Conserved microRNA targeting reveals preexisting gene dosage sensitivities that shaped amniote sex chromosome evolution

Sahin Naqvi[1, 2], Daniel W. Bellott[1], Kathy S. Lin[1, 3], & David C. Page[1, 2, 4]

[1] Whitehead Institute, Cambridge MA 02142;
[2] Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139
[3] Program in Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge MA 02139
[4] Howard Hughes Medical Institute, Whitehead Institute, Cambridge MA 02142

Mammalian X and Y chromosomes evolved from an ordinary autosomal pair. Genetic decay of the Y led to X chromosome inactivation (XCI) in females, but some Y-linked genes were retained during the course of sex chromosome evolution, and many X-linked genes did not become subject to XCI. We reconstructed gene-by-gene dosage sensitivities on the ancestral autosomes through phylogenetic analysis of microRNA (miRNA) target sites and compared these preexisting characteristics to the current status of Y-linked and X-linked genes in mammals. Preexisting heterogeneities in dosage sensitivity, manifesting as differences in the extent of miRNA-mediated repression, predicted either the retention of a Y homolog or the acquisition of XCI following Y gene decay. Analogous heterogeneities among avian Z-linked genes predicted the retention of a W homolog. Genome-wide analyses of human copy number variation indicate that these heterogeneities consisted of sensitivity to both increases and decreases in dosage. We propose a model of XY/ZW evolution incorporating such preexisting dosage sensitivities in determining the evolutionary fates of individual genes. Our findings thus provide a more complete view of the role of dosage sensitivity in shaping the mammalian and avian sex chromosomes, and reveal an important role for post-transcriptional regulatory sequences (miRNA target sites) in sex chromosome evolution.

INTRODUCTION

The mammalian X and Y chromosomes evolved from a pair of ordinary autosomes over the past 300 million years (Lahn & Page, 1999). Only 3% of genes on the ancestral pair of autosomes survive on the human Y chromosome (Bellott et al., 2010; Skaletsky et al., 2003), compared to 98% on the X chromosome (Mueller et al., 2013). In females, one copy of the X chromosome is

48 silenced by X inactivation (XCI); this silencing evolved on a gene-by-gene basis following Y
49 gene loss and compensatory X upregulation (Jegalian & Page, 1998; Ross et al., 2005), and some
50 genes escape XCI in humans (Carrel & Willard, 2005) and other mammals (Yang, Babak,
51 Shendure, & Disteche, 2010). In parallel, the avian Z and W sex chromosomes evolved from a
52 different pair of autosomes than the mammalian X and Y chromosomes (Bellott et al., 2010;
53 Nanda et al., 1999; Ross et al., 2005). Decay of the female-specific W chromosome was
54 similarly extensive, but birds did not evolve a large-scale inactivation of Z-linked genes
55 analogous to XCI in mammals (Itoh et al., 2007; Mank & Ellegren, 2009; Uebbing et al., 2015;
56 Wright, Zimmer, Harrison, & Mank, 2015). Thus, genes previously found on the ancestral
57 autosomes that gave rise to the mammalian or avian sex chromosomes have undergone
58 significant changes in gene dosage. In modern mammals, these molecular events have resulted in
59 three classes of ancestral X-linked genes representing distinct evolutionary fates: those with a
60 surviving Y homolog, those with no Y homolog and subject to XCI, and those with no Y
61 homolog but escaping XCI. In birds, two classes of ancestral Z-linked genes have arisen: those
62 with or without a W homolog. Identifying gene-by-gene properties that distinguish classes of X-
63 and Z-linked genes is thus crucial to understanding the selective pressures underlying the
64 molecular events of mammalian and avian sex chromosome evolution.

65 Emerging evidence suggests a role for gene dosage sensitivity in mammalian and avian
66 sex chromosome evolution. X- and Z-linked genes with surviving homologs on the mammalian
67 Y or avian W chromosomes are enriched for important regulatory functions and predictors of
68 haploinsufficiency compared to those lacking Y or W homologs (Bellott et al., 2014, 2017);
69 similar observations have been made in fish (White, Kitano, & Peichel, 2015) and Drosophila
70 (Kaiser, Zhou, & Bachtrog, 2011). Human X- and chicken Z-linked genes that show the

strongest signatures of dosage compensation in either lineage also show signs of dosage sensitivity as measured by membership in large protein complexes (Pessia, Makino, Bailly-Bechet, McLysaght, & Marais, 2012) or evolutionary patterns of gene duplication and retention (Zimmer, Harrison, Dessimoz, & Mank, 2016). Despite these advances, little is known regarding selective pressures resulting from sensitivity to dosage increases, as these studies either focused on haploinsufficiency or employed less direct predictors of dosage sensitivity. Furthermore, it is not known whether heterogeneities in dosage sensitivity among classes of sex-linked genes were acquired during sex chromosome evolution, or predated the emergence of sex chromosomes, as there has been no explicit, systematic reconstruction of dosage sensitivity on the ancestral autosomes that gave rise to the mammalian and avian sex chromosomes.

To assess the role of preexisting dosage sensitivities in XY and ZW evolution, we sought to employ a measure of dosage sensitivity that could be 1) demonstrably informative with respect to sensitivity to dosage increases, and 2) explicitly reconstructed on the ancestral autosomes. We focused on regulation by microRNAs (miRNAs), small noncoding RNAs that function as tuners of gene dosage by lowering target mRNA levels through pairing to the 3` untranslated region (UTR) (Bartel, 2009). The repressive nature of miRNA targeting is informative with respect to sensitivity to dosage increases, allowing for a more complete understanding of the role of dosage sensitivity in sex chromosome evolution. Both miRNAs themselves and their complementary target sites can be preserved over millions of years of vertebrate evolution, facilitating the reconstruction of miRNA targeting on the ancestral autosomes through cross-species sequence alignments. As miRNA targeting occurs post-transcriptionally, reconstruction of its ancestral state is decoupled from transcriptional regulatory mechanisms such as XCI that evolved following X-Y differentiation.

First, through genome-wide analysis of human copy number variation, we established conserved miRNA targeting as an indicator of sensitivity to dosage increases, which we used to infer heterogeneities in dosage sensitivity between the classes of present-day X- and Z-linked genes. We then reconstructed miRNA target sites on the ancestral autosomes, finding significant heterogeneity in the extent to which the autosomal precursors of X- or Z-linked genes were targeted by miRNAs. Finally, we reanalyzed publically available experimental datasets to validate that this conserved miRNA targeting is operative in living cells. Combined with results from previous studies, these findings provide the foundation for a model of X-Y and Z-W evolution in which both survival on the sex-specific Y or W chromosome and the subsequent evolution of XCI in mammals were determined by preexisting sensitivities to both decreases and increases in gene dosage. Our findings also demonstrate distinct selective pressures acting on post-transcriptional regulatory sequences that functioned to tune gene dosage both prior to and during amniote sex chromosome evolution.

RESULTS

Analysis of human copy number variation indicates conserved microRNA targeting of genes sensitive to dosage increases

We first sought to determine whether conserved targeting by microRNAs (miRNAs) correlates with sensitivity to dosage increases across the human genome. To estimate pressure to maintain miRNA targeting, we used published probabilities of conserved targeting (P_{CT} scores) for each gene-miRNA interaction in the human genome. The P_{CT} score reflects an estimate of the probability that a given gene-miRNA interaction is conserved due to miRNA targeting, obtained by calculating the conservation of the relevant miRNA target sites relative to the conservation of

the entire 3` UTR (Friedman, Farh, Burge, & Bartel, 2009). In this manner, the $P_{CT}$ score intrinsically controls for differences in 3` UTR background conservation and sequence composition due to differing rates of expression divergence and/or sequence evolution, making it comparable across both genes and miRNA families. We refer to these P_{CT} scores as "miRNA conservation scores" in the remainder of the text.

A recent study reported a correlation between these miRNA conservation scores and predicted haploinsufficiency (Pinzón et al., 2016), indicating that conserved miRNA targeting broadly corresponds to dosage sensitivity. However, such a correlation does not isolate the effects of sensitivity to dosage increases, which we expect to be particularly important in the context of miRNA targeting. We reasoned that genes for which increases in dosage are deleterious should be depleted from the set of observed gene duplications in healthy human individuals. We used a catalogue of rare genic copy number variation among 59,898 control human exomes (Exome Aggregation Consortium, ExAC)(Ruderfer et al., 2016) to classify autosomal protein-coding genes as exhibiting or lacking duplication or deletion in healthy individuals (see Methods). We compared duplicated and non-duplicated genes with the same deletion status in order to control for differences in sensitivity to underexpression. We found that non-duplicated genes have significantly higher miRNA conservation scores than duplicated genes, irrespective of deletion status (Figure 1A,B). Non-deleted genes also have significantly higher scores than deleted genes irrespective of duplication status (Supplemental Figure S1), but duplication status has a greater effect on miRNA conservation scores than does deletion status (blue vs. orange boxes, Figure 1C). Thus, conserved miRNA targeting is a feature of genes sensitive to changes in gene dosage in humans and is especially informative with regards to sensitivity to dosage increases.

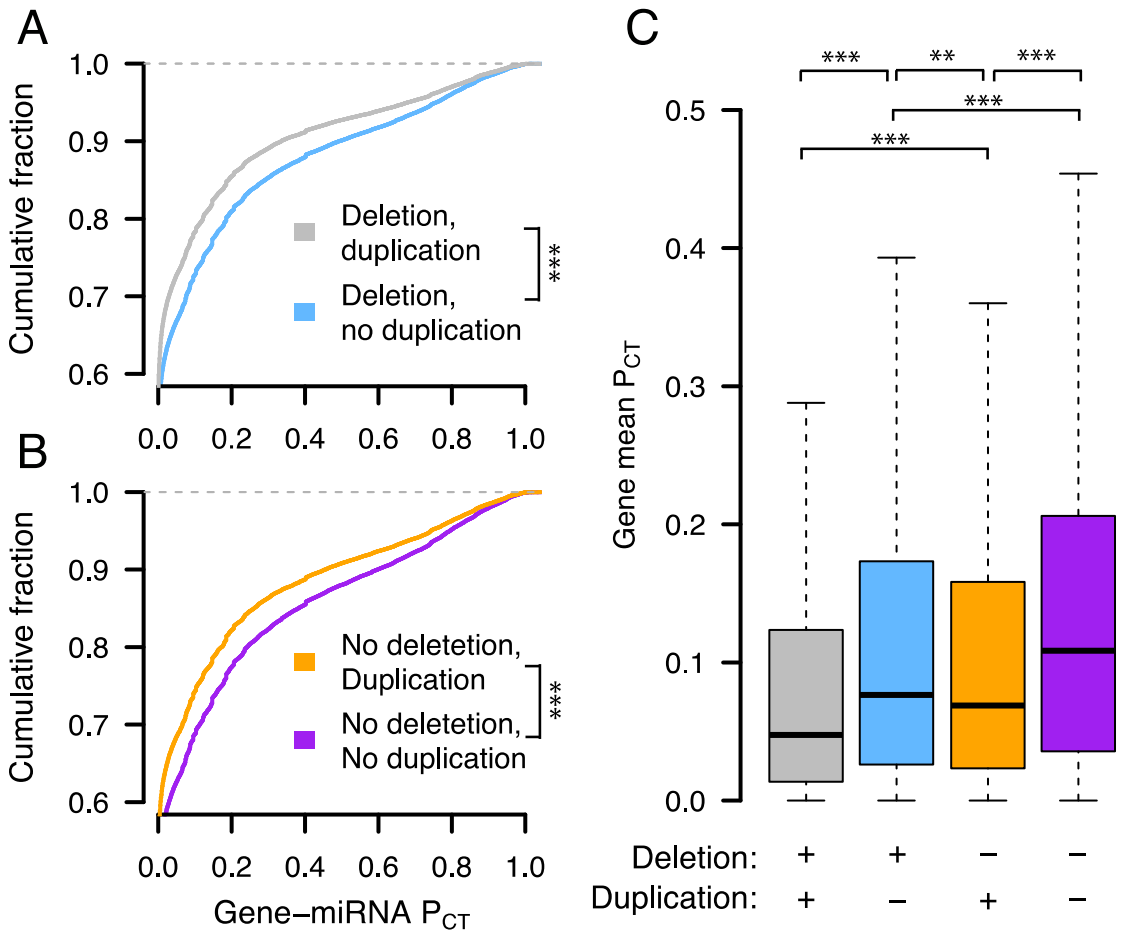


Figure 1: Conserved miRNA targeting of autosomal genes stratified by copy number variation in 59,898 human exomes. Probabilities of conserved targeting (P_{CT}) of all gene-miRNA interactions involving non-duplicated and duplicated genes, further stratified as (A) deleted (grey, n = 69,339 interactions from 4,118 genes; blue, n = 80,290 interactions from 3,976 genes) or (B) not deleted (orange, n = 51,514 interactions from 2,916 genes; purple, n = 72,826 interactions from 3,510 genes). *** $p < 0.001$, two-sided Kolmogorov-Smirnov test. (C) Mean gene-level P_{CT} scores. ** $p < 0.01$, *** $p < 0.001$, two-sided Wilcoxon rank-sum test.

X-Y pairs and X-inactivated genes have higher miRNA conservation scores than X escape genes

We next assessed whether the three classes of X-linked genes differ with respect to dosage sensitivity as inferred by conserved miRNA targeting. To delineate these classes, we began with the set of ancestral genes reconstructed through cross-species comparisons of the human X

chromosome and orthologous chicken autosomes (Bellott et al., 2014, 2017, 2010; Hughes et al., 2012; Mueller et al., 2013). We designated ancestral X-linked genes with a surviving human Y homolog (Skaletsky et al., 2003) as X-Y pairs and also considered the set of X-linked genes with a surviving Y homolog in any of eight mammals (Bellott et al., 2014) to increase the phylogenetic breadth of findings regarding X-Y pairs. A number of studies have catalogued the inactivation status of X-linked genes in various human tissues and cell-types. We used a meta-analysis that combined results from three studies by assigning a "consensus" X-inactivation status to each gene (Balaton, Cotton, & Brown, 2015) to designate the remainder of ancestral genes lacking a Y homolog as subject to or escaping XCI. In summary, we classified genes as either: 1) X-Y pairs, 2) lacking a Y homolog and subject to XCI (X-inactivated), or 3) lacking a Y homolog but escaping XCI (X escape).

We found that human X-Y pairs have the highest miRNA conservation scores, followed by X-inactivated and finally X escape genes (Figure 2A,B). The expanded set of X-Y pairs across eight mammals also has significantly higher miRNA conservation scores than ancestral X-linked genes with no Y homolog (Supplemental Figure S2). Observed differences between miRNA conservation scores are not driven by distinct subsets of genes in each class, as indicated by gene resampling with replacement (Supplemental Figure S3), and the decrease in miRNA conservation scores of X escape genes relative to X-inactivated genes and X-Y pairs is not driven by genes that escape XCI variably across individuals (Supplemental Figure S4). We also verified that these differences were not driven by artificially inflated or deflated conservation scores of certain target sites due to non-uniformity in 3` UTR conservation (Methods, Supplemental Figure S5).

Finally, we assessed whether miRNA conservation scores distinguish the three classes by providing additional information not accounted for by known factors (Bellott et al., 2014) influencing evolutionary outcomes. We used logistic regression to model, for each gene, the probability of falling into each of the three classes (X-Y pair, X-inactivated, or X escape) as a linear combination of haploinsufficiency probability (pHI) (Huang, Lee, Marcotte, & Hurles, 2010), human expression breadth (GTEx Consortium, 2015), and purifying selection, measured by the ratio of non-synonymous to synonymous substitution rates (dN/dS) between human and mouse orthologs (Yates et al., 2016). Adding miRNA conservation as an additional predictor resulted in a better model fit as measured by Aikake's information criterion (AIC), which quantifies goodness of fit while also penalizing increases in model complexity (known factors only, AIC 327.7; known factors plus miRNA conservation, AIC 321.5; lower AIC indicates better model). Based on our analyses of autosomal copy number variation (Figure 1), we attribute this improvement in model fit to the fact that miRNA conservation scores are most informative with regards to sensitivity to dosage increases. Thus, there is significant heterogeneity in dosage sensitivity, both upwards and downwards, among the three classes of ancestral X-linked genes: X-Y pairs are the most dosage-sensitive, while X-inactivated genes are of intermediate dosage sensitivity, and X escape genes are the least dosage-sensitive.

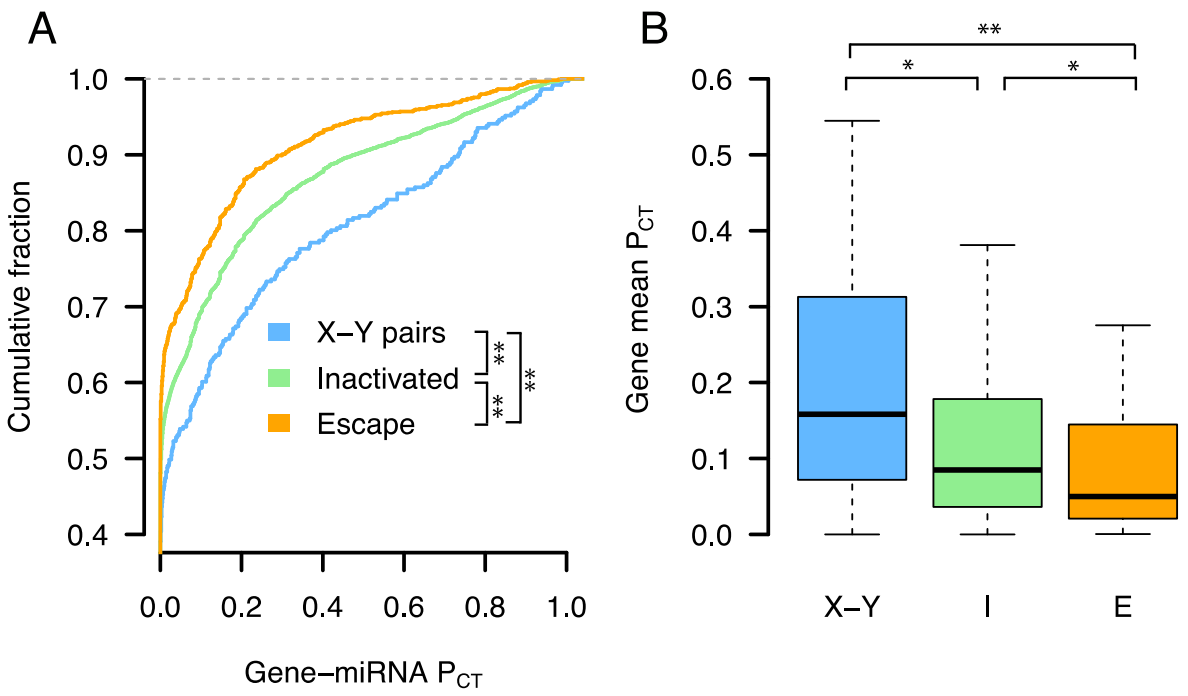


Figure 2. X-Y pairs and X-inactivated genes have higher miRNA conservation scores than X escape genes. P_{CT} score distributions of all gene-miRNA interactions involving (A) human X-Y pairs (n = 371 interactions from 15 genes), X-inactivated genes (n = 6,743 interactions from 329 genes), and X escape genes (n = 1,037 interactions from 56 genes). ** $p < 0.01$, two-sided Kolmogorov-Smirnov test. (B) Mean gene-level P_{CT} scores. * $p < 0.05$, ** $p < 0.01$, two-sided Wilcoxon rank-sum test.

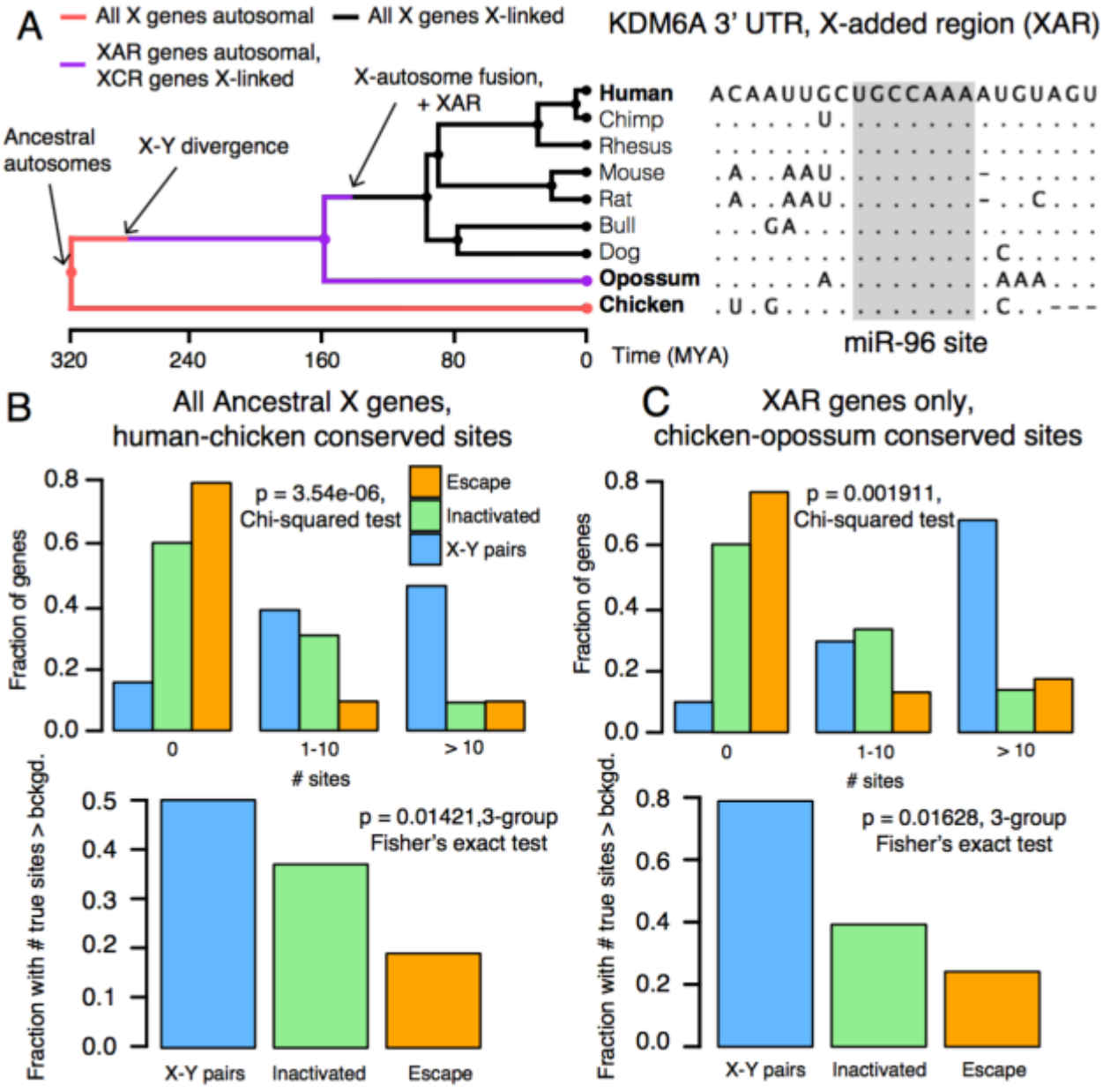
Heterogeneities in X-linked miRNA targeting were present on the ancestral autosomes

We next asked whether differences in miRNA targeting were present on the ancestral autosomes that gave rise to the mammalian X and Y chromosomes. To reconstruct the ancestral state of miRNA targeting, we first focused on miRNA target sites in the 3` UTR of human orthologs that align with perfect identity to a site in the corresponding chicken ortholog; these sites were likely present in the common ancestor of mammals and birds (Figure 3A,B). We found that X-Y pairs have the most human-chicken conserved target sites, followed by X-inactivated genes, and then X escape genes (Figure 3C, top). Unlike the miRNA conservation scores used earlier, this metric does not account for background conservation; we therefore estimated the background

conservation of each 3` UTR using shuffled miRNA family seed sequences (see Methods). X-Y pairs, X-inactivated genes, and X escape genes do differ significantly with respect to background conservation (data not shown), but these differences cannot account for the observed differences in true human-chicken conserved sites (Figure 3C, bottom). We observed similar results for the expanded set of X-Y pairs across 8 mammals (Supplemental Figure S6A).

Differences in the number of human-chicken conserved sites among the three classes of X-linked genes could be explained by heterogeneity in miRNA targeting present on the ancestral autosomes, or by ancestral homogeneity followed by different rates of target site loss during or following X-Y differentiation. To distinguish between these two possibilities, we took advantage of previous reconstructions of human sex chromosome evolution (Figure 3A) (Bellott et al., 2014), which confirmed that, following the divergence of placental mammals from marsupials, an X-autosome chromosomal fusion generated the X-added region (XAR) (Watson, Spencer, Riggs, & Graves, 1990). Genes on the XAR are therefore X-linked in placental mammals, but autosomal in marsupials such as the opossum. We limited our analysis to genes in the XAR and target sites conserved between orthologous chicken and opossum 3` UTRs, ignoring site conservation in humans; these sites were likely present in the common ancestor of mammals and birds, and an absence of such sites cannot be explained by site loss following X-Y differentiation. We observed the same pattern as with the human-chicken conserved sites, both before and after accounting for background 3` UTR conservation (Figure 3D, three gene classes; Supplemental Figure S6B, X-Y pairs across 8 mammals). These results demonstrate that the autosomal precursors of X-Y pairs and X-inactivated genes were subject to more miRNA-mediated regulation than X escape genes. Combined with our earlier results, we conclude that present-day

237 heterogeneities in dosage sensitivity on the mammalian X chromosome existed on the ancestral
238 autosomes from which it derived.
239



240
241 **Figure 3. Heterogeneities in X-linked miRNA targeting were present on the ancestral**
242 **autosomes.** (A) Example reconstruction of an ancestral miR-96 target site in the 3` UTR of
243 KDM6A, an X-linked gene in the X-added region (XAR) with a surviving Y homolog. Dots in
244 non-human species indicate identity with the human sequence, dashes gaps indicate gaps in the
245 multiple sequence alignment. (B) Distributions of sites conserved between 3` UTRs of human
246 and chicken orthologs (top) or comparisons to background expectation (bottom, see Methods) for
247 human X-Y pairs (n = 16), X-inactivated genes (n = 251), and X escape genes (n = 42). (C)

Statistics as in (B), but using sites conserved between chicken and opossum 3` UTRs only for genes in the XAR; X-Y pairs (n = 11), X-inactivated genes (n = 58), and X escape genes (n = 27).

Z-W pairs have higher miRNA conservation scores than other ancestral Z-linked genes

We next assessed whether classes of avian Z-linked genes, those with and without a W homolog, show analogous heterogeneities in sensitivity to dosage increases. We used the set of ancestral genes reconstructed through cross-species comparisons of the avian Z chromosome and orthologous human autosomes and focused on the set of Z-W pairs identified by sequencing of the chicken W chromosome (Bellott et al., 2017, 2010). To increase the phylogenetic breadth of our comparisons, we also included candidate Z-W pairs obtained through comparisons of male and female genome assemblies (4 species set) or inferred by read-depth changes in female genome assemblies (14 species set, see Methods for details) (Zhou et al., 2014). The more complete 3` UTR annotations in the human genome relative to chicken allow for a more accurate assessment of conserved miRNA targeting. Accordingly, we analyzed the 3` UTRs of the human orthologs of chicken Z-linked genes, reasoning that important properties of chicken Z-linked genes should be retained by their human autosomal orthologs.

We found that the human orthologs of Z-W pairs have higher miRNA conservation scores than the human orthologs of other ancestral Z genes (Figure 4A, B). Differences in miRNA conservation scores between Z-W pairs and other ancestral Z genes remained significant when considering the expanded sets of Z-W pairs across four and 14 avian species (Supplemental Figure S7). These differences are not driven by distinct subsets of genes, as indicated by gene resampling with replacement (Supplemental Figure S8), and cannot be accounted for by within-UTR variation in regional conservation (Supplemental Figure S9). Logistic regression models indicate that miRNA conservation scores provide additional

information not captured by known factors (Bellott et al., 2017) influencing survival of W-linked genes (known factors only model, AIC 137.8; known factors plus miRNA conservation model, AIC 127.1; lower AIC indicates better model). Together, these results indicate that the autosomal orthologs of Z-linked genes with a surviving W homolog are more sensitive to changes in dosage -- both increases and decreases -- than are genes without a surviving W homolog.

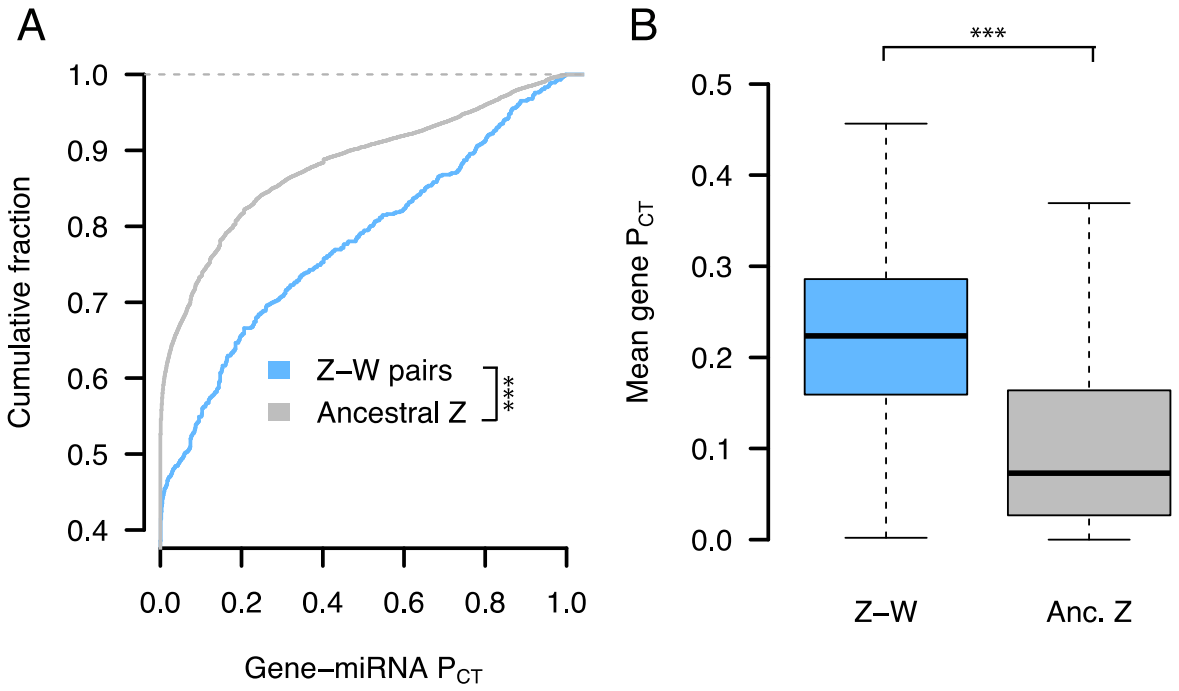


Figure 4. Z-W pairs have higher miRNA conservation scores than other ancestral Z-linked genes. P_{CT} score distributions of all gene-miRNA interactions involving the human orthologs of (A) chicken Z-W pairs (n = 832 interactions from 28 genes) and other ancestral Z genes (n = 16,692 interactions from 657 genes). *** $p < 0.001$, two-sided Kolmogorov-Smirnov test. (B) Mean gene-level P_{CT} scores. *** $p < 0.001$, two-sided Wilcoxon rank-sum test.

Heterogeneities in Z-linked miRNA targeting were present on the ancestral autosomes

We next asked whether differences in miRNA targeting between Z-W pairs and other ancestral Z-linked genes were present on the ancestral autosomes that gave rise to the avian Z and W chromosomes. We found that chicken Z-W pairs have more human-chicken-conserved miRNA

target sites than their Z-linked counterparts without surviving W homologs, both before (Figure 5C, top) and after (Figure 5C, bottom) accounting for the background conservation of each individual 3` UTR. To confirm that these differences represent ancestral heterogeneity rather than differential site loss during or following Z-W differentiation, we instead considered the number of sites conserved between human and anolis lizard, which diverged from birds prior to Z-W differentiation (Figure 5A). Chicken Z-W pairs contain an excess of human-anolis conserved miRNA target sites, both before (Figure 5D, top) and after (Figure 5D, bottom) accounting for the background conservation of each individual 3` UTR. We observed similar results with the predicted four-species (Supplemental Figure S10) and 14-species (Supplemental Figure S11) sets of Z-W pairs. Thus, the autosomal precursors of avian Z-W pairs were subject to more miRNA-mediated regulation than the autosomal precursors of Z-linked genes that lack a W homolog, indicating heterogeneity in dosage sensitivity among genes on the ancestral autosomes that gave rise to the avian Z chromosome.

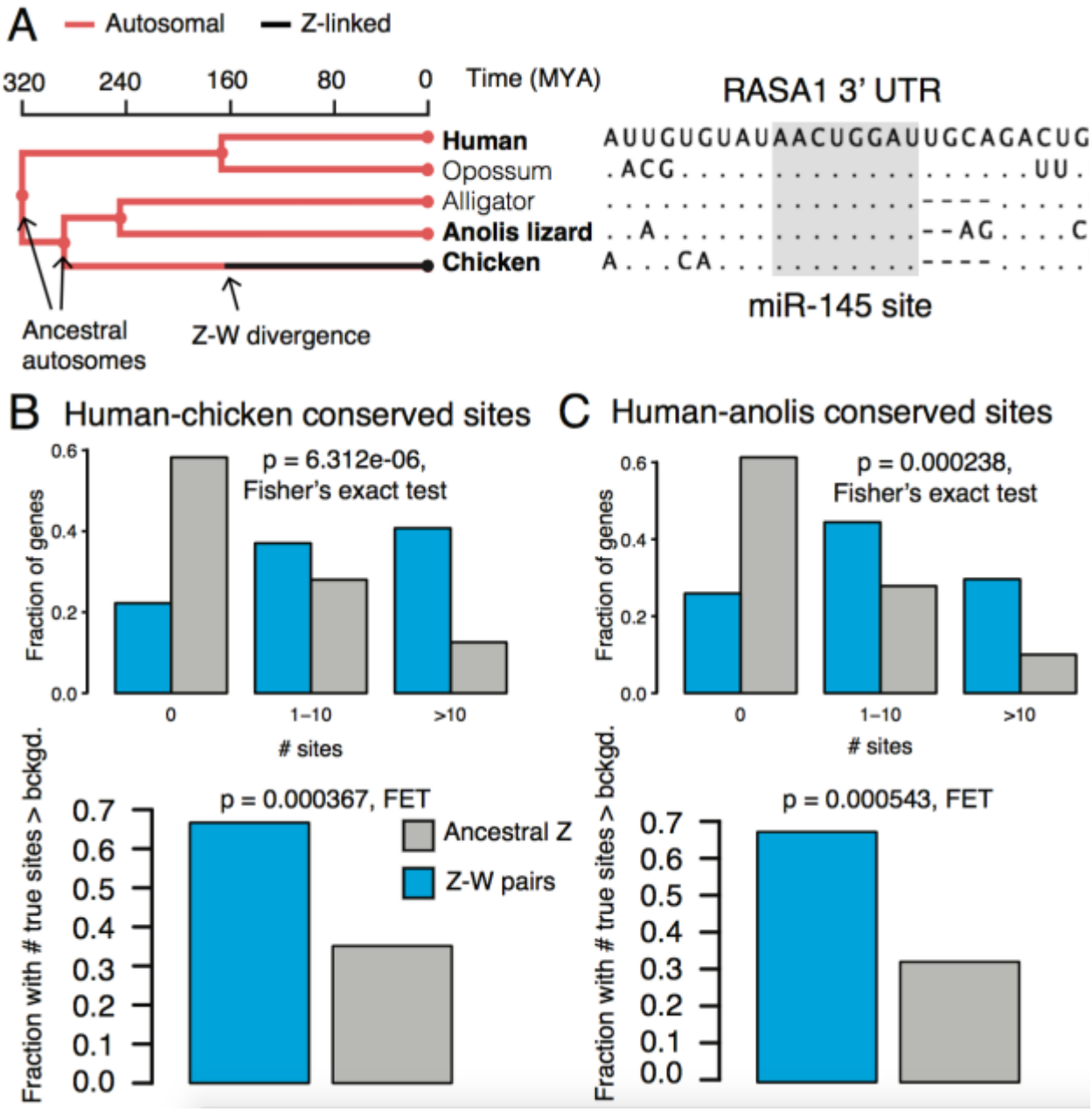


Figure 5. Heterogeneities in Z-linked miRNA targeting were present on the ancestral autosomes. (A) Example reconstruction of an ancestral miR-145 target site in the 3` UTR of RASA1, a Z-linked gene with a surviving W homolog. Example of 3` UTR sequence alignment for RASA1, a Z-linked gene with a surviving W homolog, with a target site for miR-145 highlighted in gray. (B) Numbers of sites conserved between 3` UTRs of human and chicken orthologs (top) or comparisons to background expectation (bottom) for chicken Z-W pairs (n = 27) and other ancestral Z genes (n = 578). (C) Statistics as in (B), but using sites conserved between human and anolis 3` UTRs.

Analyses of experimental datasets validate miRNA target site function

Our results to this point, which indicate preexisting heterogeneities in dosage constraints among X- or Z-linked genes as inferred by predicted miRNA target sites, lead to predictions regarding the function of these sites in vivo. To test these predictions, we turned to publically available experimental datasets consisting both of gene expression profiling following transfection or knockout of individual miRNAs, and of high-throughput crosslinking-immunoprecipitation (CLIP) to identify sites that bind Argonaute in vivo (see Methods). If the above-studied sites are effective in mediating target repression, targets of an individual miRNA should show increased expression levels or Argonaute binding following miRNA transfection, and decreased expression levels following miRNA knockout. Together, our analyses of publically available datasets fulfilled these predictions, validating the function of sites targeting ancestral X-linked genes and the autosomal orthologs of ancestral Z-linked genes in multiple cellular contexts and species (Figure 6). From the gene expression profiling data, we observed results consistent with effective targeting by a) eleven different miRNA families in human HeLa cells (Supplemental Figure S12), b) four different miRNAs in human HCT116 and HEK293 cells (Supplemental Figure S13), and c) miR-155 in mouse B and Th1 cells (Supplemental Figure S14). In the CLIP data, the human orthologs of X- or Z-linked targets of miR-124 are enriched for Argonaute-bound clusters that appear following miR-124 transfection, while a similar but non-significant enrichment is observed for miR-7 (Supplemental Figure S15). Thus, conserved miRNA target sites used to infer dosage constraints on X-linked genes and the autosomal orthologs of Z-linked genes can effectively mediate target repression in living cells.

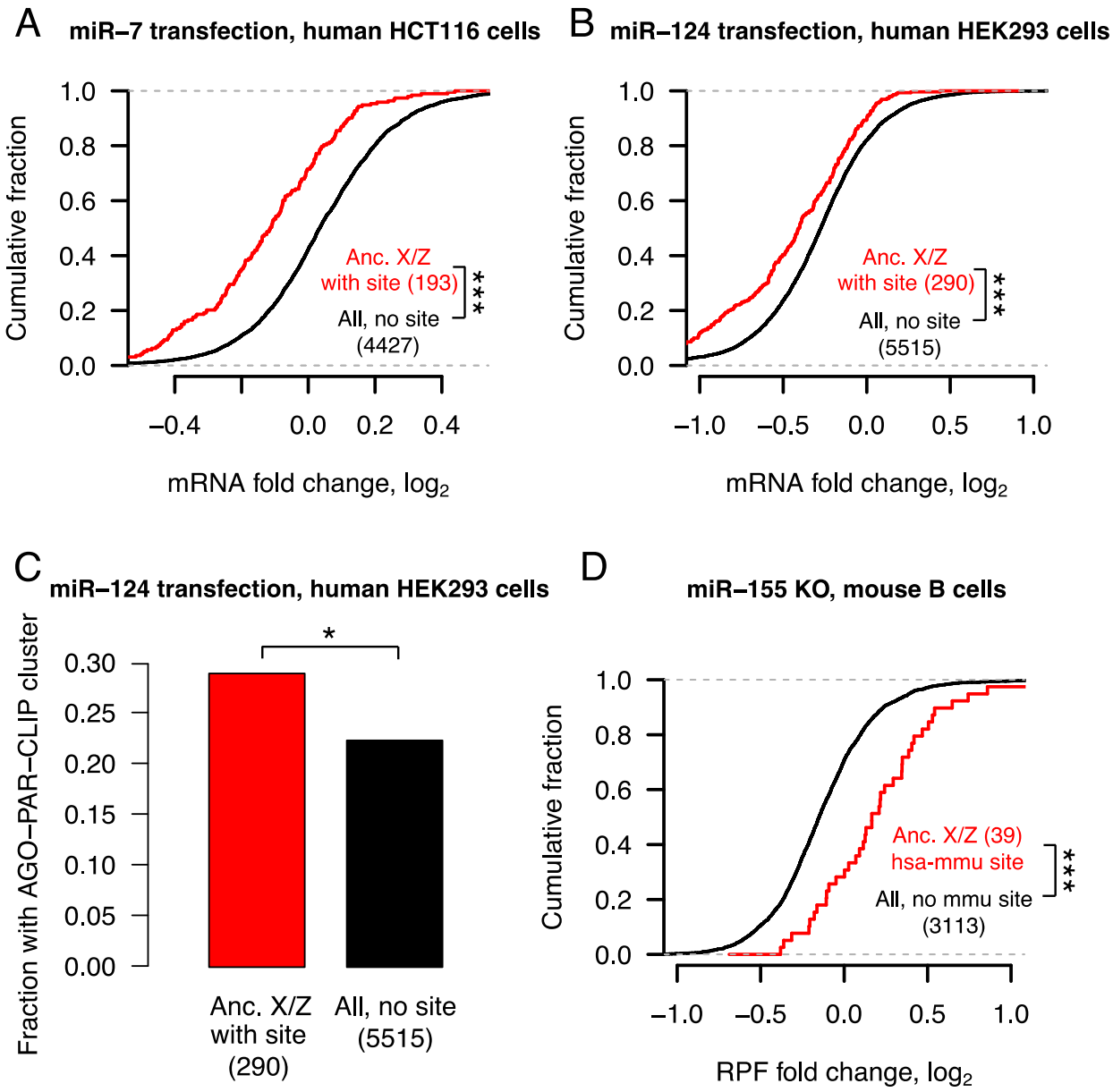


Figure 6. Analyses of experimental datasets validate miRNA target site function. Responses to transfection (A,B,C) or knockout (D) of indicated miRNAs in human (A,B,C) or mouse (D) cell-types. Each panel depicts corresponding changes in mRNA levels (A,B), in fraction of Argonaute-bound genes (C), and in mRNA stability and translational efficiency as measured by ribosome protected fragments (RPF, D). In each case, X-linked genes and the human orthologs of Z-linked genes containing target sites with an assigned P_{CT} score (red) for the indicated miRNA were compared to all expressed genes lacking target sites (black); gene numbers are indicated in parentheses. (A,B,D) *** $p < 0.001$, two-sided Kolmogorov-Smirnov test. (C) * $p < 0.05$, two-sided Fisher's exact test.

DISCUSSION

Here, through the evolutionary reconstruction of microRNA (miRNA) target sites, we demonstrate preexisting heterogeneities in dosage sensitivity among genes on the mammalian X and avian Z chromosomes. We first showed that, across all human autosomal genes, dosage sensitivity -- as indicated by patterns of genic copy number variation -- correlates with the degree of conserved miRNA targeting. We found that conserved targeting correlates especially strongly with sensitivity to dosage increases, consistent with miRNA targeting serving to reduce gene expression. Turning to the sex chromosomes of mammals and birds, genes that retained a homolog on the sex-specific Y or W chromosome (X-Y and Z-W pairs) are more dosage-sensitive than genes with no Y or W homolog. In mammals, genes with no Y homolog that became subject to XCI are more dosage-sensitive than those that continued to escape XCI following Y gene decay. We then reconstructed the ancestral state of miRNA targeting, observing significant heterogeneities in the extent of miRNA targeting, and thus dosage sensitivity, on the ancestral autosomes that gave rise to the mammalian and avian sex chromosomes. Finally, through analysis of publically available experimental datasets, we validated the function, in living cells, of the miRNA target sites used to infer dosage sensitivity. We thus conclude that differences in dosage sensitivity – both to increases and to decreases in gene dosage -- among genes on the ancestral autosomes influenced their evolutionary trajectory during sex chromosome evolution, not only on the sex-specific Y and W chromosomes, but also on the sex-shared X chromosome.

Our findings build upon previous work in three important ways. First, our analysis of miRNA-mediated repression indicates that these heterogeneities consist of sensitivities to dosage increases and decreases, whereas previous studies had either focused on sensitivity to

underexpression or could not differentiate the two. Second, our reconstruction of miRNA targeting on the ancestral autosomes provides direct evidence that heterogeneities in dosage sensitivity among classes of X- and Z-linked were preexisting rather than acquired during sex chromosome evolution. Finally, by pointing to specific regulatory sequences (miRNA target sites) functioning to tune gene dosage both prior to and during sex chromosome evolution, our study provides a view of dosage compensation encompassing post-transcriptional regulation.

Human disease studies support the claim that increased dosage of X-Y pairs and X-inactivated genes is deleterious to fitness. Copy number gains of the X-linked gene *KDM6A*, which has a surviving human Y homolog, are found in patients with developmental abnormalities and intellectual disability (Lindgren et al., 2013). *HDAC6, CACNA1F, GDI1,* and *IRS4* all lack Y homologs and are subject to XCI in humans. A mutation in the 3` UTR of *HDAC6* abolishing targeting by miR-433 has been linked to familial chondrodysplasia in both sexes (Simon et al., 2010). Likely gain-of-function mutations in *CACNA1F* cause congenital stationary night blindness in both sexes (Hemara-Wahanui et al., 2005). Copy number changes of *GDI1* correlate with the severity of X-linked mental retardation in males, with female carriers preferentially inactivating the mutant allele (Vandewalle et al., 2009). Somatic genomic deletions downstream of *IRS4* lead to its overexpression in lung squamous carcinoma (Weischenfeldt et al., 2017). Males with partial X disomy due to translocation of the distal long arm of the X chromosome (Xq28) to the long arm of the Y chromosome show severe mental retardation and developmental defects (Lahn et al., 1994). Most genes in Xq28 are inactivated in 46,XX females but escape inactivation in such X;Y translocations, suggesting that increased dosage of Xq28 genes caused the cognitive and developmental defects. We anticipate that further studies will

reveal additional examples of the deleterious effects of increases in gene dosage of X-Y pairs and X-inactivated genes.

We and others previously proposed that Y gene decay drove upregulation of homologous X-linked genes in both males and females, and that XCI subsequently evolved at genes sensitive to increased expression from two active X-linked copies in females (Jegalian & Page, 1998; Ohno, 1967). Our finding that X-inactivated genes have higher miRNA conservation scores than X escape genes is consistent with this aspect of the model. However, recent studies indicating heterogeneity in dosage sensitivity between classes of mammalian X- or avian Z-linked genes (Bellott et al., 2014, 2017; Pessia et al., 2012; Zimmer et al., 2016), combined with the present finding that these dosage sensitivities existed on the ancestral autosomes, challenge the previous assumption of a single evolutionary pathway for all sex-linked genes.

We therefore propose a revised model of X-Y and Z-W evolution in which the ancestral autosomes that gave rise to the mammalian and avian sex chromosomes contained three (or two, in the case of birds) classes of genes with differing dosage sensitivities (Figure 7A,B). For ancestral genes with high dosage sensitivity, Y or W gene decay would have been highly deleterious, and thus the Y- or W-linked genes were retained. According to our model, these genes' high dosage sensitivity also precluded upregulation of the X- or Z-linked homolog, and, in mammals, subsequent X-inactivation; indeed, their X-linked homologs continue to escape XCI (Bellott et al., 2014). For ancestral mammalian genes of intermediate dosage sensitivity, Y gene decay did occur, and was accompanied or followed by compensatory upregulation of the X-linked homolog in both sexes; the resultant increased expression in females was deleterious and led to the acquisition of XCI. Ancestral mammalian genes of low dosage sensitivity continued to escape XCI following Y decay, and perhaps did not undergo compensatory X upregulation

(Figure 6A). These genes' dosage insensitivity set them apart biologically, and evolutionarily, from the other class of X-linked genes escaping XCI -- those with a surviving Y homolog.

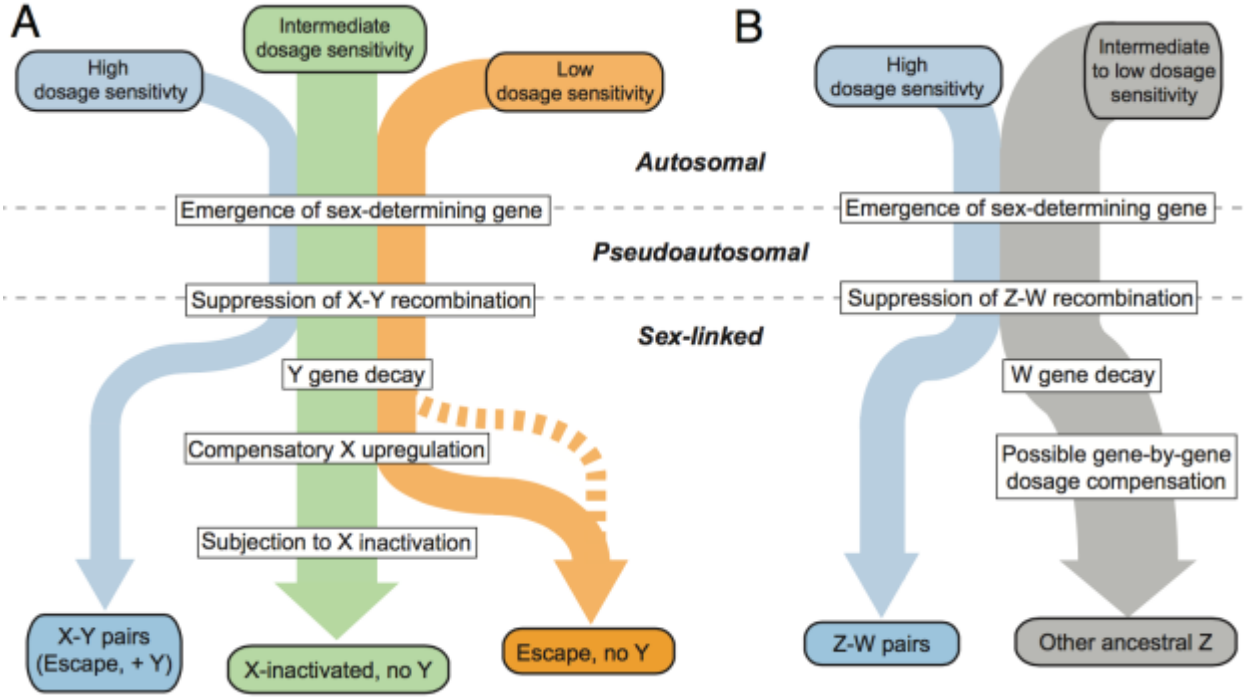


Figure 7. An evidence-based model of mammalian and avian sex chromosome evolution from ancestral autosomes. In this model, preexisting heterogeneities in dosage sensitivity determined the trajectory of Y/W gene loss in both mammals and birds, and of subsequent X-inactivation in mammals. Colored arrow widths are scaled approximately to the number of ancestral genes in each class. The dashed orange line represents the possibility that a subset of X-linked genes may have not undergone compensatory X upregulation following Y gene decay. Small black arrows adjacent to molecular evolutionary transitions indicate increases or decreases in sex-linked gene dosage

Our revised model relates preexisting, gene-by-gene heterogeneities in dosage sensitivity to the outcomes of sex chromosome evolution. However, the suppression of X-Y recombination did not occur on a gene-by-gene basis, instead initiating Y gene decay and subsequent dosage compensation through a series of large-scale inversions encompassing many genes (Lahn & Page, 1999). The timings and boundaries of these evolutionary strata varied among mammalian lineages, thus leading to unique chromosome-scale evolutionary dynamics across mammals.

These large-scale changes would have then allowed for genic selection to take place according to the preexisting dosage sensitivities outlined above. In this way, the course of sex chromosome evolution in mammals is a composite of 1) preexisting, gene-by-gene dosage sensitivities and 2) the manner in which the history of the X and Y unfolded in particular lineages via discrete, large-scale inversions.

In this study, we have focused on classes of ancestral X-linked genes delineated by the survival of a human Y homolog or by the acquisition of XCI in humans, but such evolutionary states can differ among mammalian lineages and species. In mouse, for instance, both Y gene decay (Bellott et al., 2014) and the acquisition of X-inactivation (Yang et al., 2010) are more complete than in humans or other mammals. These observations could be explained by shortened generation times in the rodent lineage, resulting in longer evolutionary times, during which the forces leading to Y gene decay and the acquisition of X-inactivation could act (Charlesworth & Crow, 1978; Jegalian & Page, 1998; Ohno, 1967). In the future, more complete catalogues of X-inactivation and escape in additional mammalian lineages would make it possible to examine whether analogous, preexisting dosage sensitivities differentiate the three classes of X-linked genes (X-Y pairs, X-inactivated genes, and X escape genes) in other species.

Previous studies have sought evidence of X-linked upregulation during mammalian sex chromosome evolution using comparisons of gene expression levels between the whole of the X chromosome and all of the autosomes, with equal numbers of studies rejecting or finding evidence consistent with upregulation (Deng et al., 2011; Julien et al., 2012; Kharchenko, Xi, & Park, 2011; Lin, Xing, Zhang, & He, 2012; Xiong et al., 2010). This is likely due to gene-by-gene heterogeneity in dosage sensitivities that resulted in a stronger signature of upregulation at more dosage sensitive genes (Pessia et al., 2012). Similarly, studies of Z-linked gene expression

463 in birds provide evidence for the gene-by-gene nature of Z dosage compensation, as measured by
464 comparisons of gene expression levels between ZZ males and ZW females (Itoh et al., 2007;
465 Mank & Ellegren, 2009; Uebbing et al., 2015), and indicate a stronger signature of dosage
466 compensation at predicted dosage-sensitive genes (Zimmer et al., 2016). By showing that such
467 dosage sensitivities existed on the ancestral autosomes and consist of sensitivity to both increases
468 and decreases, our findings highlight an additional aspect of dosage compensation that affects
469 both birds and mammals.

470 In addition to revealing similarities between mammals and birds, our study provides a
471 view of dosage compensation that highlights post-transcriptional regulatory mechanisms. The
472 majority of previous studies, by relying upon measurements of mRNA abundance, could not
473 differentiate transcriptional and post-transcriptional effects (Deng et al., 2011; Itoh et al., 2007;
474 Julien et al., 2012; Kharchenko et al., 2011; Lin et al., 2012; Xiong et al., 2010; Zimmer et al.,
475 2016). Other studies have drawn inferences regarding post-transcriptional dosage compensation
476 from measurements of extant protein abundance (Chen & Zhang, 2015; Pessia et al., 2012;
477 Uebbing et al., 2015). Our study points to specific non-coding sequences with known
478 mechanisms (microRNA target sites) functioning across evolutionary time. Furthermore, our
479 finding of greater conserved miRNA targeting of X-inactivated genes relative to X escape genes
480 shows that it is possible to predict the acquisition of a transcriptional regulatory state (XCI)
481 during sex chromosome evolution on the basis of a preexisting, post-transcriptional regulatory
482 state. Perhaps additional post-transcriptional regulatory mechanisms and their associated
483 regulatory elements will be shown to play roles in mammalian and avian dosage compensation.

484 Recent work has revealed that the sex-specific chromosome -- the Y in mammals and the
485 W in birds -- convergently retained dosage-sensitive genes with important regulatory functions

(Bellott et al., 2014, 2017). Our study, by reconstructing the ancestral state of post-transcriptional regulation, provides direct evidence that such heterogeneity in dosage sensitivity existed on the ancestral autosomes that gave rise to the mammalian and avian sex chromosomes. This heterogeneity influenced both survival on the sex-specific chromosomes in mammals and birds and the evolution of XCI in mammals. Thus, two independent experiments of nature offer empirical evidence that modern-day amniote sex chromosomes were shaped, during evolution, by the properties of the ancestral autosomes from which they derive.

METHODS

Statistics

Details of all statistical tests (type of test, test statistic, and p-value) used in this manuscript are provided in Supplemental Table S1.

Human genic copy number variation

To annotate gene deletions and duplications, we used data from the Exome Aggregation Consortium (ExAC, RRID:SCR_004068) (ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3.1/cnv/), which consists of autosomal genic duplications and deletions (both full and partial) called in 59,898 exomes (Ruderfer et al., 2016). We used the publicly available genic deletion counts but re-computed genic duplication counts using only full duplications, reasoning that partial duplications are unlikely to result in increased dosage of the full gene product. We thus required that an individual duplication fully overlapped the longest protein-coding transcript (GENCODE v19) of a gene using BEDtools (RRID:SCR_006646) (Quinlan & Hall, 2010). We removed genes flagged by ExAC as lying in known regions of recurrent CNVs. This yielded 4,118 genes within duplications and deletions, 3,976 genes within deletions but not duplications, 2,916 genes within duplications but not deletions, and 3,510 genes not subject to duplication or deletion. These gene assignments are provided in Supplemental Table S2.

X-linked gene sets

Analyses of conserved miRNA targeting based on multiple species alignments are unreliable for multicopy or ampliconic genes due to ambiguous sequence alignment between species. To avoid

such issues, we first removed multicopy and ampliconic genes (Mueller et al., 2013) from a previously published set of human X genes present in the amniote ancestor (Bellott et al., 2014). We then excluded genes in the human pseudoautosomal (PAR) regions since these genes have not been exposed to the same evolutionary forces as genes in regions where X-Y recombination has been suppressed. Of the remaining ancestral X genes, we classified the 15 genes with human Y-linked homologs as X-Y pairs. We also analyzed the larger set of 32 X-Y pairs across eight mammals (human, chimpanzee, rhesus macaque, marmoset, mouse, rat, bull, and opossum) with sequenced Y chromosomes (Bellott et al., 2014).

To classify ancestral X-linked genes without Y homologs as subject to or escaping XCI in humans, we used a collection of consensus XCI calls which aggregate the results of three studies (Carrel & Willard, 2005; A. M. Cotton et al., 2013; a. M. Cotton et al., 2014) assaying XCI escape (Balaton et al., 2015). Out of 472 ancestral X genes without a human Y homolog assigned an XCI status by Balaton et al. (Balaton et al., 2015), 329 were subject to XCI ("Subject" or "Mostly subject" in Balaton et al.), 26 displayed variable escape ("Variable escape" or "Mostly variable escape") from XCI, and 30 showed consistent escape ("Escape" or "Mostly escape"). We excluded 40 ancestral X genes with a "Discordant" XCI status as assigned by Balaton et al. In the main text, we present results obtained after combining both variable and consistent escape calls from Balaton et al. into one class, yielding the following counts: 15 X-Y pairs, 329 ancestral X genes subject to XCI, and 56 ancestral X genes with evidence of escape from XCI. We also performed analyses considering escape and variable escape genes separately. Information on X-linked genes is provided in Supplemental Table S3.

539 **Z-linked gene sets**

540 We previously refined the ancestral gene content of the avian sex chromosomes to 685 Z-linked
541 genes with human orthologs by sequencing of the chicken Z chromosome and analysis of 13
542 other avian species with published female genomes (Bellott et al., 2017). Of these 685 ancestral
543 Z genes, 28 retained a homolog on the fully sequenced chicken W chromosome. Including three
544 additional avian species in which candidate W-linked genes were ascertained by directly
545 comparing male and female genome assemblies results in a total of 78 W-linked genes. Including
546 another 10 avian species in which W-linkage was inferred by read depth changes in a female
547 genome results in a total of 157 W-linked genes. Information on Z-linked genes is provided in
548 Supplemental Table S4.

549

550 **microRNA target site P_{CT} scores**

551 Pre-calculated P_{CT} scores for all gene-miRNA family interactions were obtained from
552 TargetScanHuman v7.1 (RRID:SCR_010845)
553 (http://www.targetscan.org/vert_71/vert_71_data_download/Summary_Counts.all_predictions.tx
554 t.zip), (Friedman et al., 2009). We excluded mammalian-specific miRNA families based on
555 classifications by Friedman et al (Friedman et al., 2009) and updated in TargetScanHuman
556 v7.1(Agarwal, Bell, Nam, & Bartel, 2015). To account for gene-specific variability in the
557 number and P_{CT} score of gene-miRNA interactions within a group of genes, we sampled 1000x
558 with replacement from the same group of genes and computed the mean gene-miRNA P_{CT} score
559 for all associated gene miRNA interactions from each sampling. These 1000 samplings were
560 then used to estimate the median resampled gene-miRNA P_{CT} and 95% confidence intervals.

561

562 **Variation in within- conservation bias**

563 The P_{CT} of a given miRNA target site depends on the conservation of the site, as measured by the
564 total branch length of the phylogenetic tree containing the target sites (branch length score, BLS)
565 relative to the mean BLS of the whole 3` UTR. To address the possibility that non-uniformity in
566 the regional BLS could artificially inflate or deflate conservation scores of certain target sites, we
567 implemented a step-detection algorithm to segment 3` UTRs into regions of homogeneous BLS
568 values and calculated miRNA site conservation relative to these smaller regions. In order to call
569 steps within a 3` UTR, we computed the t-test p-value between the BLS values of the 50-nt
570 window upstream and downstream of each nucleotide position in the 3` UTR. Transitions were
571 called at a log p-value cutoff of -15. Because of noise in the BLS signal, the log p-value often
572 dips below -15 several times around each transition. If more than 1 position met the cutoff within
573 100 nucleotides of each other, we took only the one with the smaller p-value. We then computed,
574 for each miRNA site, the ratio of the mean BLS of its section to that of the entire 3` UTR; we
575 term this statistic the "within-UTR conservation bias". Values of this statistic greater than 1
576 indicate that the P_{CT} overestimates the relative conservation of a given target site, while values
577 less than 1 indicate that the P_{CT} underestimates conservation. For gene-miRNA interactions with
578 multiple sites, we used the mean within-UTR conservation bias for all sites. We also repeated
579 P_{CT} score comparisons between classes of X- and Z-linked genes with P_{CT} scores normalized by
580 the corresponding gene-miRNA within-UTR conservation bias (Supplemental Figure S5C,
581 Supplemental Figure S9B). These scores, corresponding to all gene-miRNA interactions, are
582 provided in Supplemental Table S5.
583
584

585 **Logistic regression**

586 Logistic regression models were constructed using the function 'multinom' in the R package
587 'nnet.' For Z-linked genes, this reduces to binomial logistic regression, with probabilities of a
588 binary outcome (surviving W homolog or not) modeled as a log-linear combination of predictor
589 variables. Since there are three classes of X-linked genes, class probabilities are modeled by a
590 multinomial distribution. To compare models including only known factors to ones with mean
591 P_{CT} scores as an addition predictor, we used Aikake's Information Criterion (AIC), since it
592 penalizes more complex models (such as those with P_{CT} as an additional predictor). We used
593 previously published values for known factors in the survival of Y-linked (Bellott et al., 2014)
594 and W-linked (Bellott et al., 2017) genes except for human expression breadth, which we
595 recalculated using data from the GTEx Consortium v6 data release (Consortium, 2015). Briefly,
596 kallisto was used to estimate transcript per million (TPM) values in the 10 male samples with the
597 highest RNA integrity numbers (RINs) from each of 37 tissues, and expression breadth across
598 tissues was calculated as described in (Bellott et al., 2014), using median TPM values for each
599 tissue.
600
601 **Human-chicken conserved microRNA target sites**
602 Site-wise alignment information was obtained from TargetScanHuman v7.1
603 (http://www.targetscan.org/vert_71/vert_71_data_download/Conserved_Family_Info.txt.zip). To
604 determine which target sites are present in the 3` UTRs of both human and chicken orthologs, we
605 counted, for genes with both a human and chicken ortholog, the number of miRNA interactions
606 that had at least one target site in both human and chicken. To control for gene-specific
607 background 3` UTR conservation, we generated six control k-mers for each miRNA family seed

608 sequence that were matched exactly for nucleotide and CpG content. Six was the maximum
609 number of unique control k-mers that could be generated for all sequences. We repeated the
610 above counting analysis with each of the control k-mers using scripts from TargetScan, and
611 compared, for each gene, the observed number of human-chicken-conserved miRNA interactions
612 (the observed conservation signal) to the average number from controls (the background
613 conservation). This same procedure was repeated for alternative pairs of species considered
614 (opossum-chicken and human-anolis lizard).
615
616 **Gene expression profiling and crosslinking datasets**
617 Fold-changes in mRNA expression from a compendium of small RNA (sRNA) transfections
618 (corresponding to twelve different miRNAs) in HeLa cells were obtained from Agarwal and
619 colleagues (Agarwal et al., 2015) (GSM210904, GSM37601, GSM210913, GSM210903,
620 GSM210911, GSM210898, GSM210897, GSM210897, GSM210901, GSM210909,
621 GSM119747; E-MEXP-1402(1595297513)). Further datasets describing the effects of
622 transfecting miR-103 in HCT116 cells (Linsley et al., 2007) (GSM156580), knocking down
623 miR-92a in HEK293 cells (Hafner et al., 2010) (GSM538818), transfecting miR-7 or miR-124 in
624 HEK293 cells (Hausser, Landthaler, Jaskiewicz, Gaidatzis, & Zavolan, 2009) (GSM363763,
625 GSM363766, GSM363769, GSM363772, GSM363775, GSM363778), or of knocking out miR-
626 155 in mouse B cells (Eichhorn et al., 2014)
627 (GSM1479572, GSM1479576, GSM1479580, GSM1479584), T cells (Loeb et al., 2012)
628 (GSM1012118, GSM1012119, GSM1012120, GSM1012121, GSM1012122, GSM1012123), or
629 Th1 and Th2 cells (Rodriguez et al., 2007) (E-TABM-232), processed as described in (Agarwal
630 et al., 2015), were provided by V. Agarwal. Targets for the PAR-CLIP study (Hafner et al.,

2010) were inferred from an online resource of HEK293 clusters observed after transfection of either miR-124 (http://www.mirz.unibas.ch/restricted/clipdata/RESULTS/miR124_TRANSFECTION/miR124_TRANSFECTION.html) or miR-7 (http://www.mirz.unibas.ch/restricted/clipdata/RESULTS/miR7_TRANSFECTION/miR7_TRANSFECTION.html). All fold-changes and CLIP targets are provided in Supplemental Table S6.

DATA ACCESS

Data supporting the findings of this study are available within the paper and its Supplemental information files.

Code availability

A custom Python (RRID:SCR_008394) script utilizing Biopython (RRID:SCR_007173) was used to generate shuffled miRNA family seed sequences. Identification of miRNA target site matches using shuffled seed sequences was performed using the 'targetscan_70.pl' perl script (http://www.targetscan.org/vert_71/vert_71_data_download/targetscan_70.zip). 3` UTR segmentation was performed with the 'plot_transitions.py' python script. Code is available at: https://github.com/snaqvi1990/Naqvi17-code

ACKNOWLEDGEMENTS

We thank V. Agarwal, S. Eichorn, S. McGeary, and D. Bartel for assistance with the TargetScan database and helpful discussions; A. Godfrey for updated human-chicken orthology information; and A. Godfrey, J. Hughes and H. Skaletsky for critical reading of the manuscript. This work

654 was supported by the National Institutes of Health and the Howard Hughes Medical Institute.

655 S.N. was supported under a research grant by Biogen.

656

Author contributions

S.N., D.W.B. and D.C.P designed the study. S.N. performed analyses with assistance from D.W.B. K.S.L developed and implemented the step-detection algorithm. S.N. and D.C.P wrote the paper.

DISCLOSURE DECLARATION

The authors declare no competing financial interests.

Agarwal, V., Bell, G. W., Nam, J., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs, 1–38. https://doi.org/10.7554/eLife.05005

Balaton, B. P., Cotton, A. M., and Brown, C. J. (2015). Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biology of Sex Differences*, *6*(1), 35. https://doi.org/10.1186/s13293-015-0053-7

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, *136*(2), 215–33. https://doi.org/10.1016/j.cell.2009.01.002

Bellott, D. W., Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Cho, T.-J., Koutseva, N., Zaghlul, S., Graves, T., Rock, S., Kremitzki, C., Fulton, R. S., Dugan, S., Ding, Y., Morton, D., Khan, Z., Lewis, L., … Page, D. C. (2014). Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature*, *508*(7497), 494–9. https://doi.org/10.1038/nature13206

Bellott, D. W., Skaletsky, H., Cho, T.-J., Brown, L., Locke, D., Chen, N., Galkina, S., Pyntikova, T., Koutseva, N., Graves, T., Kremitzki, C., Warren, W. C., Clark, A. G., Gaginskaya, E., Wilson, R. K., and Page, D. C. (2017). Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nature Genetics, in press*.

Bellott, D. W., Skaletsky, H., Pyntikova, T., Mardis, E. R., Graves, T., Kremitzki, C., Brown, L. G., Rozen, S., Warren, W. C., Wilson, R. K., and Page, D. C. (2010). Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature*, *466*(7306), 612–616. https://doi.org/10.1038/nature09172

Carrel, L., and Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, *434*(March), 400–404. https://doi.org/10.1038/nature03479

Charlesworth, B., and Crow, J. F. (1978). Model for evolution of Y chromosomes and dosage compensation, *75*(11), 5618–5622.

Chen, X., and Zhang, J. (2015). No X-chromosome dosage compensation in human proteomes. *Molecular Biology and Evolution*, *32*(6), 1456–1460. https://doi.org/10.1093/molbev/msv036

Consortium, G. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (New York, N.Y.)*, *348*(6235), 648–660.

Cotton, A. M., Ge, B., Light, N., Adoue, V., Pastinen, T., and Brown, C. J. (2013). Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biology*, *14*(11), R122. https://doi.org/10.1186/gb-2013-14-11-r122

Cotton, a. M., Price, E. M., Jones, M. J., Balaton, B. P., Kobor, M. S., and Brown, C. J. (2014). Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Human Molecular Genetics*, *24*(6), 1528–1539. https://doi.org/10.1093/hmg/ddu564

Deng, X., Hiatt, J. B., Nguyen, D. K., Ercan, S., Sturgill, D., Hillier, L. W., Schlesinger, F., Davis, C. a, Reinke, V. J., Gingeras, T. R., Shendure, J., Waterston, R. H., Oliver, B., Lieb, J. D., and Disteche, C. M. (2011). Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nature Genetics*, *43*(12), 1179–85. https://doi.org/10.1038/ng.948

Eichhorn, S. W., Guo, H., McGeary, S. E., Rodriguez-Mias, R. a, Shin, C., Baek, D., Hsu, S.-H., Ghoshal, K., Villén, J., and Bartel, D. P. (2014). mRNA Destabilization Is the Dominant Effect of Mammalian MicroRNAs by the Time Substantial Repression Ensues. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2014.08.028

711 Friedman, R. C., Farh, K. K.-H., Burge, C. B., and Bartel, D. P. (2009). Most mammalian
712 mRNAs are conserved targets of microRNAs. *Genome Research*, *19*(1), 92–105.
713 https://doi.org/10.1101/gr.082701.108
714 Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Berninger, P., Rothballer, A., Jr, M. A.,
715 Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., and Tuschl, T. (2010).
716 Transcriptome-wide identification of RNA-binding protein and microRNA target sites by
717 PAR-CLIP. *Cell*, *141*(1), 129–141.
718 https://doi.org/10.1016/j.cell.2010.03.009.Transcriptome-wide
719 Hausser, J., Landthaler, M., Jaskiewicz, L., Gaidatzis, D., and Zavolan, M. (2009). mRNA
720 binding of Argonaute / EIF2C − miRNA complexes and the degradation of miRNA targets
721 Relative contribution of sequence and structure features to the mRNA binding of Argonaute
722 / EIF2C – miRNA complexes and the degradation of miRNA targets. *Genome Research*,
723 2009–2020. https://doi.org/10.1101/gr.091181.109
724 Hemara-Wahanui, A., Berjukow, S., Hope, C. I., Dearden, P. K., Wu, S.-B., Wilson-Wheeler, J.,
725 Sharp, D. M., Lundon-Treweek, P., Clover, G. M., Hoda, J.-C., Striessnig, J., Marksteiner,
726 R., Hering, S., and Maw, M. a. (2005). A CACNA1F mutation identified in an X-linked
727 retinal disorder shifts the voltage dependence of Cav1.4 channel activation. *Proceedings of*
728 *the National Academy of Sciences of the United States of America*, *102*(21), 7553–7558.
729 https://doi.org/10.1073/pnas.0501907102
730 Huang, N., Lee, I., Marcotte, E. M., and Hurles, M. E. (2010). Characterising and predicting
731 haploinsufficiency in the human genome. *PLoS Genetics*, *6*(10), e1001154.
732 https://doi.org/10.1371/journal.pgen.1001154
733 Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Graves, T., Fulton, R. S., Dugan, S.,
734 Ding, Y., Buhay, C. J., Kremitzki, C., Wang, Q., Shen, H., Holder, M., Villasana, D.,
735 Nazareth, L. V, Cree, A., Courtney, L., … Page, D. C. (2012). Strict evolutionary
736 conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature*,
737 *483*(7387), 82–6. https://doi.org/10.1038/nature10843
738 Itoh, Y., Melamed, E., Yang, X., Kampf, K., Wang, S., Yehya, N., Van Nas, A., Replogle, K.,
739 Band, M. R., Clayton, D. F., Schadt, E. E., Lusis, A. J., and Arnold, A. P. (2007). Dosage
740 compensation is less effective in birds than in mammals. *Journal of Biology*, *6*(1), 2.
741 https://doi.org/10.1186/jbiol53
742 Jegalian, K., and Page, D. (1998). A proposed path by which genes common to mammalian X
743 and Y chromosomes evolve to become X inactivated. *Nature*, *394*(August), 776–780.
744 Retrieved from http://www.nature.com/nature/journal/v394/n6695/abs/394776a0.html
745 Julien, P., Brawand, D., Soumillon, M., Necsulea, A., Liechti, A., Schütz, F., Daish, T., Grützner,
746 F., and Kaessmann, H. (2012, January). Mechanisms and evolutionary patterns of
747 mammalian and avian dosage compensation. *PLoS Biology*.
748 https://doi.org/10.1371/journal.pbio.1001328
749 Kaiser, V. B., Zhou, Q., and Bachtrog, D. (2011). Nonrandom gene loss from the drosophila
750 miranda neo-Y chromosome. *Genome Biology and Evolution*, *3*, 1329–1337.
751 https://doi.org/10.1093/gbe/evr103
752 Kharchenko, P. V, Xi, R., and Park, P. J. (2011). Evidence for dosage compensation between the
753 X chromosome and autosomes in mammals. *Nature Genetics*, *43*(12), 1167-9–2.
754 https://doi.org/10.1038/ng.991
755 Lahn, B. T., Ma, N., Breg, R. W., Stratton, R., Surti, U., and Page, D. C. (1994). Xq-Yq
756 interchange resulting in supernormal X-linked gene expression in severely retarded males

with 46,XYq- karyotype. *Nature Genetics*, *8*, 362–369. https://doi.org/10.1038/ng1294-340

Lahn, B. T., and Page, D. (1999). Four Evolutionary Strata on the Human X Chromosome. *Science*, *286*(5441), 964–967. https://doi.org/10.1126/science.286.5441.964

Lin, F., Xing, K., Zhang, J., and He, X. (2012). Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. *Proceedings of the National Academy of Sciences*, *109*(29), 11752–11757. https://doi.org/10.1073/pnas.1201816109

Lindgren, A. M., Hoyos, T., Talkowski, M. E., Hanscom, C., Blumenthal, I., Chiang, C., Ernst, C., Pereira, S., Ordulu, Z., Clericuzio, C., Drautz, J. M., Rosenfeld, J. a, Shaffer, L. G., Velsher, L., Pynn, T., Vermeesch, J., Harris, D. J., … Morton, C. C. (2013). Haploinsufficiency of KDM6A is associated with severe psychomotor retardation, global growth restriction, seizures and cleft palate. *Human Genetics*, *132*(5), 537–52. https://doi.org/10.1007/s00439-013-1263-x

Linsley, P. S., Schelter, J., Burchard, J., Kibukawa, M., Martin, M. M., Bartz, S. R., Johnson, J. M., Cummins, J. M., Raymond, C. K., Dai, H., Chau, N., Cleary, M., Jackson, A. L., Carleton, M., and Lim, L. (2007). Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol*, *27*(6), 2240–2252. https://doi.org/10.1128/MCB.02005-06

Loeb, G. B., Khan, A. A., Canner, D., Hiatt, J. B., Shendure, J., Darnell, R. B., Leslie, C. S., and Rudensky, A. Y. (2012). Transcriptome-wide miR-155 Binding Map Reveals Widespread Noncanonical MicroRNA Targeting. *Molecular Cell*, *48*(5), 760–770. https://doi.org/http://dx.doi.org/10.1016/j.molcel.2012.10.002

Mank, J. E., and Ellegren, H. (2009). All dosage compensation is local: gene-by-gene regulation of sex-biased expression on the chicken Z chromosome. *Heredity*, *102*(3), 312–320. https://doi.org/10.1038/hdy.2008.116

Mueller, J. L., Skaletsky, H., Brown, L. G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren, W. C., Wilson, R. K., and Page, D. C. (2013). Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature Genetics*, *45*(9), 1083–1087. https://doi.org/10.1038/ng.2705

Nanda, I., Shan, Z., Schartl, M., Burt, D. W., Koehler, M., Nothwang, H., Grützner, F., Paton, I. R., Windsor, D., Dunn, I., Engel, W., Staeheli, P., Mizuno, S., Haaf, T., and Schmid, M. (1999). 300 million years of conserved synteny between chicken Z and human chromosome 9. *Nature Genetics*, *21*(march), 258–259. https://doi.org/10.1038/6769

Ohno, S. (1967). *Sex chromosomes and sex-linked genes*.

Pessia, E., Makino, T., Bailly-Bechet, M., McLysaght, A., and Marais, G. a B. (2012). Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(14), 5346–51. https://doi.org/10.1073/pnas.1116763109

Pinzón, N., Li, B., Martinez, L., Sergeeva, A., Presumey, J., Apparailly, F., and Seitz, H. (2016). The number of biologically relevant microRNA targets has been largely over-estimated The number of biologically relevant microRNA targets has been largely overestimated, (November), 1–11. https://doi.org/10.1101/gr.205146.116

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rodriguez, A., Vigorito, E., Clare, S., Warren, M. V, Couttet, P., Soond, D. R., Dongen, S. Van, Grocock, R. J., Das, P. P., Miska, E. A., Vetrie, D., Okkenhaug, K., Enright, A. J., Dougan, G., Turner, M., and Bradley, A. (2007). Requirement of bic/microRNA-155 for normal immune function. *Science*, *981*(April), 608–611. https://doi.org/10.1126/science.1139253

Ross, M. T., Grafham, D. V, Coffey, A. J., Scherer, S., McLay, K., Muzny, D., and Platzer, M. (2005). The DNA sequence of the human X chromosome. *Nature*, *434*(March), 325–337.

Ruderfer, D. M., Hamamsy, T., Lek, M., Karczewski, K. J., Kavanagh, D., Samocha, K. E., Exome Aggregation Consortium, Daly, M. J., MacArthur, D. G., Fromer, M., and Purcell, S. M. (2016). Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nature Genetics*, *48*(10). https://doi.org/10.1038/ng.3638

Simon, D., Laloo, B., Barillot, M., Barnetche, T., Blanchard, C., Rooryck, C., Marche, M., Burgelin, I., Coupry, I., Chassaing, N., Gilbert-Dussardier, B., Lacombe, D., Grosset, C., and Arveiler, B. (2010). A mutation in the 3′-UTR of the HDAC6 gene abolishing the post-transcriptional regulation mediated by hsa-miR-433 is linked to a new form of dominant X-linked chondrodysplasia. *Human Molecular Genetics*, *19*(10), 2015–2027. https://doi.org/10.1093/hmg/ddq083

Skaletsky, H., Kuroda-kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., … Page, D. C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, *423*, 825–838.

Uebbing, S., Konzer, A., Xu, L., Backström, N., Brunström, B., Bergquist, J., and Ellegren, H. (2015). Quantitative Mass Spectrometry Reveals Partial Translational Regulation for Dosage Compensation in Chicken. *Molecular Biology and Evolution*, *32*(10), 2716–2725. https://doi.org/10.1093/molbev/msv147

Vandewalle, J., Van Esch, H., Govaerts, K., Verbeeck, J., Zweier, C., Madrigal, I., Mila, M., Pijkels, E., Fernandez, I., Kohlhase, J., Spaich, C., Rauch, A., Fryns, J. P., Marynen, P., and Froyen, G. (2009). Dosage-Dependent Severity of the Phenotype in Patients with Mental Retardation Due to a Recurrent Copy-Number Gain at Xq28 Mediated by an Unusual Recombination. *American Journal of Human Genetics*, *85*(6), 809–822. https://doi.org/10.1016/j.ajhg.2009.10.019

Watson, J. M., Spencer, J. A., Riggs, A. D., and Graves, J. A. (1990). The X chromosome of monotremes shares a highly conserved region with the eutherian and marsupial X chromosomes despite the absence of X chromosome inactivation. *Proceedings of the National Academy of Sciences*, *87*(18), 7125–7129. https://doi.org/10.1073/pnas.87.18.7125

Weischenfeldt, J., Dubash, T., Drainas, A. P., Mardin, B. R., Chen, Y., Stütz, A. M., Waszak, S. M., Bosco, G., Halvorsen, A. R., Raeder, B., Efthymiopoulos, T., Erkek, S., Siegl, C., Brenner, H., Brustugun, O. T., Dieter, S. M., Northcott, P. A., … Korbel, J. O. (2017). Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nature Genetics*, (49), 65–74. https://doi.org/10.1038/ng.3722

White, M. a, Kitano, J., and Peichel, C. L. (2015). Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. *Molecular Biology and Evolution*, *32*(8), 1981–1995. https://doi.org/10.1093/molbev/msv078

Wright, A. E., Zimmer, F., Harrison, P. W., and Mank, J. E. (2015). Conservation of Regional Variation in Sex-Specific Sex Chromosome Regulation. *Genetics*, *201*(2), 587–598.

https://doi.org/10.1534/genetics.115.179234

Xiong, Y., Chen, X., Chen, Z., Wang, X., Shi, S., Wang, X., Zhang, J., and He, X. (2010). RNA sequencing shows no dosage compensation of the active X-chromosome. *Nature Genetics*, *42*(12), 1043–7. https://doi.org/10.1038/ng.711

Yang, F., Babak, T., Shendure, J., and Disteche, C. M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Research*, *20*(5), 614–622. https://doi.org/10.1101/gr.103200.109

Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., … Flicek, P. (2016). Ensembl 2016. *Nucleic Acids Research*, *44*(D1), D710–D716. https://doi.org/10.1093/nar/gkv1157

Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E. D., Gilbert, M. T. P., and Zhang, G. (2014). Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science*, *346*(6215), 1246338–1246338. https://doi.org/10.1126/science.1246338

Zimmer, F., Harrison, P. W., Dessimoz, C., and Mank, J. E. (2016). Compensation of Dosage-Sensitive Genes on the Chicken Z Chromosome. *Genome Biology and Evolution*, *8*(4), 1233–1242. https://doi.org/10.1093/gbe/evw075