8] Department of Obstetrics and Gynecology, McGill University, Montreal, QC H4A 3J1, Canada

## Abstract

**Background:** Sex-biased gene regulation is the basis of sexual dimorphism in phenotypes and has been studied across different cell types and different developmental stages. However, sex-biased expression of transposable elements (TEs), which represent nearly half of the mammalian genome and have the potential of influencing genome integrity and regulation, remains underexplored.

**Results:** We report a survey of gene, lncRNA, and TE expression in four organs from mice with different combinations of gonadal and genetic sex. The data show remarkable variability among organs with respect to the impact of gonadal sex on transcription with the strongest effects observed in the liver. In contrast, the X-chromosome dosage alone had a modest influence on sex-biased transcription across organs, albeit interaction between X-dosage and gonadal sex cannot be ruled out. The presence of the Y-chromosome influences TE, but not gene or lncRNA, expression in the liver. Notably, 90% of sex-biased TEs (sDETEs) reside in clusters. Moreover, 54% of these clusters overlap or reside less than 100 kb from sex-biased genes or lncRNAs, share the same sex bias, and also have higher expression levels than sDETE clusters that do not co-localize with other types of sex-biased transcripts. We test the heterochromatic sink hypothesis that predicts higher expression of TEs in XX individuals finding no evidence to support it.

**Conclusions:** Our data show that sex-biased expression of TEs varies among organs with the highest numbers of sDETEs found in the liver following trends observed for genes and lncRNAs. It is enhanced by proximity to other types of sex-biased transcripts.

**Keywords:** Biological sex, Mouse organs, Regulation of transcription, Transposable elements, LncRNA

## Background

Biological sex is an important factor in mammalian biology with sex differences observed in multiple phenotypes from physiology and metabolism to complex behaviors. In humans, male sex is a major risk factor for most types of cancers and poor outcomes following infections, whereas the female sex increases the risk of autoimmune diseases, to give a few examples [1–5]. The major contributors to biological sex differences in mammals are the sex chromosomes (XX in females and XY in males) defining the genetic sex, and the gonads defining the gonadal sex [6, 7], reviewed in [8]. The gonadal sex hormone and growth hormone signaling pathways and networks of their target genes represent the next layers of sexually dimorphic factors that contribute to gene regulation and, hence, phenotypes [9–15]. X-chromosome dosage or the presence/absence of the Y chromosome may impact gene regulation through the actions of X- or Y-linked epigenetic modifiers and transcription factors [16–20].

To understand the regulatory hierarchies governing sex-biased gene expression and get closer to identifying specific signaling pathways or transcription factors, several approaches have been used, including sex-reversed mice [16, 21], the four core genotypes (FCG) mouse model [11], mouse models addressing the roles of pituitary hormones [9, 10], testosterone and estrogen-signaling pathways [11–13], as well as mutations in genes encoding specific transcription factors and epigenetic modifiers [13, 14, 22–26]. Liver is one of the best-studied mouse organs with respect to sex-biased gene regulation. Most of the expression data from adult liver show major impacts of gonadal sex, growth hormone, testosterone, and estrogen but minor contributions from the sex chromosomes. In contrast, expression studies of certain mouse and human cell types have demonstrated the significant role of the sex-chromosome complement [16, 26–32]. Such differences likely reflect cell-type specificity of sex-biased regulation. Analyses of sex-biased expression of another group of transcripts, the long non-coding RNAs (lncRNAs) suggest a strong regulatory potential and significant contribution to sex-biased transcriptomes and sexually dimorphic phenotypes [33–37].

In most studies, the focus has been on sex-biased gene or lncRNA expression, whereas the role of biological sex in the regulation of TEs has not been examined in detail. Transposition for the majority of competent Class 1 TEs requires an RNA intermediate and, hence, active transcription [38–41]. During mammalian development, TEs normally become active in the developmental time windows when global epigenome reprogramming takes place: in primordial germ cells (PGCs) and preimplantation embryos. Expression of certain TE families facilitates the activation of the embryonic genome [42, 43]. However, after the blastocyst stage, most TEs are silenced again [44–46]. TEs may be expressed as part of lncRNAs, represent about 40% of their exonic sequences [47], and potentially contribute to their function [48]. It is worth noting that the long terminal repeats (LTRs) of many endogenous retroviruses harbor steroid hormone response elements that mediate activation of transcription in response to hormonal stimuli [49–52].

While most TEs are expected to be transcriptionally repressed in adult somatic cells, their abnormal reactivation may lead to gene dysregulation or increase the mutational burden and the risk of neoplastic transformation [40, 53–55] (reviewed in [56–59]). Indeed, global demethylation of TEs has been reported for certain types of cancers;

however, the driving forces behind TE activation that could explain sex bias in cancer risk have not been elucidated yet (reviewed in [46]).

Furthermore, the "heterochromatic sink" hypothesis provides an intriguing link between sex chromosomes and TE regulation. It has been proposed that X-chromosome inactivation (XCI), which creates a large demand for heterochromatin, could compete with the rest of the cell genome for silencing factors and causes stochastic upregulation of autosomal genes and TEs in individuals with two or more X chromosomes [60]. This may explain sexually dimorphic phenomena associated with different X-chromosome dosages in mammals, such as increased predisposition to neural tube defects in females [61] or increased variability in the expression of agouti alleles in mice [60]. Conversely, TEs may impact genome-wide transcription by sequestering heterochromatin machinery as demonstrated in *Drosophila*, where the Y chromosome that is gene-poor but TE-rich acts as a "heterochromatic sink" influencing heterochromatin distribution patterns genome-wide and contributing to sex bias in aging [62].

To the best of our knowledge, the impacts of gonadal vs genetic sex on TE regulation and its relationship with sexually dimorphic gene expression in mammalian somatic cells have not been characterized [26]. Our goal was to examine the sex-biased expression patterns of TEs, the roles of the sex chromosomes, and gonadal sex in their regulation and elucidate the relationship between TEs and sex-biased expression of other types of transcripts, i.e., genes and lncRNAs. We conducted a survey of sex-biased expression in four organs from mice with different combinations of gonadal sex and sex-chromosome complement which allows to separate the effects of gonadal sex, Y chromosome, and X-dosage. In addition, we used our expression data to test the hypothesis that the inactive X acted as a heterochromatic sink and enhanced autosomal TE or gene expression.

## Results

### Mice with different combinations of gonadal sex and sex-chromosome complement and their genetic backgrounds

To explore the contributions of gonadal sex and sex-chromosome complement in an organ-specific context, we analyzed the transcriptomes of four organs from mice with different combinations of gonadal sex and sex-chromosome complement. To generate such combinations, two mouse strains were used: the B6.$Y^{TIR}$ (Tirano) consomic mice that carry the Y chromosome from a house mouse caught in Tirano, Italy, on a C57BL/6 J genetic background [63] and mice carrying the patchy fur (*Paf*) mutation on the X chromosome [64]. About 20% of the offspring of $X^{Paf}Y$ males are females with one X chromosome (XO.F) [64], whereas about 50% of B6.$Y^{TIR}$ develop into sex-reversed females with two ovaries (XY.F) [63, 65]. The comparison of sex-reversed XY.F to XY males (XY.M) detects biased transcription driven by the gonadal sex differences, whereas comparing XX female littermates (XX.F) to XY.F detects biased transcription driven by the sex-chromosome complement. To distinguish the impact of X-chromosome dosage from that of the Y chromosome, we compared females with monosomy X (XO.F) to their $XX^{Paf}$ ($XX^{Paf}$.F) littermates originating from crosses between $X^{Paf}Y$ males and C57BL/6 J females.

Since genetic background influences gene regulation, and variability in the genetic background may obscure the effect of sex on expression [66, 67], several mice from

the experimental crosses were genotyped using the MiniMUGA version of the Mouse Universal Genotyping Array (MUGA) [68]. Mice from the B6.Y$^{TIR}$ cross carried 100% C57BL/6 J alleles on autosomes, the X chromosome, and mitochondria, while the Y chromosome was derived from the TIRANO strain. The XO.F and XX$^{Paf}$.F mice had varying contributions from the C3H/He strain in heterozygosity (ranging from 0 to 54%) with the remaining alleles derived from the C57BL/6 J strain (Additional file 1: Table S1).

RNA-seq was performed on lung parenchyma, heart, and whole brain samples from adult XX.F, XY.F, XY.M, XO.F, and XX$^{Paf}$.F mice. Liver data from our previous study were also analyzed [21]. All data were deposited to NCBI (GSE248074).

#### Organ-biased gene, lncRNA, and TE expression

As a starting point, we identified de novo lncRNAs (referred to as lncRNAs hereafter) in our dataset. In total, 9754 lncRNAs were identified in the brain, 3823 in the heart, 7885 in the liver, and 6325 in the lung (Additional file 2: Table S2). Data quality check confirmed that all three types of transcripts (genes, lncRNAs, and TE subfamilies) had comparable global expression levels (Additional file 3: Fig. S1) and expression patterns were highly similar between samples from the same organ (Additional file 3: Fig. S2). Principal component analysis (PCA) for the most variable genes, lncRNAs, and TE subfamilies showed clusters formed by organs (Fig. 1A).

To test if organ differences in TE profiles were the result of predominant organ-specific expression of certain TE subfamilies, we identified organ-biased TE subfamilies (with higher expression in one specific organ but lower in the other three) by combining Tao index $\tau$ [69, 70] and $z$-score ($\tau > 0.6$, $z$-score $> 1$) (Fig. 1B, Additional file 3: Fig. S3A, Additional file 4: Table S3). Among the four organ-biased TE subfamilies as well as the subset that was common across organs, the long terminal repeat (LTR) class subfamilies were the most numerous (Additional file 3: Fig. S3B). We further identified organ-biased TEs at the instance level and observed the highest number of organ-biased TE instances in the liver (Fig. 1C, Additional file 4: Table S3).

In summary, our data confirmed organ-specific expression of annotated genes and lncRNAs, in agreement with previous reports [71–73]. In addition, we identified organ-biased TEs at both subfamily and instance levels.

#### *Gonadal sex and sex-chromosome complement impact mouse transcriptomes in an organ-specific fashion*

To explore the impacts of gonadal sex and sex-chromosome complement on expression in different organs, PCA was performed for each type of transcript and each organ, separately (Fig. 2A, Additional file 3: Fig. S4, Fig. S5). In the gene PCA, brain samples are clustered by sex-chromosome complement and genetic background. In the liver, samples formed clusters by both gonadal sex and sex-chromosome complement. It is also notable that in the liver, principal component 1 (PC1) had the highest percentage of variance explained (64%) and seemed to be associated with gonadal sex. This suggests that not only were samples in the liver analysis clustering by gonadal sex, but that this was also the variable that explained the greatest difference between samples. In the heart and lung, samples formed clusters based on the presence or absence of the Y chromosome,

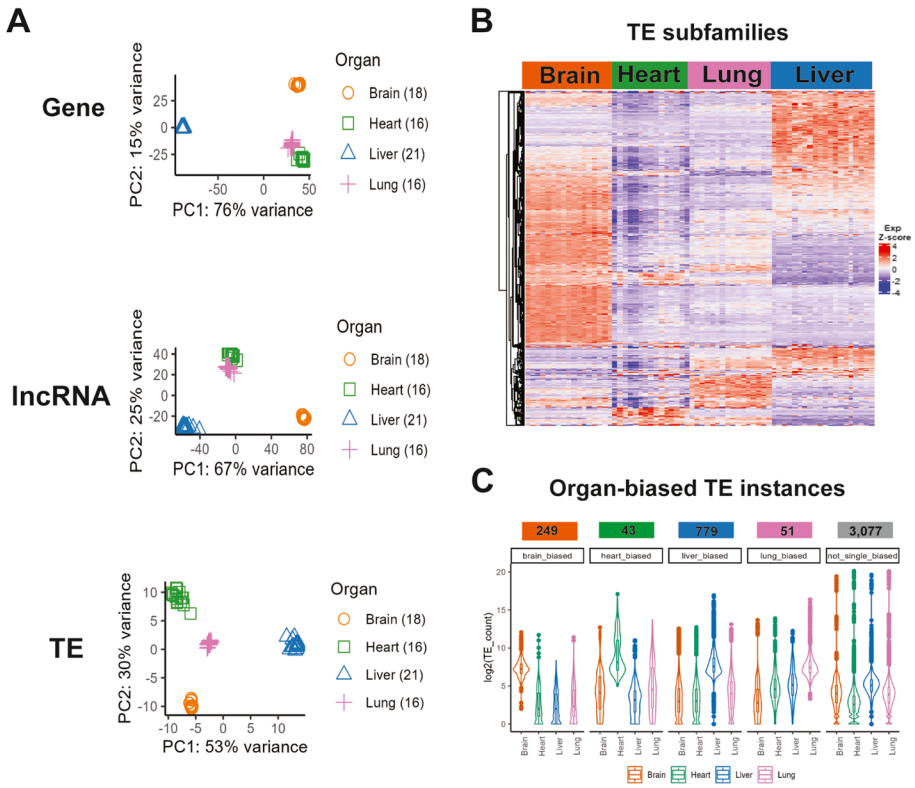
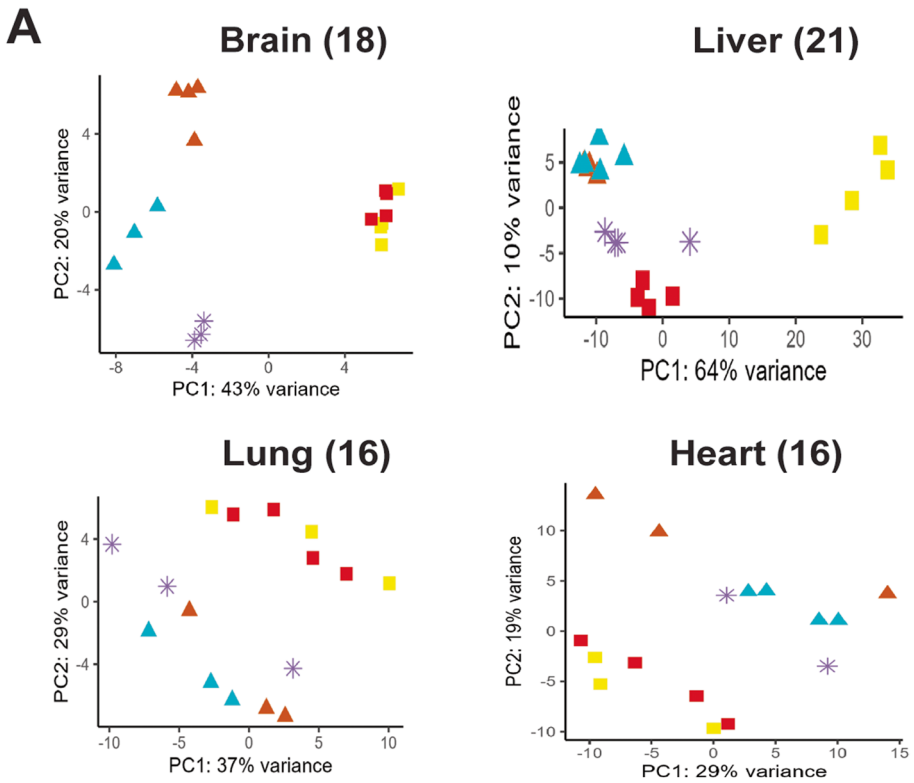



**Fig. 1** Organ-specific expression patterns. **A** PCA plot of gene (top), lncRNA (middle), and TE (bottom) expression. The top 300 genes, lncRNAs, or TE subfamilies with the highest variance across samples were used. Numbers in the parenthesis refer to the number of samples for each organ. **B** TE subfamily expression profiles show organ-specific patterns. Each column corresponds to one sample and each row—to one TE subfamily (1054 TE subfamilies in total). The *Z*-score was calculated for each TE subfamily and used to generate the heatmap. **C** Organ-biased expression levels of TE instances. Organ-biased TE instances were defined with *Z*-score > 1, $\tau > 0.6$, and average aligned reads > 100

but the variance explained as estimated by PC1 was lower (29% and 37%, respectively) than in the brain and liver (43% and 64%, respectively).

Separate analysis of X-linked genes showed samples with one and two X chromosomes falling into different clusters, as expected (Additional file 3: Fig. S4). No major effects of gonadal sex or sex-chromosome complement on autosomal gene expression were found in the brain, heart, or lung, whereas in liver samples formed clusters by gonadal sex (Additional file 3: Fig. S4). Therefore, the effect of the Y chromosome on the lung and heart transcriptomes was likely driven by Y-linked genes. Similar trends were found for lncRNAs and TEs (Additional file 3: Fig. S5).

Next, we identified sex-associated DEGs (sDEGs) (Additional file 5: Table S4), lncRNAs (slncRNAs) (Additional file 6: Table S5), and differentially expressed TEs (sDETEs) (Additional file 7: Table S6) in four pair-wise comparisons of sex/genotype groups for each organ (numbers summarized in Table 1). Overall, the numbers of sDEGs, slncRNAs, and sDETEs were concordant within organs with the highest numbers of sex-biased transcripts found in the liver. The overlap between sDEGs from different organs was small (Fig. 2B) with only five DEGs shared between all organs: X-inactive specific transcript (*Xist*) and Y-linked *Kdm5d*, *Uty*, *Ddx3y*, and *Eif2s3y*.




| | Brain | Heart | Lung | Liver |
|---|---|---|---|---|
| XO.F | 3 | 2 | 3 | 4 |
| XX.F | 4 | 3 | 3 | 4 |
| XXpaf.F | 3 | 4 | 3 | 5 |
| XY.F | 4 | 4 | 4 | 4 |
| XY.M | 4 | 3 | 3 | 4 |

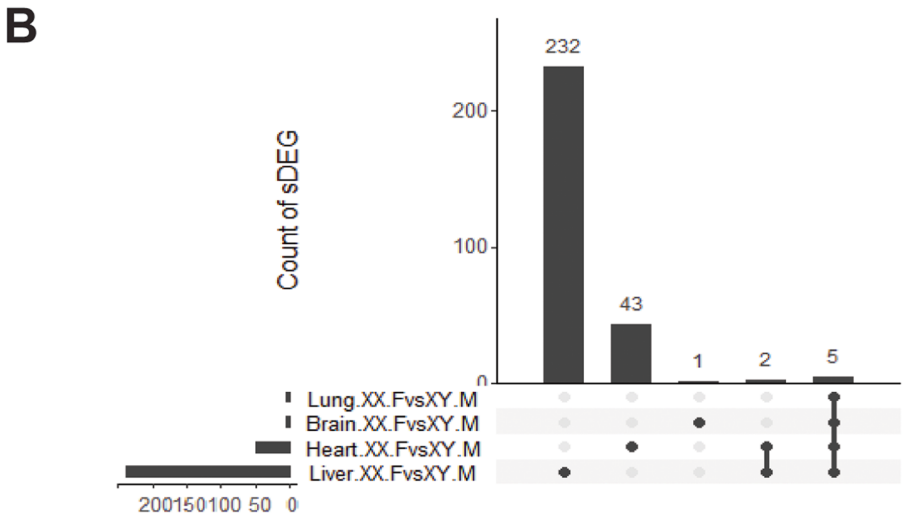



**Fig. 2** Effects of gonadal sex and sex-chromosome complement on gene expression in four organs. **A** PCA plot of gene expression in four organs. **B** Numbers of sDEGs from the XX.F vs XY.M comparison in the four organs. Numbers of DEGs in each subgroup are shown above each bar

**Table 1** Numbers of sDEG, slncRNAs, and sDETE in the four organs

| Comparison | Organ | Lung | | | Heart | | | Whole brain | | | Liver | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Chromosomes | Auto | X | Y | Auto | X | Y | Auto | X | Y | Auto | X | Y |
| **XX.F vs XY.M** | sDEG | 0 | 1 | 4 | 43 | 3 | 4 | 2 | 1 | 4 | 230 | 5 | 4 |
| | slncRNA | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 201 | 6 | 2 |
| | sDETE instance | 3 | 0 | 0 | 12 | 0 | 0 | 1 | 0 | 1 | 1403 | 16 | 3 |
| **XY.F vs XY.M** | sDEG | 6 | 1 | 0 | 33 | 2 | 0 | 0 | 0 | 0 | 174 | 2 | 0 |
| | slncRNA | 1 | 0 | 0 | 1 | 0 | 0 | 3 | 0 | 0 | 103 | 7 | 0 |
| | sDETE instance | 2 | 1 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 854 | 5 | 0 |
| **XX.F vs XY.F** | sDEG | 0 | 1 | 4 | 4 | 1 | 4 | 0 | 2 | 4 | 3 | 1 | 4 |
| | slncRNA | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 8 | 1 | 3 |
| | sDETE instance | 1 | 0 | 0 | 3 | 1 | 0 | 1 | 0 | 1 | 146 | 1 | 3 |
| **XX$^{Paf}$.F vs XO.F (X-dosage)** | sDEG | 1 | 1 | 0 | 0 | 1 | 0 | 4 | 1 | 0 | 2 | 1 | 0 |
| | slncRNA | 1 | 2 | 0 | 0 | 1 | 0 | 2 | 2 | 0 | 1 | 2 | 0 |
| | sDETE instance | 1 | 2 | 0 | 0 | 0 | 0 | 5 | 1 | 0 | 16 | 12 | 0 |
| **Y-dependent*** | sDEG | 0 | 0 | 4 | 4 | 0 | 4 | 0 | 1 | 4 | 3 | 0 | 4 |
| | slncRNA | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 8 | 0 | 3 |
| | sDETE instance | 1 | 0 | 0 | 3 | 1 | 0 | 1 | 0 | 1 | 146 | 1 | 3 |

Two sDEGs, NADPH oxidase 4 (*Nox4*) and ERBB receptor feedback inhibitor 1 (*Errfi1*), were found in the heart and liver in the XX.F vs XY.M comparisons.

sDEG, slncRNA, and sDETE cut-offs: $|\log FC| \geq 1.0$, adj $p$-value $< 0.05$. * To remove low-expressed transcripts, $\log CPM \geq 1.0$ was applied for sDEGs, and number of average aligned reads > 50 was applied for slncRNA and sDETE. Y-dependent transcripts: identified in XX.F vs XY.F but not in XX$^{paf}$.F vs XO.F

Overall, gonadal sex had the biggest effect in the liver, followed by a moderate effect in the heart but little effect on brain or lung transcriptomes (Table 1, Additional file 5: Table S4, Additional file 6: Table S5, Additional file 7: Table S6). In contrast to liver where both male- (86 sDEGs, 58 slncRNAs, and 444 sDETEs) and female-biased (87 sDEGs, 45 slncRNAs, and 410 sDETEs) autosomal expression was observed, in the heart, most gonadal-sex dependent sDEGs (32 of 33), the slncRNA and all sDETEs had higher expression in males.

The numbers of X-dosage-dependent and Y-dependent sDEGs were low and similar between organs (Table 1), suggesting that the influence of gonadal sex alone or interaction between gonadal sex and sex-chromosome complement rather than sex-chromosome complement alone accounted for most of the organ differences in sex-biased transcription. Y-linked genes *Kdm5d*, *Uty*, *Eif2s3y*, and *Ddx3y* were expressed in all four organs, which suggests that this region of the Y chromosome is transcriptionally active in all four adult organs. A notably high proportion (nearly 10%) of liver sDETEs were sensitive to the presence of the Y chromosome (identified in XX.F vs XY.F but not in XX$^{Paf}$.F vs XO.F comparison) (Table 1). Among these, 98 had higher expression in XX.F and 52 had higher expression in XY.F samples (Additional file 7: Table S6). The effect of the Y chromosome in the brain, heart, and lung was modest compared to that in the liver.

In principle, the striking organ differences with respect to sex-biased expression of TEs may be due to an increase in the expression of TEs that are sensitive to sex-associated factors in the liver and reduced expression of such TEs in the other three organs. To test such a possibility, we compared the overlap between sDETEs and organ-biased TEs. Remarkably, we found only 92 out of 1422 sDETE instances were liver-biased. Hence, liver-biased TEs do not explain the high abundance of sDETE in the liver.

We also tested the heterochromatic sink hypothesis, which predicts that genes and TEs would have higher expression in females with two X chromosomes compared to those with one X (e.g., XY.F or XO.F). We did not observe an excess of TEs with higher expression in XX females: 9 of 16 autosomal liver DETEs (Additional file 7: Table S6) had higher expression in $XX^{Paf}$.F compared to XO.F. Hence, our data from adult organs do not support the heterochromatic sink hypothesis.

We next asked if there was interaction between gonadal sex and the sex-chromosome complement. Transcription that requires both factors would be detected only in the XX.F vs XY.M but not XY.F vs XY.M or XX.F vs XY.F comparisons. Thirty-two percent of liver and 56% of heart sDEGs were unique to the XX.F vs XY.M comparison (Additional file 5: Table S4), suggesting a significant proportion of sDEGs may be regulated by both gonadal sex and sex-chromosome complement. Interaction was observed for slncRNAs and sDETEs too, where 51% of slncRNAs and 49% of sDETEs were unique to the XX.F vs XY.M comparison.

In conclusion, the impact of gonadal sex on expression was wider and most pronounced in the liver and heart, whereas the lung and whole brain showed very modest or no effect of gonadal sex. The effects of the sex-chromosome complement on autosomal expression were limited to a small number of transcripts in all four organs. Yet, the interaction between X-dosage and gonadal sex cannot be ruled out.

#### Both gonadal sex and sex-chromosome complement influence expression of X-linked genes, lncRNAs, and TEs

To examine the contributions of gonadal sex and sex-chromosome complement on X chromosomal transcription, we next analyzed the expression of X-linked genes, lncRNAs, and TEs. X-linked sDEGs, slncRNAs, and sDETEs were selected using relaxed criteria of adj deseq2 $p$-value < 0.1, either logCPM > 1.0 (for sDEG) or baseMean > 50 (for slncRNAs and sDETEs), no cut-off for logFC (Table 2, Additional file 8: Table S7). This relaxed cut-off was used because genes that escape X-inactivation do not necessarily show double the expression in individuals with two X chromosomes compared to those with one [74, 75]. The $XX^{Paf}$.F vs XO.F comparison revealed the impact of X-dosage, whereas sex-biased transcripts present in the XX.F vs XY.F but not in the $XX^{Paf}$.F vs XO.F comparison were considered to be Y-chromosome-dependent (Table 2).

In the brain, all X-linked sDEGs depended on the sex-chromosome complement. All of them, except for glycoprotein m6b (*Gpm6b*) and dystrophin, muscular dystrophy (*Dmd*), had higher expression in females with two X chromosomes, which may be due to escape from X-inactivation (Additional file 8: Table S7). In the lungs, a relatively small number of X-linked sDEGs was detected. Six out of the 7 X-linked sDEGs in the XX.F vs XY.M comparison and all those detected in XX.F vs XY.F or $XX^{Paf}$.F vs XO.F comparisons had higher expression in XX females. In the liver and heart, the influence of

**Table 2** X-linked sDEGs, slncRNAs, and sDETEs

| Comparison | Transcript type | Lung | Heart | Whole brain | Liver |
|---|---|---|---|---|---|
| **XX.F vs XY.M** | sDEG | 7 | 6 | 7 | 15 |
| | slncRNA | 1 | 1 | 0 | 13 |
| | sDETE instance | 0 | 1 | 0 | 59 |
| **XY.F vs XY.M** | sDEG | 1 | 9 | 0 | 10 |
| | slncRNA | 0 | 0 | 0 | 7 |
| | sDETE instance | 1 | 0 | 0 | 50 |
| **XX.F vs XY.F** | sDEG | 10 | 4 | 7 | 2 |
| | slncRNA | 1 | 1 | 0 | 4 |
| | sDETE instance | 0 | 1 | 0 | 18 |
| **$XX^{Paf}$.F vs XO.F (X-dosage)** | sDEG | 8 | 2 | 5 | 2 |
| | slncRNA | 1 | 1 | 3 | 5 |
| | sDETE instance | 2 | 0 | 1 | 15 |
| **Y-dependent*** | sDEG | 2 | 2 | 2 | 1 |
| | slncRNA | 0 | 0 | 0 | 3 |
| | sDETE instance | 0 | 1 | 0 | 16 |

Relaxed cut-off: adj deseq2 *p*-value <0.1, either logCPM>1.0 (for sDEG) or baseMean > 50 (for slncRNAs and sDETEs), no cut-off for logFC

* Y-dependent transcripts: identified in XX.F vs XY.F but not in $XX^{Paf}$.F vs XO.F

gonadal sex on the expression of X-linked genes was evident, which follows the trends observed for autosomal sDEGs. The impact of X-dosage was also detected (Table 2). Among genes known to escape X-inactivation, two, *Kdm5c* and *Ddx3x,* had significantly higher expression in the XX compared to XY or XO lungs and brains, but not hearts and livers, whereas *Kdm6a* and *Eif2s3x* displayed consistently higher expression in XX females across all organs. (Additional file 8: Table S7).

One gene, *Hccs* showed biased expression in three organs but only in the XO.F vs $XX^{Paf}$.F comparison with higher expression in $XX^{Paf}$.F. *Hccs* is located within the region of heterozygosity where $XX^{Paf}$ females carry C3H/HeNRj alleles on their paternal X chromosomes. Therefore, it is possible that the *Hccs* expression bias reflects allelic effects rather than escape from XCI.

Similar to autosomal sDEGs, the largest number of X-linked DETEs and slncRNAs was detected in the liver with well-pronounced effects of gonadal sex and X-chromosome dosage. Only a few X-linked sDETEs/slncRNAs were found in the brain, heart, and lung, even with a relaxed cut-off (Table 2). There was no overlap between X-linked sDETEs from the liver and those from other organs. Moreover, the Lx8b_dup13983:Lx8b:L1:LINE (genomic position chrX:169,360,167–169361279), which was the only X-linked sDETE identified in the $XX^{Paf}$.F vs XO.F comparison in both brain and lung, resides distal to *Hccs* and likely reflects the genetic make-up of the $X^{Paf}$ chromosome.

Thus, expression of X-linked genes, lncRNAs, and TEs is regulated by X-dosage in all organs and gonadal sex in the liver and heart, consistent with the trends observed for autosomal transcripts.

#### Concordant expression between sDEG/slncRNA and sDETEs

Concordance in the numbers of sDEGs, slncRNAs, and sDETEs across organs led us to hypothesize that the three types of transcripts were co-regulated. If this was the case,

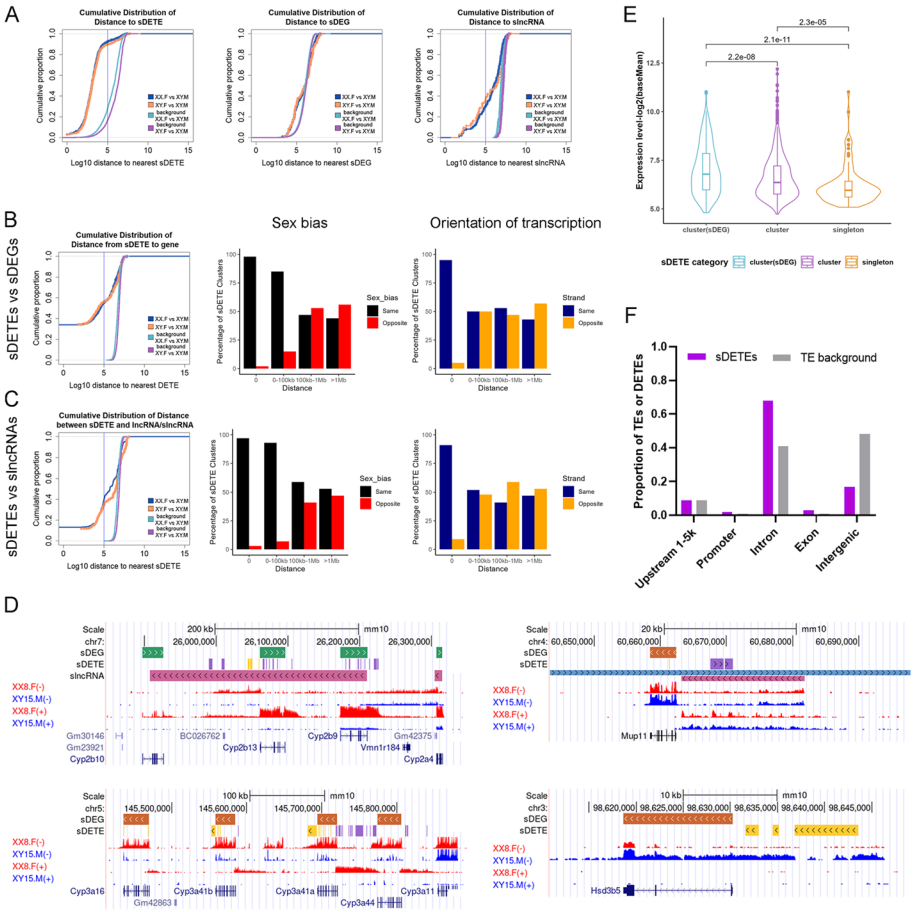



**Fig. 3** Concordant expression between sDETE and nearest sDEG or sIncRNA. **A** Cumulative distribution of distances between sDETE and the nearest sDETE (left panel), sDEG and the nearest sDEG (central panel), and sIncRNA and nearest sIncRNA (right panel). The blue vertical line refers to the distance of 100 kb. **B** Concordant expression of sDETEs/sDETE clusters and nearby sDEGs. Cumulative distribution of distance between a sDETE and the nearest sDEG (left panel); sex bias (center) and orientation of expression (right panel) in sDETE clusters and the nearest sDEG as a function of distance. **C** Concordant expression of sDETE clusters and nearby sIncRNAs, similar to **B**. **D** Examples of co-localization of sDETE clusters, sDEGs, and sIncRNAs. Examples of female-biased transcripts are shown in the two left panels, and examples of male-biased transcripts are shown on the right. Colors and arrows indicate the direction of transcription. For sDEGs, green shows the + strand, and orange indicates the − strand. For sDETEs: purple (+) and yellow (-) strands, respectively. For sIncRNAs: blue (+) and pink (−) strands. All features are shown in the context of the UCSC browser. The bottom right panel shows a putative initiating sDETE. **E** Expression levels of singleton sDETEs vs sDETEs residing in clusters. Each dot refers to one sDETE. Three groups were plotted: sDETEs that belong to clusters that overlap with sDEGs, sDETEs that belong to clusters but do not overlap with sDEGs, and sDETE singletons. **F** Genic annotations of sDETEs from the liver XX.F vs XY.M comparison vs all TEs

co-regulated transcripts would be found in closer proximity than those that are not associated by regulation. To test this hypothesis, we focused on the liver transcriptome, for which we had the largest number of sDETEs. We mapped sDEGs, sIncRNAs, and sDETEs across chromosomes and confirmed that they were scattered across all chromosomes rather than located on a few specific chromosomes (Additional file 3: Fig. S6).

Next, we tested the co-localization of sex-biased transcripts within and between the three groups (sDEGs, sIncRNAs, sDETEs) using data from the liver XX.F vs XY.M and XY.F vs XY.M comparisons (Fig. 3A). About 90% of sDETEs, 30% of sDEGs, and 35%

of slncRNAs were located within 100 kb of another element from the same category. Moreover, more than 55% of sDETEs were located within 100 kb of a sDEG and 35% of slncRNA (Fig. 3B and C, left panel). These sDETEs are referred to as "proximal" from this point on, whereas sDETEs that were located farther than 100 kb from another sex-biased transcript are referred to as "distal". To determine if different classes or families of TEs were more likely to be enriched among proximal or distal sDETEs, we inspected sDETEs from the XX.F vs XY.M comparison (Additional file 3: Fig. S7). LTRs, and more specifically ERVK were slightly over-represented among proximal sDETEs, while SINE and more specifically the Alu family were more observable among distal compared to proximal sDETEs (Additional file 3: Fig. S7). LINE and L1 were underrepresented among distal sDETEs. ERV1 and MIR were enriched among both proximal and distal sDETEs.

Several lines of evidence suggest that TEs may act as promoters and influence transcription of proximal genes [76–79]. Such putative initiating TEs would be expected to reside close (e.g., within 2 kb) to the TSS of the gene under their regulation and share the same orientation of transcription. Since we used discrete cut-offs to identify sex-biased transcripts, there was a chance that a number of low-expressed sex-biased genes were missed. Hence, we generated a comprehensive list of transcripts from a de novo transcriptome assembly using RNA-seq data from all 21 liver samples, referred as liver transcripts hereafter. Next, we identified putative initiating sDETEs that resided in close proximity to a TSS of a liver transcript. sDETEs were more likely to reside upstream of TSS of liver transcripts on the same strand than that on the opposite strand (Additional file 3: Fig. S8). Fifty-one sDETE (3.6%) either overlapped with or resided within 2 kb upstream of a TSS of liver transcripts (Fig. 3D, Additional file 3: Fig. S9, Additional file 9: Table S8). LTRs were significantly enriched compared to all sDETEs from the XX.F vs XY.M comparison: 27 out of 51 (53%) were LTRs ($p = 3.1 \times 10^{-4}$, hypergeometric test, Additional file 9: Table S8) compared to the background of all 1422 sDETEs. As for TE families, ERVK and ERV1 families were significantly enriched ($p = 0.0060$ and $p = 0.028$, respectively, hypergeometric test) (Additional file 9: Table S8).

Nevertheless, these data suggest that putative-initiating sDETEs represent a very small fraction of all sDETEs and do not explain the co-localization between sDETEs and other sex-biased transcripts.

To better understand the relationship between sDETEs, sDEGs, and slncRNAs, we merged neighboring sDETEs (distance < 100 kb) with the same sex bias into sDETE clusters using data from the XX.F vs XY.M comparison and examined the co-localization between clusters of sDETEs and sDEGs or slncRNAs. Ninety-one percent of sDETEs (1296 out of 1422) formed 173 (101 female-biased vs 72 male-biased) clusters with an average of 7.5 sDETEs per cluster (range 2 to 49) (Additional file 10: Table S9). Around 85% (414/486) of distal sDETEs and 86% (44/51) of putative initiating sDETEs resided within sDETE clusters. In total, 45.1% (78 out of 173) of sDETE clusters overlapped with sDEGs and/or slncRNAs (46 with sDEGs only, 20 with slncRNAs only, and 12 with both sDEGs and slncRNAs). The percentage increased to 54% (94 out of 173) when expanding the distance between the sDETE and the nearest sDEG/slncRNA to 100 kb. Hence, slightly more than half of sDETE clusters co-localized with sDEGs or slncRNAs. The rest did not have sex-biased transcripts of either type within a 100 kb distance. Ninety-five percent of sDEG-overlapping and 91% of slncRNA-overlapping sDETE clusters had the

same sex bias as the overlapped sDEG or slncRNA (Fig. 3B–C, right panel), respectively. The proportion of same-sex bias remained above 85% when extending the distance between the sDETE cluster and other sex-biased transcripts to 100 kb but dropped to around 50% when the distance was larger than 100 kb (Fig. 3B–C). When the direction of transcription was examined, 95% (55 out of 58) and 91% (29 out of 32) of sDETE clusters were transcribed from the same strand as the overlapping sDEG or slncRNA, respectively. Examples of sDETE clusters that co-localize with sDEG/slncRNA are shown in Fig. 3D. Worth noting, singleton sDETEs (located outside of sDETE clusters) showed significantly lower expression levels than those forming clusters, especially those clusters that overlapped with sDEGs (Fig. 3E).

TEs are often part of non-coding transcripts, and 80% of lncRNAs harbor TEs within their exons [47] To determine whether the sDETEs that overlapped sDEGs or slncRNAs were parts of their mature transcripts, we inspected the genic annotation of sDETEs from the liver XX.F vs XY.M comparison. Briefly, while 33.5% (476 out of 1422) of sDETE overlapped with sDEGs, only 3.6% (17) were exonic. Meanwhile, 12.8% (182 out of 1422) of sDETEs overlapped with slncRNAs, of which 62.6% (114) were exonic. In general, TEs tend to reside in intronic and intergenic regions; however, sDETEs were enriched in introns of overlapping coding and non-coding transcripts, but less so in intergenic regions (Fig. 3F). A possible explanation of such enrichment is that intronic sDETEs are parts of pre-mRNA molecules or intronic debris captured by the RNA-sequencing. If this were the case, one would expect that all intronic TE sequences would be equally represented among transcripts. To test this possibility, we compared the expression levels of sDEG-overlapping TEs based on their sex bias. Among the TEs with at least one aligned read, sDETEs that overlapped with sDEGs had higher expression levels than other TEs overlapping the same sDEG (Additional file 3: Fig. S10). Hence, sDETEs are not a technical artifact but *bona fide* transcripts.

In conclusion, sDETEs tended to form clusters and showed high concordance with overlapping sDEG/slncRNA. sDETE clusters located within 100 kb also showed highly concordant sex bias with neighboring sDEG/slncRNA, regardless of the strand. This suggests possible shared regulation or an impact of domains with active chromatin on TE transcription but putative initiating sDETEs were very rare. About 34% of sDETE instances and nearly 50% of sDETE clusters do not colocalize with either type of the sex-biased transcripts and therefore are independently regulated. These clusters have lower transcription levels.

## Discussion

### The contributions of sex-chromosome complement and gonadal sex to sex-biased gene, lncRNA, and TE regulation vary in different organs

Here, we conducted a survey of the sex-biased gene, lncRNA, and TE expression and compared the contributions of the sex-chromosome complement and gonadal sex in four mouse organs: lung parenchyma, heart, whole brain, and liver. For all three types of transcripts, the impact of gonadal sex was wider and most pronounced in the liver and heart, whereas the lung and whole brain transcriptomes showed very modest or no effect of gonadal sex. Our data support previously reported dominant roles of gonadal-sex-associated factors, including those from growth hormone signaling pathways, on

gene expression in the liver [9–13] and demonstrate the influence of gonadal sex on TE expression. They also suggest that the four organ differences in sex-biased expression may be largely driven by different sensitivity to factors associated with gonadal sex rather than sex-chromosome complement. The overall low numbers of sDEGs in the lung and brain are consistent with those observed in other studies of the C57BL/6 J strain [80–82]. However, the number of heart sDEGs found in our study is considerably lower than those reported by others [27]. Such differences may be associated with the use of more stringent inclusion/exclusion criteria. In addition, the slncRNAs that we identified among the de novo lncRNAs provided novel targets, beyond annotated genes, of sex-biased transcripts for further investigation. Overall, we observed a correlation between numbers of sex-biased genes, lncRNAs, and TEs among organs, which suggests common tissue-specific regulatory mechanisms affecting sex-biased transcription.

The effects of the sex-chromosome complement on autosomal gene expression were limited, mostly to four Y-linked genes that were expressed across all tested organs and a few autosomal genes, which showed organ-specific sex-biased expression. In the mouse liver, autosomal TEs but not genes or lncRNAs were sensitive to the presence of the Y chromosome. This is consistent with our previous finding of the influence of the Y chromosome on methylation levels of autosomal repetitive elements [83].

When the expression of X-linked transcripts was examined using relaxed inclusion criteria, we found a relatively small number of X-linked DEGs, slncRNAs, and DETEs. In the brain, heart and lung, most of the X-linked transcripts were sensitive to X-dosage. Most of the X-dosage-sensitive DEGs are either known or were predicted to escape X-inactivation [75, 84, 85]. Among them are the well-known escapees *Kdm6a*, *Kdm5c*, *Ddx3x*, and *Eif2s3x.* However, in the liver, gonadal sex had a stronger influence on X-linked transcription, which is consistent with the patterns observed for autosomal transcripts. Our data suggest that transcriptional regulation of X-linked genes and TEs depends on general regulatory mechanisms shared with the rest of the genome as well as the X-inactivation status and these impacts are organ-specific.

#### Sex-biased TE expression is organ-specific and associated with sex-biased gene expression

We observed variation in the repertoires and numbers of expressed TEs at both subfamily and instance levels among the four tested organs. Liver had the largest number of highly expressed TEs compared to the other three organs investigated. Organ-biased TE expression spans multiple TE classes; however, the LTR class was the most enriched among organ-biased TE subfamilies as well as subfamilies that were common between organs, suggesting that LTRs are most actively transcribed TEs in all four mouse organs.

Organ differences in TE expression were also observed when comparing between sex/genotype groups within each organ separately. Liver had by far more sDETE subfamilies and instances than the brain, heart, or lung. On autosomes, both gonadal sex and sex-chromosome complement (mainly the presence of the Y chromosome) influenced the sex-biased expression of TEs in the liver, but the impact of gonadal sex was more observable than that of the sex-chromosome complement. However, there were sex-biased transcripts that were found only in the XX.F vs XY.M comparison suggesting potential interaction between gonadal sex and sex-chromosome complement. In conclusion, the impact of gonadal sex and sex-chromosome complement on TE expression is

organ-specific, and most observable in the liver. This warrants further investigation of the tissue-specific mechanisms involved.

The similarities between gene or lncRNA and TE expression patterns with respect to sex and organ-specificity led us to examine the relationship between the three types of sex-biased transcripts in further detail. We found clustering of sDETEs and concordant regulation of sex-biased expression of TEs and genes/lncRNAs in the adult mouse liver, which is consistent with findings in other species [86, 87]. We speculate that sex-biased expression of other types of transcripts boosts the expression of neighboring TEs in two different ways: sDETEs are activated by the expression of overlapping sDEGs or slncRNAs, whereas sDETEs located within 100 kb of the nearest sDEG or slncRNA are likely the result of local active chromatin environment. The effect fades with increasing distance from the sDEG. We also speculate that such a concordance between gene, lncRNA, and TE expression may have implications for TE-driven mutagenesis in somatic cells of an adult organism.

In our dataset, we found asymmetry with respect to enrichment of certain classes of TEs among proximal vs distal sDETEs that may offer clues to the functional roles of sDETEs in sex-biased expression. LTRs were enriched among proximal but not distal sDETEs, whereas the reverse pattern was seen for SINE (Additional file 3: Fig. S7). TEs are also known to act as promoters or enhancers modifying the expression of proximal genes and causing phenotypic variation [76–79]). In our attempt to identify putative initiating sDETEs, we used a conservative approach and set the cut-off for distance to the proximal transcript at 2 kb. Only about 1% of liver sDETEs reside within 2 kb or closer to a TSS of a sDEG or a slncRNA, suggesting that they are unlikely to be the main drivers of sex-biased expression in adult mouse liver. To establish the regulatory potential of sDETEs, experiments involving genetic manipulation and directed mutagenesis would be required.

#### The heterochromatic sink hypothesis

The heterochromatic sink hypothesis predicts that XX animals should have higher expression levels of autosomal genes or TEs compared to animals with only one X [60, 88]. In our liver dataset, we found only 2 X-dosage dependent autosomal DEGs and 16 sDETEs. X-dosage-dependent autosomal DETEs show comparable proportions of XO-biased and XX-biased expression. Overall, the small numbers of X-dosage-sensitive DEGs and DETEs do not support a major genome-wide effect of X-dosage on autosomal gene regulation in either of the four adult organs. Hence, if the inactive X acts as a heterochromatic sink activating expression across autosomes, its effect is not detected in adult somatic cells using our present approaches. We speculate that in mammals, the phenomenon is either limited to early developmental stages [89], or increases the rate of stochastic epigenetic anomalies, or absent in normal conditions and requires mutations or environmental insults that would destabilize the epigenetic silencing machinery to exert its regulatory impact. However, we observe a rather high number of DETEs whose expression seems to be sensitive to the presence/absence of the Y chromosome. The mouse Y chromosome has a unique structure with at least 200 copies of the same DNA region and is also enriched for TEs [90]. Our observation warrants further experiments to test whether the Y chromosome may act as a heterochromatic sink in mouse cells.

## Conclusions

In conclusion, our results support previous findings that sex-biased expression varies among organs. They also suggest that gonadal sex may be the driving force for organ differences in sex-biased expression. It has the most pronounced impact on the adult liver and heart transcriptomes, whereas lung parenchyma and the whole brain have few sex-biased genes and TEs. We find a significant impact of sex on TE transcription. Importantly, sDETEs tend to reside in clusters with about 50% of these clusters colocalizing with other sex-biased transcripts sharing the same sex-bias.

## Methods

### Mice and crosses

C57BL/6 J mice were purchased from the Jackson Laboratory (Bar Harbor, Maine, USA).

B6.$Y^{TIR}$ consomic mice were maintained in our colony (TT) by breeding of B6.$Y^{TIR}$ males to C57BL/6 J females. B6.$Y^{TIR}$ males were crossed to wild-type C57BL/6 J females to generate $XY^{TIR}$ sex-reversed females (XY.F) and males (XY.M), as well as XX females (XX.F). Female offspring were genotyped using PCR amplification of the zinc finger protein on the Y (*Zfy*) sequence of DNA from ear punches, using the primers and conditions described in [91]. Liver (4 XX.F, 4 XY.M, and 4XY.F), lung parenchyma (3 XX.F, 3 XY.M, and 4XY.F), heart (3 XX.F, 3 XY.M, and 4XY.F), and whole brain (4 XX.F, 4 XY.M, and 4XY.F) were collected from 8-week old mice and used for RNA extraction. Gonadal sex was confirmed at the time of organ collection.

B6.C3H/HeSn-Paf mice (referred to as *Paf* from this point on) were generated by backcrossing C3H/HeSn-Paf/J carriers of the patchy fur (*Paf*) mutation purchased from the Jackson Laboratory to C57BL/6 J mice (Jackson Laboratory, Bar Harbor, Maine, USA). Males that carry the *Paf* mutation were identified based on their hair loss phenotype and crossed to C57BL/6 J females. Female offspring from these crosses were genotyped using RT-PCR for the *Xist* gene, which is expressed in XX females ($XX^{Paf}$.F) but not in XO females (XO.F) [92]. Liver (5 $XX^{Paf}$.F and 4 XO.F), lung parenchyma (3 $XX^{Paf}$.F and 3 XO.F), heart (4 $XX^{Paf}$.F and 3 XO.F), whole brain (3 $XX^{Paf}$.F and 3 XO.F) samples from 8-week old N6 and N7 XO females (XO.F) and their $XX^{Paf}$ ($XX^{Paf}$.F) littermates were collected and used for RNA extraction.

To reduce the variability that could arise from circadian rhythmicity, samples were collected always on the same time of the day between Zeitgeber time (ZT) 6 and ZT8.

All procedures were conducted in accordance with the guidelines set by the Canadian Council on Animal Care (Ottawa, Ontario, Canada) and were approved by the Animal Care Committee of the McGill University Health Center (Montreal, Quebec, Canada).

### Genotyping

DNA was extracted from mouse tissues using the DNeasy Blood & Tissue Kit (catalog # 69,596, QIAGEN, Valencia, CA) and shipped to Neogen Inc (Lincoln, NE) for genotyping using the MiniMUGA array [68]. Chromosomal sex was determined as originally described [68]. Note that MiniMUGA can accurately discriminate between XX, XY, XO, and XXY complements independently of genetic background. The contribution of different inbred strains to the genetic make-up of each sample was estimated as the percent of

the autosomal and X chromosome genome using exclusively markers that were informative for these backgrounds. IBD, or identity by descent represents regions of the genome where MiniMUGA is unable to robustly discriminate between the relevant genetic backgrounds.

#### RNA extraction and sequencing

Total RNA was extracted from mouse organs using Trizol Reagent (Thermo Fisher Scientific, MA, US) and purified using the RNeasy MinElute Cleanup Kit (Qiagen, NL).

Library preparation and RNA-sequencing (RNA-seq) were performed at the McGill University and Genome Quebec Innovation Centre for the liver and at the McGill Genome Centre for the brain, lung, and heart tissues. Paired-end sequencing was performed using a NovaSeq6000 S4 sequencer. For liver comparisons, previously generated RNA-seq data were used [21].

#### Data preprocessing

Data preprocessing and alignment was performed using the GenPipes RNA-seq pipeline [93], with default parameters applied unless mentioned otherwise. Reads were trimmed and filtered for quality, then aligned to the mouse reference genome (GRCm38) using STAR [94], and transcript abundance was estimated using HT-Seq Count (version 0.11.0 for brain, 0.6.0 for liver) [95] for gene expression. We applied the default setting of STAR [94] to retrieve only properly paired alignments and both parameters –winAnchorMultimapNmax and – outAnchorMultimapNmax were set as 100 for the optimization of assigning multi-mapped reads. Alignment BAM files were input to package TEtranscripts (version 2.0.3) [96] to obtain read counts at a TE subfamily level. TE annotation files were retrieved from (https://labshare.cshl.edu/shares/mhammelllab/www-data/TEtranscripts/TE_GTF/mm10_rmsk_TE.gtf.gz). The raw count of the TE subfamily was defined as the global sum of reads aligned to all TE instances of each subfamily. Options [-b] and [–sortByPos] were applied to run *TEcount* in the TEtranscripts package. Default parameters were applied if not mentioned otherwise. To obtain TE read counts at a locus/instance level, alignment BAM files were input to package TElocal (version 1.1.1) (https://github.com/mhammell-laboratory/TElocal). Options [-b], [–sortByPos], and default number of iterations [–iteration 100] that optimized multi-reads assignment were applied. Genes and TEs with extremely low read counts (average < 5 in all samples) were excluded to reduce noise.

#### De novo *transcriptome assembly and identification of* de novo *lncRNAs*

To obtain both known and novel transcripts in each organ, we first conducted de novo assemblies for the four organs separately. Briefly, alignment output files (i.e. sorted bam files) from STAR [94] of the GenPipes RNA-seq pipeline [93] were merged for all samples within each organ, followed by a reference-guided assembly to call transcript with Stringtie (version 2.1.4) [97]. Output files from GTF files from Stringtie (109,830 brain transcripts, 67,432 heart transcripts, 86,237 lung transcripts, and 107,807 liver transcripts) contained both known and novel transcripts. Next, the list of transcripts (GTF files) from each organ was used as input for FEElnc (version 0.1–0) [98], a computational tool used to predict lncRNAs. FEElnc identified novel transcripts that were longer than

200 bp and did not overlap with annotated genes of the reference genome (GRCm38. Ensembl102.gtf) [99], then FEELnc predicted coding potentials of input transcripts with a Random Forest model trained with intrinsic properties of user-input mRNA sequence (such as mRNA sizes, k-mer frequencies, and ORF coverage). Output files of FEElnc included lists of predicted lncRNAs and mRNAs. We curated the list of lncRNA to include the ones with a minimum of 50 aligned reads as the estimated de novo lncRNA for our dataset.

#### PCA and heatmap

Principal component analysis (PCA) was performed for the top 300 genes with the highest variance of expression level, for the top 300 lncRNA, and for the top 300 most TE subfamilies or TE instances. Heatmap was generated with ComplexHeatmap (v3.10 [100]). To visualize relative expression levels of TE subfamilies across organs, organ-relative expression z-score was calculated. To calculate the *z*-score, mean and squared root (sqrt) of variance were firstly calculated for each TE subfamily across four organs. *Z*-score = (expression of TE subfamily − mean)/sqrt(variance) was calculated for each TE subfamily.

#### Organ-biased TE analysis

To identify TE subfamily with higher expression in one specific organ but lower in all other three organs, we applied organ-relative expression *z*-score (*z*-score > 1) and Tao Index (τ > 0.6, [69]. *Z*-score > 1 was used to select TE subfamilies with relatively higher expression among the four organs, while τ > 0.6 was applied to underscore high expression in only one organ. The Tao Index $\tau$ threshold was tuned to best reflect organ differences with both good sensitivity and specificity (Additional file 3: Fig. S11). To calculate $\tau$, the mean value μ was calculated in each organ for each TE subfamily. The maximum value (*Max*) across four organs was also calculated for each TE subfamily and the Tao Index $\tau$ for each TE subfamily was then calculated with the following formula (modified from [69]). An additional parameter (average expression level across all four organs: baseMean > 500) was also applied to inspect the subset of highly expressed organ-biased TE subfamilies. Organ-biased TE instances within such highly expressed organ-biased TE subfamilies were also identified with parameters *z*-score > 1 and τ > 0.6 and baseMean > 100. We used baseMean > 100 (instead of baseMean > 500) because individual TE instances showed overall lower aligned reads than all instances combined for the TE subfamily level.

$$\tau = \frac{(Max - \mu_{liver})/Max + (Max - \mu_{brain})/Max + (Max - \mu_{heart})/Max + \left(Max - \mu_{lung}\right)/Max}{4 - 1}$$

#### Differential analysis

For all four investigated organs, sDEGs, slncRNAs, and sDETEs were identified using the DESeq2 (v.1.20.0) [101] and EdgeR (version 3.22.5) [102] packages were used. The ct-off threshold for sDEGs was set as $|\log FC| \geq 1.0$, $\log CPM \geq 1.0$, adjusted *p*-value < 0.05. Only lncRNAs and DETEs with a number of aligned reads > 50 were kept in the preprocessing step. Cut-off thresholds for both were set as $|\log FC| \geq 1.0$, adjusted *p*-value < 0.05.

#### Genic annotation of sDETE instances

The genic annotation was performed using the R package GenomicFeatures (v1.38.2, [103]). The genic annotations included 1–5 kb upstream of the transcription start site (Upstream 1–5 kb), the promoter (< 1 kb upstream of TSS), introns, and exons.

#### Visualization of the location of sDETE instances

Visualization of the chromosomal location of sDETE, sDEG, and slncRNAwas performed with R package chromoMap (version 4.1.1) [104].

#### Cumulative distribution of distances between closest sex-biased signals

Distance between the sex-biased signal and the nearest peer or signal from another sub-group was calculated using the bedtools closest (v2.29.2, [105]).

Cumulative distribution of the distance between sDETEs/sDEG/slncRNAs and the nearest sDETE/sDEG/slncRNA was calculated for the liver comparison XX.F vs XY.M.Z. To generate a random background group for sDEGs, we randomly selected 200 genes, 2000 TEs, or 2000 lncRNAs and calculated the distance between the selected genes to the nearest sDEG/sDETE/slncRNA. The generation process was repeated 10,000 times to build a background group.

#### Cluster identification

Clusters were identified for liver sDETEs for the comparison XX.F vs XY.M. Nearby sDETEs (two or more) were merged into a sDETE cluster if their distance to each other was < 100 kb. The cutoff threshold of 100 kb was chosen based on the cumulative distribution of distance between nearby sex-biased signals (as shown in Fig. 3A).

#### Cumulative distribution of distances between sDETE and transcript

To obtain the cumulative distribution of distance between 1422 sDETE and the nearest transcript in the liver, we first split sDETEs and transcripts into forward (+) and reverse (−) strands. We then used bedtools closest (v2.29.2, [105]) to calculate the distance between sDETE and liver transcripts from the same strand as well as between the ones from opposite strands. Lastly, we plotted the cumulative distribution of distance between sDETE and the nearest transcript on the same or opposite strand, respectively.

#### Identification of putative initiating sDETE

To identify liver putative initiating sDETEs, we first collected 1422 sDETEs the XX.F vs XY.M comparison in the liver. We then identified the nearest liver transcript TSS on the same strand for each sDETE with bedtools closest (v2.29.2, [105]). Among those, putative initiating sDETEs were defined as the ones overlapped with TSS of the nearest liver transcript or within 2 kb upstream of TSS. The offset value of 2 kb was an empirical choice, based on the distribution shown in Additional file 3: Fig. S6B, to handle the challenge of imperfect annotation of TE coordinates, especially for truncated TEs.

### Statistical analysis of TE class/family/subfamily enrichment

To evaluate the enrichment of TE class/family/subfamily in sDEG-proximal and sDEG-distal sDETEs, we conducted a hypergeometric test [106]. We evaluated the enrichment of TE class/family/subfamily of 1422 sDETEs from the liver XX.F vs XY.M comparison compared to all reference TEs. In addition, we also performed a hypergeometric test to evaluate the enrichment of TE class and family in putative initiating sDETEs compared to all sDETEs from the liver XX.F vs XY.M comparison.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03547-0.

Additional file 1: Table S1. Genetic background of animals from experimental crosses.

Additional file 2: Table S2. Identification of de novo LncRNAs in four organs.

Additional file 3: Fig. S1-10.

Additional file 4: Table S3. Organ-biased TE subfamilies and instances.

Additional file 5: Table S4. sDEGs in four organs.

Additional file 6: Table S5. sIncRNAs in four organs.

Additional file 7: Table S6. sDETEs in four organs.

Additional file 8: Table S7. Biased expression of X-linked genes in four organs.

Additional file 9: Table S8. Putative initiating sDETEs in liver.

Additional file 10: Table S9. sDETE clusters in liver comparisons XX.F vs XY.M and XY.F vs XY.M.

**Acknowledgements**

We are grateful for Pablo Hock, Timothy A Bell and Matt Blanchard for help processing DNA samples for genotyping. This research was enabled in part by support provided by Calcul Quebec (calculquebec.ca) and the Digital Research Alliance of Canada (alliancecan.ca).

**Peer review information**

Tim Sands was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

**Authors' contributions**

QKWZ drafted parts of the manuscript and generated most figures, performed RNA-seq analyses; KB collected material, conducted RNA work, analyzed part of the RNA-seq data and drafted the earliest version of the manuscript; JHG contributed to RNA-seq analysis; NA and EB performed tissue collection and RNA work; FPMdV oversaw MiniMUGA genotyping; TT generated the mouse models; GB oversaw the bioinformatics aspects of the project, contributed to concept development and edited the manuscript; AKN conceived and oversaw the project, writing and assembly of the manuscript. All authors edited the manuscript.

**Funding**

The study was supported by funds from the Natural Sciences and Engineering Research Council of Canada Discovery Grant program and Discovery Accelerator Supplement (to AKN) and in part by a Canada Institute of Health Research (CIHR) program grant (CEE-151618) for the McGill Epigenomics Mapping Center, which is part of the Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC) Network (to GB). G.B. is supported by a Canada Research Chair Tier 1 award and a FRQ-S Distinguished Research Scholar award.
QKWZ is a recipient of a Doctoral Training Scholarship from Fonds de recherche du Québec (FRQS) and Japan Science and Technology Agency Support for Pioneering Research Initiated by the Next Generation (JST SPRING). NA is a recipient of the RI MUHC graduate scholarship.

**Data availability**

Scripts and pre-processed files to generate figures for the paper are uploaded on GitHub (https://github.com/qzhuang8/sex-biased-TEs-paper-script), licensed under the MIT License (MIT) [107]. Bigwig and bed files for UCSC genome browser visualization of tracks are available publicly on Zenodo (https://zenodo.org/records/14955055) and licensed under a Creative Commons Attribution 4.0 International License [108].

## Declarations

**Ethics approval and consent to participate**

All procedures were conducted in accordance with the guidelines set by the Canadian Council on Animal Care (Ottawa, Ontario, Canada) and were approved by the Animal Care Committee of the McGill University Health Center (Montreal, Quebec, Canada).

**Competing interest**
The authors declare no competing interests.

Received: 30 May 2024 Accepted: 17 March 2025
Published online: 26 March 2025

**References**

1. Rubin JB, Lagas JS, Broestl L, Sponagel J, Rockwell N, Rhee G, Rosen SF, Chen S, Klein RS, Imoukhuede P, Luo J. Sex differences in cancer mechanisms. Biol Sex Differ. 2020;11:17.
2. Koti M, Ingersoll MA, Gupta S, Lam CM, Li X, Kamat AM, Black PC, Siemens DR. Sex differences in bladder cancer immunobiology and outcomes: a collaborative review with implications for treatment. Eur Urol Oncol. 2020;3:622–30.
3. Mattiuzzi C, Lippi G. Current cancer epidemiology. J Epidemiol Glob Health. 2019;9:217–22.
4. Wilkinson NM, Chen HC, Lechner MG, Su MA. Sex differences in immunity. Annu Rev Immunol. 2022;40:75–94.
5. Jackson SS, Marks MA, Katki HA, Cook MB, Hyun N, Freedman ND, Kahle LL, Castle PE, Graubard BI, Chaturvedi AK. Sex disparities in the incidence of 21 cancer types: Quantification of the contribution of risk factors. Cancer. 2022;128:3531–40.
6. Burgoyne PS, Arnold AP. A primer on the use of mouse models for identifying direct sex chromosome effects that cause sex differences in non-gonadal tissues. Biol Sex Differ. 2016;7:68.
7. Arnold AP, Cassis LA, Eghbali M, Reue K, Sandberg K. Sex hormones and sex chromosomes cause sex differences in the development of cardiovascular diseases. Arterioscler Thromb Vasc Biol. 2017;37:746–56.
8. Arnold AP. Sexual differentiation of brain and other tissues: Five questions for the next 50 years. Horm Behav. 2020;120: 104691.
9. Waxman DJ, O'Connor C. Growth hormone regulation of sex-dependent liver gene expression. Mol Endocrinol. 2006;20:2613–29.
10. Lau-Corona D, Suvorov A, Waxman DJ. Feminization of male mouse liver by persistent growth hormone stimulation: activation of sex-biased transcriptional networks and dynamic changes in chromatin states. Mol Cell Biol. 2017;37(19):e00301-17.
11. Blencowe M, Chen X, Zhao Y, Itoh Y, McQuillen CN, Han Y, Shou BL, McClusky R, Reue K, Arnold AP, Yang X. Relative contributions of sex hormones, sex chromosomes, and gonads to sex differences in tissue gene regulation. Genome Res. 2022;32:807–24.
12. Reizel Y, Spiro A, Sabag O, Skversky Y, Hecht M, Keshet I, Berman BP, Cedar H. Gender-specific postnatal demethylation and establishment of epigenetic memory. Genes Dev. 2015;29:923–33.
13. AlOgayil N, Bauermeister K, Galvez JH, Venkatesh VS, Zhuang QK, Chang ML, Davey RA, Zajac JD, Ida K, Kamiya A, et al. Distinct roles of androgen receptor, estrogen receptor alpha, and BCL6 in the establishment of sex-biased DNA methylation in mouse liver. Sci Rep. 2021;11:13766.
14. Hao P, Waxman DJ. STAT5 regulation of sex-dependent hepatic CpG methylation at distal regulatory elements mapping to sex-biased genes. Mol Cell Biol. 2021;41:e00166–e00120.
15. Conforto TL, Zhang Y, Sherman J, Waxman DJ. Impact of CUX2 on the female mouse liver transcriptome: activation of female-biased genes and repression of male-biased genes. Mol Cell Biol. 2012;32:4611–27.
16. Wijchers PJ, Yandim C, Panousopoulou E, Ahmad M, Harker N, Saveliev A, Burgoyne PS, Festenstein R. Sexual dimorphism in mammalian autosomal gene regulation is determined not only by Sry but by sex chromosome complement as well. Dev Cell. 2010;19:477–84.
17. Kido T, Lau YF. Roles of the Y chromosome genes in human cancers. Asian J Androl. 2015;17:373–80.
18. Tricarico R, Nicolas E, Hall MJ, Golemis EA. X- and Y-linked chromatin-modifying genes as regulators of sex-specific cancer incidence and prognosis. Clin Cancer Res. 2020;26:5567–78.
19. Cabrera Zapata LE, Cisternas CD, Sosa C, Garcia-Segura LM, Arevalo MA, Cambiasso MJ. X-linked histone H3K27 demethylase Kdm6a regulates sexually dimorphic differentiation of hypothalamic neurons. Cell Mol Life Sci. 2021;78:7043–60.
20. AlSiraj Y, Chen X, Thatcher SE, Temel RE, Cai L, Blalock E, Katz W, Ali HM, Petriello M, Deng P, et al. XX sex chromosome complement promotes atherosclerosis in mice. Nat Commun. 2019;10:2631.
21. Zhuang QK, Galvez JH, Xiao Q, AlOgayil N, Hyacinthe J, Taketo T, Bourque G, Naumova AK. Sex chromosomes and sex phenotype contribute to biased DNA methylation in mouse liver. Cells. 2020;9:1436.
22. Zhang Y, Laz EV, Waxman DJ. Dynamic, sex-differential STAT5 and BCL6 binding to sex-biased, growth hormone-regulated genes in adult mouse liver. Mol Cell Biol. 2012;32:880–96.
23. Chikada H, Ida K, Ando E, Inagaki Y, Sakamoto A, Kamiya A. Establishment and analysis of a mouse model that regulates sex-related differences in liver drug metabolism. Lab Invest. 2018;98:1500–11.
24. Lau-Corona D, Bae WK, Hennighausen L, Waxman DJ. Sex-biased genetic programs in liver metabolism and liver fibrosis are controlled by EZH1 and EZH2. PLoS Genet. 2020;16: e1008796.
25. Cheng MI, Li JH, Riggan L, Chen B, Tafti RY, Chin S, Ma F, Pellegrini M, Hrncir H, Arnold AP, et al. The X-linked epigenetic regulator UTX controls NK cell-intrinsic sex differences. Nat Immunol. 2023;24:780–91.
26. Itoh Y, Golden LC, Itoh N, Matsukawa MA, Ren E, Tse V, Arnold AP, Voskuhl RR. The X-linked histone demethylase Kdm6a in CD4+ T lymphocytes modulates autoimmunity. J Clin Invest. 2019;129:3852–63.
27. Deegan DF, Karbalaei R, Madzo J, Kulathinal RJ, Engel N. The developmental origins of sex-biased expression in cardiac development. Biol Sex Differ. 2019;10:46.
28. Ho B, Greenlaw K, Al Tuwaijri A, Moussette S, Martinez F, Giorgio E, Brusco A, Ferrero GB, Linhares ND, Valadares ER, et al. X chromosome dosage and presence of SRY shape sex-specific differences in DNA methylation at an autosomal region in human cells. Biol Sex Differ. 2018;9:10.

29. Richardson V, Engel N, Kulathinal RJ. Comparative developmental genomics of sex-biased gene expression in early embryogenesis across mammals. Biol Sex Differ. 2023;14:30.
30. Skakkebaek A, Nielsen MM, Trolle C, Vang S, Hornshoj H, Hedegaard J, Wallentin M, Bojesen A, Hertz JM, Fedder J, et al. DNA hypermethylation and differential gene expression associated with Klinefelter syndrome. Sci Rep. 2018;8:13740.
31. Trolle C, Nielsen MM, Skakkebaek A, Lamy P, Vang S, Hedegaard J, Nordentoft I, Orntoft TF, Pedersen JS, Gravholt CH. Widespread DNA hypomethylation and differential gene expression in Turner syndrome. Sci Rep. 2016;6:34220.
32. Werner RJ, Schultz BM, Huhn JM, Jelinek J, Madzo J, Engel N. Sex chromosomes drive gene expression and regulatory dimorphisms in mouse embryonic stem cells. Biol Sex Differ. 2017;8:28.
33. Chojnowski JL, Braun EL. An unbiased approach to identify genes involved in development in a turtle with temperature-dependent sex determination. BMC Genomics. 2012;13: 308.
34. Melia T, Waxman DJ. Sex-Biased lncRNAs Inversely Correlate With Sex-Opposite Gene Coexpression Networks in Diversity Outbred Mouse Liver. Endocrinology. 2019;160:989–1007.
35. Goldfarb CN, Waxman DJ. Global analysis of expression, maturation and subcellular localization of mouse liver transcriptome identifies novel sex-biased and TCPOBOP-responsive long non-coding RNAs. BMC Genomics. 2021;22:212.
36. Goldfarb CN, Karri K, Pyatkov M, Waxman DJ. Interplay between GH-regulated, Sex-biased Liver transcriptome and hepatic zonation revealed by single-nucleus RNA sequencing. Endocrinology. 2022;163:bqac059.
37. López-Royo T, Moreno-Martínez L, Moreno-García L, Calvo AC, Manzano R, Osta R. Sex differences on constitutive long non-coding RNA expression: modulatory effect of estradiol and testosterone in muscle cells. Andrology. 2024;12(8):1887–96.
38. Schmid CW, Jelinek WR. The Alu family of dispersed repetitive sequences. Science. 1982;216:1065–70.
39. Boeke JD, Garfinkel DJ, Styles CA, Fink GR. Ty elements transpose through an RNA intermediate. Cell. 1985;40:491–500.
40. Howard BH, Sakamoto K. Alu interspersed repeats: selfish DNA or a functional gene family? New Biol. 1990;2:759–70.
41. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. 1993;72:595–605.
42. Jachowicz JW, Bing X, Pontabry J, Bošković A, Rando OJ, Torres-Padilla ME. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat Genet. 2017;49:1502–10.
43. Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, Biechele S, Huang B, Shen X, Ramalho-Santos M. A LINE1-nucleolin partnership regulates early development and ESC Identity. Cell. 2018;174:391-405.e319.
44. Fadloun A, Le Gras S, Jost B, Ziegler-Birling C, Takahashi H, Gorab E, Carninci P, Torres-Padilla ME. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nat Struct Mol Biol. 2013;20:332–8.
45. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature. 2015;522:221–5.
46. Jansz N. DNA methylation dynamics at transposable elements in mammals. Essays Biochem. 2019;63:677–89.
47. Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 2012;13: R107.
48. Johnson R, Guigó R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. RNA. 2014;20:959–76.
49. Geisse S, Scheidereit C, Westphal HM, Hynes NE, Groner B, Beato M. Glucocorticoid receptors recognize DNA sequences in and around murine mammary tumour virus DNA. Embo j. 1982;1:1613–9.
50. Scheidereit C, Geisse S, Westphal HM, Beato M. The glucocorticoid receptor binds to defined nucleotide sequences near the promoter of mouse mammary tumour virus. Nature. 1983;304:749–52.
51. Hsu K, Lee YK, Chew A, Chiu S, Lim D, Greenhalgh DG, Cho K. Inherently variable responses to glucocorticoid stress among endogenous retroviruses isolated from 23 mouse strains. Biochim Biophys Acta Mol Basis Dis. 2017;1863:2594–600.
52. Beato M, Wright RHG, Dily FL. 90 YEARS OF PROGESTERONE: Molecular mechanisms of progesterone receptor action on the breast cancer genome. J Mol Endocrinol. 2020;65:T65-t79.
53. Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004;431:96–9.
54. Tharp ME, Malki S, Bortvin A. Maximizing the ovarian reserve in mice by evading LINE-1 genotoxicity. Nat Commun. 2020;11:330.
55. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, Santamarina M, Ju YS, Temes J, Garcia-Souto D, et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet. 2020;52:306–19.
56. Mullins CS, Linnebacher M. Human endogenous retroviruses and cancer: causality and therapeutic possibilities. World J Gastroenterol. 2012;18:6027–35.
57. Babaian A, Mager DL. Endogenous retroviral promoter exaptation in human cancer. Mob DNA. 2016;7:24.
58. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. Ten things you should know about transposable elements. Genome Biol. 2018;19:199.
59. Fueyo R, Judd J, Feschotte C, Wysocka J. Roles of transposable elements in the regulation of mammalian transcription. Nat Rev Mol Cell Biol. 2022;23:481–97.
60. Blewitt ME, Vickaryous NK, Hemley SJ, Ashe A, Bruxner TJ, Preis JI, Arkell R, Whitelaw E. An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. Proc Natl Acad Sci U S A. 2005;102:7629–34.

61. Juriloff DM, Harris MJ. Hypothesis: the female excess in cranial neural tube defects reflects an epigenetic drag of the inactivating X chromosome on the molecular mechanisms of neural fold elevation. Birth Defects Res A Clin Mol Teratol. 2012;94:849–55.
62. Brown EJ, Nguyen AH, Bachtrog D. The Y chromosome may contribute to sex-specific ageing in Drosophila. Nat Ecol Evol. 2020;4:853–62.
63. Taketo-Hosotani T, Nishioka Y, Nagamine CM, Villalpando I, Merchant-Larios H. Development and fertility of ovaries in the B6.YDOM sex-reversed female mouse. Development (Cambridge, England). 1989;107:95–105.
64. Burgoyne PS, Evans EP. A high frequency of XO offspring from X(Paf)Y* male mice: evidence that the Paf mutation involves an inversion spanning the X PAR boundary. Cytogenet Cell Genet. 2000;91:57–61.
65. Nagamine CM, Taketo T, Koo GC. Studies on the genetics of tda-1 XY sex reversal in the mouse. Differentiation. 1987;33:223–31.
66. Cowles CR, Hirschhorn JN, Altshuler D, Lander ES. Detection of regulatory variation in mouse genes. Nat Genet. 2002;32:432–7.
67. Grimm SA, Shimbo T, Takaku M, Thomas JW, Auerbach S, Bennett BD, Bucher JR, Burkholder AB, Day F, Du Y, et al. DNA methylation in mice is influenced by genetics as well as sex and life experience. Nat Commun. 2019;10:305.
68. Sigmon JS, Blanchard MW, Baric RS, Bell TA, Brennan J, Brockmann GA, Burks AW, Calabrese JM, Caron KM, Cheney RE, et al. Content and performance of the MiniMUGA genotyping array: a new tool to improve rigor and reproducibility in mouse research. Genetics. 2020;216:905–30.
69. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005;21:650–9.
70. Larouche JD, Trofimov A, Hesnard L, Ehx G, Zhao Q, Vincent K, Durette C, Gendron P, Laverdure JP, Bonneil É, et al. Widespread and tissue-specific expression of endogenous retroelements in human somatic tissues. Genome Med. 2020;12:40.
71. Kouadjo KE, Nishida Y, Cadrin-Girard JF, Yoshioka M, St-Amand J. Housekeeping and tissue-specific genes in mouse tissues. BMC Genomics. 2007;8: 127.
72. Marques AC, Ponting CP. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. Genome Biol. 2009;10: R124.
73. Xiao SJ, Zhang C, Zou Q, Ji ZL. TiSGeD: a database for tissue-specific genes. Bioinformatics. 2010;26:1273–5.
74. Balaton BP, Cotton AM, Brown CJ. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. Biol Sex Differ. 2015;6:35.
75. Berletch JB, Ma W, Yang F, Shendure J, Noble WS, Disteche CM, Deng X. Identification of genes escaping X inactivation by allelic expression analysis in a novel hybrid mouse model. Data Brief. 2015;5:761–9.
76. Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet. 1999;23:314–8.
77. Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KV, Whitelaw E. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proc Natl Acad Sci U S A. 2003;100:2538–43.
78. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet. 2010;42:631–4.
79. Todd CD, Deniz Ö, Taylor D, Branco MR. Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. Elife. 2019;8:e44344.
80. Choudhary I, Vo T, Paudel K, Patial S, Saini Y. Compartment-specific transcriptomics of ozone-exposed murine lungs reveals sex- and cell type-associated perturbations relevant to mucoinflammatory lung diseases. Am J Physiol Lung Cell Mol Physiol. 2021;320:L99-l125.
81. Fels JA, Casalena GA, Manfredi G. Sex and oestrogen receptor β have modest effects on gene expression in the mouse brain posterior cortex. Endocrinol Diabetes Metab. 2021;4: e00191.
82. Xu J, Burgoyne PS, Arnold AP. Sex differences in sex chromosome gene expression in mouse brain. Hum Mol Genet. 2002;11:1409–19.
83. Batdorj E, AlOgayil N, Zhuang QK, Galvez JH, Bauermeister K, Nagata K, Kimura T, Ward MA, Taketo T, Bourque G, Naumova AK. Genetic variation in the Y chromosome and sex-biased DNA methylation in somatic cells in the mouse. Mamm Genome. 2023;34:44–55.
84. Berletch JB, Ma W, Yang F, Shendure J, Noble WS, Disteche CM, Deng X. Escape from X inactivation varies in mouse tissues. PLoS Genet. 2015;11: e1005079.
85. Lopes AM, Arnold-Croop SE, Amorim A, Carrel L. Clustered transcripts that escape X inactivation at mouse XqD. Mamm Genome. 2011;22:572–82.
86. Treiber CD, Waddell S. Transposon expression in the Drosophila brain is driven by neighboring genes and diversifies the neural transcriptome. Genome Res. 2020;30:1559–69.
87. Dechaud C, Miyake S, Martinez-Bengochea A, Schartl M, Volff JN, Naville M. Clustering of sex-biased genes and transposable elements in the genome of the medaka fish Oryzias latipes. Genome Biol Evol. 2021;13:evab230.
88. Robinson WH, Cotton AM, Penaherrera MS, Peeters SB, Brown CJ. X chromosome inactivation. In: Naumova AK, Greenwood CM, editors. Epigenetics and Complex Traits. New York: Springer Science + Business; 2013. p. 63–85.
89. Murphy PJ, Berger F. The chromatin source-sink hypothesis: a shared mode of chromatin-mediated regulations. Development 2023;150:dev201989.
90. Soh YQ, Alföldi J, Pyntikova T, Brown LG, Graves T, Minx PJ, Fulton RS, Kremitzki C, Koutseva N, Mueller JL, et al. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell. 2014;159:800–13.
91. Amleh A, Smith L, Chen H, Taketo T. Both nuclear and cytoplasmic components are defective in oocytes of the B6.Y(TIR) sex-reversed female mouse. Dev Biol. 2000;219:277–86.

92. Kay GF, Penny GD, Patel D, Ashworth A, Brockdorff N, Rastan S. Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. Cell. 1993;72:171–82.
93. Bourgey M, Dali R, Eveleigh R, Chen KC, Letourneau L, Fillon J, Michaud M, Caron M, Sandoval J, Lefebvre F, Leveque G. GenPipes: an open-source framework for distributed and scalable genomic analyses. Gigascience. 2019;8(6):giz037.
94. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
95. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.
96. Jin Y, Tam OH, Paniagua E, Hammell M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics. 2015;31:3593–9.
97. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5.
98. Wucher V, Legeai F, Hédan B, Rizk G, Lagoutte L, Leeb T, Jagannathan V, Cadieu E, David A, Lohi H, et al. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. Nucleic Acids Res. 2017;45: e57.
99. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. Ensembl 2020. Nucleic Acids Res. 2020;48:D682-d688.
100. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics (Oxford, England). 2016;32:2847–9.
101. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11: R106.
102. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
103. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9: e1003118.
104. Anand L, Rodriguez Lopez CM. ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. BMC Bioinformatics. 2022;23:33.
105. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
106. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics. 2007;23:257–8.
107. Zhuang Q, Bauermeister K, Gálvez JH, Alogayil N, Batdorj E, Pardo Manuel de Villena F, Taketo T, Bourque G, Naumova AK. Sex-biased-TEs-paper-script. Github. 2025. https://github.com/qzhuang8/sex-biased-TEs-paper-script?tab=MIT-0-1-ov-file.
108. Zhuang Q, Bauermeister K, Gálvez JH, Alogayil N, Batdorj E, Pardo Manuel de Villena F, Taketo T, Bourque G, Naumova AK: Survey of gene, lncRNA and transposon transcription patterns in four mouse organs highlights shared and organ-specific sex-biased regulation. Zenodo; 2025. https://doi.org/10.5281/zenodo.14955055.


## Publisher's Note