



Bovine mammary gland X chromosome inactivation

C. Couldrey,¹ T. Johnson, T. Lopdell, I. L. Zhang, M. D. Littlejohn, M. Keehan, R. G. Sherlock, K. Tiplady, A. Scott, S. R. Davis, and R. J. Spelman

Research and Development, Livestock Improvement Corporation, Hamilton 3240, New Zealand

ABSTRACT

X chromosome inactivation (XCI) is a process by which 1 of the 2 copies of the X chromosomes present in female mammals is inactivated. The transcriptional silencing of one X chromosome achieves dosage compensation between XX females and XY males and ensures equal expression of X-linked genes in both sexes. Although all mammals use this form of dosage compensation, the complex mechanisms that regulate XCI vary between species, tissues, and development. These mechanisms include not only varying levels of inactivation, but also the nature of inactivation, which can range from being random in nature to driven by parent of origin. To date, no data describing XCI in calves or adult cattle have been reported and we are reliant on data from mice to infer potential mechanisms and timings for this process. In the context of dairy cattle breeding and genomic prediction, the implications of X chromosome inheritance and XCI in the mammary gland are particularly important where a relatively small number of bulls pass their single X chromosome on to all of their daughters. We describe here the use of RNA-seq, whole genome sequencing and Illumina BovineHD BeadChip (Illumina, San Diego, CA) genotypes to assess XCI in lactating mammary glands of dairy cattle. At a population level, maternally and paternally inherited copies of the X chromosome are expressed equally in the lactating mammary gland consistent with random inactivation of the X chromosome. However, average expression of the paternal chromosome ranged from 10 to 90% depending on the individual animal. These results suggest that either the mammary gland arises from 1 or 2 stem cells, or a nongenetic mechanism that skews XCI exists. Although a considerable amount of future work is required to fully understand XCI in cattle, the data reported here represent an initial step in ensuring that X chromosome variation is captured and used in an appropriate manner for future genomic selection.

Key words: RNA sequencing, X chromosome inactivation, genomic selection, cattle

INTRODUCTION

The mammalian X chromosome is a large, gene-rich chromosome. Females have 2 X chromosomes, whereas males have a single X chromosome and a small, gene-poor Y chromosome. If not for X chromosome inactivation (XCI) taking place during early embryonic development, the difference in X chromosome gene dosage between the sexes would lead to X-linked gene expression differences in males and females.

Although dosage compensation through XCI has been the subject of intense study for more than 50 yr, the overwhelming majority of research has focused on understanding its regulation in model organisms, particularly mice and *Caenorhabditis elegans*. To date, few reports have described XCI in cattle beyond investigation of gene expression from in vitro generated preimplantation embryos (De La Fuente et al., 1999; Peippo et al., 2002; Xue et al., 2002). X chromosome gene expression studies from these in vitro bovine embryos are consistent with well-studied model organisms, in that XCI occurs by the blastocyst stage of development. However, substantial developmental differences between embryos produced in vivo and those produced in vitro are well documented (Niemann et al., 2008) and care is still required when data are extrapolated to in vivo embryo development. The only in vivo data of XCI in cattle have been generated from *Bos taurus/Bos indicus* crosses (Dindot et al., 2004; Chen et al., 2016). Analysis of expression patterns from a single X-linked gene (*XIST*) indicated that XCI was occurring in a parent of origin manner in extraembryonic tissues and in a random manner in the embryo proper (Dindot et al., 2004). However, this conflicts with extensive analysis of X-linked gene expression in extraembryonic membranes where XCI appeared random in nature (Chen et al., 2016). Convincing evidence is lacking on XCI timing and mechanisms in the developing bovine embryo, fetus, or adult tissues. These conflicting results contribute to the challenge of understanding the contribution of genetic variation from the X chromosome to phenotypes

Received December 19, 2016.

Accepted March 15, 2017.

¹Corresponding author: Christine.couldrey@lic.co.nz

of interest. These challenges have largely resulted in the exclusion of this large gene-rich chromosome from genomic selection strategies.

In the absence of cattle-specific data, we must rely on XCI data from mice to infer potential mechanisms and timings for this process. In mice, 2 forms of XCI have been widely reported during embryogenesis (Deng et al., 2014). Beginning at the 4-cell stage, imprinted XCI exclusively silences the paternal X chromosome. Later, around the time of embryonic implantation (developmental d 3.5), epiblast cells of the inner cell mass, which give rise to the embryo, reactivate the paternal X chromosome and undergo a random form of XCI (Deng et al., 2014). Random XCI in the inner cell mass is permanent and therefore gives rise to patches of clonally derived cells having the same X chromosome inactivated in the adult.

During random XCI, the 2 X chromosomes theoretically have equal chances of being inactivated. Hence, the percentage of cells with the paternal X chromosome inactivated is expected to be approximately 50%. However, in practice, XCI can be skewed in one or more tissues of an individual. This skewing of XCI can occur by random chance, be mediated by genetic mechanisms (Thorvaldsen et al., 2012; Calaway et al., 2013), or be a result of an X-linked mutation that affects cell proliferation or survival (Belmont, 1996).

Although XCI silences the majority of the X chromosome, several genes are known to escape this inactivation to some extent (Valencia and Wutz, 2015). Whereas many genes that escape XCI are located within the pseudo-autosomal regions (PAR) of the X chromosome and have a homolog on the Y chromosome, others have been located outside the pseudo-autosomal regions. Although the mechanisms of escape from XCI of non-PAR genes are not fully understood, both the number and distribution of these genes across the X chromosome differ significantly between species (Valencia and Wutz, 2015).

One gene that escapes random XCI (but retains monoallelic expression) in mice is the RING finger ubiquitin ligase *Rnf12/RLIM* gene (Bach et al., 1999; Ostendorff et al., 2002). Genetic analyses in mice have demonstrated that *Rnf12/RLIM* expressed from the paternal X chromosome serves as a survival factor for milk-producing alveolar cells in mammary glands of pregnant and lactating females (Jiao et al., 2012, 2013). These studies also reported that only the paternal *Rnf12/RLIM* allele was expressed in the majority of mammary epithelial cells, leading Jiao et al. (2013) to suggest that this gene displayed an imprinted XCI pattern of expression in this cell type. However, to date no further evidence supports this suggestion in other mammals, and although many aspects of XCI appear

well conserved throughout mammals, extrapolation of *Rnf12/RLIM* regulation to other mammals should be undertaken with caution.

To date, genome-wide association studies and genomic selection strategies undertaken in cattle have for the most part not included genetic variation on the X chromosome due to challenges around Mendelian transmission and lack of data on XCI in cattle regarding whether this mechanism is random in nature or determined by parent of origin. The implications of X chromosome inheritance patterns and random/imprinted XCI in the mammary gland are considerable in terms of undertaking genomic selection for improved production performance. For example, if parent of origin XCI resulted in expression of only the paternally inherited X chromosome, selection of bulls with desirable X chromosome variation would lead to rapid improvement in production performance for all their daughters. Conversely, if XCI resulted in expression only from the maternally inherited X chromosome, improvement in production would take many generations and rely entirely on increasing the frequency of desirable alleles in the population. These implications are particularly important in the context of dairy cattle breeding where a relatively small number of bulls pass their single X chromosome on to all of their daughters. This project has undertaken analyses to understand XCI in lactating mammary glands of New Zealand dairy cattle as a first step in including X chromosome variation in genomic selection strategies.

MATERIALS AND METHODS

Ethics Statement

Animal ethics approval was granted for all animal work by the Ruakura Animal Ethics Committee, Hamilton, New Zealand. No animals were killed for this study. Mammary tissue samples were obtained by needle biopsy in accordance with protocols approved by the ethics committee (approval AEC 12845).

DNA Genotypes

To determine levels of heterozygosity in a representative sample of the New Zealand dairy population, whole genome sequencing, read mapping, and variant calling was undertaken on 556 individuals as previously described (Littlejohn et al., 2016). Variants were called across the entire genome including the X chromosome. BCFtools stats (Li et al., 2009) was used to obtain the per sample counts for heterozygous SNP at each position from the sequence reference set of variants. Proportion heterozygosity was calculated in each in-

dividual for each of the chromosomes using all 556 individuals for autosomes and the 312 females for the X chromosome.

For XCI analysis, all cows from which a mammary gland biopsy was collected were genotyped on the Illumina BovineHD BeadChip platform (Illumina, San Diego, CA). Genotypes for the sires of RNA-seq cows were obtained from either the Illumina BovineHD BeadChip ($n = 43$) or whole-genome sequencing ($n = 257$).

Genotypes for a lactating cow (subsequently referred to as the **SRAcow**) and her sire (accession numbers SRS597746 and SRS597673) were obtained from an independent data set by downloading whole genome sequence files from an existing sequence read archive previously used for multi-tissue RNA-seq (Chamberlain et al., 2015).

Illumina BovineHD BeadChip genotypes from 358 bulls and 3,307 cows were used to identify the pseudo-autosomal region in dairy cattle. The proportion of animals heterozygous at each SNP on chromosome X was calculated at each SNP for bulls and cows separately. The region for which the proportion of heterozygosity was similar between males and females was determined to be the pseudo-autosomal region.

RNA Sequencing

Mammary gland biopsies and RNA sequencing data used in this study have previously been described in detail (Littlejohn et al., 2014, 2016). Of the 406 cows sampled, a 300-cow subset was retained after filtering to remove animals greater than 3 standard deviations from the mean in the first 6 principal components as previously described (Littlejohn et al., 2016). Distribution of sire allele expression was analyzed and a further 2 animals were removed from the data set due to excess kurtosis, leaving 298 cows for analysis of XCI.

Mapped RNA sequence data generated in a previously described, independent data set of technical replicates across multiple tissues of the SRAcow (Chamberlain et al., 2015) were downloaded from the sequence read archive (SRP042639). These data were then converted to BAM (binary alignment/map) format using sam-dump from the SRA toolkit 2.5.7 (<https://github.com/ncbi/sra-tools>).

Analysis of Allele-Specific Expression

Heterozygous genotype calls located within annotated exons (Ensemble build 79) across all chromosomes were identified from BovineHD BeadChip genotyping in the 298 RNA-seq cows described above. Sire genotypes were identified either from BovineHD BeadChip genotyping

or whole genome sequencing. For the independent SRAcow, heterozygous genotype calls located within exons of all chromosomes were identified from sequence files downloaded from the sequence read archive. To allow a direct comparison of data from the SRAcow and cows in our study, the SRAcow sequence-called variants were filtered to retain only SNP corresponding to positions present on the BovineHD BeadChip. The use of genotypes derived from DNA, rather than the RNA-seq data itself is necessary for these analyses. If genotypes are determined from RNA-seq data considerable bias is introduced due to the fact that heterozygous sites are not identified where expression is highly skewed toward a single parent. Allelic depths at all heterozygous BovineHD BeadChip sites within exons were then determined from both data sets using GATK (Xue et al., 2002) SNP calling from RNA-seq BAM files.

Mapping Bias

Allelic depths from the 298 RNA-seq animal subset were combined and the proportion of reads containing the reference allele (UMD3.1; <http://bovinegenome.org/?q=node/61>) was calculated for heterozygous SNP with a combined read depth of greater than 100. The proportion of reference allele expression was examined at individual SNP, and as a chromosome average for each chromosome and also as a genome-wide average.

Paternal Chromosome Expression

To determine population trends of XCI and examine whether regions of the X chromosome were expressed in a parent of origin fashion, allelic depths determined by GATK from all animals were combined. Heterozygous SNP with a combined read depth of greater than 100 were used to determine the average expression from the paternal chromosomes across all animals at individual sites.

Animals with greater than 200 RNA-seq reads across all heterozygous positions on the X chromosome were used for individual animal analysis. The proportion of expression from the X chromosome inherited from the sire was calculated for each animal at each heterozygous position with a read depth greater than 10, using GATK-generated allelic depth. These data were further used to determine average expression from the paternal chromosome X for each individual. The entire bioinformatic pipeline described in the Materials and Methods is illustrated in Supplemental Figure S1 (<https://doi.org/10.3168/jds.2016-12490>).

The distribution of paternal chromosome expression was visually assessed by Q-Q plots and statistically tested using the Anderson-Darling normality test. Pa-

ternal chromosome expression in the autosomes was observed to follow a normal distribution, and statistical significance in the population was determined using a *t*-test for autosomal chromosomes. The distribution of average paternal X chromosome expression for each individual did not follow a normal distribution, but was not observed to be multimodal based on Hartigan's dip test. The nonparametric Wilcoxon signed-rank test was used to determine significance of paternal X chromosome expression.

Genome-Wide Association

Genome-wide association was undertaken using Illumina BovineHD BeadChip genotypes and proportion of paternal expression for individual RNA-seq cows as phenotypes. Single SNP analysis using PLINK (version 1.9; Purcell et al., 2007) and a BayesC approach using GenSel (version 4.53; Fernando and Garrick, 2010) were both tested to determine if regions of the genome were associated with paternally inherited X chromosome expression. Levels of significance from PLINK genome-wide association studies (GWAS) were compared with levels of significance produced from GWAS analysis following (1) randomization of paternal expression phenotypes (repeated 10 times) and (2) simulation of phenotypes by randomly sampling from a normal distribution (repeated 10 times).

RESULTS

Identification of the X Chromosome Pseudo-Autosomal Region

Cow Illumina BovineHD BeadChip genotypes showed highly heterozygous sites spread along the length of the X chromosome (Supplemental Figure S2; <https://doi.org/10.3168/jds.2016-12490>). In contrast, the presence of SNP reported as heterozygous in greater than 1% of the bulls was largely restricted to 2 (presumably pseudo-autosomal) regions. The first region extended from approximately 137 to 141 Mb and the second beginning at approximately 144 Mb and extending to the end of the chromosome (Supplemental Figure S2; <https://doi.org/10.3168/jds.2016-12490>). The entire region from 137 Mb to the end of the chromosome was considered to be pseudo-autosomal and excluded from further analysis.

Heterozygosity on the X Chromosome

Whole-genome sequencing of 556 animals resulted in the detection of 730,602 SNP on the non-PAR of the X chromosome. Mean proportion heterozygosity across

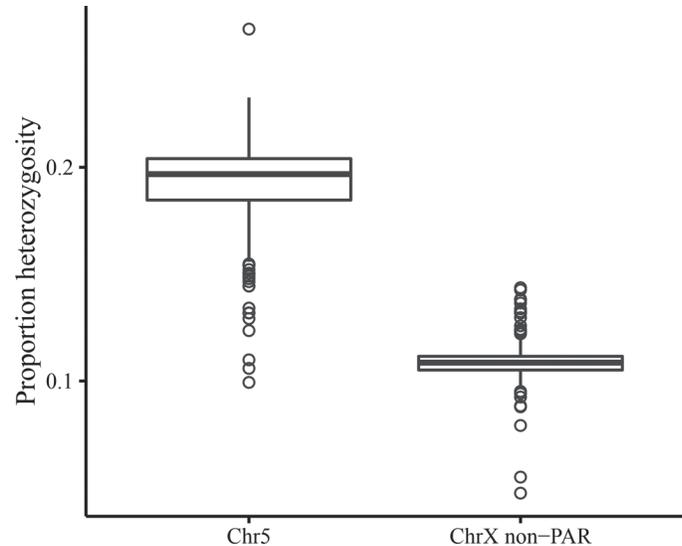


Figure 1. Proportion of heterozygosity observed in New Zealand dairy cattle from whole-genome sequence data. Heterozygosity at individual SNP on the non-pseudo-autosomal regions (PAR) of the X chromosome (ChrX) was calculated in cows, whereas a mixture of cows and bulls were analyzed to determine heterozygosity on autosomes (chromosome 5, Chr5, is illustrated as a comparison). The upper, middle, and lower lines in each plot define the upper quartile, medium, and lower quartiles, respectively; SNP with heterozygosity levels that fall outside this range are illustrated by circles.

this region in the 312 sequenced females on the X chromosome was 0.11 (SD = 0.009; Figure 1). In contrast, the mean heterozygosity across all 19,659,070 SNP detected on the 29 autosomes was 0.21 (SD = 0.012). Comparison of heterozygosity on the non-PAR of the X chromosome with the similarly sized (albeit somewhat smaller) chromosome 5 is illustrated in Figure 1. A total of 823,709 SNP were detected on chromosome 5 with a mean proportion heterozygosity of 0.19 (SD = 0.017).

RNA-seq Mapping Bias

Illumina BovineHD BeadChip analysis resulted in genotype data at sites within exons for 10,980 SNP across the 29 autosomes and 417 SNP on the X chromosome. Combining data from all animals with a total RNA-seq read depth of at least 100 across heterozygous sites within exons resulted in 72 and 6,280 informative sites on the X chromosome and all autosomes, respectively. Using these informative sites, distribution of percentage of reads mapping to the reference allele on all individual chromosomes was unimodal and centered around 50% (Supplemental Figure S3, <https://doi.org/10.3168/jds.2016-12490>, illustrates the distribution of mapping percentages at all informative sites across the genome).

Paternal Chromosome Expression

Population Distribution of Paternal Expression. Using the same 6,280 (all autosomes), 72 (X chromosome), and 337 (chromosome 5) informative SNP described in the mapping bias analysis above, parent of origin expression was examined. At the population level, lactating mammary gland gene expression did not indicate preferential expression of either the paternal or maternal allele [X chromosome: 52.05% ($P = 0.14$) paternal allele; chromosome 5: 50.22% ($P = 0.08$) paternal allele; all autosomes: 50.05% ($P = 0.08$)].

Although XCI is largely thought to occur at random in somatic tissues, imprinted X chromosome expression patterns have been reported in mice (Shin et al., 2010; Jiao et al., 2012). The average percentage of paternal expression at individual SNP across all animals collectively was close to 50% for all X chromosome SNP and therefore did not indicate obvious parent-of-origin-specific expression at any position (Figure 2). In contrast, expression at a single SNP on chromosome 5 (in the 6 animals heterozygous at this site with suitable read coverage) was more than 75% derived from the paternal allele (t -test $P = 0.04$) (Figure 2).

Individual Animal Distribution of Paternal X Chromosome Expression. The RNA-seq data were analyzed in individual animals. Average paternal expression at heterozygous SNP along the length of the X chromosome (excluding the PAR) was calculated for each animal whose minimum total RNA-seq read count across heterozygous sites for the individual was >200 . The number of SNP able to be analyzed for each individual ranged from 10 to 24. Figure 3 illustrates the frequency of average paternal expression observed across all heterozygous sites for each cow. In contrast to the results observed at a population level, the distribution of paternal expression on the X chromosome in individual animals appeared to be different to that on the autosomes (Figure 3 for chromosome 5 and Supplemental Figure S4, <https://doi.org/10.3168/jds.2016-12490>, for all chromosomes). Average gene expression for autosomal SNP for individuals is biallelic with similar levels of expression from both maternal and paternal chromosomes. By contrast, on the X chromosome of the same individuals, average expression varies from $>90\%$ maternal expression to $>90\%$ paternal expression (Figure 3).

To further investigate XCI in individuals, the proportion of paternal allele expression at each informative SNP was plotted across the X chromosome and chromosome 5 to allow comparisons between individual animals (examples presented in Supplemental Figure S5; <https://doi.org/10.3168/jds.2016-12490>). The percentage of paternal expression in individuals shows some

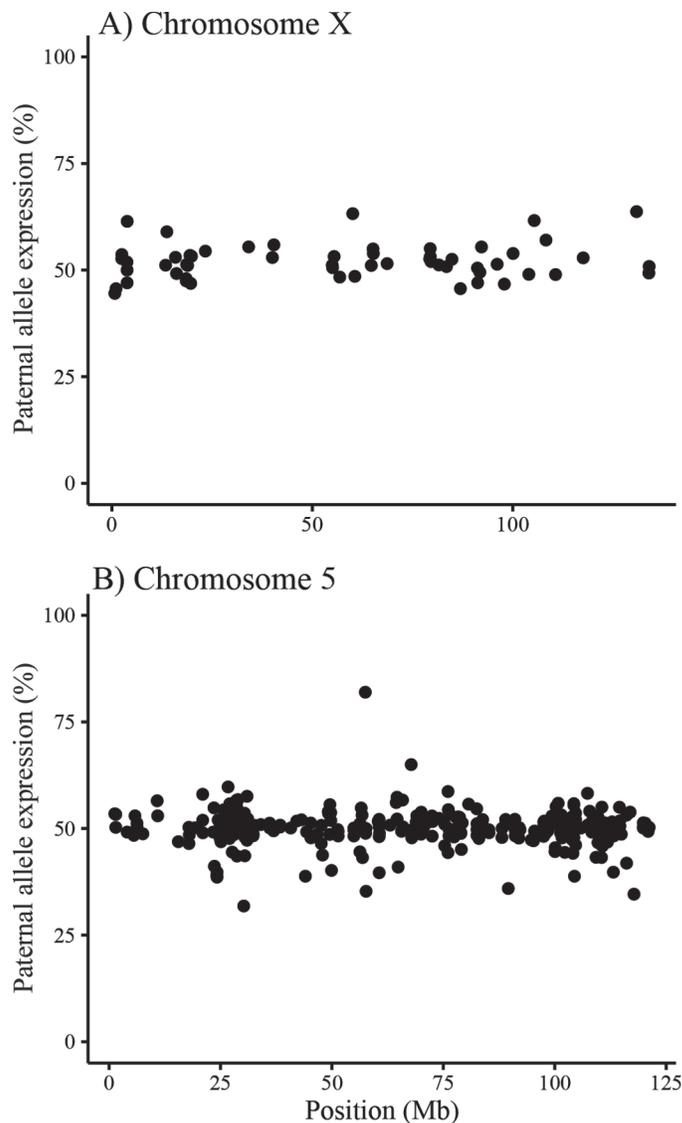


Figure 2. Parent of origin expression in the population at individual Illumina HD BeadChip (Illumina, San Diego, CA) SNP on the non-pseudo-autosomal regions (PAR) of the X chromosome (A) and chromosome 5 (B) as a comparison.

variation across the length of the chromosomes (within the range expected using a minimum read depth of 10). In spite of the variation along the length of each chromosome, chromosome 5 paternal expression is centered around 50%. In contrast, whereas paternal expression of individual SNP on chromosome X is similar along the length of the chromosome within each individual, they are centered around the chromosomal average, which ranges from 10 to 90% depending on the individual.

A gene shown to be imprinted in the mouse mammary gland, *Rfn12/RLIM* (Jiao et al., 2012), did not display consistent parent of origin expression in the bovine mammary gland (Supplemental Figure S6;

<https://doi.org/10.3168/jds.2016-12490>). Instead, parent of origin expression was consistent with the average level of paternal expression across the X chromosome in the same individual.

Analysis of sire allele expression from the SRAcow indicated high repeatability between the 3 technical replicates in all tissues (Supplemental Figure S7; <https://doi.org/10.3168/jds.2016-12490>). Considerable variation in the level of paternal expression at individual SNP was observed in each tissue on chromosomes X and 5 (Figure 4). A comparison of these 2 chromosomes

showed that average paternal expression from chromosome 5 was approximately 50% in each tissue (range 46–54%), whereas average paternal expression from chromosome X ranged from 29 to 65% with the lowest levels being observed in the lactating mammary gland (Figure 4). This level of paternal expression positions the SRAcow into the low group of paternal expression seen in the 298 RNA-seq animals (Figure 3).

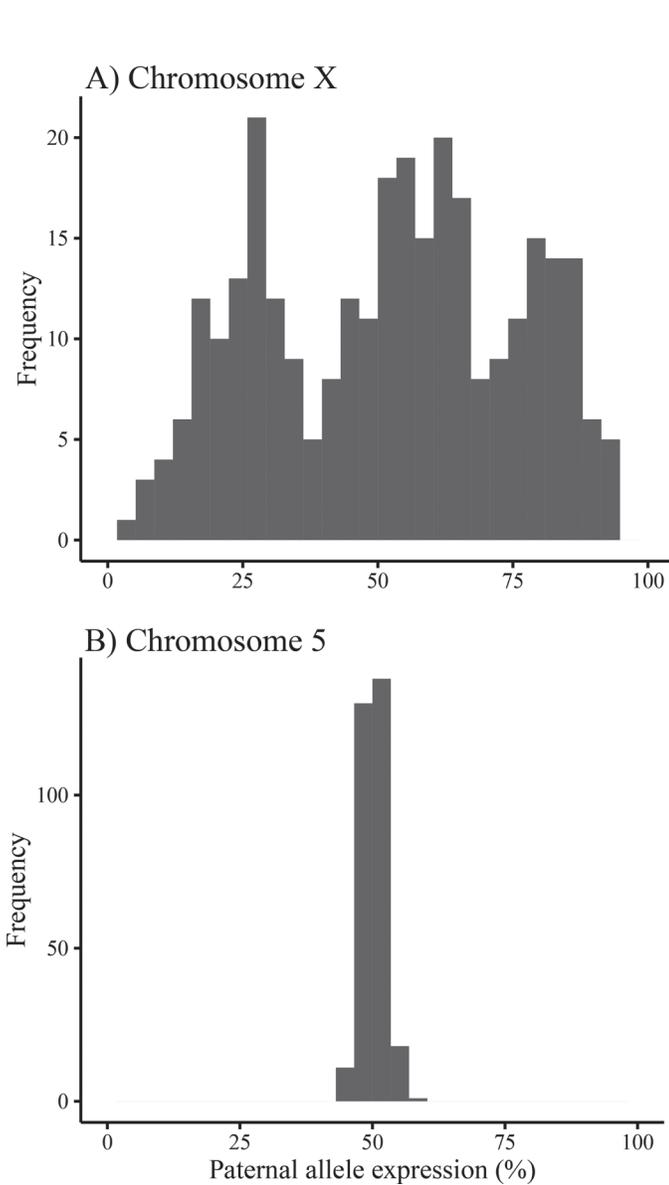


Figure 3. Frequency of mean parent of origin expression in individual animals at Illumina HD BeadChip (Illumina, San Diego, CA) SNP on the non-pseudo-autosomal regions (PAR) of the X chromosome (A) and chromosome 5 (B) as a comparison (for all chromosomes, see Supplemental Figure S4; <https://doi.org/10.3168/jds.2016-12490>).

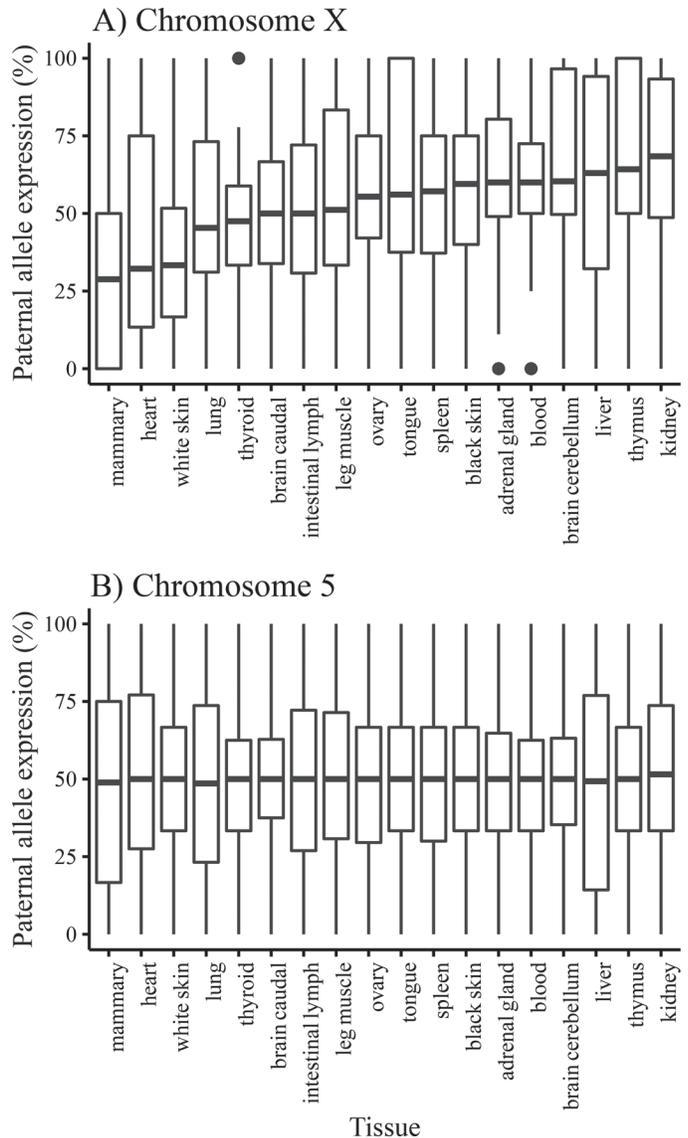


Figure 4. Parent of origin expression at individual Illumina HD BeadChip (Illumina, San Diego, CA) SNP across 18 tissues from a single cow on the non-pseudo-autosomal regions (PAR) of the X chromosome (A) and chromosome 5 (B) as a comparison. The upper, middle, and lower lines in each plot define the upper quartile, medium, and lower quartiles, respectively. Range of outliers is indicated by vertical lines and dots.

Genetic Basis of X Chromosome Inactivation

Although the distribution of the proportion of paternal X chromosome expression was not significantly different to a unimodal distribution, visually the distribution appeared trimodal in the 298 RNA-seq cows analyzed and was suggestive of a “phenotype” under genetic control. PLINK GWAS analysis using proportion of paternal expression as the phenotype indicated a genomic region on chromosome X that may contribute to parent-of-origin-specific XCI. However, analysis of the same data using a Bayesian GWAS approach did not identify any significant correlation between genotype and parent-of-origin-specific XCI. Further analysis of the PLINK approach comparing significance of individual SNP after randomization of phenotypes and simulation of random phenotypes confirmed that the PLINK GWAS peaks identified were random in nature.

DISCUSSION

Characteristics of X Chromosome in New Zealand Dairy Cattle

Analyses of SNP genotype heterozygosity in New Zealand dairy cattle for the purposes of identifying the PAR were similar to previously published reports in cattle (Jansen et al., 2013). The presence of 2 distinct regions of heterozygosity separated by a region of low heterozygosity in bulls is suggestive of 2 PAR, especially as the intervening region is frequently observed in a heterozygous state in cows. It is, however, possible that this pattern of heterozygosity observed in bulls is due to misassembly of chromosome X in UMD3.1. Analysis of the more recent Btau5.0.1 genome assembly (https://www.ncbi.nlm.nih.gov/assembly/GCA_000003205.6) does not provide any further insight as to whether these results represent 2 PAR or misassembly of the genome. Given the ambiguity of the PAR, the entire region of the X chromosome was excluded from further analysis.

The difference in the number of heterozygous SNP observed between the X chromosome and chromosome 5 is likely due, at least in part, to the hemizygous state of the X chromosome in males, which results in greater exposure to selection (Gottipati et al., 2011). Furthermore, the large half-sib population structure in New Zealand dairy cattle (due to the widespread use of AI) has resulted in a relatively small number of highly used bulls transmitting their only copy of the X chromosome to all their daughters, which is also likely to be a contributing factor to the reduced diversity of the X chromosome when compared with chromosome 5.

Given the level of sequence variation expected between the dairy cattle population in these experiments

and the cattle reference genome derived from a beef breed (Daetwyler et al., 2014), preferential mapping of sequences containing reference alleles was a potential risk. No bias was observed toward the reference allele in our Illumina HiSeq generated RNA-seq data at Illumina BovineHD BeadChip sites. This result is contradictory to a widely cited manuscript (Degner et al., 2009); however, it supports the findings of others (Panousis et al., 2014). The discrepancy in the observation of mapping bias in RNA-seq studies reported in peer-reviewed publications is likely to be in most part due to the selection of BovineHD SNP sites during the design of the SNP chip. These sites do not include indels, which appear to cause the largest proportion of mapping bias in sequence data (data not shown). Due to the absence of mapping bias in this and other studies using this data set, mapping of RNA-seq data to parent-specific genomes was not undertaken.

Is XCI Truly Random?

Analysis of average paternal expression across all animals for each individual SNP in the non-PAR region of X chromosome was strongly suggestive that, at a population level, the X chromosome is expressed equally from maternally and paternally inherited chromosomes. However, to build appropriate models for analyzing the effects of X chromosome variation in terms of production traits, a more detailed understanding of XCI is required.

In other species, evidence indicates that some X chromosome genes in the non-PAR are able to escape XCI, often in a species-specific manner (Deng et al., 2014). Our analysis revealed that, of the 1,128 genes annotated in the non-PAR region of the bovine X chromosome, approximately half were expressed in the lactating mammary gland. Of these, 72 contained one heterozygous Illumina BovineHD BeadChip present in an exon and were therefore able to be analyzed in this study. Individual animal and SNP analysis showed that none of the genes were simultaneously expressed from both copies of chromosome X, suggesting that none of these genes consistently escape XCI. Although further whole-genome sequencing-based analysis is required to be able to undertake a more comprehensive analysis of genes able to escape XCI in the lactating mammary gland, the results presented here suggest that escape from XCI is not a widespread occurrence in the bovine mammary gland. Although it is tempting to speculate on XCI based on results from other species, the lack of *in vivo* data, together with differences in escape from XCI across tissues, development, and sex (Calaway et al., 2013; Deng et al., 2014), make it difficult to com-

pare the results presented here with those from other species or to infer XCI mechanisms from other species.

One gene for which in vivo evidence of nonrandom inactivation has been reported is *Rfn12/RLIM*. However, rather than biallelic expression, *Rfn12/RLIM* is an example of an X chromosome gene that displays an imprinted pattern of expression in mice. Maternally expressed *Rfn12/RLIM* is required for XCI in the developing embryo (Shin et al., 2010), whereas paternal expression is not only predominant in, but absolutely required as a survival factor for milk-producing alveolar cells (Jiao et al., 2012). Analysis of *Rfn12/RLIM* expression in lactating mammary gland samples from cows heterozygous at an Illumina BovineHD BeadChip SNP in exon5 (X chromosome:81570597; rs110709261) identified that parent of origin expression at this position reflected parent of origin expression across the entire chromosome for individual animals, rather than consistent expression of either the maternally or paternally inherited chromosome. Further examination of other genes able to be analyzed on the X chromosome did not reveal consistent patterns of either maternal or paternal expression that would indicate genomic imprinting.

With no evidence for consistent biallelic or imprinted expression on chromosome X, the next question was, what would random XCI look like in the lactating bovine mammary gland? If XCI were completely random in the lactating mammary gland, needle biopsy samples collected for these experiments should consist of a suitably large number of cells so that some would express the paternally inherited X chromosome and others would express the maternally inherited X chromosome. We would therefore expect that, on average, both X chromosomes would be expressed approximately equally. However, in contrast to the results observed at a population level, for an individual animal the difference in distribution of parent of origin X chromosome expression is very different to that seen on autosomes. Gene expression at the majority of SNP on autosomes is similar from both maternal and paternal chromosomes, whereas X chromosome expression varies from 90% maternal expression to 90% paternal expression. These results could be explained by the manner in which the mammary gland develops. Unlike most somatic tissues, the majority of mammary gland development occurs with the onset of puberty, with only rudimentary gland development occurring during fetal development beginning when the embryo is around 1 cm in length (approximately d 32; Hurley, 2009). With random XCI occurring early in embryonic development (around the blastocyst stage in mouse and human), it is possible that all epithelial cells in a mammary gland (or at least

a single quarter of the mammary gland) originate from a single cell that underwent XCI in the embryo.

If the entire mammary parenchyma, or at least an individual quarter, is derived from a single cell, it would be expected that biopsy samples would display either 100% maternal or 100% paternal expression. However, the distribution of paternal X chromosome expression in individuals, both at individual SNP and on average across the chromosome, is not consistent with the entire gland being derived from a single cell. Rather, distribution of paternal X chromosome expression across the 298 cows being analyzed appears trimodal, with peaks around 25, 60, and 80% paternal expression. While these RNA-seq data are derived from a mixed cell population (epithelial and stromal cells), this distribution of paternal chromosomal expression is suggestive of either (1) sampling of between 1 and 3 individual lobules, each of which developed from a single X chromosome inactivated cell, or (2) genetically skewed XCI.

Technical replicates were not available for the 298 samples collected using a needle biopsy. To determine if the skewed chromosome X inactivation was due to the small sample size collected, the 3 (1 cm³) technical replicates from the SRAcow were analyzed. The high degree of similarity in paternal expression measured in the replicates suggests that the skewed level of XCI, at least in this animal, was not simply due to biased sampling of 1 or 2 lobules. This does not, however, rule out the possibility that an entire quarter, or a large proportion of a quarter, may be derived from a single cell that underwent XCI.

A genetic basis for skewed XCI has been reported in both humans and mice (reviewed by Valencia and Wutz, 2015). In mice, a genetic locus called the X controlling element (**Xce**) has been shown to influence XCI choice (Cattanach 1970). Since its discovery more than 40 yr ago, the identity of the Xce has not been clearly defined; the most precise mapping is at a 176-kb segmental duplication located 500 kb proximal to *Xist* (Calaway et al., 2013). Using GWAS, no genetic basis for nonrandom inactivation of the X chromosome based on the current assembly of the bovine genome (UMD3.1) was observed in this population of New Zealand dairy cattle. However, given that GWAS approaches using biallelic SNP are unlikely to perform well when multiple alleles are present, this analysis does not rule out the possibility for a multiallelic segmental duplication (as observed in mice) influencing XCI. We further examined the possibility of a genetic basis for skewed chromosome X inactivation using data from multiple tissues in the SRAcow. If an Xce-like locus was skewing XCI in the SRAcow, skewing would be expected to be

consistent across all tissues. X chromosome inactivation appeared to be skewed in the mammary gland of the SRACow, with the average paternal expression fitting into the “low paternal expression” peak in the trimodal distribution of RNA-seq data from 298 cows (Figure 3). However, although some variation in paternal X chromosome expression was observed across other tissues (Figure 4), average paternal expression was around 50%. Although this does not rule out genetic control of XCI in other animals, the identification of high (or low levels) of paternal XCI is not necessarily caused by an Xce-like locus. Further analysis is required in the 298 cow RNA-seq data set to determine if structural variants similar to those reported in mice (Calaway et al., 2013) may be involved in influencing XCI in dairy cattle.

CONCLUSIONS

At a population level, maternally and paternally inherited copies of the X chromosome are expressed equally in the lactating mammary gland, which is consistent with random inactivation of the X chromosome. At an individual level, mammary gland biopsies suggest that in some animals the cells collected in the sample may have all originated from either 1 or 2 stem cells that had undergone X inactivation (either maternal or paternal). However, at this stage the possibility of a genetic basis for skewing of XCI cannot be ruled out. Unlike the mouse, no parent of origin effects were observed. Although some aspects of XCI appear to be conserved across species, the differences observed indicate that a good deal of caution should be applied when extrapolating XCI results from one species to another. Further work using sequence-based genotypes to enhance resolution is required to provide a deeper understanding of XCI. Such an understanding is required before models that accurately reflect X chromosome usage can be generated for GWAS studies and for X chromosome variation to be captured in a truly informative manner in genomic prediction. To be truly informative, these GWAS and genomic prediction models would capture not only the Mendelian transmission, but also the mono-allelic expression of this chromosome in females.

ACKNOWLEDGMENTS

We acknowledge New Zealand Genomics Limited (Auckland, New Zealand) and the University of Auckland Centre for Genomics, Proteomics, and Metabolomics (Auckland, New Zealand) for genomic DNA extraction, Australian Genomics Research Facility (Melbourne, Australia) and Illumina FastTrack (San

Diego, CA) for genomic DNA sequencing, and Vivienne Bennett (LIC, Hamilton, New Zealand) for critical reading of the manuscript. Funding for this project was provided by the Ministry for Primary Industries as a Primary Growth Partnership (Wellington, New Zealand).

REFERENCES

- Bach, I., C. Rodriguez-Esteban, C. Carrière, A. Bhushan, A. Krones, D. W. Rose, C. K. Glass, B. Andersen, J. C. Izpisua Belmonte, and M. G. Rosenfeld. 1999. RLIM inhibits functional activity of LIM homeodomain transcription factors via recruitment of the histone deacetylase complex. *Nat. Genet.* 22:394–399. <https://doi.org/10.1038/11970>.
- Belmont, J. W. 1996. Genetic control of X inactivation and processes leading to X-inactivation skewing. *Am. J. Hum. Genet.* 58:1101–1108.
- Calaway, J. D., A. B. Lenarcic, J. P. Didion, J. R. Wang, J. B. Searle, L. McMillan, W. Valdar, and F. Pardo-Manuel de Villena. 2013. Genetic architecture of skewed X inactivation in the laboratory mouse. *PLoS Genet.* 9:e1003853 <https://doi.org/10.1371/journal.pgen.1003853>.
- Cattanach, B. M. 1970. Controlling elements in the mouse X-chromosome. 3. Influence upon both parts of an X divided by rearrangement. *Genet. Res.* 16:293–301.
- Chamberlain, A. J., C. J. Vander Jagt, B. J. Hayes, M. Khansefid, L. C. Marett, C. A. Millen, T. T. Nguyen, and M. E. Goddard. 2015. Extensive variation between tissues in allele specific expression in an outbred mammal. *BMC Genomics* 16:993. <https://doi.org/10.1186/s12864-015-2174-0>.
- Chen, Z., D. E. Hagen, J. Wang, C. G. Elsik, T. Ji, L. G. Siqueira, P. J. Hansen, and R. M. Rivera. 2016. Global assessment of imprinted gene expression in the bovine conceptus by next generation sequencing. *Epigenetics* 11:501–516. <https://doi.org/10.1080/15592294.2016.1184805>.
- Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R. F. Brøndum, X. Liao, A. Djari, S. C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M. N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P. J. Bowman, D. Coote, A. J. Chamberlain, C. Anderson, C. P. VanTassell, I. Hulsegge, M. E. Goddard, B. Guldbrandtsen, M. S. Lund, R. F. Veerkamp, D. A. Boichard, R. Fries, and B. J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46:858–865. <https://doi.org/10.1038/ng.3034>.
- De La Fuente, R., A. Hahnel, P. K. Basur, and W. A. King. 1999. X inactive-specific transcript (*Xist*) expression and X chromosome inactivation in the preattachment bovine embryo. *Biol. Reprod.* 60:769–775.
- Degner, J. F., J. C. Marioni, A. A. Pai, J. K. Pickrell, E. Nkadori, Y. Gilad, and J. K. Pritchard. 2009. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* 25:3207–3212. <https://doi.org/10.1093/bioinformatics/btp579>.
- Deng, X., J. B. Berletch, D. K. Nguyen, and C. M. Disteche. 2014. X chromosome regulation: Diverse patterns in development, tissues and disease. *Nat. Rev. Genet.* 15:367–378. <https://doi.org/10.1038/nrg3687>.
- Dindot, S. V., K. C. Kent, B. Evers, N. Loskutoff, J. Womack, and J. A. Piedrahita. 2004. Conservation of genomic imprinting at the *XIST*, *IGF2*, and *GTL2* loci in the bovine. *Mamm. Genome* 15:966–974. <https://doi.org/10.1007/s00335-004-2407-z>.
- Fernando, R. L., and D. J. Garrick. 2010. GenSel—User manual for a portfolio of genomic selection related analyses. Animal Breeding and Genetics, Iowa State University, Ames.
- Gottipati, S., L. Arbiza, A. Siepel, A. G. Clark, and A. Keinan. 2011. Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. *Nat. Genet.* 43:741–743. <https://doi.org/10.1038/ng.877>.

- Hurley, W. L. 2009. "Mammary Gland Development Fetal" Mammary Development Resource Library. Accessed May 23, 2015. <http://ansci.illinois.edu/static/ansc438/Mamdevelop/fetal.html>.
- Jansen, S., B. Aigner, H. Pausch, M. Wysocki, S. Eck, A. Benet-Pagès, E. Graf, T. Wieland, T. M. Strom, T. Meitinger, and R. Fries. 2013. Assessment of the genomic variation in a cattle population by re-sequencing of key animals at low to medium coverage. *BMC Genomics* 14:446. <https://doi.org/10.1186/1471-2164-14-446>.
- Jiao, B., H. Ma, M. N. Shokhirev, A. Drung, Q. Yang, J. Shin, S. Lu, M. Byron, S. Kalantry, A. M. Mercurio, J. B. Lawrence, A. Hoffmann, and I. Bach. 2012. Paternal RLIM/Rnf12 is a survival factor for milk-producing alveolar cells. *Cell* 149:630–641. <https://doi.org/10.1016/j.cell.2012.02.056>.
- Jiao, B., N. Taniguchi-Ishigaki, C. Gungör, M. A. Peters, Y. W. Chen, S. Riethdorf, A. Drung, L. G. Ahronian, J. Shin, R. Pagnis, K. Pantel, T. Tachibana, B. C. Lewis, S. A. Johnsen, and I. Bach. 2013. Functional activity of RLIM/Rnf12 is regulated by phosphorylation-dependent nucleocytoplasmic shuttling. *Mol. Biol. Cell* 24:3085–3096. <https://doi.org/10.1091/mbc.E13-05-0239>.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Littlejohn, M. D., K. Tiplady, T. A. Fink, K. Lehnert, T. Lopdell, T. Johnson, C. Couldrey, M. Keehan, R. Sherlock, C. Harland, A. Scott, R. Snell, S. Davis, and R. J. Spelman. 2016. Sequence-based association analysis reveals an *MGST1* eQTL with pleiotropic effects on bovine milk composition. *Sci. Rep.* 6:25376.
- Littlejohn, M. D., K. Tiplady, T. Lopdell, T. A. Law, A. Scott, C. Harland, R. Sherlock, K. Henty, V. Obolonkin, K. Lehnert, A. Macgibbon, R. J. Spelman, S. R. Davis, and R. G. Snell. 2014. Expression variants of the lipogenic *AGPAT6* gene affect diverse milk composition phenotypes in *Bos taurus*. *PLoS One* 9:e85757. <https://doi.org/10.1371/journal.pone.0085757>.
- Niemann, H., X. C. Tian, W. A. King, and R. S. Lee. 2008. Epigenetic reprogramming in embryonic and foetal development upon somatic cell nuclear transfer cloning. *Reproduction* 135:151–163. <https://doi.org/10.1530/REP-07-0397>.
- Ostendorff, H. P., R. I. Peirano, M. A. Peters, A. Schlüter, M. Bossenz, M. Scheffner, and I. Bach. 2002. Ubiquitination-dependent cofactor exchange on LIM homeodomain transcription factors. *Nature* 416:99–103. <https://doi.org/10.1038/416099a>.
- Panousis, N. I., M. Gutierrez-Arcelus, E. T. Dermitzakis, and T. Lapalainen. 2014. Allelic mapping bias in RNA-sequencing is not a major confounder in eQTL studies. *Genome Biol.* 15:467. <https://doi.org/10.1186/s13059-014-0467-2>.
- Peippo, J., A. Farazmand, M. Kurkilahti, M. Markkula, P. K. Basrur, and W. A. King. 2002. Sex-chromosome linked gene expression in in-vitro produced bovine embryos. *Mol. Hum. Reprod.* 8:923–929.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575. <https://doi.org/10.1086/519795>.
- Shin, J., M. Bossenz, Y. Chung, H. Ma, M. Byron, N. Taniguchi-Ishigaki, X. Zhu, B. Jiao, L. L. Hall, M. R. Green, S. N. Jones, I. Hermans-Borgmeyer, J. B. Lawrence, and I. Bach. 2010. Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. *Nature* 467:977–981. <https://doi.org/10.1038/nature09457>.
- Thorvaldsen, J. L., C. Krapp, H. F. Willard, and M. S. Bartolomei. 2012. Nonrandom X chromosome inactivation is influenced by multiple regions on the murine X chromosome. *Genetics* 192:1095–1107. <https://doi.org/10.1534/genetics.112.144477>.
- Valencia, K., and A. Wutz. 2015. Recent insights into the regulation of X-chromosome inactivation. *Adv. Genomics Genet.* 5:227–238.
- Xue, F., X. C. Tian, F. Du, C. Kubota, M. Taneja, A. Dinnyes, Y. Dai, H. Levine, L. V. Pereira, and X. Yang. 2002. Aberrant patterns of X chromosome inactivation in bovine clones. *Nat. Genet.* 31:216–220. <https://doi.org/10.1038/ng900>.