ANALYSIS

nature genetics

RNA sequencing shows no dosage compensation of the active X-chromosome

Yuanyan Xiong[1,6], Xiaoshu Chen[1,6], Zhidong Chen[1], Xunzhang Wang[1], Suhua Shi[1], Xueqin Wang[2,3], Jianzhi Zhang[4] & Xionglei He[1,5]

Mammalian cells from both sexes typically contain one active X chromosome but two sets of autosomes. It has previously been hypothesized that X-linked genes are expressed at twice the level of autosomal genes per active allele to balance the gene dose between the X chromosome and autosomes (termed 'Ohno's hypothesis'). This hypothesis was supported by the observation that microarray-based gene expression levels were indistinguishable between one X chromosome and two autosomes (the X to two autosomes ratio (X:AA) ~1). Here we show that RNA sequencing (RNA-Seq) is more sensitive than microarray and that RNA-Seq data reveal an X:AA ratio of ~0.5 in human and mouse. In *Caenorhabditis elegans* hermaphrodites, the X:AA ratio reduces progressively from ~1 in larvae to ~0.5 in adults. Proteomic data are consistent with the RNA-Seq results and further suggest the lack of X upregulation at the protein level. Together, our findings reject Ohno's hypothesis, necessitating a major revision of the current model of dosage compensation in the evolution of sex chromosomes.

The mammalian X chromosome and its much degenerated counterpart Y originated from a pair of autosomes[1]. Upon X inactivation in females, both sexes have one active allele per X-linked gene but two active alleles per autosomal gene. In 1967, it was hypothesized that X-linked genes are expressed at twice the level of autosomal genes per active allele to regain dosage balance (Ohno's hypothesis)[1]. This hypothesis is the cornerstone of the current evolutionary model of dosage compensation[1–4] and was supported by two recent microarray studies[5,6] in which the X:AA expression ratio was found to be ~1. However, this result could be an artifact of the insensitivity of microarray in detecting small expression differences between genes, because microarray was designed primarily for comparing expressions of the same genes across different conditions rather than the expressions of different genes. Direct comparison of hybridization signals from different genes, which necessarily have different probes, is often inappropriate[7–10]. Below, we demonstrate this point by reanalyzing the human and mouse microarray data[6] previously used to support Ohno's hypothesis[1] (**Supplementary Table 1**). We then show that RNA-Seq is more sensitive than microarray in detecting expression differences between genes and that RNA-Seq data reject Ohno's hypothesis.

[1]State Key Laboratory of Biocontrol, College of Life Sciences, Sun Yat-sen University, Guangzhou, China. [2]School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou, China. [3]Zhongshan Medical School, Sun Yat-sen University, Guangzhou, China. [4]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, USA. [5]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. [6]These authors contributed equally to this work. Correspondence should be addressed to X.H. (hexiongl@mail.sysu.edu.cn) or J.Z. (jianzhi@umich.edu).

Published online 21 November 2010; doi:10.1038/ng.711

RESULTS

RNA-Seq outperforms microarray in measuring gene expression

There are ~5,000 genes each represented by at least two probesets in the Affymetrix array HG-U133 Plus 2.0, the platform used in ref. 6 and one of the most comprehensive human gene expression microarrays. Ideally, hybridization intensities of different probesets targeting the same gene should be the same. However, analysis of the human liver gene expression data reported in ref. 6 shows that a large fraction of these same-gene probesets generated radically different signals. The median intensity difference between two same-gene probesets is 3.9-fold, and the estimated expression levels from different same-gene probesets vary >10-fold in 27% of genes (**Fig. 1a**), even when only reliable probesets are considered. We observed a similar pattern upon analysis of the mouse gene expression microarray data analyzed in ref. 6, which was generated on the Affymetrix GNF1M platform (**Fig. 1b**). These expression-signal discrepancies are not due to alternative splicing that could lead to different expression levels of multiple exons of the same genes because different expression signals were also generated from probesets of the same exons (**Supplementary Fig. 1**).

The newly developed RNA-Seq technique, based on high-throughput DNA sequencing, can generate digital counts of transcripts in a largely unbiased fashion[11]. When an RNA-Seq read is mapped to a gene, only the first nucleotide from the 5′ end of the aligned region is counted once. The expression level of the gene is the average number of mapped reads of all positions. This way, every nucleotide of a gene is considered independently. The reliability of RNA-Seq–based measures of gene expression level can be assessed by examining the internal consistency of different positions of the same gene. Our analysis of RNA-Seq data in refs. 12 and 13 shows that variation in expression signals estimated from RNA-Seq reads of two halves of the same transcript is much lower than that from two probesets targeting the same gene (**Fig. 1a**,**b**; $P < 2.2 \times 10^{-16}$ for both panels, Mann-Whitney U test). Further, when two genes show a twofold expression difference based on the RNA-Seq reads from one half of the transcripts, reads from the other half also show an

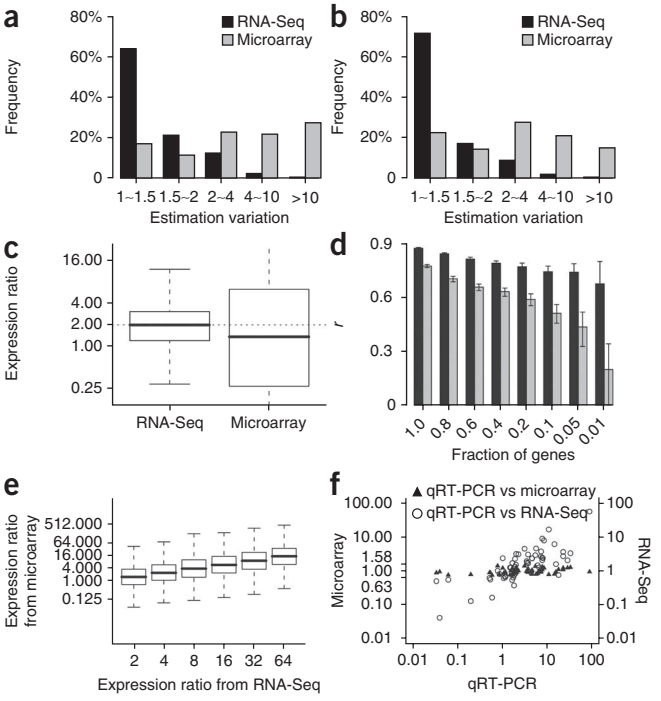


Figure 1 Comparison of gene expressions measured by microarray and RNA-Seq[6,11–13]. Human liver is considered unless otherwise noted. (**a**) Estimation variation measured by the fold difference of microarray intensities of two same-target probesets or of RNA-Seq signals from two halves of the same gene. (**b**) Identical to **a**, except that mouse liver is considered here. (**c**) Comparison of the internal consistency of RNA-Seq data and microarray data. The expression differences from one-half of the nucleotides (RNA-Seq) or a probeset (microarray) are shown for 1,000 randomly picked gene pairs each with twofold ± 0.01-fold expression difference from the other half of nucleotides (RNA-Seq) or from the other probeset (microarray). The central bold line shows the median, the box encompasses 50% of data points and the error bars include 90% of data points. (**d**) Pearson's correlation (*r*) of microarray and RNA-Seq expression signals (gray) and of RNA-Seq signals from two independent experiments (black). A certain fraction of genes (*x* axis) with the highest expression according to one of the RNA-Seq datasets are examined. Error bars show 95% confidence intervals estimated by bootstrapping. (**e**) Microarray consistently underestimates expression differences between genes. The microarray expression differences of 1,000 randomly picked gene pairs each with *x*-fold ($x = 2 \pm 0.01$, 4 ± 0.02, 8 ± 0.04, 16 ± 0.08, 32 ± 0.16, and 64 ± 0.32) RNA-Seq expression difference are shown. The central bold line shows the median, the box encompasses 50% of data points and the error bars include 90% of data points. (**f**) Relative liver expressions of 55 mouse genes, measured by RNA-Seq, microarray and qRT-PCR.

approximately twofold difference with a narrow standard deviation (s.d). ($P = 0.24$, Mann-Whitney U test), whereas we observed an average 1.36-fold difference and a much larger s.d for the corresponding microarray analysis of different probesets of the same transcripts ($P = 4.1 \times 10^{-7}$, Mann-Whitney U test) (**Fig. 1c**). Because of the huge among-gene variation in expression, the overall correlation of expression levels measured by microarray and RNA-Seq is not low (**Fig. 1d**). However, this correlation decreases drastically when only genes with similar expression levels are considered (**Fig. 1d**). By contrast, the correlation of RNA-Seq signals between two independent experiments remains reasonably high for genes with similar expression levels (**Fig. 1d**). RNA-Seq also substantially outperforms microarray in the dynamic range of expression levels over which transcripts can be accurately counted[14]. For example, an expression range spanning five orders of magnitude was achieved in a recent RNA-Seq analysis of mouse transcriptomes, in contrast to only two to three orders of magnitude in typical microarray analysis[12,14]. Thus, similar to what was recently demonstrated in the comparison between microarray and barcode analysis by sequencing (Bar-Seq) in detecting a known difference in DNA concentration[15], we predict that, compared to RNA-Seq, microarray tends to compress expression differences among genes. Indeed, we found that the median microarray expression difference is 1.48-fold for gene pairs having twofold RNA-Seq expression differences ($P = 2.5 \times 10^{-8}$, Mann-Whitney U test) (**Fig. 1e**). We confirmed the above findings by measuring the liver expressions of 55 mouse genes (**Supplementary Table 2**) using real-time quantitative RT-PCR (qRT-PCR), a method generally regarded as the gold standard for gene expression quantification. We found that the expression levels determined by RNA-Seq (Pearson's correlation $r^2 = 0.56$) were substantially better than those determined by microarray ($r^2 = 0.13$) in correlating with the expression levels determined by qRT-PCR (**Fig. 1f**). We further verified this result by an independent qRT-PCR study of 120 randomly chosen genes (**Supplementary Table 2**) from the mouse chromosome 13 (**Supplementary Fig. 2**). The comparison with qRT-PCR results also verified the aforementioned compression problem of microarray (**Fig. 1f**). This compression seriously undermines the capability of microarray in differentiating Ohno's hypothesis of twofold upregulation of X-linked genes (that is, the expression ratio between one active X chromosome and two autosomes being X:AA = 1) from no upregulation (X:AA = 0.5).

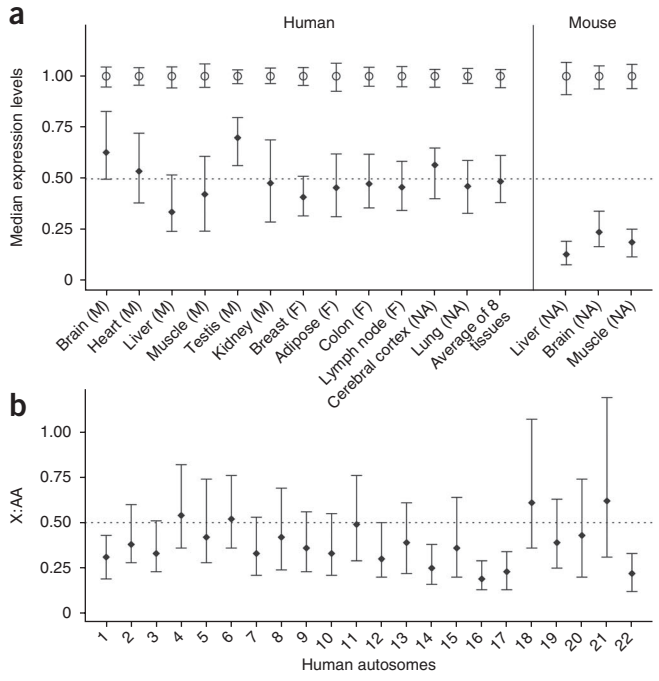


Figure 2 Comparisons of RNA-Seq gene expression levels between the X chromosome and autosomes in 12 human tissues and 3 mouse tissues[11–13,16]. (**a**) The median expression levels of X-linked genes (closed diamonds) and autosomal genes (open circles) are compared. Median expressions of autosomal genes were normalized to 1. Error bars show 95% bootstrap confidence intervals. Sex information is listed in the parantheses after the tissue names (M, male; F, female; NA, unknown). (**b**) X:AA ratios of median expressions from the human liver when X is compared to individual autosomes. Error bars show 95% bootstrap confidence intervals.

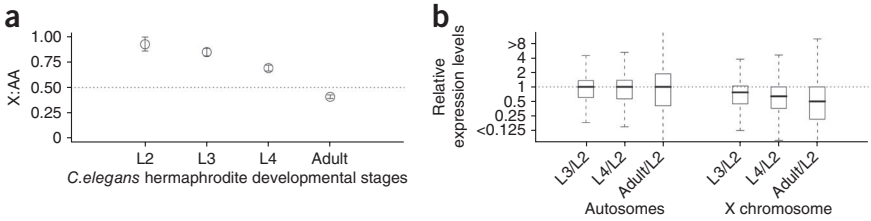


Figure 3 Comparison of RNA-Seq gene expression levels of the X chromosome and autosomes in *C. elegans*[19]. (**a**) X:AA expression ratios at four developmental stages estimated by Miller's jackknife method. Error bars show 95% confidence intervals. (**b**) Gene expression levels of later developmental stages relative to L2. The overall expressions of autosomal genes at different stages relative to L2 are largely the same, with the medians being 0.98, 0.93 and 0.97 for L3/L2, L4/L2 and adult/L2 *C. elegans*, respectively. X-linked genes show an overall approximate twofold downregulation, with the median relative expressions being 0.71, 0.55 and 0.43 for L3/L2, L4/L2 and adult/L2 *C. elegans*, respectively.

Mammalian X:AA expression ratios estimated by RNA-Seq

We used publicly available RNA-Seq data[11,15,16] to compare expression levels of X-linked and autosomal genes in 12 human tissues (**Supplementary Table 1**). The expressions of X-linked genes are significantly ($P < 0.05$) lower than those of autosomal genes in all male and female samples (**Fig. 2a** and **Supplementary Table 3**). The ratio of the median expression level of X-linked genes to that of autosomal genes (X:AA) ranges from 0.34 to 0.70 in the 12 tissues (**Supplementary Table 3**). Using Miller's jackknife method and a method modified from the Mann-Whitney U test (Online Methods) gave similar results (**Supplementary Table 3**). Furthermore, the X:AA ratio is almost always lower than 1, even when the 22 pairs of autosomes are compared individually (**Supplementary Table 4**) (see **Fig. 2b** for the result from the liver) or when different gene functional categories are considered separately (**Supplementary Table 5**). Consistent with previous findings[6], brain and testis showed higher X:AA ratios than other tissues. For each human gene, we computed the average expression in eight tissues excluding sex-specific tissues (testis and breast) and brain regions (brain and cerebral cortex) and found the overall human X:AA ratio to be 0.49 by a direct comparison of two medians, 0.64 by Miller's jackknife method, and 0.45 by the modified Mann-Whitney method (**Supplementary Table 3**). Thus, unlike the insensitive microarray method (**Supplementary Table 6**), RNA-Seq shows an X:AA ratio of ~0.5 in humans. We conducted a similar analysis of publicly available RNA-Seq data of three mouse tissues[12] and observed the X:AA ratio to be significantly ($P < 0.05$) lower than 1 in each of them (**Fig. 2a** and **Supplementary Table 3**). Of note, the X:AA ratio is generally smaller in mouse than in human (**Fig. 2a**), a pattern not previously known. A previous microarray study found an X:A ratio of ~0.5 from an equal mixture of X-bearing and Y-bearing mouse spermatids and suggested that X-linked genes are fully active but are not upregulated in spermatids[6]. Although RNA-Seq data from spermatids are unavailable at this time, given the properties of microarray shown in **Figure 1**, we predict that the true X:A ratio is much lower than 0.5 in spermatids, consistent with the known paucity of post-meiotic X expression[17].

In Ohno's hypothesis, upregulation is needed for those X-linked genes that had existed in the genome before the emergence of the X chromosome; X-linked genes that originated *de novo* on X presumably do not require upregulation. We estimated the X:AA expression ratio in human and mouse by using only those genes that have orthologous genes in chicken and found the results to be similar to those obtained from all genes (**Supplementary Table 7**). Hence, even for the X-linked genes that had existed in the genome before the emergence of the mammalian X chromosome, no doubling of expression was found.

Nematode X:AA expression ratios estimated by RNA-Seq

The nematode *C. elegans* has been subject to intense studies of dosage compensation. *C. elegans* hermaphrodites have two X chromosomes and males have one X chromosome. Dosage compensation between the two sexes is achieved by halving the expression of each X in hermaphrodites through the action of a protein complex known as the dosage compensation complex[18]. We still use X:AA to denote the expression ratio between X and autosomes because although hermaphrodites have two Xs, their total expression is equivalent to one X. Hence, without twofold upregulation of X-linked genes in both sexes, X:AA = 0.5 and gene dosage is expected to be imbalanced between X and autosomes. A previous microarray study showed an X:AA ratio of ~1 in both males and hermaphrodites[5], but the finding could again be an artifact of microarray insensitivity. Analyzing newly published *C. elegans* RNA-Seq data[19], we observed X:AA ratios of 0.92, 0.84, 0.69 and 0.41 for hermaphrodites at the L2, L3, L4 and adult stages, respectively (**Fig. 3a** and **Supplementary Table 3**). Subsequent analysis showed that the overall expression level of autosomal genes remains largely constant during the four developmental stages, whereas X-linked genes are on average downregulated by twofold from the L2 to the adult stage (**Fig. 3b**), causing the gradual decline of the X:AA ratio during development. Acquisition of corresponding RNA-Seq data throughout male development can help discern the underlying mechanism of this phenomenon.

DISCUSSION

There are three caveats in our RNA-Seq analysis and one in a previous microarray study[20] that warrant discussion. First, the Illumina sequencing that generated the RNA-Seq data used here may be biased toward certain sequences or nucleotides[21]. This potential bias can be evaluated by examining the distribution of Illumina-derived sequence reads generated by genome re-sequencing projects (DNA-Seq)[22,23]. Ideally, all nucleotides in a genome should have the same probability of being included in DNA-Seq data. We calculated for each gene an Illumina sequencing preference index (ISPI) using DNA-Seq data and draw the distributions of ISPI for human and *C. elegans* (**Supplementary Fig. 3**). On average, ISPI of an X-linked gene is 1.1 times that of an autosomal gene in human and is 0.96 times that of an autosomal gene in *C. elegans*. Our results of X:AA expression ratios remained largely unchanged when ISPI was considered (**Supplementary Table 8**).

Second, reverse transcription during complementary DNA (cDNA) library preparation is likely to be less efficient for longer transcripts, which could lead to underestimation of expression levels of genes with long transcripts. However, there is no significant difference in transcript length between X-linked genes and autosomal genes in human ($P = 0.72$; **Supplementary Fig. 4**) or mouse ($P = 0.17$; **Supplementary Fig. 4**). In *C. elegans*, transcript length is on average longer for X-linked genes than for autosomal genes ($P = 0.02$; **Supplementary Fig. 4**), but the X:AA ratios of median expression remain qualitatively unchanged when the transcript length is controlled for (**Supplementary Table 9**). The three mouse RNA-Seq datasets

were generated from libraries prepared by shearing mRNA before reverse transcription[12] and thus should be immune to the potential length bias.

Third, GC content may affect RNA-Seq results. We found that the median GC percentage is slightly, but significantly, higher for autosomal genes (46.96% ± 0.06%) than for X-linked genes (45.27% ± 0.26%) in humans ($P < 1.6 \times 10^{-7}$, Mann-Whitney U test). However, the estimate of the X:AA ratio remained unchanged even when genes of similar GC percentages were compared (**Supplementary Table 10**).

Fourth, a recent study traced the shift of X:AA expression ratios during mouse embryonic stem cell differentiation using time-course microarray data and detected an approximately twofold upregulation of the single active X chromosome[20]. The authors excluded lowly expressed genes which could not produce reliable above-background signals and compared 180 X-linked genes with ~5,100 autosomal genes. This treatment, although justifiable for comparing X:AA ratios among developmental stages, is inappropriate for measuring the absolute value of the X:AA ratio because a higher fraction of lowly expressed genes on X than on autosomes is excluded from the comparison. Gene expression approximately follows a power-law distribution, with a high proportion of lowly expressed genes[24]. For example, ~45% of genes have <1 mRNA molecule per cell under the assumption of 1 million mRNA molecules per human liver cell[19]. If we had considered only the 50% most highly expressed genes in human liver, 61.5% of X-linked and 49.6% of autosomal genes would have been excluded from our RNA-Seq data, resulting in a change of the X:AA ratio from 0.33 (**Supplementary Table 8**) to 0.74 (**Supplementary Table 11**).

To further confirm the RNA-Seq data, we examined published mass-spectrometry–based proteomic data from mouse[25] and *C. elegans*[26]. Because of the limited dataset size, we combined the mouse proteomic data from all tissues and found that they can be mapped to 12.3% of autosomal genes but only 8.7% of X-linked genes (**Supplementary Table 12**). This difference ($P = 0.002$) is consistent with a lower protein concentration for X-linked genes than for autosomal genes. Indeed, when we compared the same fraction (8.7%) of the most abundant proteins from X and autosomes, we found an X:AA ratio of 0.47 using median protein levels (**Supplementary Table 12**). For the same set of genes, the RNA-Seq–based X:AA ratio was between 0.44 and 0.47, depending on the individual tissues examined. Using 12.3% of the most abundant proteins from X and autosomes respectively resulted in a slightly lower X:AA protein expression ratio (**Supplementary Table 12**). We similarly analyzed stage-mixed *C. elegans* hermaphrodite proteomic data and found an X:AA ratio of 0.59–0.60 (**Supplementary Table 12**). These proteomic results are consistent with the RNA-Seq data and further suggest the lack of X upregulation at the protein level, but we caution that further larger proteomic analyses are required to confirm our results.

In summary, our results reject Ohno's hypothesis that the expressions of X-linked genes are doubled to mitigate gene-dose imbalance between X and autosomes, at least in mammals and nematodes. A natural question is why such an apparent imbalance has been tolerated in multiple lineages where sex chromosomes emerged independently. One potential answer is that the majority of genes on the proto-X may be insensitive to dosage such that halving the expression levels of these genes had little fitness effect. In fact, haplosufficiency is much more common than haploinsufficiency[27,28]; for example, only 3% of yeast genes show a detectable fitness impact when one allele is deleted from a diploid cell[27]. In mammals, a few X-linked housekeeping genes are known to have their functionally equivalent homologous copies retained in the non-recombining region of the Y chromosome, probably reflecting the importance of maintaining an appropriate dose of these genes[29]. For the rest of haploinsufficient genes on the proto-X, they may either relocate to autosomes or acquire upregulation. Because X-linked genes became hemizygous individually during evolution as a result of the stepwise decay of the Y chromosome[3,29], dosage balance between X and autosomes could not be solved by a chromosome-wide upregulation mechanism. Thus, even if Ohno's X-upregulation hypothesis were correct, the upregulation would have to be acquired by individual genes through evolution and would not happen immediately to transgenes put on the X chromosome. The relative prevalence of relocation and gene-specific upregulation must depend on the relative mutation rates of gene transposition and expression doubling. Although our results suggest that the latter route was uncommon, it was apparently taken by at least one mouse gene, which evolved increased (brain) expression since its recent transposition from an autosome to X[30]. Furthermore, X is subject to unusual selective pressures that led to highly tissue-specific and sex-biased gene expression[31]. For example, we found that 38% of human X-linked genes, in contrast to only 8% of autosomal genes, are highly expressed in one tissue (among the top 10%) but lowly expressed in another (among the bottom 10%) of the 12 tissues examined. The large changes of expression patterns and functions of the genes on X from those on proto-X[31] may also render the previous dosage balance between X and autosomes no longer necessary.

Due to the lack of proper RNA-Seq data, we were not able to examine dosage compensation in several other species that have been subject to microarray-based analysis. The fruit fly *Drosophila melanogaster* is known to equalize the expression of X-linked genes between the sexes by male-specific twofold upregulation of X-linked genes, which incidentally balances the gene dose between X and autosomes[4], resulting in an X:AA ratio of 1. In this case, the microarray result of X:AA ~1 (ref. 5) happens to be correct because an expression ratio of 1 cannot be further compressed by microarray. Birds and moths independently acquired a sex chromosome system in which males are ZZ and females are ZW. Recent microarray studies in birds showed that the expression level of Z-linked genes in males is consistently greater than that in females[32,33]. Subsequent quantitative RT-PCR experiments showed that their difference approximates twofold, suggesting that effective dosage compensation between sexes is not needed in birds[33]. More interestingly, these studies also showed a microarray expression ratio between Z and autosomes to be ~1 in males and ~0.8 in females[32,33]. Given the low sensitivity of microarray, these results are consistent with no dosage compensation between Z and autosomes. Intriguingly, a recent microarray study of silkworm also showed a general lack of between-sex dosage compensation, but there were no data available on Z:AA expression ratios[34].

The current evolutionary model of dosage compensation involves two steps[1–4]. First, expressions of X-linked genes are enhanced to equalize the gene dose between X and autosomes in males. Second, X upregulation results in an X-autosome dosage imbalance in females that is subsequently solved differently in different species: female X-inactivation in mammals, hermaphrodite X downregulation in *C. elegans* and restricted expression of a key component of the protein complex responsible for X upregulation to only males in *D. melanogaster*. In this model, between-sex dosage compensation is a byproduct of the two steps that balance X and autosomal gene dose. Our finding that the first step in this model never happened implies that the second step is unnecessary. Thus, new theories are

required to explain the between-sex dosage compensation observed in mammals, *C. elegans* and *D. melanogaster*. In this context, it is intriguing to note that two (birds and moths) of the five examined systems with independent origins of sex chromosomes do not need effective dosage compensation between sexes. It remains to be seen whether between-sex dosage compensation is the rule or the exception[35].

URLs. NCBI, http://www.ncbi.nlm.nih.gov/; UCSC database, http://genome.ucsc.edu/; Ensembl, http://www.ensembl.org/; BioMart, http://www.biomart.org/; Affymetrix, www.affymetrix.com/; BLAST, http://www.ncbi.nlm.nih.gov/tools/primer-blast/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. Human RNA-Seq data is deposited under the accession codes GSE12946, GSE13652 and SRA000299. Human microarray data is deposited under the accession code GSE3413. Mouse RNA-Seq data is deposited under the accession code SRA001030, *C. elegans* RNA-Seq data is depostied under the accession code SRA003622 and *C. elegans* genome resequencing data is deposited under the accession codes SRR003808 and SRR003809. All data are available from NCBI.

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

ACKNOWLEDGMENTS

We thank B.-Y. Liao for sharing his processed mouse proteomic data, Y. Jiang for assistance in statistical analysis and B.-Y. Liao, J. Lu, O. Podlaha, P. Shi, W. Qian, C.-I. Wu and Y. Xing for valuable comments. Part of the work was conducted in the laboratory of Peng Shi at the State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology. This work was supported by research grants from the US National Institutes of Health (to J.Z.), the National Natural Science Foundation of China (90717115 and 30871371 to X.H.; 11001280 to Xueqin Wang) and the Tianyuan Fund for Mathematics (10926200 to Xueqin Wang).

AUTHOR CONTRIBUTIONS

X.H. and J.Z. conceived the study. Y.X., X.C. and Z.C. produced data. X.H., X.C., Y.X., J.Z., Xunzhang Wang, S.S. and Xueqin Wang analyzed data. X.H., Xunzhang Wang and S.S. provided reagents. X.H. and J.Z. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Ohno, S. *Sex Chromosomes and Sex Linked Genes* (Springer Verlag, Berlin, Germany, 1967).
2. Payer, B. & Lee, J.T. X chromosome dosage compensation: how mammals keep the balance. *Annu. Rev. Genet.* **42**, 733–772 (2008).
3. Charlesworth, B. The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.* **6**, 149–162 (1996).
4. Straub, T. & Becker, P.B. Dosage compensation: the beginning and end of generalization. *Nat. Rev. Genet.* **8**, 47–57 (2007).
5. Gupta, V. *et al.* Global analysis of X-chromosome dosage compensation. *J. Biol.* **5**, 3 (2006).
6. Nguyen, D.K. & Disteche, C.M. Dosage compensation of the active X chromosome in mammals. *Nat. Genet.* **38**, 47–53 (2006).
7. Draghici, S., Khatri, P., Eklund, A.C. & Szallasi, Z. Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet.* **22**, 101–109 (2006).
8. Liao, B.Y. & Zhang, J. Evolutionary conservation of expression profiles between human and mouse orthologous genes. *Mol. Biol. Evol.* **23**, 530–540 (2006).
9. Fu, X. *et al.* Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics* **10**, 161 (2009).
10. Held, G.A., Grinstein, G. & Tu, Y. Modeling of DNA microarray data by using physical properties of hybridization. *Proc. Natl. Acad. Sci. USA* **100**, 7575–7580 (2003).
11. Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517 (2008).
12. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
13. Wang, E.T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
14. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
15. Smith, A.M. *et al.* Quantitative phenotyping via deep barcode sequencing. *Genome Res.* **19**, 1836–1842 (2009).
16. Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
17. Namekawa, S.H. *et al.* Postmeiotic sex chromatin in the male germline of mice. *Curr. Biol.* **16**, 660–667 (2006).
18. Meyer, B.J., McDonel, P., Csankovszki, G. & Ralston, E. Sex and X-chromosome-wide repression in *Caenorhabditis elegans*. *Cold Spring Harb. Symp. Quant. Biol.* **69**, 71–79 (2004).
19. Hillier, L.W. *et al.* Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Res.* **19**, 657–666 (2009).
20. Lin, H. *et al.* Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes. *PLoS Biol.* **5**, e326 (2007).
21. Dohm, J.C., Lottaz, C., Borodina, T. & Himmelbauer, H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **36**, e105 (2008).
22. Wang, J. *et al.* The diploid genome sequence of an Asian individual. *Nature* **456**, 60–65 (2008).
23. Hillier, L.W. *et al.* Whole-genome sequencing and variant discovery in *C. elegans*. *Nat. Methods* **5**, 183–188 (2008).
24. Nacher, J.C. & Akutsu, T. Recent progress on the analysis of power-law features in complex cellular networks. *Cell Biochem. Biophys.* **49**, 37–47 (2007).
25. Kislinger, T. *et al.* Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* **125**, 173–186 (2006).
26. Schrimpf, S.P. *et al.* Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS Biol.* **7**, e48 (2009).
27. Deutschbauer, A.M. *et al.* Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**, 1915–1925 (2005).
28. Qian, W. & Zhang, J. Gene dosage and gene duplicability. *Genetics* **179**, 2319–2324 (2008).
29. Lahn, B.T., Pearson, N.M. & Jegalian, K. The human Y chromosome, in the light of evolution. *Nat. Rev. Genet.* **2**, 207–216 (2001).
30. Adler, D.A. *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* **94**, 9244–9248 (1997).
31. Vicoso, B. & Charlesworth, B. Evolution on the X chromosome: unusual patterns and processes. *Nat. Rev. Genet.* **7**, 645–653 (2006).
32. Ellegren, H. *et al.* Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *BMC Biol.* **5**, 40 (2007).
33. Itoh, Y. *et al.* Dosage compensation is less effective in birds than in mammals. *J. Biol.* **6**, 2 (2007).
34. Zha, X. *et al.* Dosage analysis of Z chromosome genes using microarray in silkworm, Bombyx mori. *Insect Biochem. Mol. Biol.* **39**, 315–321 (2009).
35. Mank, J.E. The W, X, Y and Z of sex-chromosome dosage compensation. *Trends Genet.* **25**, 226–233 (2009).

ONLINE METHODS

Genomic data. Genome information for human (hg18), mouse (mm9) and *C. elegans* (ce6) was downloaded from the UCSC database (see URLs). There are 20,463; 23,147; and 20,176 known protein-coding autosomal or X-linked genes for human, mouse and *C. elegans*, respectively, according to the annotations in Ensembl (see URLs). The longest transcript was considered for each gene in determining gene expression levels. Sources of microarray, RNA-Seq, genome re-sequencing and proteomic data are summarized in **Supplementary Table 1**. Human and mouse genes that have chicken orthologs were retrieved from Ensembl using BioMart (see URLs).

Quantifying gene expression level using RNA-Seq data. We adopted an approach similar to that used in a recent study[36] to determine gene expression levels. Taking mouse as an example, we mapped all 25-mer RNA-Seq reads to the genome sequence. Only those reads uniquely mapped to exons were considered as valid hits for a given gene. The expression level of a gene is defined by the number of valid hits to the gene divided by the effective length of the gene, which is the total number of 25-mers in the DNA sequences of the exons of the gene that have no other matches anywhere in the genome. For comparisons between tissues or developmental stages, expression levels were normalized by dividing the total number of valid hits in the sample. Reads were mapped using SOAP[37], allowing a maximum of two mismatches per read. Use of different mismatch cutoffs or different mapping software gave similar results (**Supplementary Fig. 5** and **Supplementary Table 13**). To ensure the reliability of expression estimation, we excluded genes with effective length smaller than 100, resulting in 19,800 (human), 21,470 (mouse) and 19,507 (*C. elegans*) nuclear genes for subsequent analyses.

The Illumina sequencing preference index (ISPI) was computed using the same approach. Because the genome re-sequencing data of mouse were unavailable, we computed ISPI for human and *C. elegans* genes. Note that the human individual from which the DNA-Seq was collected is a male[22], so the obtained ISPI values for human X-linked genes were multiplied by 2.

About 15% of human X-linked genes are known to escape X inactivation[38], but we found that these genes are heavily biased toward high expression in both male and female tissues. This may have been caused by an ascertainment bias, because highly expressed genes that escape inactivation are more likely to be detected than lowly expressed genes that escape inactivation[38]. Thus, exclusion of this set of genes from X could cause underestimation of X:AA. Further, even when a gene escapes X inactivation, its expression from the allele on the inactivated X is usually much lower than that from the allele on the activated X[38]. Consistent with this result, we found comparable expression levels for these 15% of genes between male and female tissues. Thus, we did not exclude from our analysis the 15% of X-linked genes that escape inactivation.

Evaluating gene expression estimates from microarray and RNA-Seq. The microarray data we used in comparison with RNA-Seq data (**Fig. 1**) were previously generated and processed[6]. The original authors analyzed the raw data using Affymetrix software GCOS 1.1. After the elimination of background signals and genes with a low level of expression, the mean fluorescence intensity of duplicated spots representing the same gene was calculated and normalized to the mean fluorescence intensity of the whole array by the original authors[6]. We used the expression signals reported by these authors[6]. For microarray data, inferior probesets with suffixes like _x_at, _s_at and _a_at were not considered in our estimation of the fold differences of hybridization intensities between two same–target-gene probesets. To locate the target of a probeset, blastn was used to map the probe sequences obtained from Affymetrix (see URLs) to specific exons of a gene under the E-value cutoff of 1×10^{-20}. A probeset is usually composed of 16 pairs of perfect-match and mismatch probes for Affymetrix chips, and the final intensity of a probeset is derived from 16 signals generated by the 16 pairs of probes. Thus, in the analysis of RNA-Seq data, each gene was divided into 32 equal-size windows and the RNA-Seq signals from two randomly chosen nonoverlapping sets of 16 windows were compared.

Analysis of proteomic data. The processed mouse protein abundance data originally generated by Kislinger and colleagues[25] were provided by Dr. Ben-Yang Liao of National Health Research Institutes in Taiwan. The *C. elegans* protein abundance data[26] were provided by the authors. Briefly, proteins were denatured and digested using trypsin and then subjected to mass spectrometry analysis. The relative concentration of a protein was calculated by the mean abundance of its constituent amino acids in the mass spectrometric data[26]. The longest isoform was considered for genes with multiple isoforms.

Statistical analyses. Let the expression level of a gene be a random variable P and let $Q = nP$. Thus, $S_Q = nS_P$, where S_P and S_Q are the s.d. of P and Q, respectively. Because P follows approximately a power-law distribution[24], the median is a better statistic than the mean in characterizing the distribution. We thus use either median (Q)/median (P) or S_Q/S_P to estimate n.

We applied a modified Mann-Whitney U test to compare the expression levels of X-linked genes and autosomal genes. We multiplied the expression levels of all X-linked genes by a, which is a number between 0.5 and 10, and then compared these modified expression levels of X-linked genes with the original expression levels of autosomal genes using Mann-Whitney's U test. The $1/a$ value resulting in the largest P value in Mann-Whitney's U test became our estimate of the X:AA ratio. We could also find a range of a values for which the Mann-Whitney test is not significant at the 5% level (one-tail). We regarded this range of $1/a$ as the probable range of the X:AA ratio.

We also applied Miller's jackknife test to compare the variances in expression levels of X-linked and autosomal genes (an in-house R script of the Miller's jackknife test is available upon request). Of note, there are a large number of zero values (often 10%–25%) for RNA-Seq–based expression levels, which makes the distributions of gene expression levels discontinuous. More importantly, these zeros apparently carry inaccurate information because many of them will become nonzeros if more reads were sequenced. To avoid the confounding effect of zeros, we excluded equal proportions of lowly expressed genes from X and from autosomes to ensure no zeros. To keep the medians of the distributions unchanged, the same proportion of genes at the other end of the distribution (that is, the most highly expressed genes) were also deleted from both X and autosomes. Our results are robust when different proportions of genes were excluded (**Supplementary Tables 14** and **15**). All statistical analyses were performed in R.

Real-time quantitative RT-PCR (qRT-PCR). We used the RNeasy Mini Kit (Qiagen) to extract the total RNA from the liver of an eight-week-old C57BL/6 male mouse. The total RNA (450 ng) was reverse transcribed using the PrimeScript RT reagent Kit (Takara) using 50 picomoles of random hexamer primers (Takara). Genomic DNA was also extracted from the same mouse using the QIAamp DNA Micro kit (Qiagen).

Sixty-two mouse genes with poor correlation (Pearson's correlation $r^2 = 0.13$) between published RNA-Seq and microarray-measured liver expressions were selected. Primers for qRT-PCR were designed using NCBI/Primer3-BLAST (see URLs) and the primer sequences are shown in **Supplementary Table 2**. qRT-PCR was carried out on a Roche LightCycler 480 with a 10 µL reaction volume containing cDNAs corresponding to 0.1 ng, 1.0 ng or 10 ng of total RNA, depending on the RNA-Seq–based expression level of the gene, 1X SYBR Premix Ex Taq (Takara), and 1 picomole each of forward and reverse primers. The reaction was replicated three times for each of the 62 genes. The standard curve was prepared using the same approach, but with a dilution series of 0.1 ng, 1.0 ng, 5.0 ng, 10.0 ng, 50.0 ng and 100.0 ng of genomic DNA instead of cDNA; a negative control was included with water instead of cDNA. The crossing point values were determined with the second derivative max method in the supplied software of the instrument. The mean and s.d. of the crossing point values from the three technical replicates of the cDNA and the slope and correlation coefficient of the standard curve for each gene were calculated. Primer efficiency (E) for each gene was then calculated as $E = 10^{-\text{slope}}$. Among the 62 genes examined, 2 had no amplification and 5 had multiple melting points. These 7 genes were excluded from further analyses, leaving 55 genes, all having s.d. of cDNA's crossing point values lower than 0.6, E between 1.8 and 2.3, and correlation coefficients of the standard curve lower than −0.98. The relative gene expression levels were calculated from $L = 10^{\text{slope} \times \Delta CP}$, where ΔCP is the difference between the mean

doi:10.1038/ng.711

crossing point for the cDNA and the crossing point value for 1 ng of genomic DNA. *L* was then normalized to correspond to 1.0 ng of total RNA.

In a second experiment, we examined the qRT-PCR expression levels of 120 randomly chosen genes from chromosome 13 (**Supplementary Table 2**) using RNAs prepared from another eight-week-old C57BL/6 male mouse. The experimental procedure was the same as described above.

36. Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–960 (2008).
37. Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714 (2008).
38. Carrel, L. & Willard, H.F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400–404 (2005).

SUPPLEMENTARY INFORMATION FOR

RNA sequencing reveals no X-chromosome dosage compensation relative to autosomes

Y. Xiong, X. Chen, Z. Chen, X. Wang, S. Shi, X. Wang, J. Zhang & X. He

Supplementary information includes:
Table S1-S15
Figure S1-S5

Nature Genetics: doi:10.1038/ng.711

Table S1. RNA-Seq, DNA-Seq, microarray, and proteomic datasets used in this work

Species	Tissue/stage	Sex	Data type	Data accession	References
Human	Brain	Male	RNA-Seq	GSE12946	ref. 15
Human	Heart	Male	RNA-Seq	GSE12946	ref. 15
Human	Liver	Male	RNA-Seq	GSE12946	ref. 15
Human	Muscle	Male	RNA-Seq	GSE12946	ref. 15
Human	Testis	Male	RNA-Seq	GSE12946	ref. 15
Human	Adipose	Female	RNA-Seq	GSE12946	ref. 15
Human	Breast	Female	RNA-Seq	GSE12946	ref. 15
Human	Colon	Female	RNA-Seq	GSE12946	ref. 15
Human	Lymph node	Female	RNA-Seq	GSE12946	ref. 15
Human	Kidney	Male	RNA-Seq	SRA000299	ref. 11
Human	Liver	Male	RNA-Seq	SRA000299	ref. 11
Human	Cerebral cortex	N/A	RNA-Seq	GSE13652	ref. 16
Human	Lung	N/A	RNA-Seq	GSE13652	ref. 16
Mouse	Brain	N/A	RNA-Seq	SRA001030	ref. 13
Mouse	Liver	N/A	RNA-Seq	SRA001030	ref. 13
Mouse	Muscle	N/A	RNA-Seq	SRA001030	ref. 13
Worm	L2	Hermaphrodite	RNA-Seq	SRA003622	ref. 19
Worm	L3	Hermaphrodite	RNA-Seq	SRA003622	ref. 19
Worm	L4	Hermaphrodite	RNA-Seq	SRA003622	ref. 19
Worm	Adult	Hermaphrodite	RNA-Seq	SRA003622	ref. 19
Human[1]	Genome resequencing	Male	DNA-Seq	yh.genomics.org.cn/download.jsp	ref. 22
Worm[1]	Genome resequencing	Wild-type strain N2 (Bristol)	DNA-Seq	SRR003808 SRR003809	ref. 23
human	5 tissues	Both	microarray	GSE3413	ref. 6
human	8 tissues	Both	microarray	wombat.gnf.org/index.html	Su et. al., 2004, PNAS
mouse	5 tissues	Both	microarray	wombat.gnf.org/index.html	Su et. al., 2004, PNAS
Worm	Mixed stages	Wild-type strain N2 (Bristol)	Proteome	www.peptideatlas.org	ref. 26
Mouse	6 tissues	N/A	Proteome	Supplemental Data	ref. 25

[1] Only single-end reads were used for computing the ISPI of each gene.

Nature Genetics: doi:10.1038/ng.711

Table S2. Genes analyzed by qRT-PCR in our study.

Genes	Chr	RNA-seq[1]	Microarray[1]	qRT-PCR[1]	Forward primers	Reverse primers
First set of 55 genes						
ENSMUSG00000003458	1	0.62	884.65	0.52	AAGAGCGGTGCTGGTGTCCCTGAA	TGTGGAAGGAGCCAGAGTGGTCAGC
ENSMUSG00000026691	1	0.04	939	0.04	AAAGGCAAATGCTTCCACAGCAGGG	CCCAGGCCAATCACCAGGACTCGTT
ENSMUSG00000026715	1	56.25	939.5	85.04	TCATCGGTGCCTTGGGCTGTGCTA	ATCCTCCTCGGTGGCCTTCTTCCCA
ENSMUSG00000046062	1	0.64	903.25	1.74	TGGCTACTGCTTGGCCTTTGAGCAC	AAGAGACAGAGTGGGCACGGGTTGC
ENSMUSG00000002455	2	0.61	855.8	1.07	TTACTGCTGCGACACCCGGAGGACT	TGCCCAGGAACGGCTTCTTCTTCT
ENSMUSG00000014867	2	3.74	981.1	7.18	TTGCTGGTGTCCCAACCATGCG	AGCTGGCATCAAAGTGAAGGAGGGT
ENSMUSG00000055943	2	2.08	1326.1	7.63	AGATCGAGGGACGTGCGGTTGTTCC	GGAAGCCGACATGCTCTTCTCCGTC
ENSMUSG00000033308	3	3.04	1165.33	3.69	CACAGATCAGAAACCCATCGCTGCC	AGGAGGCACAGCTTATACTCGCAGG
ENSMUSG00000028419	4	0.75	1099.65	17.43	AGTGTTTGGGTTTCCTCGCAGCCG	ATGCAGTCCGTCAAGCTAGGTGGC
ENSMUSG00000028684	4	2.64	918.45	7.1	ACCTGGCAAAGGACCCAGCTTTCC	GCCGTTGTCGAGTAAGGGTGATGGC
ENSMUSG00000033379	4	2.05	1233	27.64	TGGACGCGGTTTAGCTCAGGTCCCA	TGCAGGCCCAGAAGGCCACAAAGA
ENSMUSG00000019179	5	4.49	1395.93	3.08	TGGCTTATGCTGGAGCCCGCTTTG	CCCAAGAGCAAGGGCGTAGAGAAGT
ENSMUSG00000029482	5	0.83	930.95	1.93	TTGCGCTCTCCGCTCTCAGCCAAA	ACCTGGCACTCCATGATCTCCTCGC
ENSMUSG00000034245	6	0.74	947.75	1.8	ACCAGAAGCGCACAGCCCGTATT	AGACGCAGGGAAACTCAGGCCCAA
ENSMUSG00000034456	6	4.58	798.85	5.57	TGTGCTGAATGCTGGTCGTCGGTA	ATCCCGCCAAGCCCAGAGGTGACAA
ENSMUSG00000072772	6	6.15	805.05	14.19	GCGCATGGACGAGGCTAGAGACAA	TTGGATCTGTGTGGCTACGGGCA
ENSMUSG00000030647	7	3.32	1269.7	32.08	GGTCAAACCTCTCGTGGAACCTGCC	GCCCAACAAGCCCATGTAGACAAGC
ENSMUSG00000030894	7	1.01	1405.7	0.78	TGAGTTCATGCGCCTATTCGGTGGC	GGCCTCGCCCTTGCTTTCCAACAAC
ENSMUSG00000039018	7	0.5	878.2	0.03	GGCCAGCGCCTGTGACCAAATAGA	GCCTGTGAGAACCTTCACTTTCCGC
ENSMUSG00000045948	7	1.77	946.4	3.53	CACCCTGAACCAGATGCACCGCTTA	GCGGAGTTGGGCTTCTTCGGCTTTC
ENSMUSG00000015994	8	0.81	775.8	0.86	TACCGGCGAGTGTATAGCAGCGAGG	TGGACAGCTTTAGCGCACAGGCA
ENSMUSG00000031782	8	2.36	797.4	4.5	TTAACTGGTACACCCGCCGTGCAG	GCGCCACGTATCTTCAAAGTCTGGG
ENSMUSG00000074064	8	1.94	986.1	3.43	TACCGGCAACCCTGTTCACGAGAG	ACCAGGCACACAGCCTCATTAGTGG
ENSMUSG00000019039	9	0.71	855.35	1.15	GAGCACCTCGACCGAAACACTGAGA	ACCAGTTCCTTCAGACACAGTCGGC
ENSMUSG00000032018	9	3.64	1224.43	22.15	CCTGGTTCTCAGTGCCGCCGATTA	TGTCGGATGATGTTGTCCTCGGGC
ENSMUSG00000015890	10	3.17	1253.45	2.02	TGGTGGAGTGCAAGAGCGGATATGG	CGCAGTAAGTGGCTGAGAGGCTGA
ENSMUSG00000049858	10	1.7	1180.85	1.66	TGGTTCCTGGTGTAGTAGGTGCCCG	AGCCTTTGTAATCCCGCCTCTGCC
ENSMUSG00000069520	10	1.05	915.8	2.09	ACAGCAGTGGGACTTGCCTCCA	ACCCAGCCACACCACCAAATGCAA
ENSMUSG00000000594	11	2.45	1093.35	7.59	ATCTTGAGCAGTGGTGGGAAGCGCC	GGGCACCGGAACACAGAAGAAGAGG
ENSMUSG00000020477	11	2.9	1360.55	5.36	AACTGTGCCCTCCAAGGTCGTGT	TTGGTCTCCTGCACCTCTGTGACC
ENSMUSG00000039640	11	2.71	1085.1	22.06	TCAAAGCCAACGTCGCCAAAGCTGA	ACCATGCCTCACTCCAGAACCACGG
ENSMUSG00000002804	12	0.56	888.05	0.54	GGCCTTCGCAGACAACCCTGACAT	AGGGCAGGAGGATGAGTGGATGGGA
ENSMUSG00000021067	12	0.16	748.2	0.59	AGCCTCCACCCATCACGTATCAGC	GCTCGGGCATAAACCTGAAGCCAGT
ENSMUSG00000021069	12	8.96	1282.78	7.62	GGCTCAGGCATGGAAACCCTTGGGA	CGTCTGGGTGTGCTCTACTCTTCCG
ENSMUSG00000006717	13	2.38	1235.85	7.84	CGGGCTTGCAGACTTGACCTTCCA	CATGCTGCTCATCGTGGACGCAAA
ENSMUSG00000021607	13	0.88	978.75	4.33	TCAGTTCTCTGTGGCCGTCCCTTG	GCCCACGCCTCTTCACCTTGTAGCA
ENSMUSG00000038175	13	0.13	782.2	0.2	TGTGTCAGCGATGGAGAACTACGGC	ACAGATCGAGATGCCTTCAGGTCCG
ENSMUSG00000002332	14	3.1	1010.95	2.26	AGCGCCTATGAAGGGCCAAGTGTG	AGCTCTAAGCGTATCCAGATGGCGG
ENSMUSG00000021273	14	0.65	936.88	1.15	TCTAAACACAGGTGGGAGCGGGAGA	TCCGGGTGGCCTAGACACTTGACGA
ENSMUSG00000016541	15	1.27	1004.35	1.54	TGCAGTGGGTGGTGTATGCTGTGC	TCAGTAGGGATGCGTCTGCCAAGCC
ENSMUSG00000022574	15	2.36	1214.65	1.02	CAAGGCTGGGCAGGAACTGAGGGTT	ATGGCTCCACTTGGGCTGGCTTCA
ENSMUSG00000006998	16	1.46	839.55	6.04	ACGAGACAAGACACCCGTGCAGTCC	ATCCCTTCGCTCCTTGCCACTCGAC
ENSMUSG00000022742	16	1.26	1028.15	6.5	AGGAGGCCGGAGGACATGAAGACCA	AAGTCGGCAACGCCATCTACCTGC
ENSMUSG00000054604	16	0.37	838.5	1.81	TGCATCGCTTCAGTGCAACAGTCC	TCTTCACATGCCGGGACAGGAACG
ENSMUSG00000024181	17	1.58	930.95	15.33	ACTGGAAGAAGAACGGACGCCGA	TGCTGGGACTCTGGAGGGTAGTGA
ENSMUSG00000040048	17	7.87	1222.18	8.34	ACCGTCGAGTGCCAGACATCACA	GTCCCTTCTCCACTGCATCTCAGCC
ENSMUSG00000040356	17	0.55	756.45	0.06	AGCCAGGACGGTGGTGTTTGACT	CGGCCCTTACAGAGAAGGATGACGG

Nature Genetics: doi:10.1038/ng.711

ENSMUSG00000006050	18	2.34	1042.45	4.55	GGAGCTGTCCTGCTACTGGTGTGGA	GTCCTCCAATGCCTGTTCCAGAGGT
ENSMUSG00000024587	18	2.27	923.65	1.9	AGCTGTGACTTCTGGGAACTGGTGG	TCATGTGCCGGTTGTTGAGCTGGA
ENSMUSG00000024853	19	1.25	761.65	1.47	GAGCATCCCGAACCTCCCAAAGGA	TCGGCGCTCCTATCTCTGCCAATC
ENSMUSG00000034371	19	16.62	783.95	10.64	ATGCACCACCCTCATCGGACTGGA	TGGCAGCACGGCTATGGGTAGAAC
ENSMUSG00000002010	X	1.69	905.95	3.73	CGTAGCAGGAGTGGTGGAGAGCTTG	TGCACAGCCGTCACTTTCTTACGCC
ENSMUSG00000025151	X	1.33	879.7	1.91	GGCACGTAATGCCAAGGAGATGCCC	TGATTTGCCTGTGGGTGCTGCCTG
ENSMUSG00000031299	X	1.52	1108.8	2	CCCAGTTTGCCACGGCTGATCCTGA	GGCACCACGCACTTCAAAGGGAGGA
ENSMUSG00000044206	X	0.54	747.45	1.64	GGTGAAATGGCTGGTAAGACACGGC	CGGCCTCTGTACTTTGCCTGCTGGA

Second set of 120 genes

ENSMUSG00000000253	13	0.02	42.35	0.0054	CGGATGAACGTGGCTCTCCTGCTGA	GGGCAAGACCGTGGAAGTGCCTTA
ENSMUSG00000001504	13	0	88.95	0	AGGCTTTCAGTGTGGCAGTGGCGTC	GTGCGGCCTATCCATACCAGCTCA
ENSMUSG00000001542	13	0.61	333.68	0.5615	GATGACCCTGTCCCTGTATGGCCTC	GTCAACGCCCATGAACCCTGCAA
ENSMUSG00000005583	13	0.01	151	0.0197	TCTGAGTTTGTCCGGCTCTCGTGCG	TGAGCGTGCTGTGCGACTGTGAGA
ENSMUSG00000006711	13	0.02	144.85	0.0007	GGGCCTGCGAACCACTCATAGAGA	ATCCGTGATGCCGTGGACTACCCA
ENSMUSG00000006717	13	2.38	1235.85	1.1122	CATGCTGCTCATCGTGGACGCAAA	CGGGCTTGCAGACTTGACCTTCCA
ENSMUSG00000019726	13	0.04	12.75	0.0972	TTAGTGAAAGGCGGCATCCTGACCA	TGGAGGAGGAGTATCGCAAGGGAGC
ENSMUSG00000021147	13	0.06	170.75	0.0778	TTCCCACCAACCAGCCAGTCAGCA	TGTCCTACTGTCCGAGTCCCACTGA
ENSMUSG00000021196	13	0.04	32.75	0	CCAGAAGTCCGCTCCACTCCTTTCG	AACCAATCTGTGCGTGATCGGCGGG
ENSMUSG00000021216	13	0	213.75	0.0005	GGCATTGTACGGTTCTACCACGGCA	GGAGGAGGCACAGGGTCGGGATTTA
ENSMUSG00000021314	13	0	233.3	0	TGCAGACAGCGAGGCATCCACAGCA	GCGGGTGAGAAGCAGGAAGTGGCTA
ENSMUSG00000021356	13	0	246.6	0	TGTCGTCCGGGTAGGGAAACAGGA	CGGGACATCCTCGTGAAAGAGCTGA
ENSMUSG00000021359	13	0	77.28	0	GCGTTAAGACACTCGGGCGGTGAGA	AAGGACAACCTCTTCGGTGGCGTGG
ENSMUSG00000021363	13	0	109.5	0	CCGTTCAGAGGAAGATGGCTTTGGC	TCCAAGTCGGTCAGGTACTGGGCTC
ENSMUSG00000021365	13	0.04	52.65	0.0402	GCTCAGAACCGTCACTGCTACCGCA	AAGCGAGTACACACATCCGGGCTCC
ENSMUSG00000021366	13	0.04	277.55	0.0407	TTCTGACGCTGGGCATGGATCTGC	AGGGTTGCTCTGCTCCTCTGCGTTC
ENSMUSG00000021367	13	0.01	346.2	0	ACCCAGATGATGTCCAGGTGGCAG	AGCTGAGAATGGAGTGCAGAGCCC
ENSMUSG00000021374	13	0.17	238.5	0.1798	ACTGGACACTCCCAAGAGCCCACAG	GGCACTGGTGTGAAGCGAGCCCTTA
ENSMUSG00000021385	13	0.11	267.7	0.0546	ACTCCTCCAGGTATTGCTCAACCCG	ACCCTCCAGGTGCAGATGTTGGAC
ENSMUSG00000021396	13	0	49.85	0	ACATCGGCTGCTTCCACCCAGTTC	GCTGTCATCACCAACAAAGGGCGG
ENSMUSG00000021400	13	0.63	188.18	0.2103	TCTCATTGGCCTCCAGTAGCGAGCG	GGCTGACAAGATGCGTCCAGACACG
ENSMUSG00000021416	13	0.69	31.95	0	TGCCCGCTCTGGTGGTCTTCATAG	AGCCCTGGAAAGGATGCCGAGAGA
ENSMUSG00000021418	13	0.28	180.55	0.1249	GCAGAAGACAGCACCAAGCCAGTCA	TCAGGAGCATCAGCCCAAAGTGGC
ENSMUSG00000021423	13	0.15	211.9	0.1098	ACTGTTGCAGGCCGTGTGCTTG	AGCTGCCCTCCTTGTGTGGATTCTG
ENSMUSG00000021428	13	0.07	311.5	0.0811	ACTGCCGCCATTCCAGGTACAACCC	ACGCTGAGACCGATGAGGAAGACGA
ENSMUSG00000021448	13	0	223.45	0.0012	TGGGCAAGCTCAGGGTTTCCTCTCC	AGGGTCTTTGACAGCATCAGCCACC
ENSMUSG00000021458	13	0.04	129.2	0.013	AGCCGGGAGAAAGGGTGTGTTCCCA	TGTGCCTTGAGGGTGCCTGCCAAGA
ENSMUSG00000021466	13	0.11	322.98	0.0808	GCACCCTGCTGTGCTTCGTATTGCC	TCCAAGTGTCGTCCGGTTTGCCGTG
ENSMUSG00000021469	13	0	41.9	0	GGTTCAGAGAGCTGGAGAACTCGGC	AAGACCAACCGGAAGCCACGCACA
ENSMUSG00000021474	13	1.26	623.65	0.7241	TCCGGCCTCCAGCACTTGATCTTG	TGCAAGCCAGCATCCAAAGGACTCA
ENSMUSG00000021481	13	0.04	7.95	0.0163	AAGTCAGAGACAGCAGGGTCCGCCA	GGCAAGAGACACAGGAAGCAGGAGA
ENSMUSG00000021485	13	0.01	65.2	0.0083	TGCTGTGTGAGTAGCTGTGCTCCCG	TGTGGAGAGCCTGGTGTTTGGGAC
ENSMUSG00000021494	13	0.4	57.9	0.1578	CCGGGAGGCTGGCTTCTTGTCTTGA	GGTCAGCAACATTGGCCGCAAGGA
ENSMUSG00000021495	13	0.09	551.6	0.044	TCCATGCTCTTTCGGCAGCTCTGG	GGACTGGTGTTGCAGGGTGAGAAGC
ENSMUSG00000021499	13	0.01	18.35	0.0011	AGCCTGTTGACCTCCTCTTGCTGGC	TGAGCGGGATCTGACGTTGGAGAG
ENSMUSG00000021501	13	0.46	509.9	0.127	CGCGGTACTTCCATCTTCCTGGCT	AGGGCTTCTCACCACGGCCTAGAAC
ENSMUSG00000021518	13	0.79	730.05	0.3068	TCCTCTCGGGTCTAGGCACTCTGGA	ACAGCCCACCGAAGCATTCAAGCA
ENSMUSG00000021534	13	0	171.3	0	TGGTGCCTTGGTGGGATGTAGCTG	TCCGAAACCCTGCGTGGCAAAC
ENSMUSG00000021540	13	0.07	209.45	0.3034	AGGAGCGTTGTTGGGTTGGTGGA	GGAACCTGAGCCACAATGAACCGCA
ENSMUSG00000021541	13	0	120.5	0.0003	TCCGGTGGAAGTGAGACGTTGTGC	ACTGCGAGGTTCATGGCTTTCCTCA

Nature Genetics: doi:10.1038/ng.711

ENSMUSG00000021557	13	0.09	164.38	0.058	AGCTTCCCAAGATGTACCGTCCGCC	GCCTCGGGCTTGAACGCAGGAATGA
ENSMUSG00000021559	13	0.31	535.65	0.296	AGGTGGTGAGACGGAGCCACGAACA	TGAGCGTAAGGAGCCGCAGCATGA
ENSMUSG00000021573	13	0.07	59.6	0.0497	CCAGCTTTGCCCTTGCCCTTTCCTG	TGTCACGGCTCACAGACACAAGCA
ENSMUSG00000021577	13	6.33	1975.1	2.4108	GTCCGATCAGCCACACAGCAACACC	CAGCGTGCATTTGGTGGACAGAGCC
ENSMUSG00000021587	13	0	114.8	0.0014	TCGGCTGTTCACCATCAAGCCTGC	GCAGCATCTGGTTGTCTGGACCTCT
ENSMUSG00000021589	13	0.04	159.8	0.0501	TGGCCGCGTTCAGCAACAAGCTAC	AGGGCGACACGTTCCACCAGGACAA
ENSMUSG00000021591	13	1.1	474.65	0.6883	CGCACTGGTGTTGTTAGTGGCTGTG	TCGTGTTCATCAAGCCCACCTGCCC
ENSMUSG00000021595	13	1.1	112.05	0.524	TCCAGGTCCGTTGGTGGGAACATGG	ACTAGAGACGGGCAGTGGTTTGCAG
ENSMUSG00000021597	13	0.02	78.05	0.014	GCAGATCCACCTCTGGACAACGCT	ATGCTGGCTGGACGCCTTTGCAT
ENSMUSG00000021598	13	0.35	600.15	0	TGACGATGATGCCGAGCTGCCGAA	GGCGATGGCGGAGAAGTTTGACCA
ENSMUSG00000021600	13	0	170.6	0.0004	TCATGGCAGGGCTTACAGCAGCGA	TTGGCGTTCTAGCGGTGTCTGGGA
ENSMUSG00000021604	13	0	14.3	0.0003	TCGTCGCAGATACACCCGTTGGAC	TCTGGTCCTTGGCACACACAGCCA
ENSMUSG00000021607	13	0.88	978.75	1.0503	AGCAGGAACGAGGAGAAGACTCGGG	TGCTTGTACGGAGCGTGGTGGCTTC
ENSMUSG00000021608	13	0.02	223.9	0.0043	GGAGGAATGTGGAGCCAAGGTGAGG	ATGTGGTTTGCTGGCGGCTTCC
ENSMUSG00000021610	13	1.44	1513.85	1.116	TGCAGGCAGGGAGGAACCATCAA	GCCGAGAAGAAGCCATCAAACGCCC
ENSMUSG00000021611	13	0.02	26.2	0.0129	CACCTTGCAGAGACAGGCCCGAAGA	AGATGCCTTGAACACCAGCCCACCG
ENSMUSG00000021614	13	0	102.25	0.0009	ATCCGACAGCCAGCCGTAATCGCA	TCGAGGAGGCCGAAGCAGAGTGTA
ENSMUSG00000021622	13	0	220.25	0.0008	TCTCCACCTCCCTTCTCTCTGCTCG	TGTTGTCCTCTCGGGTACGCACTGG
ENSMUSG00000021629	13	0.44	206.43	0.3001	ACGTGCTCTGTGCTCTCCTGGTGT	TGTGCCCGCTATGGGTCCTTTCCCA
ENSMUSG00000021636	13	0.12	162.8	0.1458	TTTGCCCTTGCTGGAGAGGCTGGA	GCCGGTAAGACGCTTTGTTCCCGA
ENSMUSG00000021638	13	0.17	142.55	0.1715	TAGTGGTCAGGGTCCGTCCTCTTGC	AGCCTCGGTACAGCAGCAATGGTAA
ENSMUSG00000021647	13	0	12.4	0	AGCAACAGGGAAAGAGCCCATCCGC	AGTGCTTGTGAAGGGACGACAGCC
ENSMUSG00000021666	13	0.46	194.8	0.1898	GGTGCAGAAGAAGACAGGCTCTGGG	AGCCGGACGAGGAGAAAGGGAGCAT
ENSMUSG00000021670	13	0.57	291.9	0	TCTCACCACCTTGGCTGGAATGACG	GCAGTCAGTGGGAACTATTGCACCG
ENSMUSG00000021676	13	1.87	299.08	1.3033	AGCTTGTTGGAGGCTGCGTGTTGC	CATTGTGGCTCCCGATGGCTTTGAC
ENSMUSG00000021680	13	0	132.55	0.0003	TCGGTGACATGCTGGCCTTGTGTC	TCAGTGTGGGCGAAGGTGTGGGAA
ENSMUSG00000021681	13	0.28	107.4	0.1457	CGGCGTCCTCCTGTCTATGGCTAAC	AGTCAGTGTGCTCAGGCTTCCGCTC
ENSMUSG00000021686	13	0.4	1187.7	0.2493	AGATGCAGGGCTGTCTTGGGAAGC	CAGTACGCCTGTCTTCGTGCCAAC
ENSMUSG00000021696	13	0	205.95	0.0048	AAGGGCTTCCGGTTCTCCATGAGC	TGTCCTCGCCTCTGCCACAAACCA
ENSMUSG00000021697	13	0	429.05	0.0008	AGCCACACGCTCTTCTATGGAGGTC	TCCCGGTGCATCCTGTGCTCTAAGG
ENSMUSG00000021701	13	0.09	138.6	0.2365	AAGGCTGCTGCGGTCGCTTCTTCT	GCTCCTGCGGACTATCACCTACCAG
ENSMUSG00000021703	13	0.22	286.9	0.2	ACGCAGGGTGAGATGGCTACCAGA	TCCGTTGCTGTTGGAGGCTTGGCT
ENSMUSG00000021706	13	0.04	76.28	0	AGACTAGAGTTGGAGGTCCCGCCGT	GGCCAGCTTCTGCCTCAGCATTACC
ENSMUSG00000021754	13	0.04	268.25	0.0759	TGTCTCTTCGCTGGGTATGACGGC	CCGTTCCGGGCAGTACAAGCAAAC
ENSMUSG00000021756	13	0.69	389.25	0.5538	AGACGGCCCAGGTGTGACTTTGT	CCGCAGTGAAGGACTGGCTCCTGAA
ENSMUSG00000021759	13	0.26	301.5	0.1072	AGCTGTTGCCTTTGAAGCGGGCCA	ACACGACTCTCCATGAAACCGCCA
ENSMUSG00000021760	13	0.02	198.55	0.0173	AGCTGTTGGTTCTCGGCTTCAGGA	GGAGCCTTTCGCTGCCTACCCATTA
ENSMUSG00000025868	13	3.7	1654.05	2.1814	TGGCTACGACCGTGAAACCCTGTGC	TGCCCTCACCTATGGCCTTTACTGC
ENSMUSG00000025869	13	0.71	217.35	0.5645	TCCGACGAGCATTCCGGTTCAGACG	GCGAGATGCCCAAGGCTAAGGGAAA
ENSMUSG00000025876	13	0.01	18.55	0.0083	TGCTTTGCTGGGCTTGATGCTGAC	ATCTTGCTGCTGCTGGTCCTCGTCC
ENSMUSG00000032621	13	0.09	218.1	0.0531	AACGCTTTGGGTTTGATCCCGCC	ACAGCACAATGGGAACTGCCAACCA
ENSMUSG00000033781	13	0.29	533	0.2474	ACGTGTAGTCCGAGGGCTTCTTGCC	GGGCCAATGTGAACGCAGCCAAA
ENSMUSG00000034152	13	0.39	22.65	0.2654	TGTCCCGTTTGTTCCCACTGTCCA	TCCTGCAAGCCCACCGGAAGCTAA
ENSMUSG00000034334	13	0.09	75.05	0.0533	ACTTTGCTGCTCCCATTTGGCCCG	AGTCAGAGCGTCCATGCTGTTCCTG
ENSMUSG00000034488	13	0	75	0	TTGTGACTGGTCGTTATGGCCGGA	AGATCACTGCCTCCAGCGTCTTCAG
ENSMUSG00000034525	13	0.07	45.95	0	ACAGCAGGGCTCAGTGTGTTCCCA	AAGCAGGCAAGACCAAGGAGGGTGC
ENSMUSG00000034617	13	0.06	50.9	0.0264	AGTCTTCGATCCGACCCTGCTGACC	TGCCATGAAGCCGGAGTTGTTGC
ENSMUSG00000034686	13	0	187.55	0	GGGCACGCAAGTTCTGCTCCCTTA	ATCGTCCTGCTCTGCTGCTTCTGC
ENSMUSG00000034751	13	0.01	31.48	0.0146	TGGAACTGCTGATGGTGCTGGCTGG	TGAGGCTGTCCCTTGCTCTGGAGAC
ENSMUSG00000034789	13	0.64	139.88	0.2046	ACGCTTGACCATTCCAGCCCAGACA	AGGTGTGGATTTGAGCCAGAAGGCA

Nature Genetics: doi:10.1038/ng.711

ENSMUSG00000034928	13	0.45	660.45	0.7322	TCAACCATGAACGGGCTCTGCTGG	TGCTGCCCAGACCATCAAACCTCTC
ENSMUSG00000035248	13	0.33	456.35	0.3351	TGCAGTTCTGGGTGCTTGCTGCTG	TGTGAGGGTTTGGAGGATGCCACA
ENSMUSG00000035493	13	0.53	463.05	0	AGGATGCGGTTCAAGGTCTCGGCA	ACTCAATACCGTGCTGGAGGGCGA
ENSMUSG00000035711	13	0.02	178.15	0.0054	AACAGGTGGCTTCGCTCAGTGGTGG	TATCTGTCCTGCCTGCGGATGGCGA
ENSMUSG00000035834	13	0.08	230.5	0.0786	TGGGTCCAAACGCTCGAAGCTGA	GCGACGACAACATGGATGAAGCCAC
ENSMUSG00000036006	13	0.01	14.15	0.016	GCCTCGGCCACACTACTCTCCAGAA	ACCATGAGCCCAAGTCCCACCCAA
ENSMUSG00000036211	13	0	28.8	0.0014	TTGCCCTTGCCCTTGTCGTTCCCA	ACAGCCGTATCAAGCTGGCCCTCA
ENSMUSG00000036376	13	0.11	366.3	0.1403	TGGTCCTCGACATTAGCCAGCCGTG	ATTCACCCAGCGTCCCACTGAGCA
ENSMUSG00000037851	13	0.35	379.7	0.3665	TGTCAACGTGAAACCCGCTCTGGTG	CCTCCCTTTGCTACTGGACTGCCTC
ENSMUSG00000037933	13	0.08	100.05	0.0332	ACTGGGTGTTGGTGAGGACGTTGCG	ACAAGAAGGTGGCTGCTGATGGTGA
ENSMUSG00000038025	13	0.15	81.1	0.081	TGCTCCTCGTAGTCCTCCAGATCCA	GGTGGCAACAAAGGCACAGGCAAGC
ENSMUSG00000038042	13	0.01	104.53	0.0046	AGCCACAGTTCAGAGGCTACAGGG	GAGCCAGTGCCGCCCAACTTCACAA
ENSMUSG00000038152	13	0.02	107.2	0.0251	GCTGCAAAGAGAACGCGGCACTGA	CCGGCTACACCTCAGCCAAGCGTAA
ENSMUSG00000038267	13	0.48	238.58	0.3262	AGCTGTAGAAGTGAGGCCAGGGCA	GTCTGGTAGTCAAGTTCCTCGGGCG
ENSMUSG00000038518	13	0.05	182.83	0.0546	TGCCATGTGCTTGTTTGCCCAGC	ACGTGAAGAAGGAAGTGCCCGAGCG
ENSMUSG00000038546	13	0.42	371.2	0.7321	TCTGCGAGAGGCCGATGTAGCTGA	TACCCGGCTGTGGATGAGCAGGAGA
ENSMUSG00000038732	13	0	179.05	0.0046	TTGGGCTTTGCGGCTTGCATGA	ACGCTCTCCAAGTCCTTTCCCGTC
ENSMUSG00000039109	13	0	50.95	0.003	GAAGACTGAACGGAAAGTCGCACCG	AAAGAGCTGCAAAGCGGAAAGTGGG
ENSMUSG00000039182	13	0.3	270.8	0.1275	TGGCTTTCAGGGTGTCTCCTGTCCC	ACCAGCCTTCTGAGCGTGCCTTGA
ENSMUSG00000039309	13	0	166.68	0	CCGGAGAAGCCATGAAGACCAGGAG	CGCATTACCACGCAAGAGGTTGGG
ENSMUSG00000041236	13	0.54	459.2	0.2449	ACACGTCACTCAACAGCACTGGCG	TGCTAAGAACCGTGGGCCTGGAAG
ENSMUSG00000041297	13	0.16	209.03	0.1733	GCATGGGAGGCAAAGGTAATGGCGG	AGCCACTCACACCAAGCACAGGAGC
ENSMUSG00000046957	13	0	102.8	0.0005	ACGGGCGTTGAAGCGGTTATTCC	TAGCGGAAGCAGACAAGACGGACGC
ENSMUSG00000049625	13	0.02	417.1	0.017	AGCTCTTGCGGGTCAGCACCTTGA	GGCTGTTGGTTGGGCGAGGACAGAA
ENSMUSG00000050244	13	0.12	245.1	0.0972	GTTCACTGCTTTGGTCTGGACGCTG	GGAGCTTCTTGGGATTGACCCTGCC
ENSMUSG00000053181	13	0	28.6	0.0006	AACGCCCGAGGAAGGAGCCACTGAA	AAGGGAGACCCGAGCAAGACCTGGA
ENSMUSG00000056257	13	0	265.45	0	GGCACTCCTGGCTAGATGCAGTCCT	AAGAGTCCACAGCGGTCAGGGATCA
ENSMUSG00000056749	13	0.35	75.85	0	ACCTCTGACACATCGGAGAGCGAGC	TTGTGGACGAGCATGAGCCTGCGA
ENSMUSG00000063529	13	0	33.3	0	CCTCTTCAGTGGCTCTCCTTCCTGC	TGACCGGCTTCTGCCCACAGCTAA
ENSMUSG00000071451	13	0.56	24.2	0	AGGTTGCGTAGATGCGGCGTTG	TTCAGTGCAAGGCTGTGGGAGCA

[1] Expression signals. Note that different methods use different expression level units.

Nature Genetics: doi:10.1038/ng.711

Table S3. Comparison of RNA-Seq gene expression levels of X-linked and autosomal genes

Tissues/stages	*P-value*[1]	X:AA (median)	X:AA[2] (Miller's jackknife)	X:AA[3] (Mann-Whitney)
Human				
(734 X-linked genes and 19066 autosomal genes)				
Brain	4.1E-04	0.63	0.79 (0.73, 0.84)	0.67 (0.59, 0.77)
Heart	1.6E-07	0.56	0.63 (0.59, 0.67)	0.51 (0.45, 0.59)
Liver	8.9E-12	0.34	0.53 (0.49, 0.57)	0.36 (0.31, 0.42)
Muscle	1.4E-08	0.42	0.52 (0.48, 0.56)	0.41 (0.36, 0.48)
Testis	8.1E-05	0.70	0.77 (0.69, 0.85)	0.71 (0.63, 0.83)
Kidney	1.0E-09	0.49	0.75 (0.70, 0.80)	0.50 (0.43, 0.59)
Breast	3.3E-12	0.42	0.58 (0.55, 0.63)	0.43 (0.38, 0.50)
Adipose	1.0E-09	0.45	0.57 (0.53, 0.61)	0.43 (0.37, 0.50)
Colon	8.3E-10	0.47	0.63 (0.58, 0.68)	0.45 (0.40, 0.53)
Lymph node	6.1E-12	0.47	0.56 (0.52, 0.60)	0.43 (0.38, 0.50)
Cerebral cortex	3.5E-07	0.58	0.76 (0.71, 0.82)	0.57 (0.50, 0.67)
Lung	1.6E-16	0.42	0.56 (0.52, 0.60)	0.42 (0.37, 0.48)
Average[4]	1.9E-15	0.49	0.64 (0.59, 0.70)	0.45 (0.40, 0.53)
Mouse				
(811 X-linked genes and 20659 autosomal genes)				
Liver	1.3E-28	0.13	0.21 (0.20, 0.24)	0.18 (0.16, 0.21)
Brain	2.5E-21	0.25	0.47 (0.44, 0.50)	0.27 (0.24, 0.32)
Muscle	6.0E-24	0.20	0.34 (0.32, 0.37)	0.24 (0.21, 0.28)
C. elegans				
(2729 X-linked genes and 16778 autosomal genes)				
L2	6.3E-09	1.29	0.92 (0.85, 0.99)	1.25 (1.11, 1.43)
L3	3.7E-01	0.90	0.84 (0.81, 0.88)	0.95 (0.83, 1.11)
L4	1.0E-14	0.54	0.69 (0.66, 0.72)	0.65 (0.59, 0.71)
Adult	3.0E-23	0.57	0.41 (0.39, 0.43)	0.56 (0.50, 0.63)

[1] From Mann-Whitney's U test of the equality of expression levels between X and autosomes.

[2] Numbers in parentheses show 95% confidence intervals.

[3] Numbers in parentheses show probable ranges (see Online Methods).

[4] Average of eight non-brain non-sex-specific tissues.

Nature Genetics: doi:10.1038/ng.711

Table S4. X:AA ratios when individual autosomes are considered. Bootstrap is used to compare the median expressions, with 95% confidence intervals shown in parentheses.

	Gene no.	Brain(M)	Heart(M)	Liver(M)	Muscle(M)	Testis(M)	Kidney(M)	Breast(F)	Adipose(F)	Colon(F)	Lymph node(F)	Cerebral cortex(NA)	Lung(NA)	Average[1]
chr1	2006	0.68(0.55,0.97)	0.54(0.36,0.77)	0.31(0.19,0.43)	0.39(0.25,0.55)	0.69(0.53,0.87)	0.49(0.33,0.69)	0.39(0.3,0.56)	0.53(0.34,0.72)	0.47(0.35,0.67)	0.45(0.28,0.59)	0.58(0.44,0.74)	0.4(0.27,0.55)	0.46(0.35,0.6)
chr2	1269	0.6(0.43,0.92)	0.56(0.38,0.85)	0.38(0.28,0.6)	0.44(0.25,0.66)	0.77(0.61,0.98)	0.43(0.3,0.61)	0.52(0.37,0.73)	0.53(0.33,0.75)	0.51(0.35,0.69)	0.57(0.41,0.75)	0.59(0.43,0.72)	0.46(0.34,0.63)	0.52(0.43,0.63)
chr3	1077	0.54(0.4,0.73)	0.43(0.28,0.58)	0.33(0.23,0.51)	0.38(0.23,0.59)	0.77(0.58,0.91)	0.39(0.24,0.53)	0.39(0.3,0.51)	0.4(0.24,0.56)	0.5(0.38,0.64)	0.49(0.31,0.63)	0.55(0.41,0.67)	0.41(0.29,0.56)	0.47(0.37,0.61)
chr4	744	1(0.71,1.39)	0.83(0.54,1.3)	0.54(0.36,0.82)	0.85(0.51,1.3)	1.14(0.91,1.47)	0.47(0.33,0.74)	0.88(0.67,1.23)	0.93(0.63,1.49)	0.89(0.58,1.24)	0.87(0.59,1.27)	0.96(0.7,1.28)	0.71(0.46,1.01)	0.71(0.52,0.95)
chr5	858	0.55(0.41,0.77)	0.46(0.29,0.68)	0.42(0.28,0.74)	0.49(0.3,0.79)	0.94(0.7,1.17)	0.5(0.36,0.74)	0.44(0.34,0.6)	0.49(0.31,0.71)	0.53(0.38,0.76)	0.61(0.37,0.76)	0.58(0.43,0.71)	0.56(0.35,0.76)	0.55(0.44,0.72)
chr6	1068	0.84(0.57,1.26)	0.62(0.37,1.09)	0.52(0.36,0.76)	0.64(0.39,0.93)	1.04(0.75,1.34)	1.36(0.78,1.95)	0.67(0.49,0.89)	0.7(0.4,0.96)	0.71(0.52,1.02)	0.62(0.41,0.83)	0.84(0.57,1.03)	0.54(0.33,0.8)	0.64(0.47,0.87)
chr7	941	0.63(0.44,0.82)	0.6(0.37,0.84)	0.33(0.21,0.53)	0.44(0.25,0.67)	0.79(0.6,0.95)	0.47(0.34,0.64)	0.41(0.27,0.56)	0.47(0.28,0.71)	0.48(0.32,0.65)	0.49(0.35,0.66)	0.52(0.38,0.67)	0.46(0.31,0.64)	0.57(0.39,0.7)
chr8	687	0.65(0.48,0.92)	0.54(0.31,0.89)	0.42(0.24,0.69)	0.45(0.25,0.68)	0.86(0.67,1.04)	0.56(0.34,0.81)	0.49(0.36,0.7)	0.54(0.31,0.71)	0.51(0.36,0.79)	0.51(0.32,0.66)	0.57(0.42,0.81)	0.53(0.34,0.74)	0.56(0.42,0.7)
chr9	795	0.67(0.45,0.88)	0.45(0.32,0.68)	0.36(0.23,0.56)	0.38(0.25,0.61)	0.59(0.46,0.75)	0.5(0.32,0.69)	0.38(0.29,0.49)	0.42(0.26,0.66)	0.4(0.27,0.57)	0.44(0.3,0.59)	0.56(0.36,0.73)	0.38(0.27,0.57)	0.50(0.36,0.64)
chr10	772	0.63(0.44,0.91)	0.55(0.34,0.79)	0.33(0.21,0.55)	0.36(0.21,0.66)	0.72(0.55,0.87)	0.46(0.32,0.7)	0.5(0.33,0.71)	0.62(0.33,0.92)	0.65(0.44,0.94)	0.6(0.35,0.84)	0.58(0.4,0.75)	0.49(0.33,0.73)	0.48(0.35,0.6)
chr11	1299	0.89(0.65,1.37)	0.81(0.49,1.36)	0.49(0.29,0.76)	0.66(0.41,1)	0.78(0.62,1.02)	0.68(0.41,1.02)	0.45(0.31,0.65)	0.56(0.36,0.85)	0.53(0.37,0.82)	0.59(0.41,0.87)	0.73(0.54,0.96)	0.55(0.28,0.76)	0.56(0.38,0.74)
chr12	1037	0.56(0.4,0.73)	0.5(0.33,0.82)	0.3(0.2,0.5)	0.37(0.23,0.57)	0.67(0.5,0.83)	0.42(0.24,0.67)	0.39(0.31,0.57)	0.4(0.25,0.56)	0.43(0.29,0.6)	0.44(0.26,0.57)	0.57(0.38,0.73)	0.41(0.27,0.58)	0.46(0.36,0.59)
chr13	338	0.63(0.39,0.91)	0.52(0.32,0.72)	0.39(0.22,0.61)	0.44(0.26,0.73)	1.03(0.67,1.28)	0.44(0.29,0.63)	0.64(0.48,0.84)	0.75(0.49,1.1)	0.73(0.43,1.1)	0.71(0.43,1.04)	0.7(0.49,1.03)	0.65(0.37,0.88)	0.65(0.44,0.87)
chr14	624	0.51(0.37,0.72)	0.42(0.24,0.61)	0.25(0.16,0.38)	0.29(0.18,0.45)	0.56(0.4,0.74)	0.38(0.27,0.51)	0.34(0.24,0.48)	0.35(0.26,0.51)	0.42(0.29,0.58)	0.38(0.26,0.5)	0.47(0.3,0.55)	0.4(0.31,0.55)	0.46(0.34,0.59)
chr15	601	0.63(0.43,0.82)	0.5(0.31,0.73)	0.36(0.2,0.64)	0.34(0.19,0.54)	0.65(0.49,0.82)	0.46(0.29,0.69)	0.4(0.29,0.59)	0.42(0.29,0.62)	0.43(0.31,0.62)	0.48(0.3,0.69)	0.53(0.35,0.7)	0.45(0.3,0.64)	0.52(0.34,0.66)
chr16	824	0.43(0.33,0.59)	0.45(0.28,0.65)	0.19(0.13,0.29)	0.22(0.11,0.32)	0.36(0.26,0.46)	0.38(0.26,0.57)	0.2(0.15,0.26)	0.24(0.14,0.34)	0.23(0.17,0.31)	0.24(0.17,0.34)	0.34(0.26,0.42)	0.24(0.17,0.34)	0.31(0.24,0.38)
chr17	1181	0.55(0.42,0.76)	0.4(0.25,0.57)	0.23(0.13,0.34)	0.25(0.13,0.38)	0.42(0.33,0.51)	0.41(0.28,0.63)	0.23(0.18,0.32)	0.24(0.15,0.33)	0.26(0.14,0.34)	0.28(0.18,0.38)	0.41(0.31,0.5)	0.29(0.21,0.4)	0.35(0.29,0.45)
chr18	282	0.71(0.44,1.03)	0.8(0.47,1.46)	0.61(0.36,1.07)	0.71(0.31,1.22)	1.1(0.8,1.34)	0.68(0.35,1.03)	0.68(0.47,0.94)	0.71(0.51,1.01)	0.87(0.6,1.23)	0.86(0.5,1.31)	0.67(0.44,0.94)	0.6(0.38,1.01)	0.81(0.48,1.19)
chr19	1390	0.73(0.52,1.03)	0.72(0.44,1.06)	0.39(0.25,0.63)	0.6(0.4,0.89)	0.56(0.42,0.72)	0.6(0.41,0.83)	0.37(0.23,0.52)	0.39(0.22,0.59)	0.41(0.26,0.53)	0.34(0.21,0.49)	0.53(0.38,0.72)	0.33(0.22,0.49)	0.48(0.34,0.62)
chr20	568	0.65(0.4,1.04)	0.69(0.48,1.02)	0.43(0.2,0.74)	0.53(0.24,1.12)	0.46(0.33,0.58)	0.78(0.39,1.23)	0.37(0.26,0.53)	0.41(0.24,0.65)	0.39(0.29,0.54)	0.39(0.27,0.62)	0.52(0.37,0.67)	0.38(0.26,0.64)	0.52(0.39,0.72)
chr21	232	1.03(0.66,2.49)	1.18(0.6,3.3)	0.62(0.31,1.19)	1.37(0.46,5.41)	1.55(0.82,2.17)	0.81(0.4,1.53)	0.76(0.38,1.38)	1.03(0.36,2.6)	1.15(0.52,1.96)	1.08(0.45,2.09)	1.06(0.58,1.74)	0.74(0.46,1.82)	0.91(0.56,1.23)
chr22	473	0.42(0.29,0.56)	0.39(0.24,0.77)	0.22(0.12,0.33)	0.27(0.16,0.48)	0.35(0.25,0.47)	0.43(0.29,0.72)	0.21(0.16,0.3)	0.22(0.14,0.33)	0.26(0.17,0.4)	0.22(0.14,0.31)	0.28(0.21,0.37)	0.27(0.17,0.4)	0.33(0.25,0.49)

[1] Average of eight non-brain non-sex-specific tissues.

Nature Genetics: doi:10.1038/ng.711

Table S5. X:AA ratios for human genes of the same functional categories based on Gene Ontology (GO) classifications. Only GO categories with at least 50 X-linked and 50 autosomal genes are compared. Bootstrap is used to compare the median expressions with 95% confidence intervals shown in the parentheses.

				# of genes (A)	# of genes (X)	Brain(M)	Heart(M)	Liver(M)	Muscle(M)	Testis(M)	Kidney(M)	Breast(F)	Adipose(F)	Colon(F)	Lymph node(F)	Cerebral cortex(NA)	Lung(NA)	Average[1]
GO:0003674	molecular_function	GO:0005488	binding	12053	451	0.62(0.47,0.78)	0.55(0.39,0.68)	0.45(0.31,0.56)	0.35(0.2,0.52)	0.5(0.39,0.62)	0.45(0.32,0.61)	0.38(0.29,0.59)	0.43(0.26,0.58)	0.42(0.27,0.53)	0.42(0.35,0.5)	0.52(0.4,0.68)	0.4(0.28,0.51)	0.4(0.3,0.49)
GO:0003674	molecular_function	GO:0003824	catalytic activity	5383	199	0.66(0.49,0.84)	0.6(0.44,0.88)	0.68(0.33,0.91)	0.56(0.32,0.92)	0.45(0.36,0.64)	0.54(0.36,0.73)	0.48(0.29,0.65)	0.61(0.36,0.85)	0.5(0.33,0.71)	0.54(0.34,0.81)	0.63(0.42,0.77)	0.6(0.37,0.75)	0.51(0.38,0.74)
GO:0003674	molecular_function	GO:0030528	transcription regulator activity	1451	55	0.52(0.18,1.76)	0.39(0.14,1.02)	0.66(0.03,1.27)	0.32(0.06,0.89)	0.86(0.52,1.38)	0.74(0.22,1.29)	0.48(0.19,1.24)	0.27(0.07,0.63)	0.39(0.11,0.65)	0.44(0.12,0.67)	0.69(0.28,1.27)	0.36(0.1,0.79)	0.36(0.15,0.95)
GO:0003674	molecular_function	GO:0005215	transporter activity	1394	55	0.54(0.15,1.16)	0.8(0.2,1.73)	0.44(0.08,1.2)	0.92(0.18,1.78)	0.53(0.24,1.13)	0.38(0.11,0.75)	0.42(0.25,1.52)	0.74(0.35,1.64)	0.65(0.24,1.09)	0.73(0.24,1.15)	0.39(0.14,1.26)	0.66(0.15,1.45)	0.31(0.16,0.65)
GO:0005575	cellular_component	GO:0044464	cell part	13898	553	0.64(0.49,0.81)	0.56(0.36,0.72)	0.39(0.23,0.5)	0.41(0.22,0.55)	0.57(0.47,0.69)	0.48(0.33,0.59)	0.36(0.28,0.51)	0.43(0.29,0.57)	0.47(0.3,0.54)	0.46(0.34,0.55)	0.52(0.41,0.63)	0.4(0.29,0.49)	0.47(0.33,0.53)
GO:0005575	cellular_component	GO:0043226	organelle	7816	377	0.54(0.38,0.72)	0.39(0.26,0.44)	0.33(0.22,0.49)	0.28(0.15,0.51)	0.45(0.36,0.61)	0.42(0.28,0.55)	0.37(0.17,0.5)	0.33(0.2,0.45)	0.36(0.26,0.47)	0.34(0.25,0.42)	0.5(0.38,0.7)	0.36(0.27,0.5)	0.38(0.3,0.45)
GO:0005575	cellular_component	GO:0044422	organelle part	3801	148	0.73(0.54,0.88)	0.54(0.32,0.87)	0.44(0.29,0.75)	0.6(0.38,0.76)	0.44(0.28,0.66)	0.5(0.34,0.71)	0.66(0.33,0.81)	0.55(0.41,0.96)	0.53(0.35,0.92)	0.51(0.33,0.89)	0.64(0.44,0.87)	0.55(0.34,0.81)	0.47(0.33,0.76)
GO:0005575	cellular_component	GO:0032991	macromolecular complex	2462	85	0.48(0.24,0.73)	0.36(0.19,0.54)	0.47(0.28,0.67)	0.36(0.13,0.58)	0.32(0.23,0.49)	0.44(0.25,0.67)	0.3(0.17,0.45)	0.36(0.17,0.49)	0.37(0.21,0.46)	0.34(0.21,0.58)	0.39(0.17,0.56)	0.43(0.25,0.52)	0.31(0.2,0.44)
GO:0008150	biological_process	GO:0009987	cellular process	10987	397	0.66(0.49,0.92)	0.62(0.41,0.72)	0.5(0.34,0.79)	0.45(0.32,0.66)	0.52(0.41,0.64)	0.57(0.39,0.73)	0.43(0.28,0.61)	0.52(0.37,0.69)	0.51(0.26,0.61)	0.45(0.33,0.53)	0.56(0.4,0.75)	0.45(0.31,0.6)	0.47(0.41,0.57)
GO:0008150	biological_process	GO:0065007	biological regulation	7181	275	0.62(0.38,0.82)	0.47(0.25,0.73)	0.4(0.27,0.73)	0.38(0.24,0.72)	0.59(0.41,0.68)	0.4(0.2,0.59)	0.34(0.28,0.55)	0.38(0.21,0.6)	0.35(0.24,0.57)	0.36(0.19,0.47)	0.53(0.3,0.64)	0.34(0.25,0.47)	0.4(0.26,0.54)
GO:0008150	biological_process	GO:0008152	metabolic process	7002	266	0.61(0.47,0.77)	0.52(0.42,0.68)	0.55(0.4,0.76)	0.52(0.28,0.77)	0.49(0.33,0.67)	0.53(0.42,0.68)	0.44(0.3,0.67)	0.46(0.28,0.67)	0.47(0.31,0.62)	0.43(0.34,0.54)	0.54(0.37,0.7)	0.45(0.33,0.62)	0.48(0.38,0.6)
GO:0008150	biological_process	GO:0032502	developmental process	2635	110	0.97(0.28,1.5)	0.58(0.26,1.21)	0.37(0.18,0.74)	0.46(0.17,1.44)	0.56(0.32,1.03)	0.54(0.13,0.9)	0.46(0.27,0.87)	0.47(0.23,0.82)	0.51(0.25,1.12)	0.34(0.15,0.53)	0.79(0.35,1.35)	0.4(0.17,0.73)	0.47(0.2,0.65)
GO:0008150	biological_process	GO:0051234	establishment of localization	2714	100	0.33(0.12,0.58)	0.46(0.08,0.67)	0.2(0.08,0.42)	0.3(0.16,0.58)	0.34(0.21,0.51)	0.22(0.09,0.36)	0.24(0.12,0.49)	0.29(0.07,0.5)	0.3(0.11,0.47)	0.37(0.13,0.49)	0.23(0.11,0.38)	0.32(0.16,0.48)	0.22(0.12,0.43)
GO:0008150	biological_process	GO:0032501	multicellular organismal process	2200	94	1.35(0.31,3.55)	0.96(0.12,2.37)	0.69(0,1.77)	0.62(0.13,1.81)	0.89(0.6,1.35)	0.57(0.27,1.8)	1.07(0.26,1.8)	0.81(0.15,2.05)	0.49(0.19,1.65)	0.61(0.32,1.26)	1.25(0.63,2.5)	0.94(0.44,1.35)	0.59(0.24,0.98)
GO:0008150	biological_process	GO:0050896	response to stimulus	2770	93	0.76(0.52,2.31)	0.6(0.22,1.15)	0.21(0.1,0.53)	0.27(0.05,1.01)	0.88(0.31,1.25)	0.24(0.06,0.94)	0.42(0.19,0.86)	0.28(0.08,0.78)	0.35(0.15,0.98)	0.3(0.14,0.55)	0.93(0.46,1.72)	0.45(0.23,0.8)	0.3(0.12,0.54)

[1] Average of eight non-brain non-sex-specific tissues.

Nature Genetics: doi:10.1038/ng.711

Table S6. Comparison of expression levels of X-linked and autosomal genes using microarray data. The mean is similar between X and autosomes in various tissues, consistent with the observations in ref. 6. But, the variance-based comparison (Miller's jackknife method), which is expected to be more robust against data noise, revealed significant difference (X:AA < 0.75) in ~50% of tissues. We here considered only tissues with corresponding RNA-Seq data.

Platform	Tissue	P[1]	X:AA[2] (mean)	X:AA (median)	X:AA[3] (Mann-Whitney)	X:AA[4] (Jackknife)
Human HG-U133 Plus 2.0	Male brain	0.28	0.99	0.92	0.95 (0.78,1.11)	1.32 (1.02,1.70)
Human HG-U133 Plus 2.0	Male liver	**1.9E-05**	0.93	0.70	0.79 (0.70,0.88)	0.90 (0.64,1.28)
Human HG-U133 Plus 2.0	Female brain	0.11	0.99	0.85	0.92 (0.79,1.05)	1.27 (0.99,1.64)
Human HG-U133 Plus 2.0	Female heart	**1.7E-03**	0.97	0.84	0.85 (0.74,0.96)	1.02 (0.69,1.51)
Human HG-U133 Plus 2.0	Male heart	**1.7E-03**	0.97	0.83	0.87 (0.76,0.97)	0.65 (0.46,0.92)
Human U133A /GNF1H	Lymph node	**2.2E-03**	0.87	0.82	0.86 (0.75,0.97)	1.11 (0.61,2.02)
Human U133A /GNF1H	Brain	0.05	0.91	0.83	0.90 (0.76,1.03)	1.02 (0.73,1.42)
Human U133A /GNF1H	Lung	**7.4E-04**	0.84	0.77	0.83 (0.71,0.95)	0.67 (0.47,0.95)
Human U133A /GNF1H	Hear	**0.02**	0.89	0.85	0.90 (0.78,1.01)	0.47 (0.35,0.63)
Human U133A /GNF1H	Liver	**1.7E-03**	0.84	0.84	0.85 (0.74,0.96)	0.40 (0.28,0.58)
Human U133A /GNF1H	Kidney	**0.02**	0.88	0.89	0.89 (0.77,1.01)	0.65 (0.46,0.91)
Human U133A /GNF1H	Testis	**0.02**	0.89	0.85	0.89 (0.77,1.01)	0.72 (0.53,0.98)
Human U133A	Muscle	0.28	0.96	0.94	0.967 (0.87,1.06)	0.68 (0.36,1.30)

Nature Genetics: doi:10.1038/ng.711

/GNF1H						
Mouse GNF1M array	Frontal cortex	**3.2E-03**	0.97	0.84	0.86 (0.75,0.97)	1.14 (0.89,1.46)
Mouse GNF1M array	Cerebral cortex	**2.2E-04**	0.96	0.80	0.83 (0.72,0.94)	1.00 (0.79,1.28)
Mouse GNF1M array	Cerebellum	**7.8E-04**	0.97	0.82	0.84 (0.73,0.95)	1.01 (0.79,1.29)
Mouse GNF1M array	Muscle	**5.5E-06**	0.96	0.75	0.80 (0.70,0.90)	0.68 (0.46,1.02)
Mouse GNF1M array	Liver	**7.4E-07**	0.95	0.77	0.79 (0.70,0.88)	0.57 (0.37,0.86)

[1] From Mann-Whitney's U test of the equality of expression levels between X and autosomes.
[2] Data were $\log_2$-transformed before the mean expression is calculated.
[3] Numbers in parentheses show probable ranges (see Online Methods).
[4] Numbers in parentheses show 95% confidence intervals.

Nature Genetics: doi:10.1038/ng.711

Table S7: X:AA expression ratios of human and mouse genes with orthologs in chicken.

Tissues	X:AA (median)	X:AA[1] (Miller's jackknife)	X:AA[2] (Mann-Whitney)
Human			
(411 X-linked genes vs 12426 autosomal gene)			
Brain	0.52	0.61 (0.56,0.66)	0.53 (0.45,0.63)
Heart	0.54	0.58 (0.53,0.64)	0.48 (0.42,0.56)
Liver	0.40	0.46 (0.42,0.50)	0.37 (0.32,0.43)
Muscle	0.38	0.47 (0.43,0.51)	0.37 (0.31,0.45)
Testis	0.46	0.45 (0.39,0.51)	0.44 (0.38,0.53)
Kidney	0.49	0.72 (0.65,0.79)	0.49 (0.42,0.59)
Breast	0.37	0.53 (0.48,0.59)	0.42 (0.36,0.50)
Adipose	0.39	0.48 (0.44,0.53)	0.38 (0.33,0.45)
Colon	0.43	0.49 (0.45,0.55)	0.42 (0.36,0.50)
Lymph node	0.45	0.45 (0.41,0.50)	0.41 (0.36,0.48)
Cerebral cortex	0.44	0.62 (0.55,0.70)	0.47 (0.40,0.56)
Lung	0.46	0.54 (0.49,0.59)	0.41 (0.36,0.48)
Average[3]	0.47	0.61 (0.51,0.72)	0.43 (0.37,0.53)
Mouse			
(426 X-linked genes vs 12504 autosomal gene)			
Liver	0.17	0.36 (0.31,0.41)	0.18 (0.15,0.22)
Brain	0.24	0.39 (0.35,0.44)	0.25 (0.22,0.30)
Muscle	0.25	0.40 (0.36,0.45)	0.25 (0.21,0.31)

[1] Numbers in parentheses show 95% confidence intervals.
[2] Numbers in parentheses show probable ranges (see Online Methods).
[3] Average of eight non-brain non-sex-specific tissues.

Nature Genetics: doi:10.1038/ng.711

Table S8. ISPI-corrected RNA-Seq gene expression levels of X chromosome and autosomes[1]

Tissues/stages	X:AA (median)	X:AA[2] (Miller's jackknife)	X:AA[3] (Mann-Whitney)
Human			
Brain	0.57	0.68 (0.63, 0.73)	0.57 (0.50, 0.67)
Heart	0.48	0.55 (0.52, 0.59)	0.45 (0.40, 0.53)
Liver	0.33	0.45 (0.42, 0.49)	0.31 (0.27, 0.36)
Muscle	0.39	0.49 (0.45, 0.54)	0.37 (0.32, 0.43)
Testis	0.62	0.61 (0.55, 0.67)	0.63 (0.56, 0.71)
Kidney	0.43	0.67 (0.63, 0.72)	0.43 (0.38, 0.50)
Breast	0.38	0.52 (0.48, 0.56)	0.39 (0.34, 0.45)
Adipose	0.41	0.50 (0.46, 0.54)	0.36 (0.31, 0.42)
Colon	0.46	0.53 (0.49, 0.58)	0.39 (0.34, 0.45)
Lymph node	0.40	0.48 (0.45, 0.53)	0.38 (0.33, 0.43)
Cerebral cortex	0.51	0.68 (0.62, 0.74)	0.50 (0.43,0.59)
Lung	0.39	0.47 (0.44, 0.51)	0.35 (0.31, 0.40)
Average[4]	0.44	0.55 (0.50, 0.60)	0.38 (0.33, 0.45)
C. elegans			
L2	1.37	0.95 (0.89, 1.03)	1.43 (1.25, 1.67)
L3	0.92	0.86 (0.82, 0.90)	1.00 (0.91, 1.11)
L4	0.56	0.71 (0.68, 0.74)	0.67 (0.63, 0.71)
Adult	0.61	0.42 (0.40, 0.43)	0.57 (0.53, 0.63)

[1] ISPI-correction cannot be made to mouse results due to the lack of mouse DNA-Seq data.
[2] Numbers in parenthesis show 95% confidence intervals.
[3] Numbers in parentheses show probable ranges (see Online Methods).
[4] Average of eight non-brain non-sex-specific tissues.

Nature Genetics: doi:10.1038/ng.711

Table S9. X:AA ratios of median expressions for *C. elegans* hermaphrodites with certain transcript lengths.

Stages	All genes	1.0-1.5 kb	1.5-2.0 kb	2.0-2.5 kb	2.5-3.0 kb	3.0-3.5 kb
L2	1.37	2.04	1.09	1.01	0.78	1.11
L3	0.92	1.36	0.81	0.78	0.54	0.60
L4	0.56	0.70	0.53	0.61	0.41	0.40
Adult	0.61	0.85	0.55	0.44	0.30	0.25

Nature Genetics: doi:10.1038/ng.711

Table S10. ISPI-calibrated gene expressions of 350 X-linked and 7868 autosomal genes with comparable GC% (40-50% in coding regions).

Tissues/stages	X:A (median)	X:A (Miller's jackknife)[1]	X:A ** (Mann-Whitney)[2]
Brain	0.48(0.30,0.75)	0.61(0.55,0.67)	0.45(0.38,0.56)
Heart	0.36(0.14,0.57)	0.55(0.50,0.60)	0.32(0.26,0.40)
Liver	0.24(0.10,0.48)	0.39(0.35,0.44)	0.23(0.20,0.29)
Muscle	0.29(0.14,0.51)	0.44(0.38,0.52)	0.26(0.22,0.33)
Testis	0.56(0.42,0.72)	0.70(0.61,0.81)	0.61(0.50,0.77)
Kidney	0.36(0.22,0.49)	0.61(0.55,0.67)	0.36(0.30,0.43)
Breast	0.32(0.19,0.47)	0.65(0.57,0.73)	0.34(0.29,0.42)
Adipose	0.31(0.15,0.48)	0.49(0.43,0.56)	0.29(0.24,0.36)
Colon	0.37(0.17,0.51)	0.52(0.46,0.58)	0.33(0.28,0.40)
Lymph node	0.35(0.19,0.49)	0.54(0.48,0.60)	0.31(0.26,0.38)
Cerebral cortex	0.41(0.23,0.66)	0.66(0.59,0.73)	0.41(0.34,0.50)
Lung	0.30(0.17,0.51)	0.50(0.46,0.54)	0.31(0.26,0.37)
Average	0.41(0.25,0.55)	0.60(0.54,0.67)	0.36(0.30,0.45)

[1] Numbers in parenthesis show 95% confidence intervals.
[2] Numbers in parenthesis show probable ranges (see Online Methods).

Nature Genetics: doi:10.1038/ng.711

Table S11. X:AA ratios calculated as in ref. 6 and ref. 20 with 50% of the most highly expressed genes of each tissue considered.

Tissues	X:AA (median)	X:AA (mean)
Brain	0.90	0.99
Heart	0.78	0.79
Liver	0.74	0.78
Muscle	0.65	0.74
Testis	0.73	0.82
Kidney	0.92	0.97
Breast	0.81	0.81
Adipose	0.87	0.88
Colon	0.90	0.92
Lymph node	0.89	0.96
Cerebral cortex	1.04	1.05
Lung	0.86	0.85
Average[1]	0.81	0.87

[1] Average of eight non-brain non-sex-specific tissues.

Nature Genetics: doi:10.1038/ng.711

Table S12. Numbers of genes detected by proteomic data and X:AA protein expression ratios.

	# of known protein-coding genes	# of genes with at least 1 hit in the proteomic data	Fraction of genes with at least 1 hit	*P –value* (chi-square test)[1]	X:AA protein expression ratio (median)[2]
Mouse					
X chromosome	991	86	8.70%	0.002	0.47 (0.26, 0.58)[3]
Autosomes	22156	2734	12.30%		0.38 (0.18, 0.65)[4]
C. elegans					
X chromosome	2801	1401	50%	0.3	0.59 (0.53, 0.67)[3]
Autosomes	17375	9021	52%		0.60 (0.52, 0.67)[4]

[1] Test of no difference between X and autosomes in the fraction of genes with at least 1 hit.

[2] In the parentheses are bootstrap 95% confidence intervals.

[3] Based on the fraction of X-linked genes with at least 1 hit and the same fraction of autosomal genes.

[4] Based on the fraction of autosomal genes with at least 1 hit and the same fraction of X-linked genes.

Nature Genetics: doi:10.1038/ng.711

Table S13. Total numbers of mapped reads in autosomes (A) and the X chromosome, based on different software or different mismatch cutoffs. The human liver data were used. Numbers in parentheses are ratios relative to that of soap-v3. There is no significant difference in ratio between A and X (P >0.5).

	eland-v2[1]	soap-v1[2]	soap-v2[3]	soap-v3[4]
A	8241529 (95.98%)	7441372 (86.66%)	8370009 (97.48%)	8586585
X	105447 (95.88%)	95425 (86.77%)	106937 (97.24%)	109976

[1] Two mismatches are allowed in ELAND mapping

[2] One mismatch is allowed in SOAP mapping

[3] Two mismatches are allowed in SOAP mapping

[4] Three mismatches are allowed in SOAP mapping

Nature Genetics: doi:10.1038/ng.711

Table S14. X:AA expression ratios determined by Miller's jackknife method are robust when different proportions of genes are excluded from the two ends of the expression distributions.

Tissues/ stages	Minimal[1]	30%[1]	35%[1]	40%[1]	45%[1]
Human					
Brain	0.68(0.63,0.73)	0.67(0.63,0.72)	0.66(0.62,0.71)	0.67(0.62,0.71)	0.78(0.70,0.86)
Heart	0.55(0.52,0.59)	0.55(0.52,0.59)	0.56(0.53,0.60)	0.62(0.58,0.67)	0.66(0.58,0.73)
Liver	0.45(0.42,0.49)	0.45(0.42,0.49)	0.44(0.41,0.48)	0.43(0.40,0.46)	0.46(0.41,0.52)
Muscle	0.49(0.45,0.54)	0.49(0.45,0.54)	0.46(0.42,0.50)	0.43(0.39,0.47)	0.41(0.37,0.46)
Testis	0.61(0.55,0.67)	0.56(0.53,0.60)	0.58(0.54,0.62)	0.57(0.52,0.62)	0.55(0.49,0.61)
Kidney	0.67(0.63,0.72)	0.62(0.58,0.67)	0.60(0.56,0.64)	0.61(0.56,0.66)	0.59(0.54,0.66)
Breast	0.52(0.48,0.56)	0.51(0.47,0.56)	0.47(0.44,0.51)	0.43(0.39,0.48)	0.39(0.35,0.43)
Adipose	0.50(0.46,0.54)	0.48(0.44,0.52)	0.46(0.43,0.49)	0.48(0.44,0.51)	0.46(0.41,0.52)
Colon	0.53(0.49,0.58)	0.49(0.45,0.53)	0.44(0.41,0.48)	0.45(0.42,0.49)	0.46(0.42,0.52)
Lymph node	0.48(0.45,0.53)	0.45(0.41,0.48)	0.41(0.39,0.44)	0.44(0.41,0.47)	0.46(0.42,0.51)
Cerebral cortex	0.68(0.62,0.74)	0.59(0.55,0.64)	0.57(0.53,0.61)	0.55(0.51,0.59)	0.55(0.50,0.60)
Lung	0.47(0.44,0.51)	0.47(0.44,0.50)	0.49(0.46,0.53)	0.49(0.46,0.53)	0.53(0.48,0.59)
Average[2]	0.55(0.50,0.60)	0.50(0.47,0.54)	0.48(0.45,0.52)	0.47(0.44,0.50)	0.48(0.42,0.54)
Mouse					
Liver	0.21(0.20,0.24)	0.19(0.17,0.21)	0.19(0.17,0.21)	0.17(0.15,0.19)	0.14(0.12,0.15)
Brain	0.47(0.44,0.50)	0.44(0.40,0.48)	0.37(0.34,0.41)	0.32(0.30,0.35)	0.32(0.30,0.36)
Muscle	0.34(0.32,0.37)	0.32(0.30,0.35)	0.29(0.27,0.32)	0.28(0.26,0.31)	0.24(0.22,0.27)
C.elegans					
L2	0.95(0.89,1.03)	0.96(0.93,0.99)	0.94(0.91,0.97)	0.91(0.87,0.95)	0.87(0.82,0.92)
L3	0.86(0.82,0.90)	0.77(0.74,0.79)	0.77(0.74,0.80)	0.77(0.74,0.80)	0.79(0.74,0.83)
L4	0.71(0.68,0.74)	0.59(0.57,0.61)	0.58(0.56,0.60)	0.56(0.54,0.59)	0.57(0.54,0.61)
Adult	0.42(0.40,0.43)	0.42(0.40,0.43)	0.46(0.44,0.47)	0.500.48,0.52)	0.48(0.45,0.51)

[1] Proportion of genes removed. "Minimal" refers to the minimal proportion removed to ensure that X and autosomes contain no zero-expression genes, which is used in the analysis presented in the main text.

[2] Average of eight non-brain non-sex-specific tissues.

Nature Genetics: doi:10.1038/ng.711

Table S15. X:AA ratios determined by the Modified Mann-Whitney U test are robust when different proportions of genes are excluded from two ends of the expression distributions.

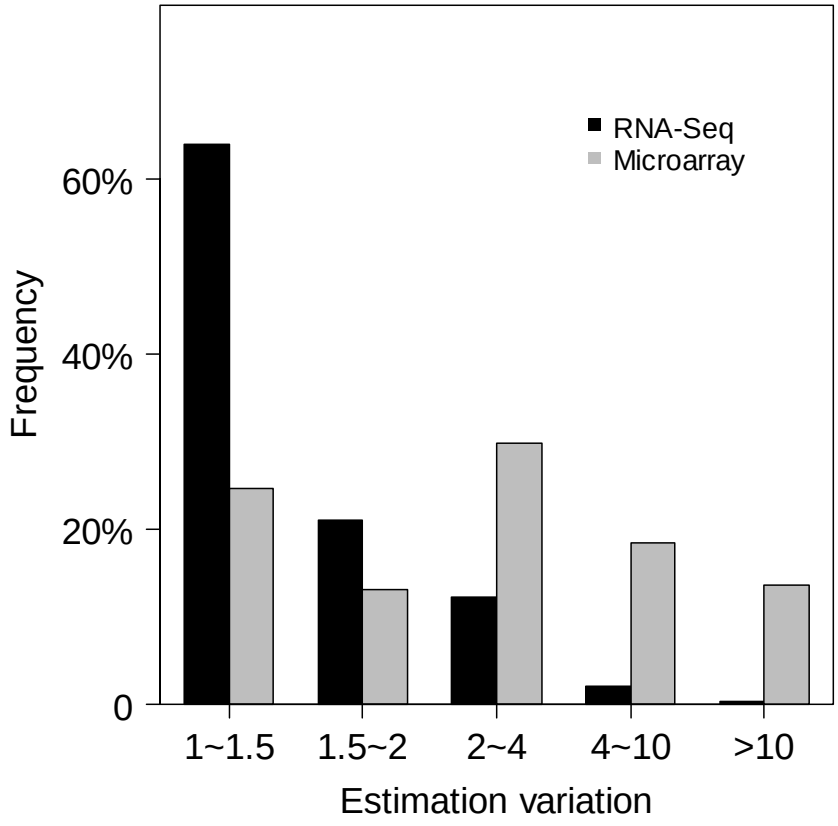
Tissues/stages	Minimal[1]	30%[1]	35%[1]	40%[1]	45%[1]
Human					
Brain	0.57(0.50,0.67)	0.59(0.53,0.67)	0.57(0.53,0.63)	0.57(0.53,0.63)	0.57(0.53,0.63)
Heart	0.45(0.40,0.53)	0.44(0.40,0.50)	0.47(0.42,0.53)	0.48(0.43,0.53)	0.48(0.43,0.53)
Liver	0.31(0.27,0.36)	0.30(0.26,0.36)	0.30(0.27,0.34)	0.32(0.29,0.36)	0.32(0.29,0.36)
Muscle	0.37(0.32,0.43)	0.36(0.31,0.42)	0.36(0.32,0.42)	0.36(0.32,0.40)	0.37(0.34,0.40)
Testis	0.63(0.56,0.71)	0.61(0.56,0.67)	0.61(0.56,0.67)	0.61(0.56,0.67)	0.59(0.56,0.63)
Kidney	0.43(0.38,0.50)	0.43(0.38,0.48)	0.43(0.38,0.48)	0.43(0.40,0.48)	0.43(0.40,0.45)
Breast	0.39(0.34,0.45)	0.38(0.34,0.43)	0.38(0.34,0.42)	0.38(0.34,0.42)	0.38(0.36,0.40)
Adipose	0.36(0.31,0.42)	0.36(0.31,0.42)	0.37(0.33,0.42)	0.38(0.34,0.43)	0.39(0.37,0.42)
Colon	0.39(0.34,0.45)	0.40(0.36,0.45)	0.41(0.37,0.45)	0.42(0.38,0.45)	0.43(0.40,0.45)
Lymph node	0.38(0.33,0.43)	0.38(0.34,0.43)	0.39(0.36,0.43)	0.40(0.37,0.43)	0.41(0.38,0.43)
Cerebral cortex	0.50(0.43,0.59)	0.50(0.45,0.56)	0.50(0.45,0.56)	0.49(0.45,0.53)	0.50(0.48,0.53)
Lung	0.35(0.31,0.40)	0.36(0.32,0.40)	0.36(0.32,0.40)	0.36(0.33,0.40)	0.37(0.34,0.40)
Average[2]	0.38(0.33,0.45)	0.41(0.37,0.45)	0.42(0.38,0.45)	0.43(0.40,0.45)	0.43(0.40,0.45)
Mouse					
Liver	0.18(0.16,0.21)	0.17(0.15,0.20)	0.16(0.14,0.19)	0.15(0.13,0.16)	0.14(0.13,0.14)
Brain	0.27(0.24,0.32)	0.26(0.23,0.29)	0.25(0.23,0.28)	0.25(0.23,0.27)	0.25(0.23,0.26)
Muscle	0.24(0.21,0.28)	0.22(0.20,0.26)	0.22(0.20,0.24)	0.21(0.19,0.23)	0.19(0.18,0.21)
C.elegans					
L2	1.43(1.25,1.67)	1.33(1.25,1.43)	1.33(1.25,1.43)	1.33(1.25,1.43)	1.33(1.25,1.43)
L3	1.00(0.91,1.11)	1.00(0.91,1.11)	1.00(0.91,1.11)	0.95(0.91,1.00)	0.95(0.91,1.00)
L4	0.67(0.63,0.71)	0.59(0.56,0.63)	0.59(0.56,0.63)	0.59(0.56,0.63)	0.57(0.56,0.59)
Adult	0.57(0.53,0.63)	0.59(0.56,0.63)	0.61(0.56,0.67)	0.63(0.59,0.67)	0.63(0.59,0.67)

[1] Proportion of genes removed. "Minimal" refers to the minimal proportion removed to ensure that X and autosomes contain no zero-expression genes, which is used in the analysis presented in the main text.

[2] Average of eight non-brain non-sex-specific tissues.

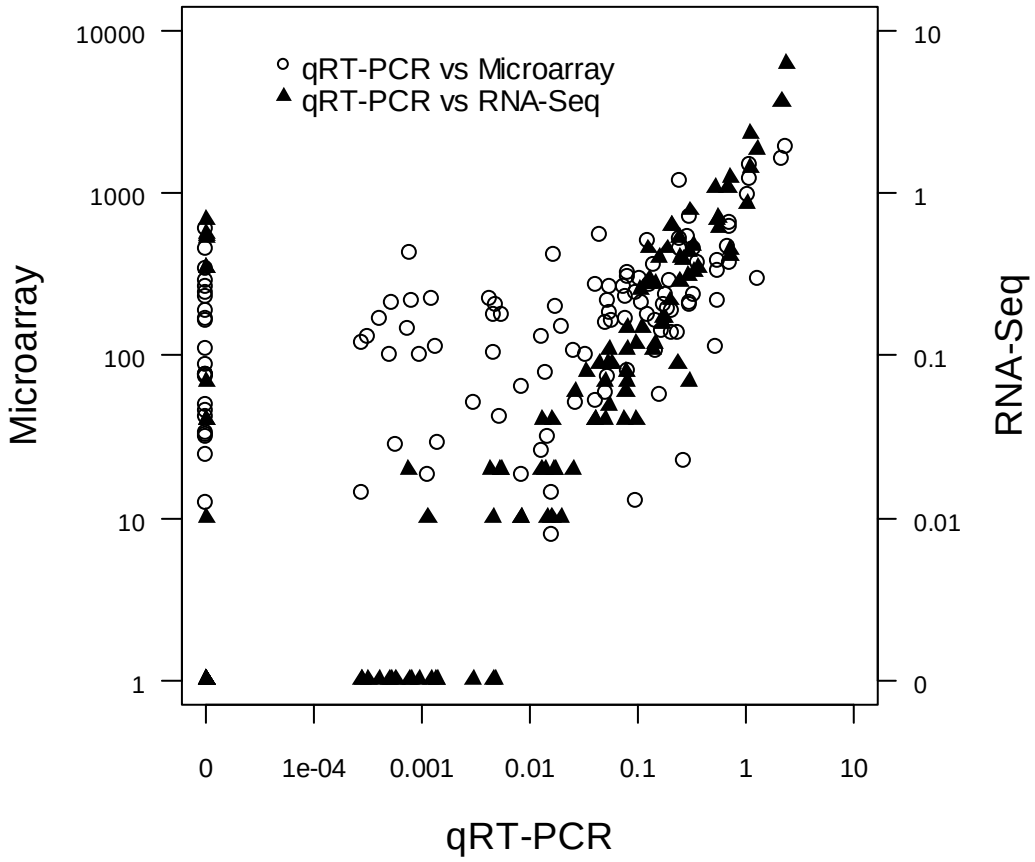
Nature Genetics: doi:10.1038/ng.711

Fig. S1. Comparison of gene expressions measured by microarray and RNA-Seq. Frequency distribution of estimation variation, which is shown by the fold difference of microarray intensities of two randomly chosen probesets targeting the same exons or fold difference of RNA-Seq signals from two halves of a transcript. In the analysis of RNA-Seq data, each transcript is divided into 32 equal-size windows and the RNA-Seq signals from two randomly chosen non-overlapping sets of 16-windows are compared. Both microarray and RNA-Seq data are from the human liver.



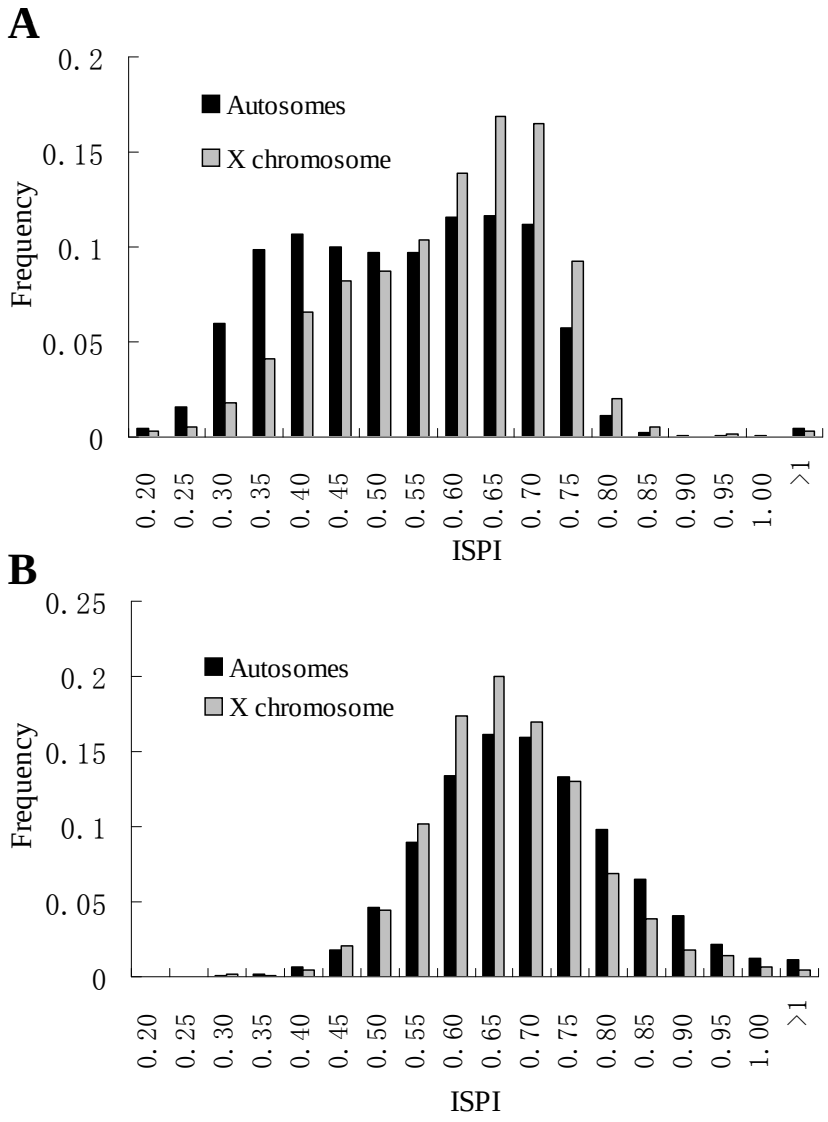
Nature Genetics: doi:10.1038/ng.711

Fig. S2. Relative expression levels in liver among 120 mouse genes, measured by RNA-Seq, microarray, and qRT-PCR. The genes are chosen randomly from chromosome 13 irrespective of their expression signals from RNA-Seq or microarray. Spearman's rank correlation between qRT-PCR and microarray $\log_{10}$ signals is 0.51 ($P = 1.8\times10^{-9}$) and that between qRT-PCR and RNA-Seq is 0.76 ($P < 2.2\times10^{-16}$). Note that different methods use different signal units.



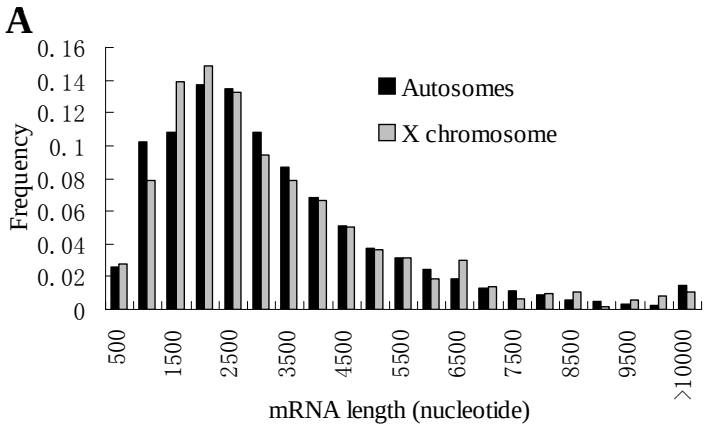
Nature Genetics: doi:10.1038/ng.711

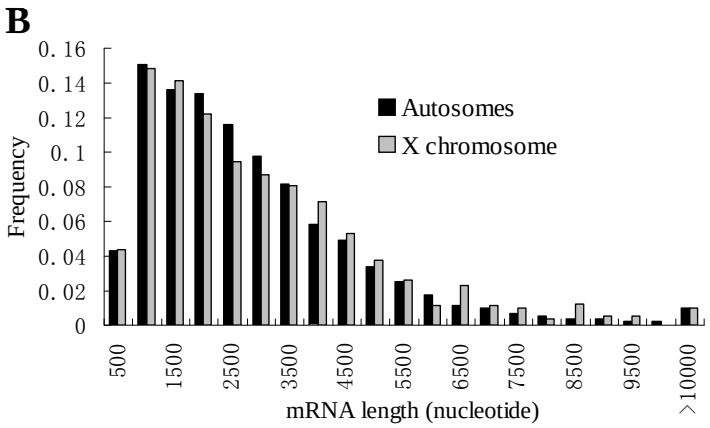
Fig. S3. Comparison of ISPI between X-linked and autosomal genes for **(A)** human and **(B)** *C. elegans*. ISPI cannot be calculated for mouse due to the lack of appropriate mouse DNA-Seq data.

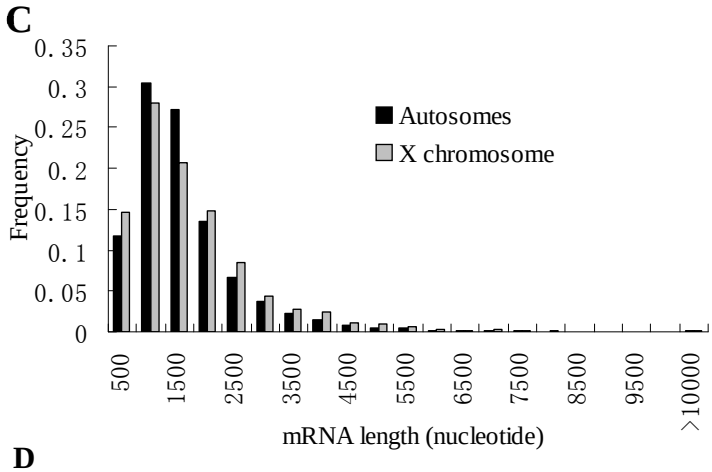


Nature Genetics: doi:10.1038/ng.711

Fig. S4. Distributions of the transcript lengths for **(A)** human, **(B)** mouse, and **(C)** *C. elegans* genes and (**D**) the comparison of transcript lengths between X-linked (X) and autosomal (A) genes for the 3 organisms.





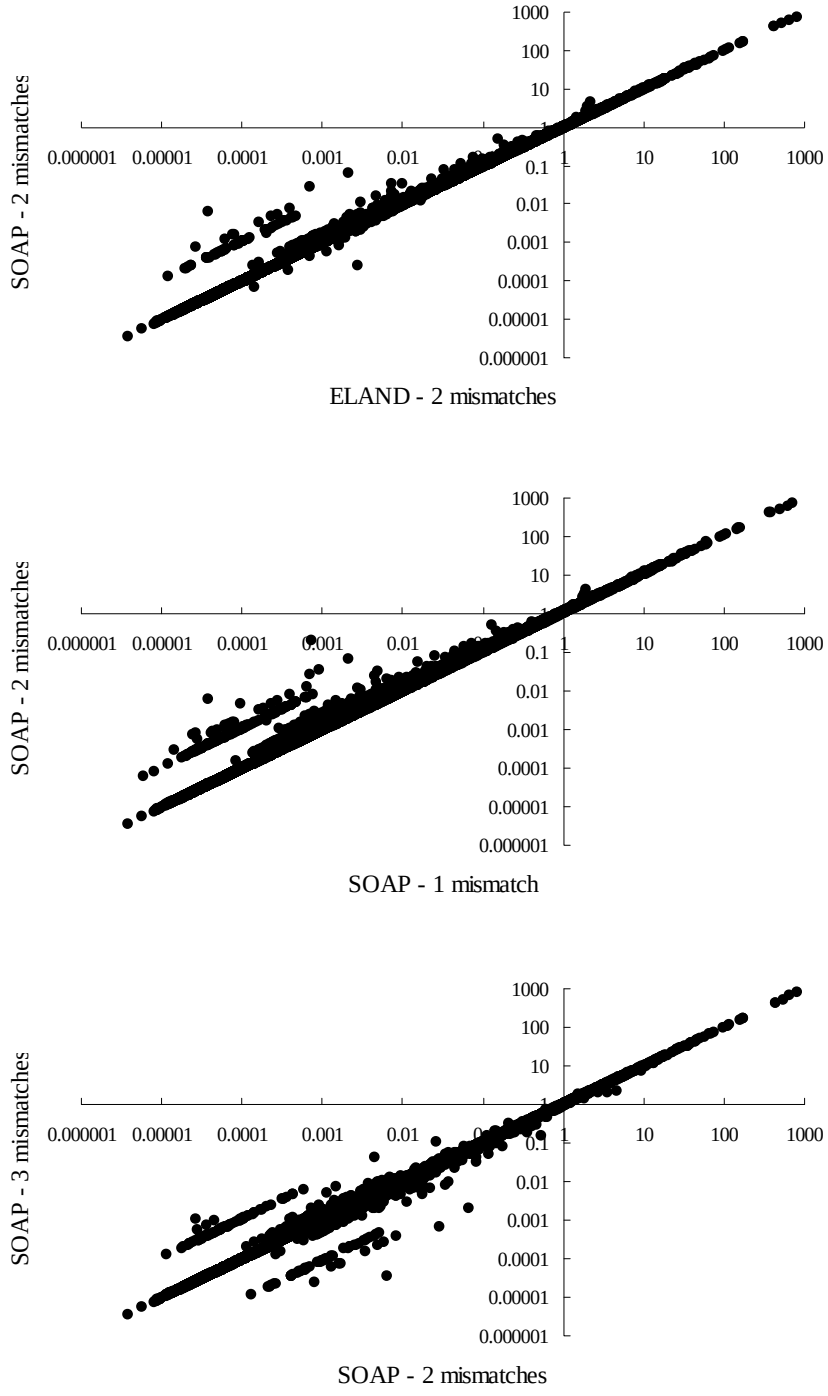


D

	P [1]	Median of X	Median of A	Mean of X±SD	Mean of A±SD
Human	0.72	2379	2467	2994±80	3002±17
Mouse	0.17	2238	2152	2760±74	2630±15
C.elegans	0.02	1167	1081	1489±25	1394±10

[1] Two-tailed Mann-Whitney U test for transcript length difference between X-linked genes (X) and autosomal genes (A).

Fig. S5. Relative gene expression levels of human liver estimated from RNA-Seq are not affected by using different read mapping software (SOAP and ELAND) or different mismatch cutoffs. Pearson correlation coefficients of the three comparisons are all larger than 0.995. Each dot represents a gene.



Nature Genetics: doi:10.1038/ng.711