

# Dosage Compensation of the X Chromosomes in Bovine Germline, Early Embryos, and Somatic Tissues

Jingyue (Ellie) Duan<sup>1</sup>, Wei Shi<sup>2</sup>, Nathaniel K. Jue<sup>3</sup>, Zongliang Jiang<sup>4</sup>, Lynn Kuo<sup>2</sup>, Rachel O'Neill<sup>5</sup>, Eckhard Wolf<sup>6</sup>, Hong Dong<sup>7</sup>, Xinbao Zheng<sup>7</sup>, Jingbo Chen<sup>7,\*</sup>, and Xiuchun (Cindy) Tian<sup>1,\*</sup>

<sup>1</sup>Department of Animal Science, University of Connecticut, Storrs, CT

<sup>2</sup>Department of Statistics, University of Connecticut, Storrs, CT

<sup>3</sup>School of Natural Sciences, California State University, Monterey Bay, CA

<sup>4</sup>School of Animal Science, Louisiana State University, Agricultural Center, Baton Rouge, LA

<sup>5</sup>Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT

<sup>6</sup>Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Germany

<sup>7</sup>Institute of Animal Science, Xinjiang Academy of Animal Sciences, Urumqi, Xinjiang, P.R. China

\*Corresponding authors: E-mails: chenjb126@126.com; xiuchun.tian@uconn.edu.

Accepted: December 12, 2018

**Data availability:** The data sets analyzed during the current study are available in the NCBI BioProject: PRJNA254699: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA254699>; PRJNA228235: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA228235>; PRJNA243569: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA243569>; PRJNA243569: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA243569>; PRJNA298914: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA298914>; PRJNA268096: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA268096>; PRJEB6377: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB6377>; PRJNA273164: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA273164>.

## Abstract

Dosage compensation of the mammalian X chromosome (X) was proposed by Susumu Ohno as a mechanism wherein the inactivation of one X in females would lead to doubling the expression of the other. This would resolve the dosage imbalance between eutherian females (XX) versus male (XY) and between a single active X versus autosome pairs (A). Expression ratio of X- and A-linked genes has been relatively well studied in humans and mice, despite controversial results over the existence of upregulation of X-linked genes. Here we report the first comprehensive test of Ohno's hypothesis in bovine preattachment embryos, germline, and somatic tissues. Overall an incomplete dosage compensation ( $0.5 < X:A < 1$ ) of expressed genes and an excess X dosage compensation ( $X:A > 1$ ) of ubiquitously expressed "dosage-sensitive" genes were seen. No significant differences in X:A ratios were observed between bovine female and male somatic tissues, further supporting Ohno's hypothesis. Interestingly, preimplantation embryos manifested a unique pattern of X dosage compensation dynamics. Specifically, X dosage decreased after fertilization, indicating that the sperm brings in an inactive X to the matured oocyte. Subsequently, the activation of the bovine embryonic genome enhanced expression of X-linked genes and increased the X dosage. As a result, an excess compensation was exhibited from the 8-cell stage to the compact morula stage. The X dosage peaked at the 16-cell stage and stabilized after the blastocyst stage. Together, our findings confirm Ohno's hypothesis of X dosage compensation in the bovine and extend it by showing incomplete and over-compensation for expressed and "dosage-sensitive" genes, respectively.

**Key words:** Ohno's hypothesis, X dosage compensation, preattachment embryos, bovine.

## Introduction

Gene dosage is the number of copies of a given gene in cells of an organism and can be manifested by the amount of its products (Ercan 2015). Maintenance of proper gene dosage is essential in functional cellular networks such as in

embryogenesis, and fetus development. Aneuploidy such as monosomy or trisomy is an abnormal change in the dosage of chromosomes and is generally detrimental to the organism (Holtzman et al. 1992). For example, aneuploidy accounts for 46.3% of spontaneous abortions in humans (Hassold et al.

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

1980). Small changes in dosage of single genes can lead to many diseases (Hurles et al. 2008) and the onset of tumorigenesis (Gordon et al. 2012). However, monosomy of the X chromosome in mammalian males is well tolerated, although over a thousand genes important for both sexes are located on X (Ercan 2015). Susumu Ohno hypothesized that to prevent the deleterious effects of haploinsufficiency in males, a compensatory mechanism involving the doubling expression of X-linked genes must occur (Ohno et al. 1959). This, however, at the same time could cause a quadruple dosage of X in females. By transcriptionally silencing one of the two X chromosomes in females the dosage of the X chromosome between males and females is balanced (Veitia and Potier 2015). Meanwhile, this also balances the gene dosage between sex chromosome and autosome pairs (A) in both sexes (Ohno 1966).

Although X chromosome inactivation (XCI) has been observed in all mammalian species studied to date (Ohno et al. 1959; Lyon 1961; Heard et al. 1997), dosage compensation by doubling the expression of X-linked genes has only been studied in very few species and is still heavily debated (Nguyen and Disteche 2006; Xiong et al. 2010; Deng et al. 2011; He et al. 2011; Kharchenko et al. 2011). Dosage compensation is determined by calculating the ratio of averaged expression value of X-linked genes to that of the autosomes (X:A ratio). The ratio X:A  $\approx$  1 indicates the doubling of X gene expression, while X:A of 0.5 rejects Ohno's hypothesis. Two previous microarray studies fully supported dosage compensation in both humans and mice (Gupta et al. 2006; Nguyen and Disteche 2006). However, the first RNA sequencing (RNA-seq) study of humans and mice (Xiong et al. 2010) claimed that microarray-based expression was not suitable for comparing expression levels of different genes, and reported X:A of approximately 0.5, that is, a lack of dosage compensation. Subsequently, the same RNA-seq data were re-analyzed with all low and non-expressed genes removed because such genes are enriched on X chromosomes and thus could skew the comparison. Results from such filtering verified the hypothesis (Deng et al. 2011; Kharchenko et al. 2011). Since then, Ohno's hypothesis has been tested with many different analysis approaches including comparing the ratio between the modern X to the proto XX using 1:1 orthologs between humans and chickens (Lin et al. 2012) or comparing only genes coding for large proteins as the "dosage-sensitive housekeeping genes" (Pessia, Engelstädter, et al. 2014). It has been found that different experimental platforms, analysis methods, and cut-off values all influenced the dosage compensation results. Thus, more questions than answers are presented on the evolution of dosage compensation of sex chromosomes (He and Zhang 2016).

While debates persist over dosage compensation, XCI has been observed in all mammalian species studied to date (Okamoto et al. 2011). In the bovine, XCI was proposed by De La Fuente et al. (1999), and confirmed by Xue et al. (2002). In early bovine embryos imprinted XCI was observed at the

morula stage (Ferreira et al. 2010) and random XCI occurred between the blastocyst and elongation stages (Bermejo-Alvarez et al. 2011). Although a recent study reported incomplete X dosage compensation in bovine fat, liver, muscle, and pituitary gland (Ka et al. 2016), further studies are needed for bovine early embryos and germ cells.

Here we report the first comprehensive test of Ohno's hypothesis in the compensatory upregulation of the X chromosome in bovine embryos, germline, and a vast array of somatic tissues using seven RNA-seq data sets (three from bovine preattachment embryos and four from somatic tissues), including immature and mature oocytes, in vivo and in vitro embryos up to the blastocyst stage (days 1–7; day 0 = standing estrus), conceptuses (embryos and associated extra-embryonic membranes) from day 7 to day 19, two adult female-specific tissues, eight adult female and male somatic tissues. Using median expression of the X:A ratio with its 95% bootstrap confidence intervals, we report incomplete compensatory upregulation of expressed X-linked genes and complete dosage compensation of "dosage-sensitive" genes. Our data thus fully support Ohno's hypothesis in that the compensatory upregulation of X chromosome expression affects "dosage-sensitive" genes in bovine developing embryos, germline cells, and female/male tissues.

## Materials and Methods

### Paralog Analysis

To determine the unique or nonunique mapping strategy, we first calculated paralog enrichment on each bovine chromosome. Paralogs were identified on BioMart (Ensembl genome browser: <http://useast.ensembl.org/biomart/>; ensemble genes 85) and defined as genes with greater than 70% amino acid identity. This minimal cut-off for paralogs was a result of our previous study, which determined that 70% was the best match of the BioMart search algorithm for the identification of X-linked multigene families (Jue et al. 2013). The total number of genes on each chromosome was calculated using bovine genome reference annotation UMD3.1 ([http://useast.ensembl.org/Bos\\_taurus/Info/Annotation?redirect=no](http://useast.ensembl.org/Bos_taurus/Info/Annotation?redirect=no)). Enrichment of paralogs gene number for each autosome was calculated by Fisher's exact test compared with that on X chromosome (table 1).

### RNA-seq Data Sets and Read Trimming

Raw FASTQ files were obtained from NCBI GEO database (table 2). A total of seven data sets including (1) in vivo developed matured oocytes and 2-cell to blastocyst stage preimplantation embryos (Jiang et al. 2014), (2) immature oocyte, in vitro developed matured oocytes, 4-cell, 8-cell, 16-cell, and blastocyst embryos (Graf et al. 2014), (3) conceptuses at days 7, 10, 13, 16, and 19 (embryos and associated extra-embryonic membranes) (Mamo et al.

**Table 1**

The Enrichment of Paralogs on Bovine Autosomes and the X Chromosome

Chromosome	Total Number of Genes	Number (%) of Paralogs	P-Value	Chromosome	Total Number of Genes	Number (%) of Paralogs	P-Value
1	985	167 (17)	5.65 e-18	16	710	129 (18)	6.56 e-13
2	1,021	229 (22)	1.88 e-08	17	665	149 (22)	6.55 e-07
3	1,372	314 (23)	7.24 e-09	18	1,236	207 (17)	1.20 e-20
4	855	222 (26)	0.00031	19	1,347	303 (22)	2.13 e-09
5	1,323	336 (25)	1.51 e-05	20	384	91 (24)	0.00027
6	692	156 (23)	6.65 e-07	21	731	221 (30)	0.10
7	1,396	377 (27)	0.00046	22	608	110 (18)	6.40 e-12
8	829	230 (28)	0.0059	23	785	264 (34)	0.61
9	602	146 (24)	6.45 e-05	24	347	98 (28)	0.049
10	1,074	316 (29)	0.033	25	766	102 (13)	7.22 e-24
11	1,047	192 (18)	1.63 e-15	26	437	82 (19)	5.63 e-09
12	414	98 (24)	0.00018	27	274	73 (27)	0.022
13	850	185 (22)	1.31 e-08	28	355	78 (22)	3.08 e-05
14	571	135 (24)	2.76 e-05	29	705	186 (26)	0.0012
15	1,050	387 (37)	0.97	X	1,128	374 (33)	
Genome average	819	199 (24)	1.46 e-05				

2012), (4) female endometria and corpora lutea (CL) (Moore et al. 2016), (5) female somatic tissues of brain, liver, muscle, and kidney (Chen et al. 2015), (6) male somatic tissues of fat, muscle, hypothalamus, duodenum, liver, lung, and kidney (PRJEB6377), and (7) female and male somatic tissues of fat, liver, muscle, and pituitary gland (Seo et al. 2016).

All RNA-seq raw reads were downloaded from NCBI using sratoolkit (version 2.5.0; [http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit\\_doc&f=std#s-2](http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit_doc&f=std#s-2)). The sequence read archive (sra) format files were converted to fastq format by fastq-dump (version 2.5.0; [http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit\\_doc&f=fastq-dump](http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit_doc&f=fastq-dump)). Quality trimming and control were conducted as follows before mapping to the reference genome. First, Trimmomatic (version 0.33; <http://www.usadellab.org/cms/?page=trimmomatic>) was applied to removing the universal sequencing adaptors of SOLiD and Illumina in respective data sets with a minimum Phred score of 20 and minimal length of 30 bp. Subsequently, read quality was examined using FastQC (version 0.11.3; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The summary of the numbers of reads in each sample after trimming is presented in [supplementary table S1, Supplementary Material online](#). The average number of read input for mapping across all samples is 22,814,027.

### Mapping and Transcript Assembly

RNA-seq read mapping and transcript assembly were performed based on the following pipeline. Trimmed RNA-seq reads were aligned to Ensemble bovine reference genome assembly UMD3.1.1 using Hisat2 version 2.0.5 aligner (Pertea et al. 2016). Transcript splice site detection was used

and both unique and nonuniquely mapped reads were kept for the subsequent analysis. The percentages of nonuniquely mapped reads for each sample are summarized in [supplementary table S1, Supplementary Material online](#), and the averaged overall mapping rate was 83.5%. IsoEM version 1.1.5 (Nicolae et al. 2011) was used to quantify gene expression to transcripts per kilobase million (TPM) using default parameters. For nonuniquely mapped reads, IsoEM assign fractions of the multiple aligned reads to each location using an expectation maximization algorithm (Nicolae et al. 2011). Expressed genes were defined as expression level TPM > 1. “Dosage-sensitive” genes were selected as ubiquitously expressed genes (TPM > 1) throughout all somatic samples or embryonic sample.

### RNA-seq Data Set Overview

Matrices of gene expression TPM for the embryo (data sets 1–3) and somatic tissues (data sets 4–8) were processed separately in R to identify ubiquitous genes. Correlation plots and unsupervised hierarchical clustering were conducted in R for quality control and identification of biologically distinct subgroups. Outliers in the biological replicates were removed for the downstream analysis.

The chromosome-wide gene expression distributions were isolated by gene locations on each chromosome in all samples using  $\log_2$ -transformed TPM (TPM > 1), the boxplots for the distribution were made in R.

### GO Analysis

Gene ontology enrichment analysis was performed in DAVID (Huang et al. 2009) and 245 and 7,603 genes on the X and

autosomes, respectively, were found as ubiquitous in the somatic tissue data sets. Similarly, 117 and 3,947 genes on the X and autosomes, respectively, were found as ubiquitous in the embryo data sets. The *P*-values in top 10 biological processes were plotted using plotly (<https://plot.ly>) in R. Pie charts for biological processes were generated as described by The Gene Ontology Consortium (Gene Ontology Consortium 2015).

### X:A Ratio Calculation

When calculating the X:A ratio, we applied the pairwiseCI package in R (Schaarschmidt and Gerhard 2015) to obtain a 95% confidence interval for the ratio of the median of X to the median of A as in a previous study (Sangrithi et al. 2017). It is based on 1,000 bootstrap replicates where sampling from the original data was done with replacement and stratified by the group variables. Bootstrapping (Efron and Tibshirani 1994) was used because it is simple to apply and does not require any distribution assumptions.

## Results

### Overview of the RNA-seq Data Sets and Paralog Analysis

We used seven bovine RNA-seq data sets, three embryonic, and four somatic, generated by us (Jiang et al. 2014) and downloaded from NCBI (table 2). In total, we have 40 samples including 19 embryos and 21 tissues from all data sets. Pearson correlation and unsupervised hierarchical clustering (supplementary fig. S1, Supplementary Material online) show that replicates within each tissue or embryonic stage clustered closely to each other, suggesting even though the data were obtained from different studies, the data were replicable and reliable. Because we only compared the X:A ratio and gene expression within each data set instead of across data sets, we were able to use the data from different studies and experimental platforms after data normalization. Paralogs are homologous genes within the same genome created by gene duplication (Gevers et al. 2004). This is one potential way to achieve X dosage compensation because X lacks a homolog in males (Jue et al. 2013). We found an approximately 1.4-fold increase in the percentage of paralogs (>70% amino acid identity [Jue et al. 2013]) on the bovine X chromosome (33%) compared with the genome averages (24%; table 1). This enrichment is significantly higher (table 1;  $P < 0.05$  by Fisher's exact test) than to that of most chromosomes, with the exception of chromosomes 15, 21, and 23 ( $P = 0.97$ , 0.10, and 0.61, respectively), suggesting the potential roles of paralogs in X dosage compensation. Such paralog enrichment on X has also been observed in humans (32% vs. 17%; 1.9-fold) and mice (51% vs. 35%; 1.5-fold) (Jue et al. 2013). This information demonstrated that unique mapping as performed in a previous RNA-seq study (Xiong et al. 2010) is not appropriate because many paralogs will be excluded from the analysis, and potentially skewing the X:A comparison. Thus,

we applied the “nonunique” mapping strategy for reads mapping. Reads that aligned to multiple locations (such as paralog gene family) in the reference genome or had alternative splice junctions were kept in all subsequent analysis. This resulted in a total of 959 X-linked and 20,316 autosomal protein coding genes.

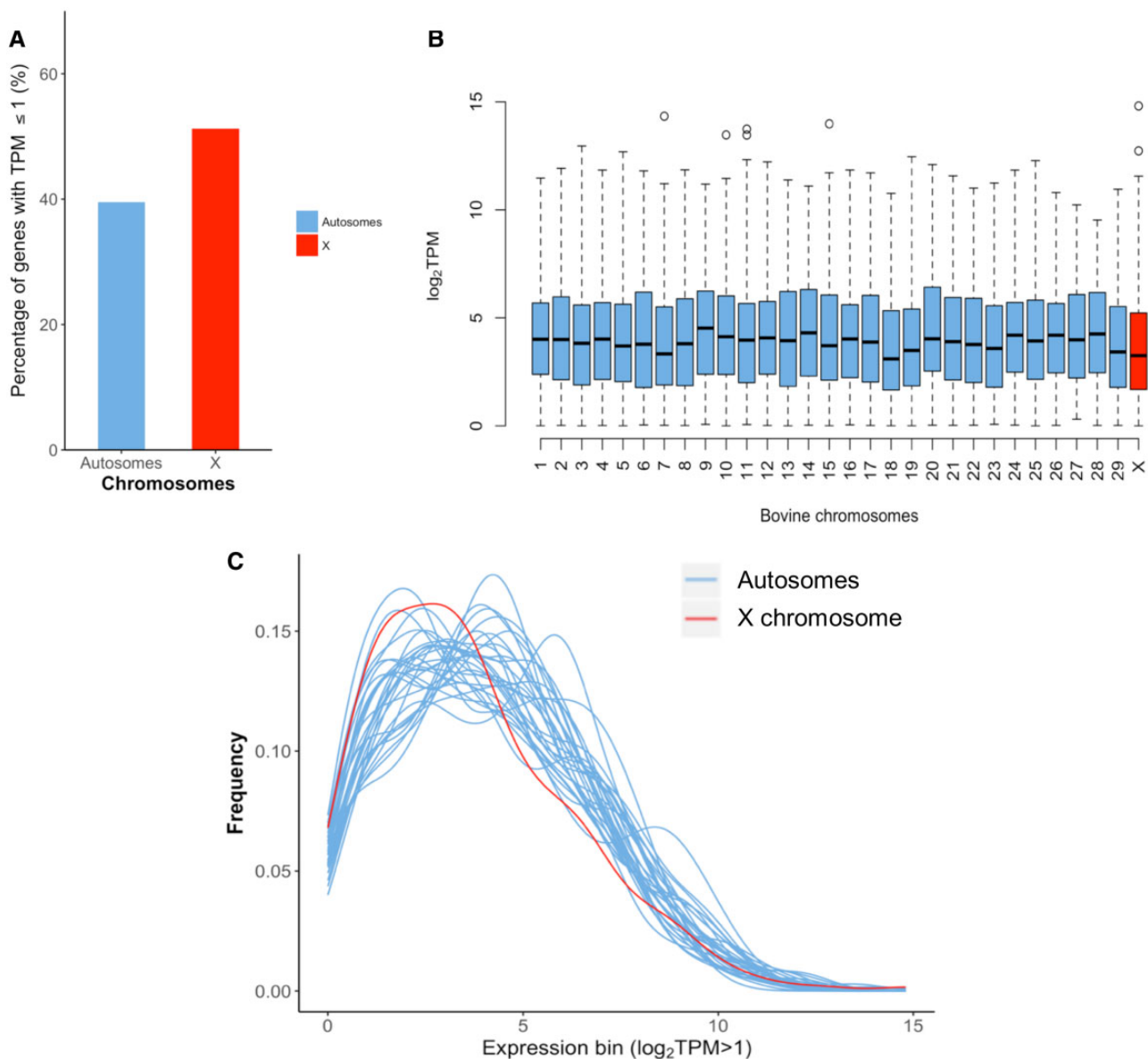
Deng et al. (2011) suggested that the low and nonexpression values in RNA-seq data may result from background noise of sequencing and read mapping. Inclusion of such values would be inappropriate and strongly influence the results. Furthermore, when we calculated the percentages of X-linked and autosomal genes with low transcript per million (TPM) values (fig. 1A and supplementary table S2, Supplementary Material online), we found that on the average the X chromosome has 11.7% more genes with  $TPM \leq 1$  than autosomes. Somatic tissues including endometrium, fat, liver, and muscle had the most enrichment of lowly expressed X-linked genes (supplementary table S2, Supplementary Material online). Therefore, we used a cut-off of  $TPM > 1$  as expressed genes to remove data bias. After this filtering, 468 X-linked genes and 12,288 autosomal genes on average were used for X:A ratio calculation. The numbers of expressed genes ( $TPM > 1$ ) on X chromosome or autosomes in each sample are listed in supplementary table S2, Supplementary Material online.

### Ranges of Gene Expression of All Chromosomes

We then investigated the gene expression profiles of all chromosomes to determine whether the transcriptional outputs from X chromosome were comparable to those of each autosome pairs. We performed  $\log_2$  transformation of TPM to normalize the data distribution. The TPM distribution of X-linked genes was not significantly different from those of all autosome pairs in 11 out of 40 samples using all expressed genes, and 31 out of 40 samples using ubiquitously expressed genes ( $P > 0.05$ , by two-sided Kolmogorov–Smirnov test, fig. 1B and supplementary table S3, Supplementary Material online). Such distributions were also demonstrated by kernel density estimation (fig. 1C). These observations demonstrate that regardless of the number of X chromosomes, the expression levels of ubiquitously expressed genes on X were comparable to those of each autosome pair in all samples, suggesting dosage compensation.

### X Chromosome Upregulation in Adult Somatic Tissues

To determine the X chromosome dosage compensation in adult cattle tissue, we analyzed RNA-seq data sets (table 2) for two female-specific tissues, endometrium and corpus luteum, and other somatic tissues from both males and females including the brain (hypothalamus), liver, kidney, muscle, fat, pituitary gland, lung, and duodenum. Overall, the X:A ratios of these tissues were in the range of 0.5–1, suggesting upregulation of the expression from the X chromosome, yet the



**Fig. 1.**—Expression ranges of genes on the X and autosome pairs in the bovine. (A) the averaged percentages of lowly expressed genes ( $\text{TPM} \leq 1$ ) on the X and autosomes in all samples. (B) A representative (in vivo matured oocytes) boxplot showing that the range and median expression levels of X-linked genes (red) ( $\text{TPM} > 1$ ) were similar to those of each autosome pairs (blue) in the bovine. (C) A representative (in vivo matured oocyte) Kernel density plot showing that the distribution of X-linked gene (red) expression ( $\text{TPM} > 1$ ) was similar to those of each autosome pairs (blue) in the bovine.

dosage compensation is incomplete (fig. 2A and B). Specifically, the liver gave the highest X:A ratio (1.01) in females and showed complete compensation, followed by the pituitary gland (0.91). These data suggest that the X chromosome expression was enriched for activities in these tissues. In contrast, fat, muscle, endometrium, and the lung gave relatively low but incomplete compensation X:A ratios (0.64–0.72), indicating less X chromosome activities. Furthermore, we compared the X chromosome expression distribution between males and females in common somatic tissues and observed no significant difference

( $P > 0.05$ , by two-sided Kolmogorov–Smirnov test, [supplementary table S4, Supplementary Material online](#)), except in muscle. The X:A ratio of common tissues between sexes was also not significantly different ( $P = 0.45$ , by paired  $t$ -test after log transformation).

Although all somatic tissues we analyzed had upregulated expression of X-linked genes which support Ohno's hypothesis, the confidence intervals of X:A ratios did not encompass 1 in most of the samples. As suggested in the previous studies "dosage-sensitive" genes with housekeeping functions were more likely affected by dosage imbalance and were

**Table 2**

The Raw FASTQ Files Generated by Us and Downloaded from NCBI GEO Database

	Tissue Type (Replicates)	Breed/Subspecies	Number of Samples	Library Type	BioProject ID	Reference
1	In vivo MII oocytes and embryos: 2-, 4-, 8-, 16-, 32-cell, CM, and BL ( $n = 2$ )	Holstein	8	Single-read SOLiD	PRJNA254699	Jiang et al. (2014)
2	In vitro GV and MII oocytes and embryos: 4-, 8-, 16-cell, and BL ( $n = 3$ )	German Simmental (♀) and Brahman (♂) cross	6	Single-read Illumina	PRJNA228235	Graf et al. (2014)
3	In vivo conceptuses: days 7 ( $n = 6$ ), 10 ( $n = 7$ ), 13 ( $n = 5$ ), 16 ( $n = 5$ ), 19 ( $n = 5$ )	Charolais and Limousin cross	5	Single-read Illumina	PRJNA243569	Mamo et al. (2012)
4	Female-specific tissue: endometria ( $n = 12$ ), corpora lutea ( $n = 14$ )	Holstein	2	Paired-end Illumina	PRJNA298914	Moore et al. (2016)
5	Female somatic tissues: brain, liver, muscle, and kidney ( $n = 4$ )	<i>Bos indicus</i> × <i>Bos taurus</i>	4	Single-read Illumina	PRJNA268096	Chen et al. (2015)
6	Male somatic tissues: fat, muscle, hypothalamus, duodenum, liver, lung, and kidney (pools of 7–14 animals)	<i>Bos taurus</i>	7	Paired-end Illumina	PRJEB6377	PRJEB6377, 2014
7	Female and male somatic tissues: fat, liver, muscle, and pituitary gland ( $n = 5$ )	Hanwoo (Korean cattle)	8	Paired-end Illumina	PRJNA273164	Seo et al. (2016)

upregulated (Pessia, Makino, et al. 2012). When nondosage-sensitive genes were included in the X:A calculation, the X:A ratio were likely lower (Sangrithi et al. 2017). Therefore, we further selected ubiquitously expressed genes (TPM > 1) throughout all somatic samples. Gene ontology analysis showed strong evidence that these ubiquitously expressed genes had housekeeping roles (supplementary fig. S2, Supplementary Material online), such as translation, RNA transcription, protein transport, and cellular and metabolic process (supplementary fig. S2, Supplementary Material online). A total of 245 and 7,603 genes on the X chromosome and autosomes, respectively, were included as ubiquitously expressed genes in the recalculation of the X:A ratios (fig. 2C and D). The confidence intervals encompassed 1 and the medians were greater than 1 in most of the samples. Brain and its specific regions such as the pituitary gland and hypothalamus had the highest X:A ratios (1.20, 1.28, and 1.22, respectively), consistent with previous reports in other species (Nguyen and Disteche 2006; Deng et al. 2011).

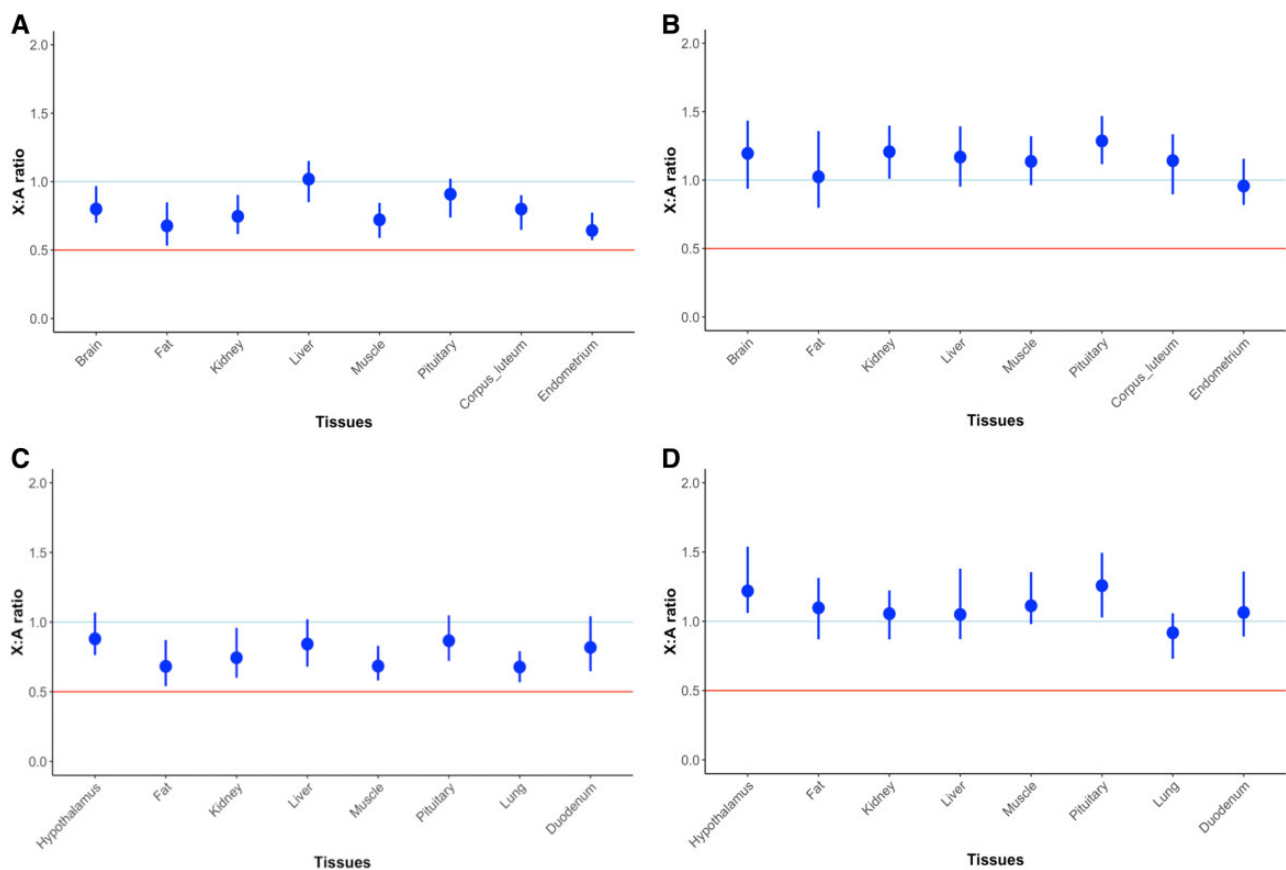
### X Chromosome Upregulation in Immature and Mature Oocytes

Germinal vesicle stage (immature) oocytes are arrested at the diplotene stage of the first prophase of meiosis (Pro I) (Mehlmann 2005), and contain a duplicated genome (XXXX:AAAA) and two active X chromosomes (Fukuda

et al. 2015). The matured oocytes, on the other hand, are arrested at the second metaphase of meiosis (MII) (Li and Albertini 2013) and are haploid (1 N) although each homologous chromosome contains two sister chromatids/complements of DNA (2 C; XX:AA). Our analysis included immature and both in vivo and in vitro matured oocytes (Graf et al. 2014; Jiang et al. 2014). First, we identified expressed genes (TPM > 1) and ubiquitously expressed genes across all preimplantation samples. Fewer expressed ubiquitous genes were found in these samples than in somatic tissues but similar gene ontology terms (supplementary figs. S3 and S4, Supplementary Material online). A total of 117 X-linked and 3,947 autosomal genes were used as ubiquitous genes for X:A ratio calculation. Compared with expressed genes (TPM > 1, X:A ≈ 0.75, fig. 3C), ubiquitously expressed genes had higher X:A ratios in immature and mature oocytes at 1 and 0.87, respectively (fig. 3D). Taken together, our analyses reveal a higher X:A ratio for ubiquitous genes, and a balanced X to autosome expression in immature diploid oocytes and an incomplete balance in mature haploid oocytes.

### X Chromosome Upregulation in Preimplantation Embryos

X inactivation and reactivation happen in cycles during early embryonic development. In mice, the zygote contains an inactive X from the sperm but three X chromosome reactivation events occur subsequently: (1) embryonic genomic activation



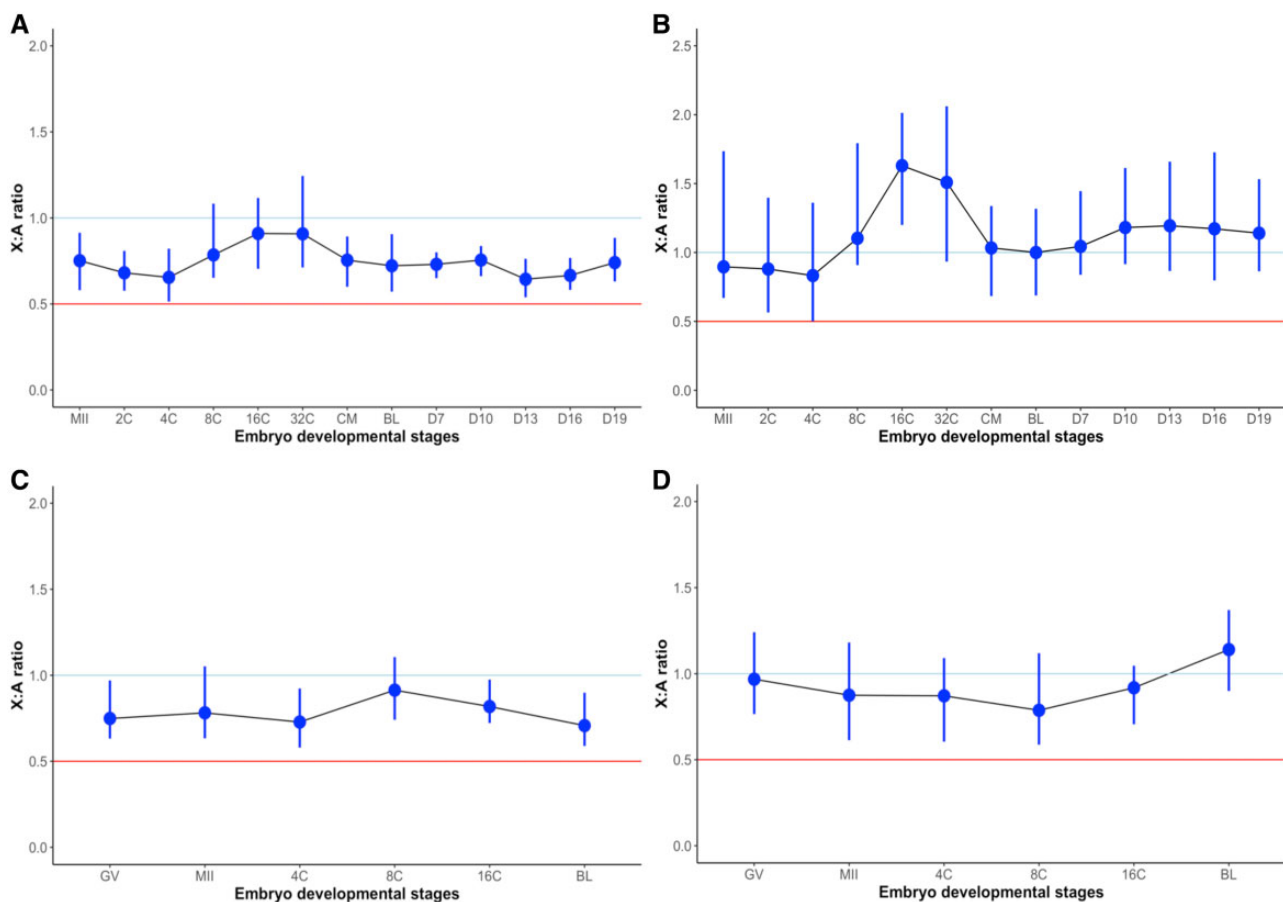
**FIG. 2.**—Median X:A ratios with 95% confidence intervals of bovine female (A, B) and male (C, D) somatic tissues. The X:A ratios were calculated using expressed (TPM > 1) genes (A, C) and ubiquitously expressed (TPM > 1 and present in all somatic data sets; B, D) genes. Blue line: X:A = 1, complete dosage compensation; red line: X:A = 0.5, no dosage compensation.

(EGA) at 2-cell, (2) pluripotency establishment in inner cell mass (ICM) of blastocyst, and (3) primordial germ cell generation in the genital ridge (Ohata and Wutz 2013). Using the bovine X-linked monamine oxidase type A (*MAOA*), Ferreira et al. (2010) demonstrated that transcripts from both the maternal and paternal *MAOA* were present in embryos at the 4-, 8-, to 16-cell, and blastocyst stages, while only the maternal transcripts were present in compact morula. These data revealed that XCI occurred in an imprinted fashion in the morula stage in the bovine and the paternal X was reactivated at the blastocyst stage. A more permanent random XCI was observed between the blastocyst and early elongation stages by analyzing seven X-linked genes in day14 embryos (Bermejo-Alvarez et al. 2011). However, the expression dynamics of an individual gene cannot represent the activity of the whole X chromosome, global transcript analysis will generate a more definitive conclusion on X inactivation-reactivation dynamics.

We therefore analyzed RNA-seq data from preattachment embryos. In *in vivo* produced preimplantation embryos, we observed an incomplete dosage compensation ( $0.5 < X:A < 1$ ) using expressed genes and an excess of

compensation ( $X:A > 1$ ) from the 8-cell to compact morula stages using ubiquitous genes (fig. 3A and B). X dosage slightly decreased after fertilization, with the lowest X:A ratio seen at the 4-cell stage, indicating that the sperm brought in an inactive X chromosome to the matured oocyte. The X:A ratio started to increase from the 4- to 8-cell stage, coincident with embryonic genome activation (EGA) for bovine *in vivo* embryos (Jiang et al. 2014), suggesting that EGA activates both paternal and maternal genome and has a more profound effect on the X chromosome. The increased X:A ratio exhibited excess compensation from the 8-cell to compact morula stage. A sharp decrease of X:A ratio was then observed between early (32-cell) and compact morula stages, corresponding to the first observed inactivation of the paternal X chromosome in bovine embryos. The X:A subsequently stabilized from days 7 to 19 of gestation, corresponding to random XCI between the blastocyst and early elongation stages.

Using expressed genes, *in vitro* produced bovine preimplantation embryos present a similar X:A dynamics from fertilization to the 8-cell stage (fig. 3C and D). However, no excess dosage compensation was observed using ubiquitous genes at any stage but the blastocyst, suggesting a deviation



**Fig. 3.**—Median X:A ratios with 95% confidence intervals of bovine in vivo (A, B) and in vitro (C, D) produced oocytes and pre- and postimplantation embryos. The X:A ratios were calculated using expressed (TPM > 1) genes (A, C) and ubiquitously expressed (TPM > 1 and present in all embryo data sets; B, D). GV, germinal vesicle; MII, metaphase of second meiosis; CM, compact morula; BL, blastocyst. Blue line: X:A = 1, complete dosage compensation; red line: X:A = 0.5, no dosage compensation.

of X compensation from in vivo embryos. These observations are consistent with recent findings of aberrant X regulation in in vitro produced human and mouse embryos (Tan et al. 2016).

### Effect of PAR and Putative XCI-Escaping Genes on X Chromosome Upregulation

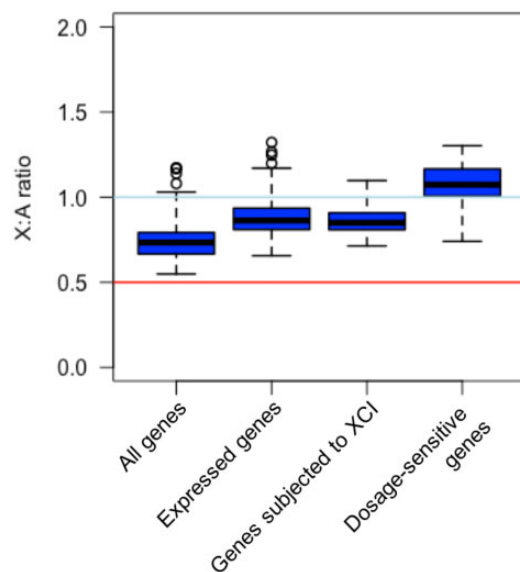
Pseudoautosomal regions (PARs) contain homologous genes between the X and Y chromosomes, and are important in homologous chromosome pairing and recombination during male meiosis (Das et al. 2009). A total of 20 PAR genes (supplementary table S3, Supplementary Material online) have been characterized in bovine, sheep, goats, and other ruminants (Raudsepp and Chowdhary 2015). Most PAR genes are known to escape XCI in humans (Helena Mangs and Morris 2007). Moreover, approximately 15% and 3% of non-PAR X-linked genes in humans and mice, respectively, are known to escape XCI (Berletch et al. 2011). In the bovine, 55 such X-linked genes (supplementary table S5, Supplementary Material online) were classified as candidates that escape

XCI (Ka et al. 2016). To tease out the effects of these bi-allelically expressed PAR genes and XCI-escaping genes, we plotted X:A ratios in the categories of “all genes,” “expressed genes,” “genes subjected to XCI (excluding PAR genes and putative XCI-escaping genes),” and “dosage-sensitive genes” (fig. 4). The X:A ratios for “all genes” had the median value closer to 0.5, while the ratios for “expressed genes” were closer to 1. When we excluded PARs and putative XCI-escaping genes from “expressed genes,” the outliers of extremely high X:A ratios disappeared, suggesting bi-allelically expressed X-linked genes did contribute to the high X:A ratios. Moreover, “dosage-sensitive genes” maintained the highest X:A ratios (> 1), further confirming the hyperexpression nature of this subgroup of genes.

### Discussion

In this study, we determined the X chromosome dosage profiles in four chromosome scenarios in the bovine. Immature oocytes represent diploid germline with duplicated genome/





**FIG. 4.**—Boxplot of the X:A medians for all data sets. Genes were categorized in “all genes,” “expressed (TPM > 1) genes,” “genes subjected to XCI (excluding PAR genes and putative XCI-escaping genes),” and “dosage-sensitive (ubiquitously expressed) genes.” Blue line: X:A = 1, complete dosage compensation; red line: X:A = 0.5, no dosage compensation.

four complements of DNA (XXXX:AAAA); mature oocyte represent haploid germline with duplicated genome/two complements of DNA (XX:AA); bovine preimplantation embryos at various stages represent the gradual change from two active X (XaXa:AA) to one inactive X chromosome (XaXi:AA); female and male somatic tissues contain diploid cells with one already inactivated X (XaXi:AA or XY:AA). Our analyses showed incomplete compensation ( $0.5 < X:A < 1$ ) of X chromosome to autosome pairs in all scenarios for expressed genes (TPM > 1) and excess compensation ( $X:A > 1$ ) for “dosage-sensitive” genes in somatic tissues and certain stages in early embryos. These findings suggest that X dosage upregulation occurs in bovine germlines, preattachment embryos, and somatic tissues analyzed here.

Our results in bovine are consistent with previous findings in other mammalian species using similar strategies of data filtering (Deng et al. 2011; Pessia, Makino, et al. 2012). However, studies applying different threshold criteria on RNA-seq data generated conflicting results for mammalian X dosage compensation. Xiong et al. (2010) included genes with low and no expression and reported X:A ratio close to 0.5. Whereas many follow-up studies reanalyzing the same RNA-seq data after removing the “noise” concluded hyperactivation of X on expressed genes (Deng et al. 2011), especially dosage-sensitive ones (genes encoding protein complex of seven or more members and having housekeeping roles) (Pessia, Makino, et al. 2012). Thus, dosage compensation, unlike XCI, has been proposed to be a local process with hyper-expression by only dosage-sensitive genes (Pessia,

Engelstädter, et al. 2014). In our study, we applied two gene selection methods in order to identify “dosage-sensitive” genes: expressed genes that are TPM > 1 and ubiquitous genes whose TPM’s are > 1 in all somatic samples or in all embryo samples. We found similar results as in humans and mice that incomplete dosage compensation was incomplete in most scenarios when expressed genes were used, while ubiquitous genes had a higher X dosage, implying that the expression of a group of X-linked bovine genes are collectively more than doubled. Furthermore, this level of upregulation could not be globally applied to all genes on the X chromosome.

The previous study by Xiong et al. (2010) filtered out reads that mapped to multiple locations of the genome and those that spanned splice junctions. Their unique mapping approach resulted in X:A ratio close to 0.5 and refuted the X dosage compensation conclusion. Multiple groups (Kharchenko et al. 2011; Pessia, Makino, et al. 2012) have re-analysis the data from Xiong et al. by different mapping strategy and reached opposite conclusions. Jue et al. (2013) showed that mapping strategies can significantly impact the conclusions regarding X:A ratios. The “unique” mapping strategy can remove reads from close paralogs which affects the X chromosome substantially (Jue et al. 2013). Considering this point, here we used nonunique mapping strategy. We used the Hisat2 software that by default reports both uniquely and multimapped reads Hisat2 is known to have the highest correctly multimapped reads compared with other aligners (Kim et al. 2015).

Bovine female somatic tissues had a slightly higher X:A ratio than the same tissues of males. This difference, however, was not significantly different ( $P = 0.45$ ). X chromosome expression distribution was also similar between sexes ( $P > 0.05$ ). These demonstrated that the expression of the X chromosome is balanced between males and females although they have different numbers of the X chromosome. The slightly higher X:A ratio in females may be a result of genes that escape XCI (Couldrey et al. 2017). Although it is unclear how many genes escape XCI in the bovine, a previous report documented a few X-linked genes escaping XCI in a mosaic fashion in the bovine (Yen et al. 2007).

The brain tissue had the highest X:A ratio compared with the other tissues, consistent with previous RNA-seq studies in several mammalian species (Nguyen and Distèche 2006; Deng et al. 2011). This could be related to the fact that many genes related to brain functions, such as *MAOA*, are located on the X chromosome (Zechner et al. 2001). On the contrary, X dosage of other bovine somatic tissues including fat, liver, muscle, and the pituitary gland was previously determined as incompletely compensated (Ka et al. 2016), which is consistent with results in humans, mice, and ours.

Germ cells and the developing embryos undergo drastic epigenetic changes. We observed balanced expression of the X chromosome with that of the autosomes in diploid

immature oocytes and incomplete balance in haploid matured oocytes. These data are consistent with those by Fukuda et al. who also showed higher dosage compensation of X chromosome in immature than matured oocytes in the human and mouse (Fukuda et al. 2015). However, X:A ratio slightly decreased after fertilization probably due to the sperm brings in some mRNAs of autosomal origin in addition to an inactive X (Huyhn and Lee 2003). Because early embryonic development is primarily dependent on stored maternal mRNAs and proteins which gradually degrade until EGA (Memili and First 2000), a consistent decrease and then increase of X:A before and after EGA, respectively, were observed. The timing of the changes in X:A in *in vivo* embryos, however, was one cell cycle earlier than their *in vitro* counterparts due to the timing difference between these two types of embryos, which are 4–8 and 8–16 cell stages, respectively (Telford et al. 1990; Misirlioglu et al. 2006; Kues et al. 2008; Graf et al. 2014; Jiang et al. 2014).

## Conclusion

In conclusion, our study shows the upregulation of X chromosome in four bovine genome scenarios, supporting a balanced expression between a single active X and autosome pairs. However, deviating from Ohno's theory, dosage compensation to rescue X haploinsufficiency appears to be an incomplete process for expressed genes but a complete process for "dosage-sensitive" genes. Removal of PAR genes and those putatively escape XCI eliminated the outliers of extremely high X:A ratios. In addition, the switch from imprinted XCI at the compact morula stage to random XCI at the blastocyst stage may happen so rapidly that the potential transient state of two active X chromosomes could not be captured in the current data set with limited time points. Whether a transient two active X state occurs or not during blastulation requires frequent sampling and further study. Lastly, no relative X expression difference was observed between bovine female and male somatic tissues, suggesting Ohno's hypothesis balanced the overall X between sexes.

## Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

Funding for this project was supported by the USDA National Institute of Food and Agriculture (NIFA) grants USDA-ARS: 1265–31000–091–02 S; NIFA-USDA: W2171/3171; National

Natural Science Foundation of China: 30760167; Department of Education of Xinjiang Uygur Autonomous Region Scholarship: 2016–3–0036.

## Author Contributions

X.C.T., J.C. and J.D. designed the study; Z.J., H.D., X.Z., and E.W. provided the data set; J.D., W.S., N.J., L.K. and R.O. analyzed the data; and J.D. and X.C.T. wrote the manuscript. All authors read and approved the final manuscript.

## Literature Cited

- Berlitch JB, Yang F, Xu J, Carrel L, Distche CM. 2011. Genes that escape from X inactivation. *Hum Genet.* 130(2):237–245.
- Bermejo-Alvarez P, Rizos D, Lonergan P, Gutierrez-Adan A. 2011. Transcriptional sexual dimorphism in elongating bovine embryos: implications for XCI and sex determination genes. *Reproduction* 141(6):801–808.
- Chen Z, et al. 2015. Characterization of global loss of imprinting in fetal overgrowth syndrome induced by assisted reproduction. *Proc Natl Acad Sci U S A.* 112(15):4618–4623.
- Couldrey C, et al. 2017. Bovine mammary gland X chromosome inactivation. *J Dairy Sci.* 100(7):5491–5500.
- Das PJ, Chowdhary BP, Raudsepp T. 2009. Characterization of the bovine pseudoautosomal region and comparison with sheep, goat, and other mammalian pseudoautosomal regions. *Cytogenet Genome Res.* 126:139–147.
- De La Fuente R, Hahnel A, Basrur PK, King WA. 1999. X inactive-specific transcript (Xist) expression and X chromosome inactivation in the pre-attachment bovine embryo. *Biol Reprod.* 60(3):769–775.
- Deng X, et al. 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nat Genet.* 43(12):1179–1185.
- Efron B, Tibshirani RJ. 1994. *An Introduction to the Bootstrap*. Boca Raton, Florida : Chapman and Hall/CRC CRC Press.
- Ercan S. 2015. Mechanisms of X chromosome dosage compensation. *J Genomics.* 3:1–19.
- Ferreira AR, et al. 2010. Allele-specific expression of the MAOA gene and X chromosome inactivation in *in vitro* produced bovine embryos. *Mol Reprod Dev.* 77(7):615–621.
- Fukuda A, Tanino M, Matoba R, Umezawa A, Akutsu H. 2015. Imbalance between the expression dosages of X-chromosome and autosomal genes in mammalian oocytes. *Sci Rep.* 5:14101.
- Gene Ontology Consortium. 2015. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43:D1049–D1056.
- Gevers D, Vandepoele K, Simillon C, Van de Peer Y. 2004. Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends Microbiol.* 12(4):148–154.
- Gordon DJ, Resio B, Pellman D. 2012. Causes and consequences of aneuploidy in cancer. *Nat Rev Genet.* 13(3):189–203.
- Graf A, et al. 2014. Fine mapping of genome activation in bovine embryos by RNA sequencing. *Proc Natl Acad Sci U S A.* 111(11):4139–4144.
- Gupta V, et al. 2006. Global analysis of X-chromosome dosage compensation. *J Biol.* 5(1):3.
- Hassold T, et al. 1980. A cytogenetic study of 1000 spontaneous abortions. *Ann Hum Genet.* 44(2):151–178.
- He X, et al. 2011. He *et al.* reply. *Nat Genet.* 43(12):1171–1172.
- He X, Zhang J. 2016. X-chromosome dosage compensation. In: eLS. John Wiley & Sons Ltd, Chichester. <http://www.els.net> [doi: 10.1002/9780470015902.a0026517].

- Heard E, Clerc P, Avner P. 1997. X-chromosome inactivation in mammals. *Annu Rev Genet.* 31(1):571–610.
- Helena Mangs A, Morris BJ. 2007. The human pseudoautosomal region (PAR): origin, function and future. *Curr Genomics.* 8(2):129–136.
- Holtzman DM, et al. 1992. Dysregulation of gene expression in mouse trisomy 16, an animal model of Down syndrome. *EMBO J.* 11(2):619–627.
- Huang DW, Sherman BT, Lempicki RA. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37(1):1–13.
- Hurles ME, Dermitzakis ET, Tyler-Smith C. 2008. The functional impact of structural variation in humans. *Trends Genet.* 24(5):238–245.
- Huynh KD, Lee JT. 2003. Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. *Nature* 426(6968):857–862.
- Jiang Z, et al. 2014. Transcriptional profiles of bovine in vivo pre-implantation development. *BMC Genomics* 15(1):756.
- Jue NK, et al. 2013. Determination of dosage compensation of the mammalian X chromosome by RNA-seq is dependent on analytical approach. *BMC Genomics* 14(1):150.
- Ka S, et al. 2016. Status of dosage compensation of X chromosome in bovine genome. *Genetica* 144(4):435–444.
- Kharchenko PV, Xi R, Park PJ. 2011. Evidence for dosage compensation between the X chromosome and autosomes in mammals. *Nat Genet.* 43(12):1167–1169.
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 12(4):357–360.
- Kues WA, et al. 2008. Genome-wide expression profiling reveals distinct clusters of transcriptional regulation during bovine preimplantation development in vivo. *Proc Natl Acad Sci U S A.* 105(50):19768–19773.
- Li R, Albertini DF. 2013. The road to maturation: somatic cell interaction and self-organization of the mammalian oocyte. *Nat Rev Mol Cell Biol.* 14(3):141–152.
- Lin F, Xing K, Zhang J, He X. 2012. Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. *Proc Natl Acad Sci U S A.* 109(29):11752–11757.
- Lyon MF. 1961. Gene Action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* 190(4773):372–373.
- Mamo S, Rizos D, Lonergan P. 2012. Transcriptomic changes in the bovine conceptus between the blastocyst stage and initiation of implantation. *Anim Reprod Sci.* 134(1–2):56–63.
- Mehlmann LM. 2005. Stops and starts in mammalian oocytes: recent advances in understanding the regulation of meiotic arrest and oocyte maturation. *Reproduction* 130(6):791–799.
- Memili E, First NL. 2000. Zygotic and embryonic gene expression in cow: a review of timing and mechanisms of early gene expression as compared with other species. *Zygote* 8(1):87–96.
- Misirlioglu M, et al. 2006. Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. *Proc Natl Acad Sci U S A.* 103(50):18905–18910.
- Moore SG, et al. 2016. Differentially expressed genes in endometrium and corpus luteum of Holstein cows selected for high and low fertility are enriched for sequence variants associated with fertility. *Biol Reprod.* 94:19.
- Nguyen DK, Disteche CM. 2006. Dosage compensation of the active X chromosome in mammals. *Nat Genet.* 38(1):47–53.
- Nicolae M, Mangul S, Măndoiu II, Zelikovskiy A. 2011. Estimation of alternative splicing isoform frequencies from RNA-Seq data. *Algorithms Mol Biol.* 6(1):9.
- Ohhata T, Wutz A. 2013. Reactivation of the inactive X chromosome in development and reprogramming. *Cell Mol Life Sci.* 70(14):2443–2461.
- Ohno S. 1966. Sex chromosomes and sex-linked genes. Springer-Verlag Berlin Heidelberg. Salmon Tower Building, New York City. [www.springer.com/us/book/9783642881800](http://www.springer.com/us/book/9783642881800) (Accessed February 19, 2018).
- Ohno S, Kaplan WD, Kinosita R. 1959. Formation of the sex chromatin by a single X-chromosome in liver cells of *Rattus norvegicus*. *Exp Cell Res.* 18(2):415–418.
- Okamoto I, et al. 2011. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* 472(7343):370–374.
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* 11(9):1650–1667.
- Pessia E, Engelstädter J, Marais GAB. 2014. The evolution of X chromosome inactivation in mammals: the demise of Ohno's hypothesis? *Cell Mol Life Sci.* 71(8):1383–1394.
- Pessia E, Makino T, Bailly-Bechet M, McLysaght A, Marais GAB. 2012. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proc Natl Acad Sci U S A.* 109(14):5346–5351.
- Raudsepp T, Chowdhary BP. 2015. The eutherian pseudoautosomal region. *CGR* 147:81–94.
- Sangrithi MN, et al. 2017. Non-canonical and sexually dimorphic X dosage compensation states in the mouse and human germline. *Dev Cell.* 40(3):289–301.e3.
- Schaarschmidt F, Gerhard D. 2015. Confidence intervals for two sample comparisons. R package version 0.1–15.
- Seo M, et al. 2016. Comprehensive identification of sexually dimorphic genes in diverse cattle tissues using RNA-seq. *BMC Genomics* 17:81.
- Tan K, et al. 2016. Impaired imprinted X chromosome inactivation is responsible for the skewed sex ratio following in vitro fertilization. *Proc Natl Acad Sci U S A.* 113(12):3197–3202.
- Telford NA, Watson AJ, Schultz GA. 1990. Transition from maternal to embryonic control in early mammalian development: a comparison of several species. *Mol Reprod Dev.* 26(1):90–100.
- Veitia RA, Potier MC. 2015. Gene dosage imbalances: action, reaction, and models. *Trends Biochem Sci.* 40(6):309–317.
- Xiong Y, et al. 2010. RNA sequencing shows no dosage compensation of the active X-chromosome. *Nat Genet.* 42(12):1043–1047.
- Xue F, et al. 2002. Aberrant patterns of X chromosome inactivation in bovine clones. *Nat Genet.* 31(2):216–220.
- Yen ZC, Meyer IM, Karalic S, Brown CJ. 2007. A cross-species comparison of X-chromosome inactivation in Eutheria. *Genomics* 90(4):453–463.
- Zechner U, et al. 2001. A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? *Trends Genet.* 17(12):697–701.

Associate editor: Soojin Yi