



Published in final edited form as:

*Nat Genet.* ; 43(12): 1179–1185. doi:10.1038/ng.948.

## Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*

Xinxian Deng<sup>1</sup>, Joseph B Hiatt<sup>2</sup>, Di Kim Nguyen<sup>1</sup>, Sevinc Ercan<sup>3</sup>, David Sturgill<sup>4</sup>, LaDeana W Hillier<sup>5</sup>, Felix Schlesinger<sup>6</sup>, Carrie A Davis<sup>6</sup>, Valerie J Reinke<sup>7</sup>, Thomas R Gingeras<sup>6</sup>, Jay Shendure<sup>2</sup>, Robert H Waterston<sup>2</sup>, Brian Oliver<sup>4</sup>, Jason D Lieb<sup>8</sup>, and Christine M Disteche<sup>1,9</sup>

<sup>1</sup>Department of Pathology, University of Washington School of Medicine, Seattle, Washington, USA

<sup>2</sup>Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA

<sup>3</sup>Department of Biology and Center for Genomics and Systems Biology, New York University, New York, New York, USA

<sup>4</sup>Section of Developmental Genomics, Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, US National Institutes of Health, Bethesda, Maryland, USA

<sup>5</sup>Genome Sequencing Center, Washington University School of Medicine, Saint Louis, Missouri, USA

<sup>6</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA

<sup>7</sup>Department of Genetics, Yale University, New Haven, Connecticut, USA

<sup>8</sup>Department of Biology, Carolina Center for the Genome Sciences and Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

<sup>9</sup>Department of Medicine, University of Washington School of Medicine, Seattle, Washington, USA

© 2011 Nature America, Inc. All rights reserved.

Correspondence should be addressed to C.M.D. (cdistech@u.washington.edu), B.O. (oliver@helix.nih.gov) or J.D.L. (jlieb@bio.unc.edu).

**Accession numbers.** Human RNA-Seq data are available at EMBL-EBI (E-MTAB-513) and the GEO database (GSE26284 and GSE16921). Mouse RNA-Seq data are available at the GEO database (GSE22131) and the SRA database (SRA001030 and SRA012213), and RNA-seq data from mouse PGK12.1 ES cells are available at the GEO database (GSE30690). Mouse ChIP-chip array data for PolII-S5p are available at the GEO database (GSE30689). *C. elegans* RNA-Seq data are available from modENCODE (SRA003622 and SRA008646), and *C. elegans* expression array data are available at the GEO database (GSE20136). *D. melanogaster* RNA-seq data from the testes of wild-type and *bam* mutant organisms are available at the GEO database (GSE16960).

Note: Supplementary information is available on the Nature Genetics website.

### AUTHOR CONTRIBUTIONS

X.D., C.M.D., J.D.L. and B.O. conceived the project and wrote the manuscript. X.D., J.B.H., D.K.N., F.S., C.A.D., T.R.G., J.S., C.M.D. and B.O. analyzed the mammalian data; R.H.W., L.W.H., J.D.L., V.J.R. and S.E. analyzed the *C. elegans* data; B.O. and D.S. analyzed the *Drosophila* data; X.D., J.B.H. and J.S. performed or analyzed the RNA-seq and ChIP-seq data from mouse ES cells.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

## Abstract

Many animal species use a chromosome-based mechanism of sex determination, which has led to the coordinate evolution of dosage-compensation systems. Dosage compensation not only corrects the imbalance in the number of X chromosomes between the sexes but also is hypothesized to correct dosage imbalance within cells that is due to monoallelic X-linked expression and biallelic autosomal expression, by upregulating X-linked genes twofold (termed ‘Ohno’s hypothesis’). Although this hypothesis is well supported by expression analyses of individual X-linked genes and by microarray-based transcriptome analyses, it was challenged by a recent study using RNA sequencing and proteomics. We obtained new, independent RNA-seq data, measured RNA polymerase distribution and reanalyzed published expression data in mammals, *C. elegans* and *Drosophila*. Our analyses, which take into account the skewed gene content of the X chromosome, support the hypothesis of upregulation of expressed X-linked genes to balance expression of the genome.

---

The evolution of a specific set of sex-determining chromosomes has led, in many species, to the existence of a single X chromosome in males and two in females. Unchecked, this mechanism of sex determination would result both in unequal X-linked gene expression between the sexes and in a naturally occurring ‘X aneuploidy’ in males. Chromosomal aneuploidy, caused by loss or gain of large regions of the genome, results in disturbances in the overall balance of gene expression networks, leading to reduced fitness and mortality<sup>1,2</sup>. Thus, mechanisms to maintain balanced expression of the genome are vital<sup>3</sup>. To compensate for haploinsufficiency of X-linked genes in males and to restore equal expression between the sexes, dosage-compensation mechanisms evolved, which differ between organisms. In *D. melanogaster*, expression of most X-linked genes is increased about twofold selectively in males to achieve balanced expression with the autosomes and to equalize expression with females<sup>4,5</sup>. In mammals and *C. elegans*, two processes have evolved: (i) a mechanism to equalize X-linked gene expression between the sexes by either silencing one X chromosome in mammalian females (X inactivation)<sup>6</sup> or repressing both X chromosomes in *C. elegans* hermaphrodites<sup>7</sup> and (ii) a mechanism hypothesized to upregulate expression from the X chromosome in both sexes<sup>8,9</sup>. Whereas the doubling of X-linked gene expression in *Drosophila* has been well documented, evidence for X-linked gene upregulation in mammals and *C. elegans* is relatively recent. Measurements of *Ctcf4-2* expression in different species of mice and analyses of expression microarray data provided the first evidence for the upregulation of genes on the X chromosome<sup>8-10</sup>. In XX or XY cells, the ratio of X-linked to autosomal expression as measured by microarrays was ~1. This held true in several tissues from multiple mammalian species, as well as in *C. elegans* and *Drosophila*<sup>8,9</sup>.

A recent study concluded that, in contrast to expression array data, RNA sequencing data do not support dosage compensation of the active X chromosome in mammals and adult *C. elegans*, thereby rejecting Ohno’s hypothesis<sup>11</sup>. Based on reanalyses of published RNA-seq data, the authors calculated that X:autosome (X:A) median expression ratios were ~0.5 in humans and 0.3 in mice, whereas in *C. elegans*, ratios declined from 1.0 in embryos to 0.4 in adults. However, as shown here, these ratios are strongly influenced by the preponderance of reproduction-related genes on the X chromosome that are silent in somatic tissues<sup>12</sup> and by special aspects of X chromosome regulation in germ cells<sup>13-15</sup>. Our reanalyses of published RNA-seq data and data from new experiments, which take into consideration the skewed gene content and regulation of the X chromosome, uphold the idea that there is compensation of gene expression between the X chromosome and autosomes in both sexes of mammals, *C. elegans* and *Drosophila*.

## RESULTS

### RNA-seq data confirm X upregulation in mammals

To determine whether X-linked genes were expressed at twice the level of autosomal genes per active allele in mammals, we first investigated whether the X chromosome and autosomes had similar expression profiles. We suspected that the skewed gene content of the mammalian X chromosome would affect the distribution of gene expression because reproduction-related genes have low or no expression in somatic tissues<sup>16–18</sup>. The distribution of X-linked and autosomal gene expression was determined in multiple human and mouse tissues and cell lines by the reanalysis of a subset of the same RNA-seq data sets analyzed previously<sup>11</sup>, in addition to the evaluation of newly released RNA-seq data sets (Supplementary Table 1). In all human and mouse samples examined, the frequency of genes with no expression (0 fragments per kilobase of exon per million mapped fragments (FPKM)) was significantly higher on the X chromosome than on autosomes ( $P < 0.05$ , by Fisher's exact test; Supplementary Fig. 1a and Supplementary Table 2). As demonstrated in the following section, many of these are multicopy genes expressed in the testis.

The distributions of gene expression were similar between the X chromosome and autosomes in brain and five other human tissues ( $P > 0.05$ , by Kolmogorov-Smirnov test), even when including all genes with any evidence of expression ( $>0$  FPKM) (Fig. 1a, Supplementary Fig. 1b and Supplementary Table 2). This finding was also demonstrated by a computational doubling of X-linked expression values, which produced a theoretical curve shifted far to the right of autosomal expression values, rather than the overlapping curves that would be anticipated if expression on the X chromosome was not upregulated (Fig. 1a). A cutoff of  $>0$  FPKM includes some genes that are actually silent but that record reads as a result of biological noise or noise from the process of sequencing and read-mapping<sup>19</sup>. For genes with more robust expression ( $\geq 1$  FPKM), the distributions of X-linked and autosomal expression became statistically indistinguishable in 12 of 16 human tissues, consistent with the upregulation of expressed X-linked genes ( $P > 0.05$ , by Kolmogorov-Smirnov test; Supplementary Table 2). Additional RNA-seq data from four human ENCODE cell lines confirmed results obtained in human tissues (Supplementary Fig. 2a and Supplementary Table 2). Computational doubling of X-linked expression values also shifted the X-chromosome curve to the right of the autosomal curve. We then compared gene expression from individual chromosomes. Box plots showed no significant differences between the X chromosome and each human autosome ( $P > 0.05$ , one-way ANOVA test for genes with  $>0$  FPKM; Fig. 1b). In addition, the average distribution of X-linked expression in 41 human lymphoblastoid cell lines<sup>20</sup> was not significantly different from autosomal expression ( $P > 0.05$  for 15 of 22 autosomes, by Kolmogorov-Smirnov test for genes with  $>0$  FPKM; Fig. 1c).

The calculated X:A median expression ratios based on RNA-seq data were strongly dependent on inclusion or removal of genes with no or low expression (Fig. 1d and Supplementary Table 2). Comparisons between 41 male and female lymphoblastoid cell lines<sup>20</sup> also revealed variability between individual samples, albeit no apparent significant sex differences ( $P > 0.05$ , by Student's *t*-test; Fig. 1d and Supplementary Fig. 2b). Given the skewed X:A distribution of genes with spatially or temporally restricted expression, for example reproduction-related genes, it is clear that X:A median expression ratios are distorted and that a better approach is to compare complete distributions of X-linked and autosomal gene expression, as shown above (Fig. 1). The advantage of using distributions was also seen by calculating X:A median expression ratios for genes separated into 16 bins, each containing the same number of X-linked genes (33–59, dependent on tissue) and autosomal genes (933–1,396) (Fig. 1e and Supplementary Fig. 2b). Note that brain has the highest X:A median expression ratio, as previously reported<sup>9</sup>. For all tissues, the X:A

median expression ratios were variable for bins containing genes with low expression and gradually increased from ~0.7 to ~1.0 for bins containing genes with expression measurements  $\geq 1$  FPKM. Thus, the inclusion of non-expressed or very weakly expressed genes that result from the skewed gene content of the X chromosome masks the upregulation of X-linked genes.

Likewise, in mouse tissues, the distributions of X-linked and autosomal expression were similar for genes with  $\geq 1$  FPKM in brain and for genes with  $\geq 1$  FPKM in other tissues ( $P > 0.05$ , by Kolmogorov-Smirnov test; Fig. 2a, Supplementary Fig. 3 and Supplementary Table 2). There was no significant difference between the X chromosome and 14 of 19 mouse autosomes ( $P > 0.05$ , one-way ANOVA test for genes with  $> 0$  FPKM; Fig. 2b). In mouse as well, the calculated X:A median expression ratios based on RNA-seq data were strongly dependent on the inclusion or removal of genes that were expressed weakly or not expressed (Fig. 2c). The combined analyses of human and mouse tissues and cell lines not only reveal a high frequency of X-linked genes with no or low expression but also support the idea that most expressed genes on the single active X chromosome are upregulated to achieve a similar level of expression to that of autosomes present in two copies.

### The mammalian X chromosome is enriched in tissue-specific genes

Tissue-specific genes implicated in sexual reproduction have accumulated on the mammalian X chromosome. The nonrandom distribution of genes with gonad and brain expression would have a profound effect on the evaluation of Ohno's hypothesis. Reproduction-related genes on the human X chromosome include 13 multicopy families of cancer-testis antigen genes and represent at least 10% of the X-linked genes<sup>17,21,22</sup>. Genes expressed in brain are also abundant on the human X chromosome<sup>23–25</sup>. On the mouse X chromosome, about 18% of genes are included in 33 multicopy gene families representing a total of 273 genes that are mainly expressed in postmeiotic male germ cells<sup>18</sup>. Genes expressed in premeiotic male germ cells and somatic cells of the testis, as well as in female-specific tissues such as placenta and the ovary, also predominate on the mouse X chromosome<sup>16,26</sup>. Thus, it was not unexpected that the two tissues that had the lowest percentage of non-expressed X-linked genes were testis (16%) and brain (22%), as compared to other tissues (23–40%) ( $P < 4 \times 10^{-7}$ , by Fisher's exact test between testis and other tissues; Supplementary Fig. 1a and Supplementary Table 2).

In human testis, the distributions of gene expression remained significantly different between the X chromosome and autosomes, even for well-expressed genes with  $> 2$  FPKM ( $P = 1 \times 10^{-4}$ , by Kolmogorov-Smirnov test; Supplementary Table 2). This finding suggested that a specific set of genes is expressed in the testis. We conducted pairwise analyses to compare the testis to other human tissues by plotting the distribution of testis expression for the subsets of X-linked and autosomal genes that were not expressed (0 FPKM) in a given somatic tissue (liver, adrenal gland and lung). A higher proportion of these genes were X linked rather than autosomal, confirming the preponderance of reproduction-related genes on the X chromosome ( $P < 1 \times 10^{-7}$ , by Kolmogorov-Smirnov test; Fig. 1f and Supplementary Fig. 4). For example, of 3,066 genes with 0 FPKM in human liver and  $> 0$  FPKM in human testis, 192 were X linked (22% of the total number of X-linked genes assayed) and 2,874 were autosomal (14% of the total number of autosomal genes assayed) ( $P = 4 \times 10^{-9}$ , by Fisher's exact test; Supplementary Table 3). Comparisons between somatic tissues did not show this trend (Supplementary Fig. 4). Six of the top ten functional categories (false discovery rate (FDR) score  $2 \times 10^{-8}$ ) obtained by Gene Ontology analyses of the subset of X-linked genes with no or very low expression (0–0.1 FPKM) in human somatic tissues were reproduction-related. Similarly, more than 93% of known reproduction-related genes<sup>16,18</sup> were not expressed in mouse somatic tissues (Supplementary Table 4). We conclude that reproduction-related genes contribute to X-

linked genes that are not expressed or are expressed at low levels in somatic tissues of mammals. When we correct for this biased distribution, RNA-seq data are strongly consistent with the predictions of Ohno's hypothesis of similar expression levels between the X chromosome and autosomes.

### RNA PolII-S5p occupancy is higher on X-linked genes

If upregulation of the X chromosome occurs at the level of transcription, one would expect higher occupancy by active forms of RNA polymerase II on the X chromosome. To test this idea, we used an undifferentiated female embryonic stem (ES) cell line (PGK12.1) with two active X chromosomes and an X:A expression ratio of 1.4 (ref. 27). Because the X:A expression ratio should theoretically be 2 if X-linked expression were doubled by dosage-compensation mechanisms, the observed lower ratio suggests that dampening of X-linked expression occurs in PGK12.1 cells, thus allowing the survival of this particular cell line<sup>28</sup>. Most female ES cell lines lose one X chromosome, likely as a result of the deleterious effects of high X-linked expression, which is also the probable cause of lethality in mouse embryos engineered to retain two active X chromosomes<sup>29</sup>.

Chromatin immunoprecipitation combined with DNA array (ChIP-chip) analyses of PGK12.1 cells showed, in two separate experiments, that the form of RNA polymerase II that is phosphorylated at Ser5 (PolII-S5p), which is specifically associated with active transcriptional initiation<sup>30</sup>, was more highly enriched at the 5' end of genes on the X chromosome than for autosomal genes (Fig. 3a and Supplementary Fig. 5). RNA-seq analyses confirmed higher levels of expressed X-linked genes in this cell line (Fig. 3b). When we sorted genes into nine bins based on expression, X-linked genes in the upper four bins had higher PolII-S5p occupancy relative to autosomal genes (Fig. 3b). As expected, no such difference between the X chromosome and autosomes was apparent among genes with low or no expression. Scatter plots confirmed that PolII-S5p occupancy depended on expression levels (Fig. 3c) as was shown in another system<sup>31</sup>. This RNA-independent evidence strongly supports upregulation of expressed X-linked genes and provides a plausible mechanism for mediating increased expression in mammals.

### RNA-seq data in *C. elegans* somatic cells confirm X upregulation in both sexes

In *C. elegans*, Xiong *et al.* observed X:A median expression ratios of 0.92, 0.84, 0.69 and 0.41 for hermaphrodites at the L2, L3, L4 and adult stages, respectively<sup>11</sup>. These declining ratios were interpreted to indicate that the *C. elegans* X chromosomes do not undergo compensatory upregulation in later developmental stages<sup>11</sup>. However, the progressive reduction in X:A expression ratio values during development is caused by a change in the ratio of somatic to germ cells in the organism.

The anatomy of *C. elegans* has been fully described, and the precise lineage of each somatic cell, which does not vary from animal to animal, is known<sup>32–35</sup>. When the embryo hatches, there are 558 somatic cells, a subset of which undergo further division to yield approximately 700 cells by the third larval stage (L3), and a total of exactly 959 somatic cells are present in a mature hermaphrodite. Germline development commences after the animal hatches, most rapidly during the late L3 and L4 stages. There are approximately 10 germ cells by the end of L2, 100 by the end of L3, 1,000 by the end of L4 and 2,000 in a mature XX hermaphrodite adult<sup>36</sup>. Thus, the ratio of germ cells to somatic cells increases as the animal develops. This is relevant because both X chromosomes are transcriptionally silenced in germ cells<sup>13</sup>. Therefore, if the true X:A expression ratio in somatic cells were 1 (consistent with upregulation of X-linked genes to balance gene expression), we would expect measured X:A expression ratios to be approximately 0.98 in L2, 0.87 in L3, 0.48 in L4 and 0.32 in adults. These ratios are rough estimations, yet they very closely approximate

the reported ratios (0.92, 0.84, 0.69 and 0.41)<sup>11</sup>. These data are therefore consistent with known *C. elegans* biology and have no bearing on Ohno's hypothesis.

To compare expression between the X chromosome and autosomes in somatic cells, where genes on the X chromosome are known to be expressed, we performed RNA-seq on *glp-1 (q224)* XX L4 animals, which do not undergo germline proliferation. Consistent with the previous analysis<sup>11</sup>, the X:A expression ratio was 0.51 in wild-type L4 animals with a developing germline but was 0.99 in XX L4 animals lacking a germline (Table 1 and Fig. 4). This result indicates that the lower X:A expression ratio in wild-type XX animals was indeed due to the presence of the silenced X chromosomes in the germline.

Having established equalized levels of X-linked and autosomal expression in XX somatic cells throughout development (Fig. 4), we next tested whether an animal with a single X chromosome also showed equal X-linked and autosomal expression. First, we note that a wild-type XO L4 male (also subject to the germline proliferation issue described above) has an X:A expression ratio of 0.51, identical to that of an XX L4 hermaphrodite (Fig. 4 and Table 1). Second, using microarrays, we found that the mean and median X:A expression ratio in L3 animals that are karyotypically male was 0.98 (see Online Methods), indicating that X-linked genes are upregulated in XO animals.

### Proteomics data do not contradict Ohno's hypothesis

We also re-examined published mouse proteomics data<sup>11,37</sup>, again taking X-chromosome gene content into account in our analysis. We found a lower proportion of proteins encoded by X-linked genes represented in the data set (10.8%) compared to autosomal genes (17.5%). This is fully consistent with observations that a large proportion of X-linked genes are not expressed in somatic tissues. If only the top 10.8% of proteins from autosomal genes are used in calculating the X:A protein ratio, whereas all proteins from X-linked genes are included (also 10.8%, but including very low-abundance proteins), we also obtain a low X:A median protein ratio of 0.38 (ref. 11). However, if we use a threshold based on protein abundance for calculating the X:A protein ratio, in this case only considering autosomal and X-linked genes for which the corresponding proteins are found at least once in the data set, the X:A median protein ratio is 1.07 ( $P = 0.73$ , Mann-Whitney test). We conclude that the very limited mouse proteomics data available support the idea of similar protein levels from the X chromosome and autosomes.

The issue described above in the analysis of the mammalian proteomics data also applies to the prior analysis of the *C. elegans* proteomics data<sup>11</sup>. Furthermore, the *C. elegans* proteomics data were obtained from mixed-stage animals, meaning that a large proportion of the assayed cells were germ cells and subject to the same effect described above in the RNA-seq analysis. Therefore, analysis of these samples is not appropriate for testing X-linked gene upregulation.

### RNA-seq data confirm dosage compensation in *Drosophila*

As in mammals and *C. elegans*, there is biased gene content on the *Drosophila* X chromosome, such that there are fewer genes with male-biased expression on that chromosome<sup>38</sup>. Although to a lesser extent than in *C. elegans*, the *Drosophila* X chromosome is also inactivated in the late male germline<sup>14</sup>. Both of these effects reduce overall expression measured from the X chromosome. In a previous microarray study, we used sex reversal, which concomitantly results in the accumulation of mitotic germline cells, to examine X-linked expression in isolation from the effects of sex-biased gene distribution and X inactivation. We found that both somatic and germline cells have dosage compensation<sup>8</sup>.

We reanalyzed RNA-seq data from wild-type testes, which are expected to show decreased X-linked expression, and from the testes of *bam* mutants, which contain tumors of early mitotic male germ cells<sup>39</sup>. This analysis allowed us to remove the effects of X inactivation and most sex-biased expression. As expected, in wild-type testis, we found that expression of X-linked genes was significantly lower than that of autosomal genes (X:A median expression ratios 0.64–0.71;  $P < 4.4 \times 10^{-6}$ , by Kolmogorov-Smirnov test; Fig. 5). However, in testes from *bam* mutants, expression from the X chromosome was much more similar to that from autosomes (X:A median expression ratios 0.87–0.92;  $P > 0.01$  for 3 of 4 autosomes, by Kolmogorov-Smirnov test; Fig. 5). These results indicate that there is X chromosome dosage compensation in early mitotic cells in the *Drosophila* germline.

## DISCUSSION

Our reanalysis of RNA-seq data and data from new experiments are consistent with balanced expression from the X chromosome and autosomes in mammals, *C. elegans* and *Drosophila*. When considering expressed genes, there is strong evidence of X-linked gene upregulation in somatic tissues in both mammals and *C. elegans*. This finding is similar to what has been observed in somatic tissues of *Drosophila*<sup>4,5</sup>. We also confirm that there is balanced expression of the genome, even in the *Drosophila* germline that lacks the canonical compensatory MSL complex<sup>8</sup>. Furthermore, in female mouse ES cells with two active X chromosomes, we observe a greater occupancy by RNA polymerase II of expressed X-linked genes relative to autosomal genes, consistent with increased X-linked gene expression in mammals<sup>27</sup>.

We arrive at a different conclusion from Xiong *et al.*<sup>11</sup> for the following reasons. First, the X chromosomes of many species have a remarkably unusual gene content, and genes with no or very low expression in somatic tissues are especially abundant on the X chromosome, as a result of the accumulation of genes with tissue-restricted expression on the sex chromosomes (for example, testis, ovary or brain)<sup>12,16–18,26,40,41</sup>. The sex chromosomes are uniquely enriched in genes involved in sexual reproduction because of evolutionary selection mechanisms that favor location on the X or Y chromosomes<sup>12</sup>. Our reanalysis of RNA-seq data of the distribution of gene expression in human and mouse tissues and cell lines confirm that the X chromosome is enriched in genes with low expression in somatic tissues but high expression in reproduction-related organs, especially the testis. We propose that any dosage-compensation mechanism to equalize the output between the X chromosome and autosomes is expected to target only genes that are appreciably expressed, as there would be no need for such a mechanism to work on genes that are silent or strongly repressed under the cellular conditions assayed. This theory has been confirmed in *Drosophila*, in which the MSL complex mainly assembles at expressed genes to enhance transcription elongation and thus increase expression from the male X chromosome<sup>42–46</sup>. Similarly, in *C. elegans*, the condensin-like dosage-compensation complex binds to the promoters of expressed genes on the X chromosomes in XX hermaphrodites<sup>47</sup>. As shown here, inclusion of X-linked genes with very low or no expression in analyses of expression data<sup>11</sup> causes a shift to low median X:A expression ratio values. Second, in the *C. elegans* germline, the X chromosome is largely silenced<sup>13</sup>. The declining X:A expression ratio during development<sup>11</sup> is therefore a reflection of the decreasing ratio of somatic to germ cells as the germline develops in the third and fourth larval stages and does not relate to dosage compensation. We demonstrate this by showing that the distribution of X-linked and autosomal gene expression is very similar in near-adult XX animals lacking a germline.

X chromosome-wide silencing not only occurs in germ cells of *C. elegans*<sup>13</sup> but also takes place in male germ cells of mammals<sup>16</sup> and *Drosophila*<sup>14</sup>. This complex regulation affects X-linked expression in certain tissues or cell types. In addition, not all X-linked genes are

dosage compensated: some genes escape upregulation in *Drosophila*<sup>8</sup>, and some genes escape X inactivation in mammals<sup>48</sup>. These exceptional genes cause sex-specific differences in gene expression, which can vary between tissues, adding another layer of complexity to the regulation of X-linked genes<sup>49</sup>.

In conclusion, when taking into account genes implicated in sexual reproduction and unique regulatory features of the X chromosome in particular cell types such as germ cells, Ohno's hypothesis that X-linked genes are overexpressed per active allele relative to autosomal genes to regain dosage balance is supported by RNA-seq analyses and by RNA-independent results in mammals, *C. elegans* and *Drosophila*.

## URLs

MORGEN, <http://www.mouseatlas.org/data/mouse/>; R, <http://www.r-project.org/>.

## ONLINE METHODS

### ChIP-chip analysis of PGK12.1 ES cells

The female mouse ES cell line PGK12.1 (a gift from N. Brockdorff)<sup>50</sup> was maintained in standard ES medium with 1,000 U/ml leukemia inhibitory factor (LIF) (Millipore) on mouse embryonic fibroblast (MEF) feeders. We verified that two active X chromosomes were present by karyotyping and RNA-FISH for *Xist* (inactive X-specific transcripts), which showed the absence of a *Xist* cloud. ChIP was performed using 5 µg of rabbit polyclonal antibody to RNA polymerase II phosphorylated at Ser5 (PolII-S5p, Abcam) on fixed chromatin prepared from cell pellets as described<sup>51</sup>. A 10% aliquot of chromatin was saved as the input fraction. Immunoprecipitated chromatin and input fractions were treated with 0.2 M NaCl at 65 °C overnight to remove cross-links before DNA purification using the QIAquick PCR purification kit (Qiagen). An aliquot was subjected to PCR for the gene encoding the housekeeping factor β-actin to confirm the specificity of immunoprecipitation compared to a control without antibody. ChIP DNA and input DNA were amplified using the GenomePlex Complete Whole Genome Amplification Kit (Sigma) before labeling with Cy5 and Cy3, respectively. Array hybridizations were performed at the Genomics Resource Center (Fred Hutchinson Cancer Research Center, Seattle, Washington, USA) using NimbleGen mouse 2.1 M mouse promoter arrays. GFF ratios ( $\log_2$  ChIP: input ratio) and peak files were generated using NimbleScan software. Average enrichment profiles at the 5' ends of genes were calculated by end analysis<sup>52</sup>. A total of 16,141 autosomal genes and 647 X-linked genes were separately rank-ordered based on their expression levels by RNA-seq analysis and divided into nine bins (bin 1 containing all genes with 0 FPKM). Two independent biological replicates showed a high correlation (Supplementary Fig. 5).

### RNA-seq data and analyses in mammals

To generate expression data for PGK12.1 ES cells by RNA-seq, purified mRNA was fragmented into ~200 bases and reverse-transcribed with random hexamers into cDNA to prepare a library for Illumina sequencing, by using the RNA-Seq Sample Prep kit (Illumina) as described<sup>51</sup>. Two technical replicates of the same library were prepared for sequencing. Sequences (total 200 million (M) 36-bp single-end reads) were aligned to the mouse reference genome (mm9) using TopHat<sup>53</sup> with default parameters, except that the "min-isoform-fraction" flag was set to zero, "-no-novel-juncs" was specified to constrain the search to annotated exon-exon boundaries and the RefSeq mouse gene annotation (mm9) was used. Alignment resulted in 90 M mapped reads. FPKM values were obtained using Cufflinks with default parameters and "min-isoform-fraction" set to zero<sup>54</sup>.



Analysis of published RNA-seq data sets was based on data deposited in public databases (Supplementary Table 1). Hi-seq data were obtained from Illumina for 16 human tissues. The 50-bp paired-end reads had originally been analyzed as single-end reads using ELAND and CASAVA1.6 (ref. 55) to generate reads per kilobase of exon per million mapped reads (RPKM) values. Because the reads were analyzed as single-end reads, RPKM units are equivalent to FPKM units. A subset of the human tissues was also reanalyzed by TopHat and Cufflinks, which yielded similar expression profiles. In addition, removal of the pseudoautosomal genes did not alter the expression profiles. Data for the four ENCODE diploid cell lines was analyzed using Cufflinks with Gencode Annotation, version 3c. FPKM values computed by Cufflinks were averaged from two biological replicates, each sequenced to ~200 M paired-end 76-bp reads. Data for the 41 lymphoblastoid cell lines had been previously processed to generate RPKM (FPKM) values based on single-end reads<sup>20</sup>. The same RNA-seq data for mouse brain, liver 1 and muscle that were analyzed by Xiong *et al.*<sup>11</sup> were reanalyzed using TopHat and Cufflinks as described above for PGK12.1 ES cells. Data for the mouse liver 2 sample from MORGEN (<http://www.mouseatlas.org/data/mouse/>) and for mouse E15 whole brain, preoptic area (POA) and prefrontal cortex (PFC)<sup>56</sup> were also reanalyzed as described above for PGK12.1 ES cells. For simplicity, all data are presented as FPKM values. Fisher's exact tests, one-way ANOVA tests, Kolmogorov-Smirnov tests (two-sided) and 95% bootstrap confidence intervals were calculated in R.

### C. elegans RNA-seq

The data sets and methods used for obtaining and analyzing the RNA-seq data have been described<sup>57,58</sup>. Expression values used in those calculations are depth of coverage per million reads (dcpm)<sup>57</sup>.

### D. melanogaster RNA-seq

Raw RNA-seq data from the testes of wild-type and *bam* mutant organisms<sup>39</sup> were obtained from GEO (GSE16960). These data consist of 30-bp single-end reads, with multiple lanes pooled together per sample. Read-mapping was performed using Tophat<sup>53</sup> against a *D. melanogaster* reference (UCSC Dm3, FlyBase release 5 assembly, excluding "chrUextra"), turning off minimum isoform fraction filtering and setting the minimum intron size to 42 bp (the size of the smallest annotated intron). Unique mapping was performed using Tophat version v.1.1.4 (ref. 53) with Bowtie v.0.12.7 (ref. 59). Gene-level abundance measures were calculated using Cufflinks<sup>53</sup>, by turning off minimum isoform fraction filtering and supplying an annotation reference to quantify against, with all other options set at default. The alignment reference used was from ENSEMBL release 60 (November 2010), file name *Drosophila\_melanogaster*.BDGP5.25.60.gtf, accessed from [ftp://ftp.ensembl.org/pub/current\\_gtf/](ftp://ftp.ensembl.org/pub/current_gtf/). The Cufflinks version used was v.0.9.3, with SAM tools v.0.1.9 (ref. 60). Gene-level results are reported in units of FPKM. All downstream computation, including Mann-Whitney tests, was performed in R.

### C. elegans RNA microarray

We used XO L3 animals with a genotype of *her-1(e1520) sdc-3(y126) V; xol-1(y9) X* that were transformed into a hermaphrodite for ease of growth and culture<sup>61</sup>. RNA was purified by TRIzol (Invitrogen) extraction and Qiagen RNeasy kit. Total RNA was labeled and hybridized to single-color expression arrays at Roche NimbleGen. The normalization and data processing were performed using NimbleScan software, which normalizes intensity using Robust Multichip Average (RMA) and combines data from three probes per gene into one value per gene (call). The RMA calls were log<sub>2</sub> transformed and used for analysis. The raw and processed data are available from the GEO database (GSE20136).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

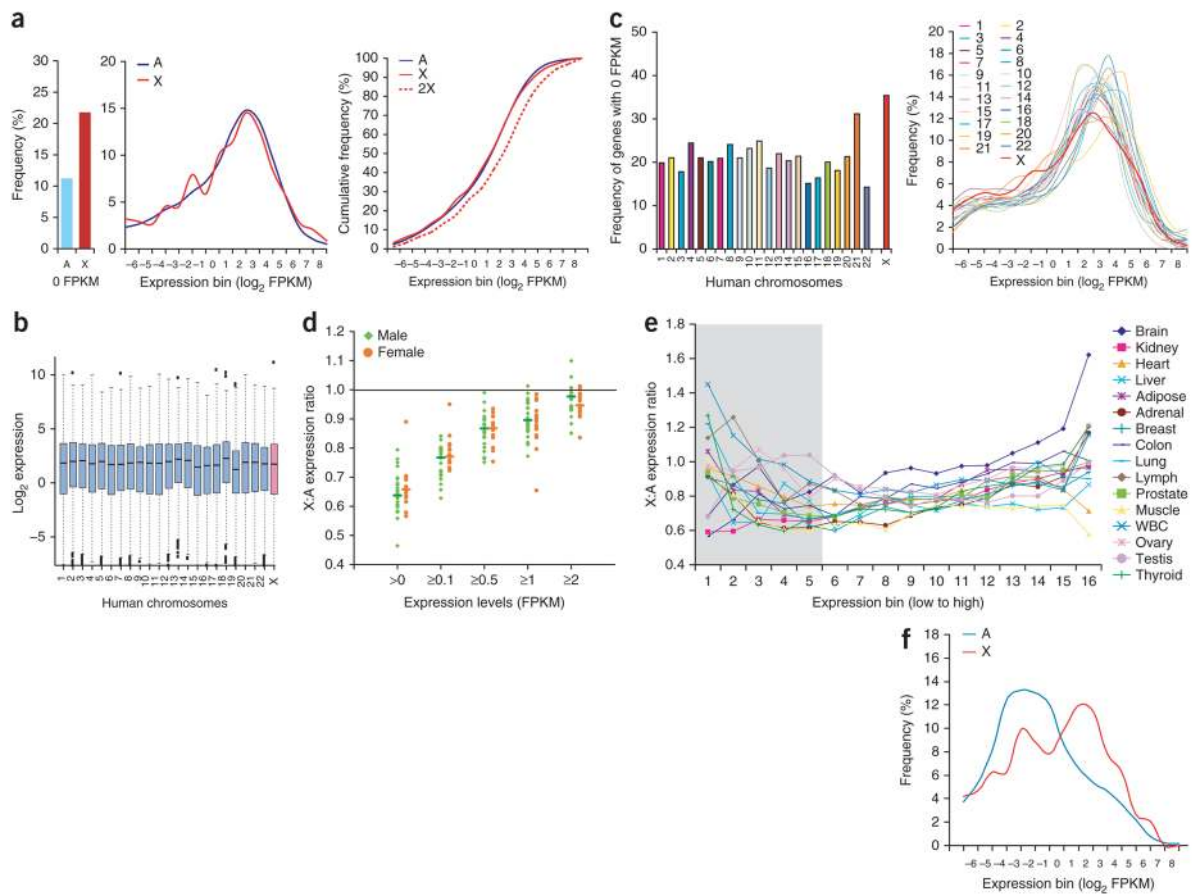
We thank A. Nelson and C. Ware (University of Washington) for expert assistance with ES cell culture and N. Brockdorff (Oxford University) for the female ES cell line PGK12.1. We thank R. Beyer (University of Washington) and X. Deng (University of Wisconsin–Madison) for help with statistical analyses and F. Yang (University of Washington) for helpful discussions. We thank I. Khrebtukova (Illumina) for RNA-seq data on human tissues. This work was supported by grants from the US National Institutes of Health GM079537 (C.M.D.), modENCODE grants HG004270 (J.D.L.), HG004263 (R.H.W.) and AG039173 (J.B.H.), ENCODE Transcriptome Project grant HG004557 (T.R.G.) and by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases (B.O.), the William H. Gates III Endowed Chair of Biomedical Sciences (R.H.W.) and a fellowship from the Achievement Rewards for College Scientists (J.B.H.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## References

- Birchler JA, Riddle NC, Auger DL, Veitia RA. Dosage balance in gene regulation: biological implications. *Trends Genet.* 2005; 21:219–226. [PubMed: 15797617]
- Veitia RA, Bottani S, Birchler JA. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet.* 2008; 24:390–397. [PubMed: 18585818]
- Deng X, Disteche CM. Genomic responses to abnormal gene dosage: the X chromosome improved on a common strategy. *PLoS Biol.* 2010; 8:e1000318. [PubMed: 20186268]
- Gelbart ME, Kuroda MI. *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development.* 2009; 136:1399–1410. [PubMed: 19363150]
- Straub T, Becker PB. Dosage compensation: the beginning and end of generalization. *Nat Rev Genet.* 2007; 8:47–57. [PubMed: 17173057]
- Lyon MF. Gene action in the X-chromosome of the mouse (*Mus musculus* L). *Nature.* 1961; 190:372–373. [PubMed: 13764598]
- Meyer BJ. Targeting X chromosomes for repression. *Curr Opin Genet Dev.* 2010; 20:179–189. [PubMed: 20381335]
- Gupta V, et al. Global analysis of X-chromosome dosage compensation. *J Biol.* 2006; 5:3. [PubMed: 16507155]
- Nguyen DK, Disteche CM. Dosage compensation of the active X chromosome in mammals. *Nat Genet.* 2006; 38:47–53. [PubMed: 16341221]
- Adler DA, et al. Evidence of evolutionary up-regulation of the single active X chromosome in mammals based on *Clc4* expression levels in *Mus spretus* and *Mus musculus*. *Proc Natl Acad Sci USA.* 1997; 94:9244–9248. [PubMed: 9256467]
- Xiong Y, et al. RNA sequencing shows no dosage compensation of the active X-chromosome. *Nat Genet.* 2010; 42:1043–1047. [PubMed: 21102464]
- Vicoso B, Charlesworth B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet.* 2006; 7:645–653. [PubMed: 16847464]
- Kelly WG, et al. X-chromosome silencing in the germline of *C. elegans*. *Development.* 2002; 129:479–492. [PubMed: 11807039]
- Hense W, Baines JF, Parsch J. X chromosome inactivation during *Drosophila* spermatogenesis. *PLoS Biol.* 2007; 5:e273. [PubMed: 17927450]
- Turner JM. Meiotic sex chromosome inactivation. *Development.* 2007; 134:1823–1831. [PubMed: 17329371]
- Khil PP, Smirnova NA, Romanienko PJ, Camerini-Otero RD. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat Genet.* 2004; 36:642–646. [PubMed: 15156144]
- Ross MT, et al. The DNA sequence of the human X chromosome. *Nature.* 2005; 434:325–337. [PubMed: 15772651]

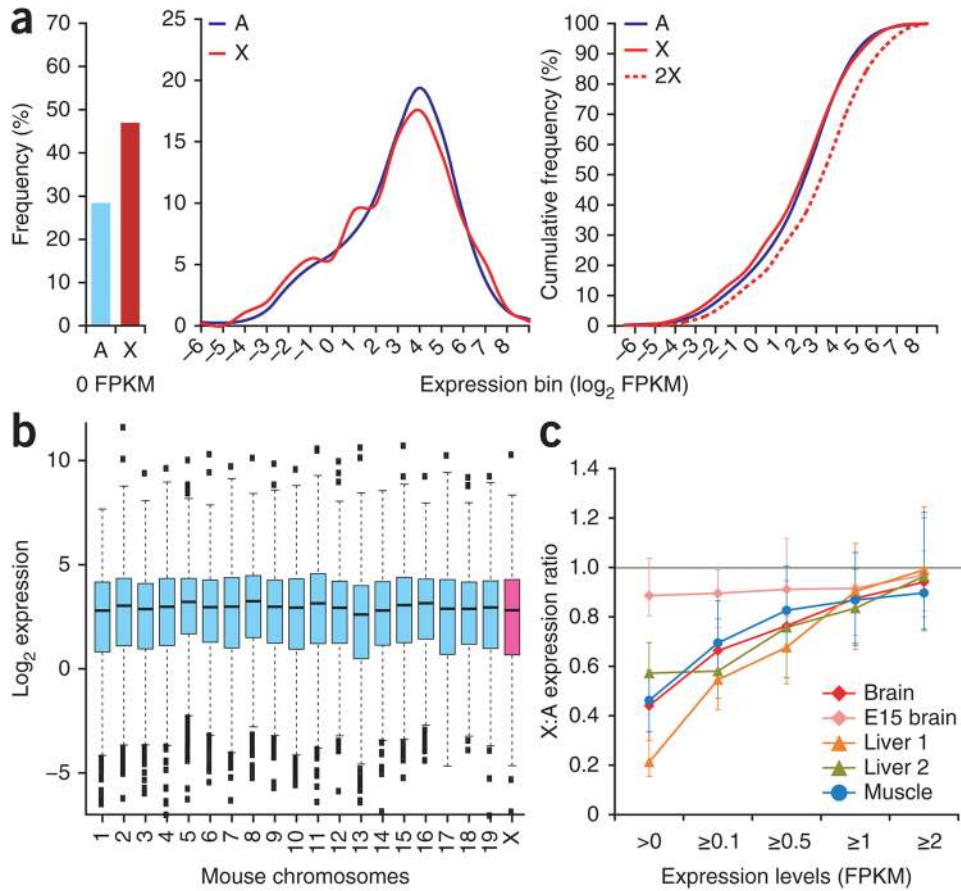
18. Mueller JL, et al. The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet.* 2008; 40:794–799. [PubMed: 18454149]
19. Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. *Nat Methods.* 2009; 6:S22–S32. [PubMed: 19844228]
20. Cheung VG, et al. Polymorphic cis- and trans-regulation of human gene expression. *PLoS Biol.* 2010; 8:e1000480. [PubMed: 20856902]
21. Saifi GM, Chandra HS. An apparent excess of sex- and reproduction-related genes on the human X chromosome. *Proc Biol Sci.* 1999; 266:203–209. [PubMed: 10097393]
22. Hofmann O, et al. Genome-wide analysis of cancer/testis gene expression. *Proc Natl Acad Sci USA.* 2008; 105:20422–20427. [PubMed: 19088187]
23. Skuse DH. X-linked genes and mental functioning. *Hum Mol Genet.* 2005; 14(Spec No 1):R27–R32. [PubMed: 15809269]
24. Ropers HH. X-linked mental retardation: many genes for a complex disorder. *Curr Opin Genet Dev.* 2006; 16:260–269. [PubMed: 16647850]
25. Stevenson RE, Schwartz CE. X-linked intellectual disability: unique vulnerability of the male genome. *Dev Disabil Res Rev.* 2009; 15:361–368. [PubMed: 20014364]
26. Wang PJ, McCarrey JR, Yang F, Page DC. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet.* 2001; 27:422–426. [PubMed: 11279525]
27. Lin H, et al. Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes. *PLoS Biol.* 2007; 5:e326. [PubMed: 18076287]
28. Zvetkova I, et al. Global hypomethylation of the genome in XX embryonic stem cells. *Nat Genet.* 2005; 37:1274–1279. [PubMed: 16244654]
29. Mizuno H, Okamoto I, Takagi N. Developmental abnormalities in mouse embryos tetrasomic for chromosome 11: apparent similarity to embryos functionally disomic for the x chromosome. *Genes Genet Syst.* 2002; 77:269–276. [PubMed: 12419899]
30. Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* 2006; 20:2922–2936. [PubMed: 17079683]
31. Goldberg AD, et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell.* 2010; 140:678–691. [PubMed: 20211137]
32. Byerly L, Cassada RC, Russell RL. The life cycle of the nematode *Caenorhabditis elegans*. I. Wild-type growth and reproduction. *Dev Biol.* 1976; 51:23–33. [PubMed: 988845]
33. Sulston JE, Schierenberg E, White JG, Thomson JN. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev Biol.* 1983; 100:64–119. [PubMed: 6684600]
34. Lewis JA, et al. Cholinergic receptor mutants of the nematode *Caenorhabditis elegans*. *J Neurosci.* 1987; 7:3059–3071. [PubMed: 3668616]
35. Wood WB. Determination of pattern and fate in early embryos of *Caenorhabditis elegans*. *Dev Biol (N Y)* 1985. 1988; 5:57–78. [PubMed: 3077982]
36. Crittenden SL, Leonhard KA, Byrd DT, Kimble J. Cellular analyses of the mitotic region in the *Caenorhabditis elegans* adult germ line. *Mol Biol Cell.* 2006; 17:3051–3061. [PubMed: 16672375]
37. Kislinger T, et al. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell.* 2006; 125:173–186. [PubMed: 16615898]
38. Parisi M, et al. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science.* 2003; 299:697–700. [PubMed: 12511656]
39. Gan Q, et al. Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. *Cell Res.* 2010; 20:763–783. [PubMed: 20440302]
40. Rice WR. Sexually antagonistic genes: experimental evidence. *Science.* 1992; 256:1436–1439. [PubMed: 1604317]
41. Zechner U, et al. A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? *Trends Genet.* 2001; 17:697–701. [PubMed: 11718922]
42. Hamada FN, Park PJ, Gordadze PR, Kuroda MI. Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes Dev.* 2005; 19:2289–2294. [PubMed: 16204180]

43. Gilfillan GD, et al. Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev.* 2006; 20:858–870. [PubMed: 16547172]
44. Zhang Y, et al. Expression in aneuploid *Drosophila* S2 cells. *PLoS Biol.* 2010; 8:e1000320. [PubMed: 20186269]
45. Alekseyenko AA, Larschan E, Lai WR, Park PJ, Kuroda MI. High-resolution ChIP-chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome. *Genes Dev.* 2006; 20:848–857. [PubMed: 16547173]
46. Larschan E, et al. X chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. *Nature.* 2011; 471:115–118. [PubMed: 21368835]
47. Ercan S, et al. X chromosome repression by localization of the *C. elegans* dosage compensation machinery to sites of transcription initiation. *Nat Genet.* 2007; 39:403–408. [PubMed: 17293863]
48. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature.* 2005; 434:400–404. [PubMed: 15772666]
49. Berletch JB, Yang F, Disteché CM. Escape from X inactivation in mice and humans. *Genome Biol.* 2010; 11:213. [PubMed: 20573260]
50. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. Requirement for *Xist* in X chromosome inactivation. *Nature.* 1996; 379:131–137. [PubMed: 8538762]
51. Yang F, Babak T, Shendure J, Disteché CM. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res.* 2010; 20:614–622. [PubMed: 20363980]
52. Mito Y, Henikoff JG, Henikoff S. Genome-scale profiling of histone H3.3 replacement patterns. *Nat Genet.* 2005; 37:1090–1097. [PubMed: 16155569]
53. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009; 25:1105–1111. [PubMed: 19289445]
54. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010; 28:511–515. [PubMed: 20436464]
55. Hosseini P, Tremblay A, Matthews BF, Alkharouf NW. An efficient annotation and gene-expression derivation tool for Illumina Solexa datasets. *BMC Res Notes.* 2010; 3:183. [PubMed: 20598141]
56. Gregg C, et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science.* 2010; 329:643–648. [PubMed: 20616232]
57. Hillier LW, et al. Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Res.* 2009; 19:657–666. [PubMed: 19181841]
58. Gerstein MB, et al. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science.* 2010; 330:1775–1787. [PubMed: 21177976]
59. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10:R25. [PubMed: 19261174]
60. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25:1754–1760. [PubMed: 19451168]
61. Davis TL, Meyer BJ. SDC-3 coordinates the assembly of a dosage compensation complex on the nematode X chromosome. *Development.* 1997; 124:1019–1031. [PubMed: 9056777]

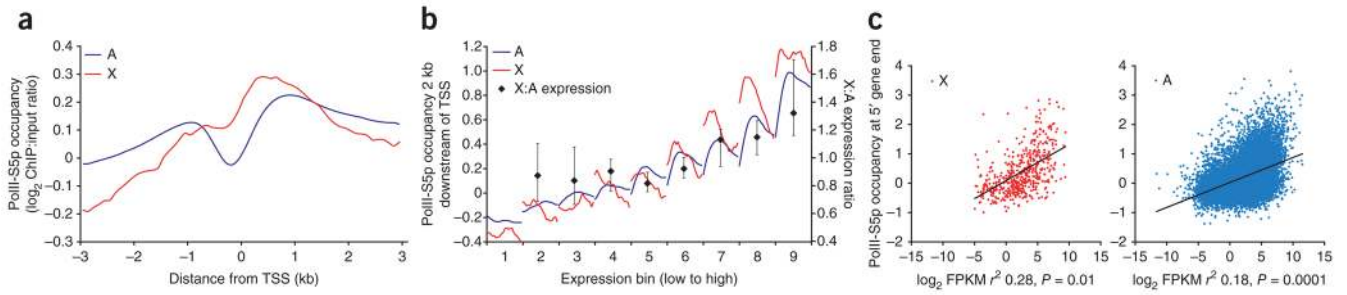


**Figure 1.**

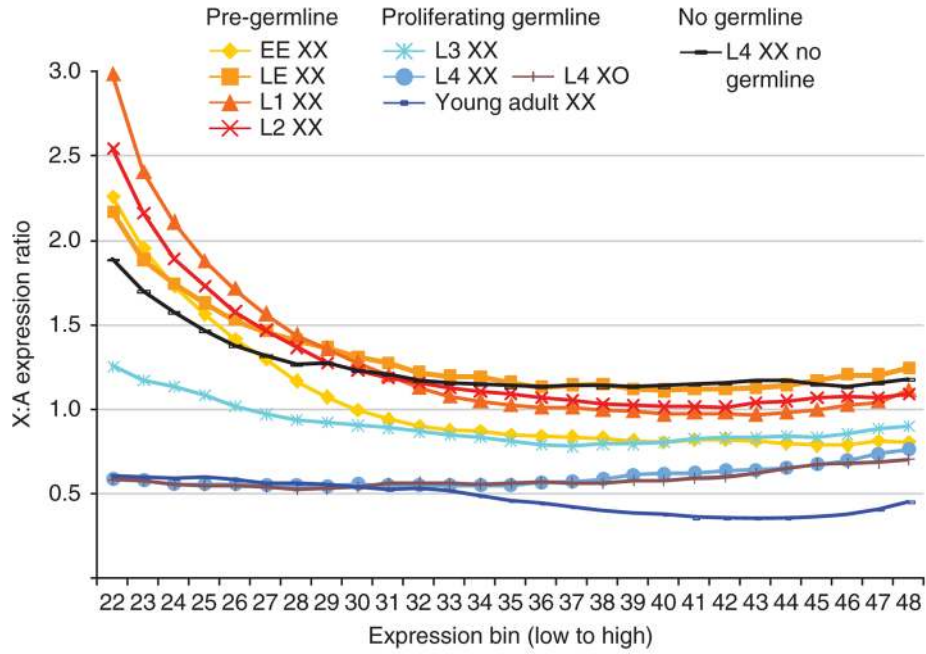
Distributions of gene expression are similar between the X chromosome and autosomes in human, except for reproduction-related X-linked genes not expressed in somatic tissues. **(a)** Expression distributions are similar for X-linked (red) and autosomal genes (blue) in human brain based on RNA-seq data ( $P = 0.71$ , by Kolmogorov-Smirnov test). Left, frequency of genes with 0 FPKM; center, histograms of X-linked and autosomal expression distributions; right, cumulative frequencies for genes with  $>0$  FPKM. A theoretical curve (dotted red line) generated by doubling X-linked expression does not result in equal X-linked and autosomal distributions ( $P = 1 \times 10^{-12}$ , by Kolmogorov-Smirnov test). **(b)** Box plots of expression of genes with  $>0$  FPKM are similar from each human chromosome in brain. **(c)** Left, the frequency of genes with 0 FPKM is significantly higher on the X chromosome than each autosome except for chr. 21 in 41 lymphoblastoid cell lines ( $P < 0.05$ , by Fisher's exact test). Right, histograms of expression distribution for genes with  $>0$  FPKM on each human chromosome. **(d)** X:A median expression ratios increase depending on FPKM cutoffs ( $>0$ ,  $\geq 0.1$ ,  $\geq 0.5$ ,  $\geq 1$  and  $\geq 2$ ). Ratios calculated for individual lymphoblastoid cell lines (17 female, orange and 24 male, green) reveal variability between lines but no differences between males and females ( $P > 0.05$ , by Student's *t*-test). **(e)** X:A median expression ratios calculated after separately rank-ordering X-linked and autosomal genes with  $>0$  FPKM by dividing them into 16 bins based on expression. Bins 1–5 contain genes with  $<1$  FPKM (shaded). **(f)** Pairwise comparison of the distribution of gene expression in the testis versus liver. Histograms of expression distribution for the subsets of X-linked (X) and autosomal (A) genes expressed in testis ( $>0$  FPKM) but not expressed (0 FPKM) in liver.



**Figure 2.** Distributions of gene expression are similar between the X chromosome and autosomes in mouse tissues. (a) Expression distributions are similar for X-linked (red) and autosomal genes (blue) in mouse brain (embryonic day 15 (E15) brain) based on RNA-seq data ( $P=0.04$ , by Kolmogorov-Smirnov test). Left, frequency of genes with 0 FPKM; center, histograms of X-linked and autosomal expression distributions; right, cumulative frequencies for genes with >0 FPKM. A theoretical curve (dotted red line) generated by doubling X-linked expression does not result in equal X-linked and autosomal distributions ( $P=1 \times 10^{-13}$  for E15 brain, by Kolmogorov-Smirnov test). (b) Box plots of expression of genes with >0 FPKM are similar from each mouse chromosome in E15 brain ( $P>0.05$  for expression comparison of the X chromosome versus 14 of 19 autosomes, one-way ANOVA test). (c) Median X:A expression ratios increase depending on FPKM cutoffs (>0, ≥0.1, ≥0.5, ≥1 and ≥2). Reanalysis of RNA-seq data for mouse brain, liver and muscle. Error bars show 95% bootstrap confidence intervals.

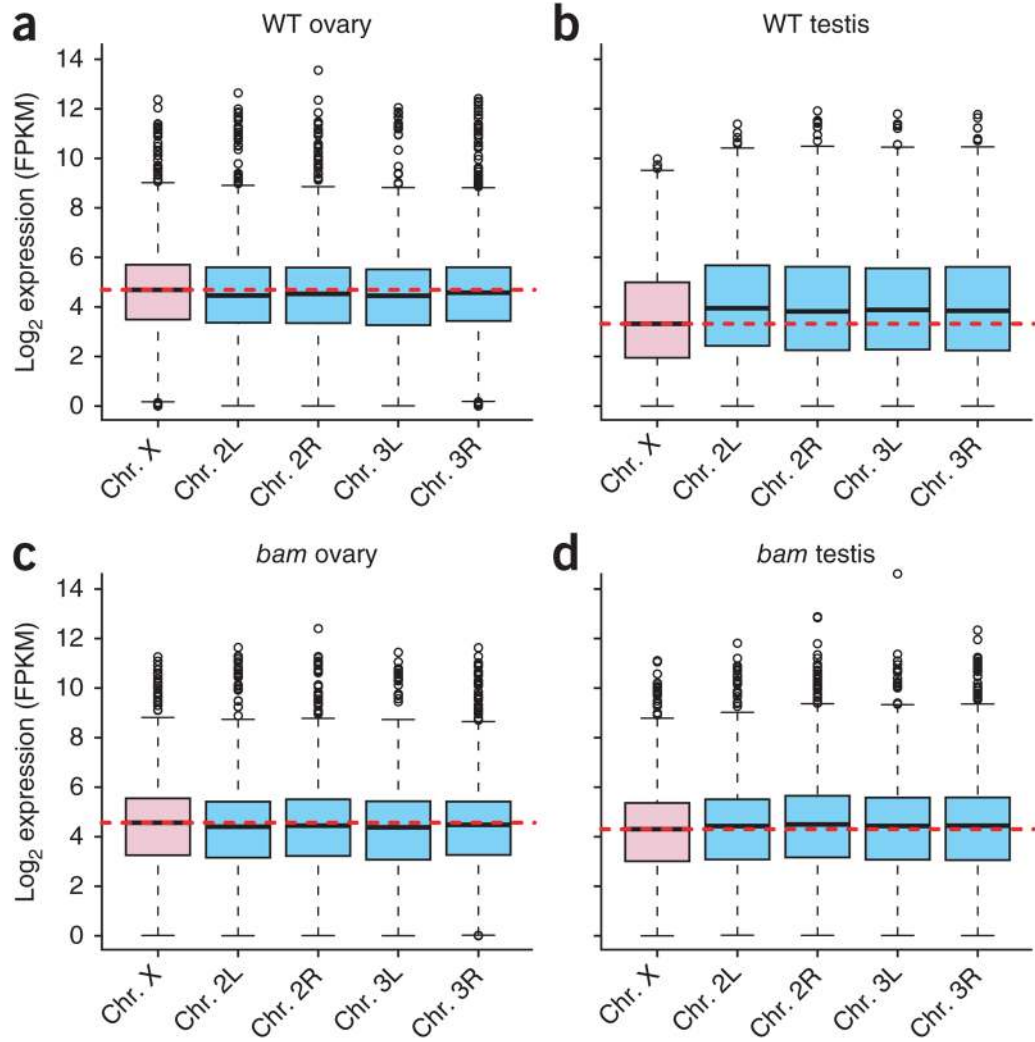


**Figure 3.** Expressed X-linked genes are enriched in RNA PolII-S5p. **(a)** Average PolII-S5p occupancy at the 5' end of 647 X-linked genes (red) is higher in a 1-kb window downstream of the transcription start site (TSS) compared to 16,141 autosomal genes (blue) in undifferentiated female ES cells. Average occupancy ( $\log_2$  ChIP: input ratio) is plotted using a 500-bp sliding window (100-bp interval) 3 kb up- and downstream of the TSS. **(b)** Snapshot of PolII-S5p occupancy in a 2-kb window downstream of the TSS relative to gene expression. Occupancy is higher on X-linked genes (red) compared to autosomal genes (blue) for expressed genes (bins 6–9) but not for weakly- or non-expressed genes (bins 1–5). X-linked and autosomal genes sorted in nine bins based on expression as determined by RNA-seq. Bin 1 contains 124 X-linked and 2,905 autosomal genes with 0 FPKM; bins 2–9 each contain 63 X-linked and 1,596 autosomal genes. X:A median expression ratios are shown in black for bins 2–9. Same ChIP-chip analysis as in **a**. Error bars show 95% bootstrap confidence intervals. **(c)** Gene expression is correlated to PolII-S5p occupancy at the 5' end of genes. Scatter plots of average PolII-S5p occupancy in a 1-kb region downstream of the TSS against X-linked (red) and autosomal (blue) gene expression in  $\log_2$  FPKM. Same ChIP-chip analysis as in **a**.



**Figure 4.** X:A expression ratios in adult *C. elegans* result from the presence of germ cells in which the X chromosomes are silenced. We rank-ordered X-linked genes according to their expression values and divided them into 55 bins of 50 genes each. Bin 1 contained the lowest-expressed genes, and bin 55 the highest. We likewise ranked the autosomal genes and divided them into 55 bins. The X:A ratio of expression levels were calculated for each bin. Genes with zero reads were included in the binning. Bins 22–48, which represent the middle of the distribution of expression values, are plotted. This ordering removes the effect of very highly or very weakly expressed genes (for example, in early embryo, bins 1–21 account for 0.3% of X-linked expression and 0.06% of autosomal expression). Weakly expressed genes (for example, bins 22–31) may have higher X:A expression ratios if dosage-compensation mechanisms work most efficiently on appreciably expressed genes.





**Figure 5.**

X chromosome dosage compensation in early mitotic cells in the *Drosophila* germline. Distributions of gene-level expression values in  $\log_2$  FPKM are shown by chromosome arm ( $\geq 1$  FPKM). (a–d) Red dashed lines indicate the median value for genes on the X chromosome for ovary (a) and testis (b) from wild-type *Drosophila* and for ovary (c) and testis (d) from *bam* mutant *Drosophila*.

Table 1

X:A:autosome expression ratios in *C. elegans*

Stage	Counting all genes <sup>a</sup>				Excluding genes with zero expression value <sup>b</sup>				Ratio			
	No. X genes	Median	No. A genes	Median	Ratio	No. X genes	Median	X zeros <sup>c</sup>		A zeros	No. A genes	Median
EE	2,756	0.12	16,984	0.11	1.08	2,238	0.25	19%	25%	12,728	0.41	0.61
LE	2,756	0.19	16,984	0.15	1.32	2,260	0.36	18%	25%	12,702	0.39	0.91
L1	2,756	0.18	16,912	0.14	1.25	2,162	0.32	22%	28%	12,108	0.42	0.76
L2	2,756	0.26	16,932	0.20	1.31	2,270	0.41	18%	24%	12,840	0.45	0.91
L3	2,756	0.17	16,984	0.19	0.90	2,236	0.31	19%	22%	13,272	0.41	0.75
L4	2,756	0.13	16,984	0.25	0.53	2,328	0.22	16%	17%	14,032	0.42	<b>0.51</b>
L4 male	2,756	0.12	16,984	0.24	0.51	2,536	0.16	8%	9%	15,454	0.30	<b>0.51</b>
L4 soma	2,756	0.05	16,984	0.04	1.20	2,432	0.08	12%	16%	14,318	0.08	<b>0.99</b>
YA	2,756	0.11	16,908	0.19	0.55	2,232	0.20	19%	21%	13,370	0.43	0.46

X, X-linked genes; A, autosomal genes; EE, early embryo; LE, late embryo; L1–L4, larval stages 1–4; YA, young adult. All stages except L4 male represent XX animals. Bold font highlights three critical X:A median expression ratios from animals of the same developmental stage but harboring XX with germline (L4), XO with germline (L4 male) and XX without a germline (L4 soma).

<sup>a</sup>X:A expression ratios calculated including genes with an expression value of 0.

<sup>b</sup>X:A expression ratios calculated excluding genes with an expression value of 0. In *C. elegans*, excluding genes with zero reads decreases the X:A median expression ratio because a greater proportion of genes with no expression in the soma are located on autosomes.

<sup>c</sup>The percentage of X-linked and autosomal genes with zero counts.