Title: A gene-by-gene mosaic of dosage compensation strategies on the human X chromosome

Authors: Adrianna K. San Roman[1], Alexander K. Godfrey[1,2], Helen Skaletsky[1,3], Daniel W. Bellott[1], Abigail F. Groff[1], Laura V. Blanton[1], Jennifer F. Hughes[1], Laura Brown[1,3], Sidaly Phou[1,3], Ashley Buscetta[4], Paul Kruszka[4]†, Nicole Banks[4,5], Amalia Dutra[6], Evgenia Pak[6], Patricia C. Lasutschinkow[7], Colleen Keen[7], Shanlee M. Davis[8], Nicole R. Tartaglia[8,9], Carole Samango-Sprouse[7,10], Maximilian Muenke[4]†, and David C. Page[1,2,3]*

Affiliations:

[1]Whitehead Institute; Cambridge, MA 02142, USA.

[2]Department of Biology, Massachusetts Institute of Technology; Cambridge, MA 02139, USA.

[3]Howard Hughes Medical Institute, Whitehead Institute; Cambridge, MA 02142, USA.

[4]Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda; MD 20892, USA.

[5]Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health; Bethesda, MD 20892 USA.

[6]Cytogenetics and Microscopy Core, National Human Genome Research Institute, National Institutes of Health; Bethesda, MD 20892 USA.

[7]Focus Foundation; Davidsonville, MD 21035, USA.

[8]Department of Pediatrics, University of Colorado School of Medicine; Aurora, CO 80045, USA.

[9]Developmental Pediatrics, eXtraOrdinarY Kids Program, Children's Hospital Colorado, Auroroa, CO 80011, USA

[10]Department of Pediatrics, George Washington University; Washington, DC 20052, USA; Department of Human and Molecular Genetics, Florida International University, Miami, FL 33199, USA.

*Corresponding author. Email: dcpage@wi.mit.edu

†Present addresses: GeneDx, Gaithersburg, MD 20877, USA (P.K.); American College of Medical Genetics and Genomics, Bethesda, MD 20814, USA (M.M.).

Abstract: Dosage compensation in humans – ensuring the viability and fitness of females, with two X chromosomes, and males, with one – is thought to be achieved chromosome-wide by heterochromatinization of one X chromosome during female development. We reassessed this through quantitative gene-by-gene analyses of expression in individuals with one to four X chromosomes, tolerance for loss-of-function mutations, regulation by miRNAs, allele-specific

expression, and the presence of homologous genes on the Y chromosome. We found a mosaic of dosage compensation strategies on the human X chromosome reflecting gene-by-gene differences in multiple dimensions, including sensitivity to under- or over-expression. These insights enrich our understanding of Turner, Klinefelter, and other sex chromosome aneuploidy syndromes, and of sex-chromosome-mediated effects on health and disease in euploid males and females.

One-Sentence Summary: The human X chromosome displays several modes of dosage compensation, tailored to the qualities of individual genes.

Main Text: Sex chromosomes evolved from ordinary autosomes independently in many metazoan lineages as a special solution to the problem of sex determination, i.e., how it is decided that an embryo will develop as a male or female. However, differentiating a pair of autosomes into X and Y chromosomes, and the accompanying decay or neofunctionalization of Y-linked genes, creates potentially lethal imbalances in X-linked gene expression. Preventing these imbalances by maintaining the ancestral expression levels of dosage-sensitive genes ensures the viability and fitness of both sexes. This is accomplished through dosage compensation, studied most intensely in *Drosophila melanogaster*, *Caenorhabditis elegans*, *Mus musculus*, and humans.

Studies in *D. melanogaster* and *C. elegans* support the operation in these species of uniform, X-chromosome-wide mechanisms of dosage compensation. In *D. melanogaster*, dosage compensation is accomplished by developmentally up-regulating genes on the single X Chr in males to match the output of the two X Chrs in females (*1*). In *C. elegans*, dosage compensation is achieved by developmentally down-regulating genes on the two X Chrs of XX hermaphrodites to match the output of the single X Chr in XO males (*2*). In mammals, dosage compensation is also assumed to operate on a chromosome-wide (as opposed to a gene-by-gene) basis, through heterochromatinization and transcriptional silencing ("inactivation") of one X Chr in XX females, mediated by the long-noncoding RNA *XIST (3)*. However, we identified four reasons to question whether in humans a single chromosome-wide mechanism, in this case X chromosome inactivation (XCI), secures the viability and fitness of both sexes.

First, if XCI alone accomplishes dosage compensation on the human X chromosome, we would expect that individuals with only one X chromosome (45,X) would be viable, as in *D. melanogaster*, *C. elegans*, and *M. musculus (4-6).* However, a second sex chromosome – either

X or Y – is essential for in utero viability in humans, where 99% of 45,X fetuses abort spontaneously. The rare survivors likely have a mixture of 45,X cells and cells with a second sex chromosome as a result of mitotic nondisjunction in early development (*7, 8*). These rare survivors display a constellation of anatomic features known as Turner syndrome (*9, 10*).

Second, molecular analyses have revealed that up to one quarter of all human X-linked genes are transcribed from heterochromatinized ("inactive") X Chrs (*11*). These genes have been interpreted as roguish outliers that "escape" or "partially escape" XCI, but they could also be viewed as critical evidence that a one-size-fits-all mechanism cannot explain dosage compensation in humans.

Third, special circumstances in *D. melanogaster* and *C. elegans* may have favored the evolution of single, chromosome-wide mechanisms in those species. In *D. melanogaster*, the absence of recombination during male meiosis (*12*) may have provided an opportunity for evolutionarily nascent X and Y chromosomes to differentiate precipitously (*13*); Y-chromosome genes are expressed only in spermatogenic cells (*14*). In *C. elegans,* males (XO) possess no Y chromosome at all. In mammals, by contrast, persistent X-Y crossing-over preserves a "pseudoautosomal" region of identity between Chr X and Y (*15*). These ongoing connections between the X and Y Chrs during human evolution may have been incompatible with the acquisition of a single, chromosome-wide mechanism of dosage compensation.

Fourth, while XCI operates in both mice and humans, it provides a more effective and thorough solution to the problem of dosage compensation in mice than it does in humans. Specifically, XO mice are fully viable and do not exhibit the Turner syndrome phenotype observed in 45,X ("XO") humans (*5*). By comparison with the human X Chr, few genes are transcribed from heterochromatinized X Chrs in mice (i.e., few genes "escape" XCI (*16*)).

Similarly, by comparison with the human Y Chr, which contains a dozen genes that are widely expressed in somatic tissues (*17*) and are postulated to contribute to male viability (*18*), fewer mouse Y Chr genes are expressed in somatic tissues (*19*). In sum, in both male and female mice, the sex chromosomes' somatic roles are delegated and restricted to the "active" X Chr to a degree not seen in humans.

These considerations suggest that in contrast with *C. elegans*, *D. melanogaster*, and *M. musculus*, multiple solutions to the problem of dosage compensation might be operative in humans. Fortunately, there is an abundance of recent genetic and epigenetic data and tools with which to examine this question on a gene-by-gene basis in humans. We combined this growing body of gene-by-gene analytics with newly-derived metrics arising from quantitative examination of X- (and Y-) linked gene expression in cells derived from individuals with one to four X chromosomes and zero to four Y chromosomes.

Results and Discussion:

Dosage sensitivity differs widely among human X chromosome genes

We reasoned that gene-by-gene differences in organismal sensitivities to gene dosage might drive the human X Chr to evolve several distinct dosage compensation strategies. To test this, we investigated gene-by-gene metrics of natural selection against under- or over-expression. To assess sensitivity to under-expression, we used a measure of the observed frequency of loss-of-function (LoF) variants in human populations (the LoF observed/expected upper fraction (LOEUF) score, (*20*)). Natural selection should cull LoF variants in genes whose precise dosage is important for organismal fitness, resulting in fewer LoF variants than expected, while such variants should accumulate in genes whose under-expression has little or no effect on fitness. To

measure sensitivity to over-expression, we examined the probability of conserved targeting by microRNAs (miRNAs; P_{CT} score (*21*)), which repress expression by binding to a gene's 3' untranslated region (*22*). Genes whose over-expression is deleterious should maintain their miRNA binding sites via natural selection, while genes whose over-expression is inconsequential to fitness should show less conservation of these sites.

We compared dosage sensitivities for genes in different regions of the sex chromosomes (**Fig. 1A**). The X and Y chromosomes originated as a pair of ordinary autosomes, differentiating over a period of ~300 million years, and they remain identical at the pseudoautosomal region (PAR) on their distal short arms (*15, 23, 24*). (A small second PAR with four genes exists at the tips of the long arms, but rather than being preserved through evolution and retaining autosome-like features, it was acquired recently in humans (*25*), and is addressed in Supplemental Text.) The non-pseudoautosomal regions of the X (NPX) and Y (NPY) have diverged in structure and gene content and no longer exchange genetic material (*26, 27*). Despite this lack of crossing-over, NPX and NPY have retained 17 ancestral genes as homologous "NPX-NPY pairs" that have diverged in sequence and function to various degrees. (An additional two genes were acquired during a transposition event from Chr X to Y in humans (*18, 26*)).

We find that, as a group, PAR genes readily accumulate LoF variants and display less conservation of miRNA targeting than other sex chromosome genes and autosomal genes (**Fig. 1B**, **S1**; **Table S1**). This indicates that the expression levels of PAR genes are not under strong selection. Indeed, homozygous LoF mutations have been reported for 3 of 15 PAR genes, suggesting that they are dispensable (*20*). The one exception is *SHOX,* whose copy-number has been linked to height in individuals with sex chromosome aneuploidy (*28-30*).

On average, single-copy NPY genes have LOEUF scores and P_{CT} distributions similar to those of autosomal genes (**Fig. 1B, S1, Table S2**). On the other hand, NPY's "ampliconic" genes, characterized by many nearly-identical copies and testis-specific expression (*26*), display more LoF variants and less miRNA conservation (**Fig. S1**).

By contrast with PAR and NPY genes, NPX genes have significantly fewer LoF variants than autosomal genes, likely due to hemizygous selection in males (**Fig. 1B, S1, Table S1**). (As with the NPY, multicopy and ampliconic genes on NPX show more LoF variants and less miRNA conservation than autosomal genes [**Fig. S1**].) Compared to NPX genes without a Y homolog, NPX genes with NPY homologs have lower LOEUF and higher P_{CT} distributions, are more broadly expressed across the body (as surveyed across tissues in the GTEx project (*31*), **Fig. 1C**), and are more likely to be required for cell viability (defined as a "common essential" gene across cancer cell lines in the Broad Dependency Map (*32*); **Fig. 1D**).

In sum, our gene-by-gene analyses of dosage sensitivities across the sex chromosomes reveal heterogeneities that could drive distinct dosage compensation strategies. If distinct strategies exist on the human X Chr, we expect that they should be revealed by examining gene expression dynamics across a range of X Chr copy numbers. We reasoned that these strategies might also be affected by the expression dynamics of homologous genes on Chr Y. To explore these ideas, we examined the effects of X and Y chromosome copy number on gene expression in cells from individuals with sex chromosome aneuploidies. In parallel, we analyzed cells with trisomy 21, the most common autosomal aneuploidy and the cause of Down syndrome (*33*), allowing us to compare the effects of sex chromosome copy number against such effects on an autosome.

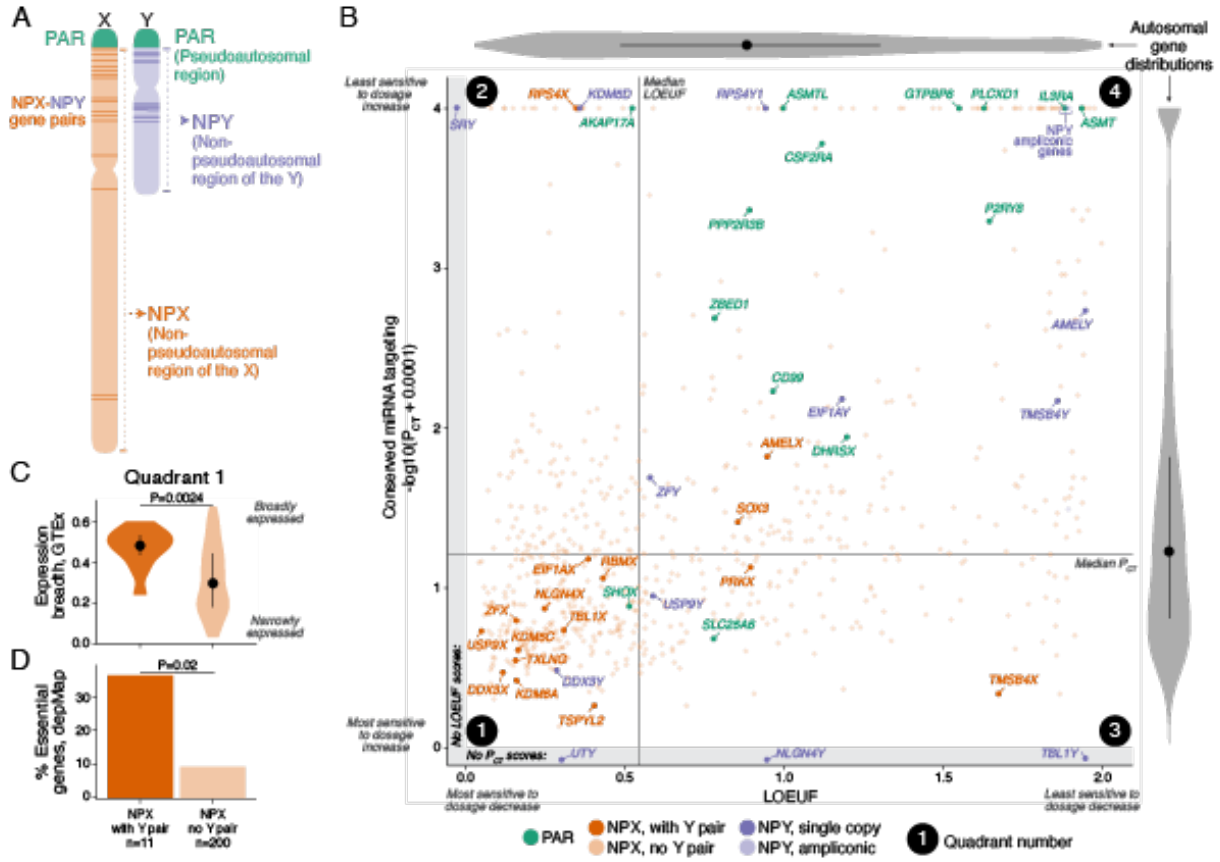

Figure 1. Heterogeneity in dosage sensitivity among X chromosome genes. (A) Anatomy of the human sex chromosomes. **(B)** Scatterplot of LOEUF versus transformed P_{CT} score for 674 sex chromosome genes with data for both measures, divided into quadrants by median values. All single-copy PAR and NPX-NPY genes are labelled. NPX-NPY genes with data for only one measure are plotted in the grey bars at the left and bottom. For comparison, we include violin plots with median (dot) and interquartile range (whiskers) of autosomal genes. **(C)** Violin plots of expression breadth across GTEx tissues for NPX genes in quadrant 1 with or without a NPY homolog. P-value, Wilcoxon rank sum test. **(D)** Percentage of cell-essential genes (depMap "common essential" genes) among quadrant 1 NPX genes with or without a NPY homolog. P-value, Fisher Exact test.

Serializing X chromosome copy number in two human cell types

We generated or received Epstein Barr Virus-transformed B cell lines (lymphoblastoid cell lines, LCLs) and/or primary dermal fibroblast cultures from 145 individuals with naturally occurring variation in the number of Chr X, Y, or 21 (**Fig. 2A**). After culturing cells under identical conditions, we profiled gene expression by RNA-sequencing (RNA-seq) in 112 LCL samples and 62 fibroblast samples in all (some individuals contributed both blood and skin samples; **Table 1 and Table S3**). Our sampling spanned individuals with one to four X chromosomes, zero to four Y chromosomes, and two or three copies of chromosome 21 (**Fig. 2B**). We assessed gene expression using 100-base-pair paired-end RNA sequencing, aiming for a depth of >50 million reads per sample (median = 78 million reads). To examine the reproducibility of our results, we sequenced independently-derived LCLs or fibroblast cultures from the same individual and found that they were highly correlated (**Fig. S2**). By our estimates, studying more individuals with sex chromosome aneuploidy would not substantially increase power in this analysis (**Fig. S3, Methods**).

Table 1. Samples included in aneuploidy analysis*

Karyotype	# LCLs	# Fibroblasts
45,X	17	17
46,XX	22	12
46,XY	17	11
47,XXX	7	3
47,XXY	11	12
47,XYY	10	4
48,XXXX	1	0

48,XXXY	4	1
48,XXYY	3	0
49,XXXXY	12	1
49,XYYYY	2	1
47,XX,+21	4	0
47,XY,+21	2	0
Total:	*112*	*62*

*Excludes technical replicates (additional cell lines derived from the same individual), which were only used for quality control.

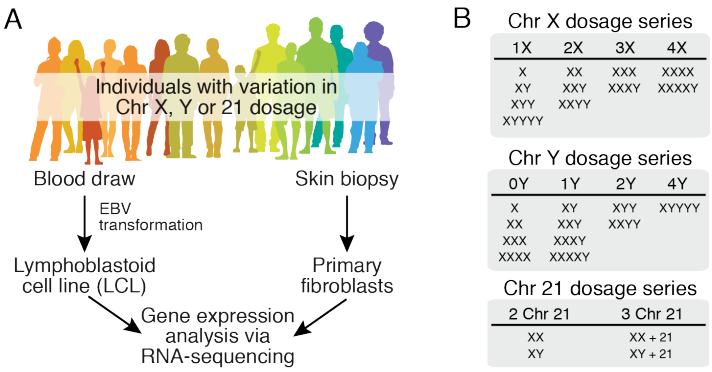

Figure 2. Gene expression analysis of aneuploid cells. (A) Collection and processing of samples from human subjects with variation in number of Chr X, Y, or 21. **(B)** We analyzed RNA-seq data from individuals with one to four copies of Chr X, zero to four copies of Chr Y, and two or three copies of Chr 21.

A metric of gene expression from supernumerary chromosomes

We anticipated that each additional chromosome generates an additive increase in expression of some or all of that chromosome's genes. Thus, we modeled sex chromosome gene expression as a linear function of the number of additional X Chr, Y Chr, and batch (**Methods**). For Chr 21, we

modeled gene expression as a linear function of Chr 21 copy number, sex chromosome complement (XX or XY), and batch (**Methods**). We included all genes whose median expression in euploid samples was at least 1 transcript per million (TPM), resulting in 424 NPX genes, 11 NPY genes, 11 PAR genes, and 95 Chr 21 genes analyzed in LCLs and/or fibroblasts (**Table 2**).

To compare the effects of changes in Chr X, Y, or 21 copy number between genes, we developed metrics that we refer to as ΔE_X, ΔE_Y, or ΔE_{21}, respectively. To calculate ΔE, we used the coefficients estimated from our linear models and divided the change in expression per additional Chr X, Y, or 21 (slope of regression – $ß_X$, $ß_Y$, or $ß_{21}$) by the expression from the first copy of Chr X or Y, or two copies of Chr 21 (average intercept across batches, $ß_0$) (**Fig. 3A,C,G, Tables S4-8, Methods**). $\Delta E=0$ indicates no change in expression with additional chromosomes, $\Delta E<0$ indicates a decrease in expression, and $\Delta E>0$ reflects an increase in expression, with $\Delta E=1$ indicating that each additional chromosome contributes an amount of expression equal to that of the first for ΔE_X and ΔE_Y. $\Delta E_{21}=1$ indicates that the third copy of Chr 21 contributes an amount equal to the mean contribution of the first two copies.

We began by evaluating PAR genes, all of which increased in expression and had ΔE_X and ΔE_Y values close to 1 (**Fig. 3B,D, S4A-B**). This implies, first, that PAR genes are expressed on each additional copy of Chr X or Y; and second, that expression from each additional Chr X or Y is roughly equal to the expression from the first (e.g., *ZBED1,* **Fig. 3E, S4C**). We observed modestly greater increases in PAR gene expression with additional Y chromosomes than with additional X chromosomes, especially in LCLs (**Fig. 3F, S4D**). This Y-vs-X effect was most pronounced for *CD99*, located near the PAR-NPX/Y boundary (**Fig. S5A**), consistent with suggestions that PAR gene expression is somewhat attenuated by spreading of heterochromatinization on X Chrs (*34*). For NPY genes, we analyzed samples with one or more

Y Chrs to observe expression differences, if any, between the first and additional Y Chrs (**Methods**). Like PAR genes, all NPY genes increased significantly, with ΔE_Y values close to 1, consistent with near-equal expression from each copy of Chr Y (e.g., *KDM5D*; **Fig. 3C-D, S4B,E, S5B;** full results in **Tables S6-7**). We next examined Chr 21 gene expression as a function of Chr 21 copy number (**Table S8**). Similar to PAR and NPY genes, most expressed Chr 21 genes increased in expression, with ΔE_{21} values close to 1 (e.g., *HLCS*; **Fig. 3G-H, S5C**), and no genes significantly decreased in expression. In sum, our analysis of PAR, NPY, and Chr 21 indicates that most genes have nearly identical expression from each copy of their respective chromosomes.

In contrast to PAR, NPY, and Chr 21, NPX genes were not identically expressed from each chromosome. Most genes did not significantly change in expression with additional copies of Chr X and had $\Delta E_X \approx 0$, consistent with expression from a cell's first X chromosome and repression on all others. 67.6% of expressed NPX genes in LCLs and 85.7% in fibroblasts fit this pattern (e.g., *PRPS2*, **Fig. 3I, S4F,** full results in **Tables S4-5**). The other NPX genes showed a wide range of ΔE_X values between -0.33 and 1.3 (e.g., *KDM5C, CA5B,* and *F8*, **Fig. 3A,J,K, S4G-I, S5D-E**). This included two long non-coding RNAs (lncRNAs) involved in XCI: *XIST*, the long non-coding RNA that acts in *cis* to transcriptionally repress X chromosomes from which it is expressed (*3, 35*), and *JPX,* an activator of *XIST (36, 37)* (**Fig. 3L-M, S4J-K**). *XIST* was the only NPX gene exclusively expressed in cells with two or more X chromosomes.

These results provide quantitative insight into the transcriptional response to increasing chromosome copy number across different regions of Chr X, Y, and 21. First, there appears to be no mechanism attenuating (or otherwise altering) the expression of genes on additional copies of Chr 21 or Y (including PAR and NPY). Second, nearly all genes on the second and subsequent X

chromosomes show either fully ($\Delta E_X=0$) or partially ($0<\Delta E_X<1$) attenuated expression. Only two NPX genes, *XIST* and *PUDP*, and a subset of PAR genes, showed near equal expression with each additional X chromosome ($\Delta E_X=1$) in both LCLs and fibroblasts. This might be expected for *XIST* given its role in repressing these supernumerary chromosomes. Third, X chromosome expression dynamics transcend cell types. We found that ΔE values were highly correlated in LCLs and fibroblasts for genes expressed in both cell types (Pearson r=0.84; **Fig. 3N**).

Table 2. Number of genes* assessed in expression analyses

Genomic region	LCLs	Fibroblasts	Union
NPX	352	399	424
NPY	11	10	11
PAR	11	9	11
Chr 21	95	NA	95

*Expressed (transcripts per million >= 1) protein-coding genes and lncRNAs known to be involved in XCI.

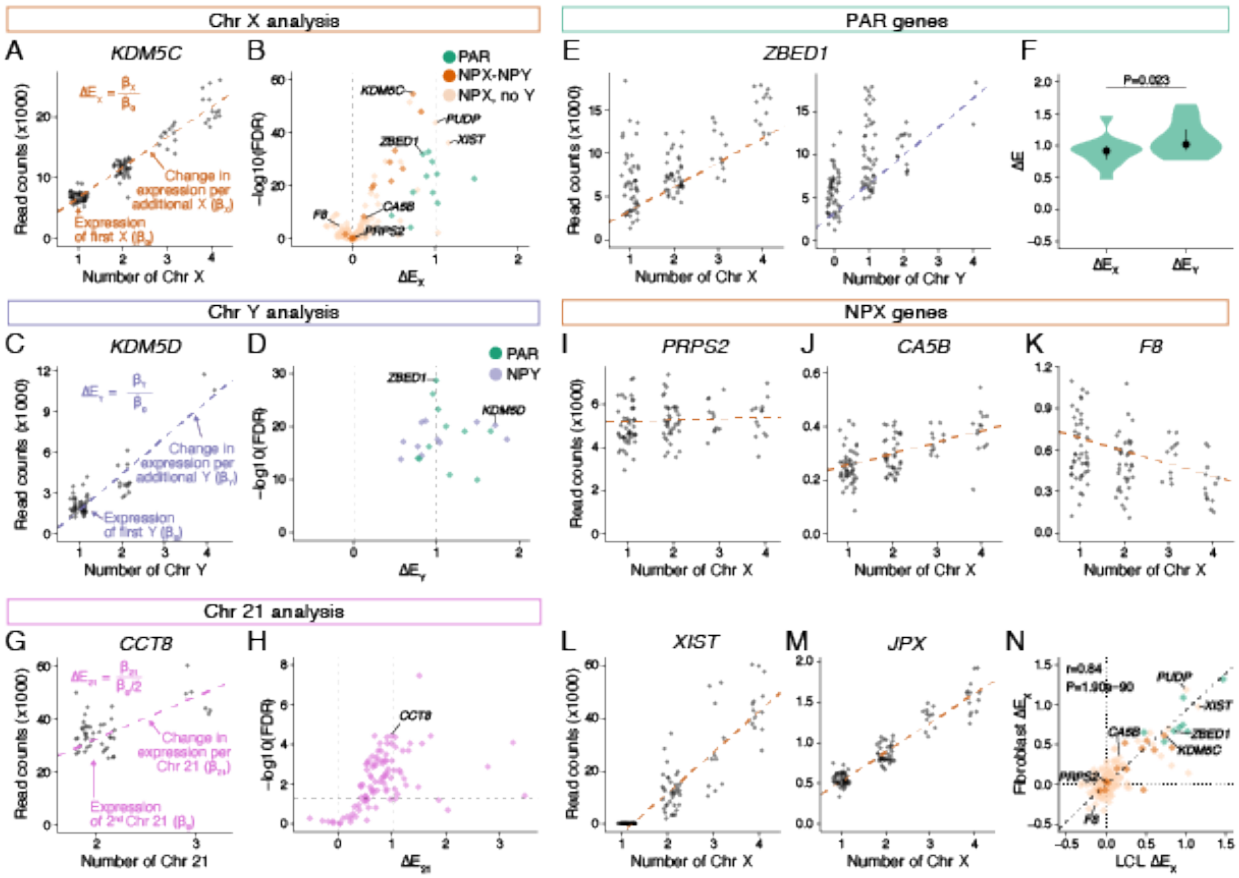

Figure 3. Calculating ΔE values for genes on Chr X, Y, and 21. Scatterplots and regression lines of individual genes as a function of Chr X **(A, E, I-M),** Y **(C, E)**, or 21 **(G)** copy number in LCLs. ΔE calculations from linear regressions are indicated. Volcano plots of ΔE_X **(B)**, ΔE_Y **(D)**, or ΔE_{21} **(H)** versus significance show differences in the overall response of genes on Chr X, Y, and 21 to increasing copy number. **(F)** Violin plots, with median (dots) and interquartile ranges (whiskers) indicated, comparing ΔE_X and ΔE_Y for PAR genes in LCLs. P-values, paired t-test. **(N)** Scatterplot comparing ΔE_X in LCLs and fibroblasts for Chr X genes. Deming regression line and Pearson correlation are indicated. See **Fig. S4** for corresponding plots for fibroblasts.

Monoallelic expression of one third of Chr X genes with ΔE_X>0

NPX genes that significantly increase in expression with Chr X copy number ($\Delta E_X>0$) may do so by transcribing the second or additional copies of Chr X, or by transcriptional modulation of the allele on the first ("active") copy without contribution from additional copies of Chr X (**Fig. 4A**). To discriminate between these two mechanisms, we identified heterozygous single nucleotide polymorphisms (SNPs) in our RNA-seq reads to distinguish between Chr X alleles in cells with two X chromosomes (**Table S9**). In 46,XX individuals, the choice to inactivate either the maternal X (X^M) or paternal X (X^P) is random, resulting in a mixture of cells in which the "active X" (Xa) is either X^M or X^P. Because of this, all transcripts in bulk RNA-seq data, even those expressed only from Xa, typically appear biallelic (**Fig. S6**). However, we identified 18 LCL and 4 fibroblast samples with "skewed XCI" whereby most cells in the culture had the same Xa, allowing us to distinguish transcription from Xa only (monoallelic) versus both Xa and Xi (biallelic; **Methods, Fig. S7**). As a control, we analyzed SNPs on Chr 8, which has a comparable number of expressed genes as Chr X, and where we invariably observed similar numbers of reads from each allele (**Fig. S8, Table S10**).

For each sample with skewed XCI, we computed the ratio of reads from each SNP allele ("allelic ratio", AR), and averaged them across exonic SNPs for each gene. We required AR data from at least two (fibroblast) or three (LCL) samples per gene, and calculated the median AR across individuals for 136 genes in LCLs and 84 genes in fibroblasts (**Fig. S9, Table S11-12**). To supplement our data, we incorporated AR data from three additional studies (*34, 38, 39*), allowing us to evaluate allelic expression for 369 Chr X genes expressed in LCLs or fibroblasts (**Fig. S10, Table S13**). Of the 80 Chr X genes whose expression significantly increased with additional X Chrs in either cell type ($\Delta E_X>0$) and had AR data, 51 had evidence of biallelic expression (denoted as biallelic in at least half of the datasets), including PAR genes and NPX

genes with NPY partners (we include *XIST* in this category, as, like the biallelic genes, it is expressed from Xi; **Fig. 4B**). In contrast, 29 genes had strong evidence of monoallelic expression (denoted as monoallelic in the majority of studies), supporting the existence of two distinct gene regulatory mechanisms among genes with $\Delta E_X>0$ (**Fig. 4B**). Compared to genes with biallelic expression, genes with monoallelic expression had lower ΔE_X values and were less likely to have $\Delta E_X>0$ in both cell types. We suspect that these genes are transcriptionally upregulated by other genes whose expression increases with additional copies of Chr X.

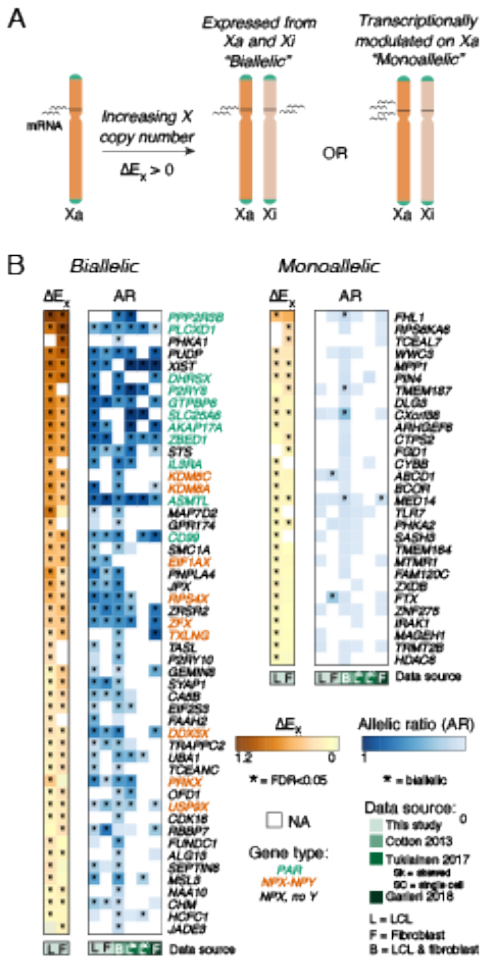

Figure 4. For genes with $\Delta E_X > 0$, allele-specific expression distinguishes transcription from both Xa and Xi from transcriptional upregulation of Xa alone. (A) Schematic of two scenarios that could result in $\Delta E_X > 0$. **(B)** Heatmaps of ΔE_X and AR for genes with $\Delta E_X > 0$ in

LCLs or fibroblasts, ordered by mean ΔE_X; 9 genes lacked AR data and are not shown. We required that the majority of evidence across studies support monoallelic expression to assign genes to that category.

Through the lens of sex chromosome evolution: ΔE_X, strata, and NPY homology

We next considered how the ΔE_X and allele-specific analyses intersect with our understanding of sex chromosome evolution. Sex chromosome differentiation was initiated when the proto-Y chromosome acquired the male sex-determining gene, *SRY* (**Fig. 5A**). Inversions on the evolving Y chromosome suppressed crossing-over with the X chromosome in stepwise fashion, creating and adding to the diverging NPX and NPY at the expense of the PAR, which continued to engage in crossing-over in XY males and XX females (*15, 23, 24*). The genetically isolated NPY lost all but 3% of the genes present on the ancestral autosomes, while the NPX, which continued to engage in crossing-over in XX females, retained 98% of the ancestral genes (*18, 26*).

We first asked whether the timing of NPY homolog loss during sex chromosome evolution impacts a NPX gene's ΔE_X. We grouped protein-coding NPX genes that have no NPY homologs in humans or other mammals (*18*) by evolutionary strata, which demarcate the time of their divergence from Chr Y. Genes in strata 1-3, which formed more than 97 million years ago, before the eutherian ancestor, had median $\Delta E_X \approx 0$, and most were expressed monoallelically (**Fig. 5B**). In contrast, genes in stratum 4, which formed more than 44 million years ago, before the simian ancestor, or stratum 5, which formed before the divergence of the macaque and human lineages, 32-34 million years ago, had a wide range of ΔE_X values, and most were expressed biallelically. Thus, NPX genes that lost their NPY partners in the more distant past are more fully attenuated than those that lost their NPY partners in the past 97 million years.

While most NPX genes lost their NPY homologs, the most dosage-sensitive NPX genes retain their NPY homologs today in humans or other mammals (*18*). Using sequencing data from the Y chromosome of other mammals, we inferred how long ago these NPY homologs were lost in the human lineage, and asked whether this divergence time is correlated with ΔE_X. NPX genes with NPY homologs in marsupials but not eutherian mammals (dating their loss to 97-176 million years ago) had $\Delta E_X=0$ and were mostly monoallelically expressed (**Fig. 5C**). In contrast, NPX genes with NPY homologs in eutherian mammals but not humans (all but one dating their loss in the human lineage to between the eutherian and simian ancestors 44-97 million years ago) had median $\Delta E_X \approx 0.2$, and most had evidence of biallelic expression.

Finally, we examined NPX genes with Y homologs in humans (**Fig. 5C, Table S14** (*40*)) and found that they fell into two groups. In the first group were NPX genes with single-copy NPY homologs that are widely and robustly expressed; they had median $\Delta E_X \approx 0.5$ and strong evidence of biallelic expression. The second group of NPX-NPY genes had $\Delta E_X=0$ and monoallelic expression. These genes included *TBL1X* and *TMSB4X,* whose NPY partners' expression is significantly reduced across the body, and *RBMX* and *TSPYL2,* whose multi-copy NPY partners are exclusively expressed in the testis *(17, 26)*).

In sum, the timing of NPY homolog loss, and the conservation or divergence of those homologs, differentiates various classes of NPX genes. While most NPX genes that lost Y homologs and reside in the older evolutionary strata (1-3) have become fully attenuated on supernumerary X chromosomes, this is true of only a small percentage of genes in the younger evolutionary strata (4-5). NPX genes with NPY homologs in eutherian mammals or humans are only partially attenuated regardless of their evolutionary strata: for example, *RPS4X*, with $\Delta E_X>0.4$, diverged from *RPS4Y* more than 176 million years ago.

Some dosage-insensitive genes adjacent to NPX-NPY pairs are epigenetic hitchhikers

While investigating NPX genes with $\Delta E_X>0$, we noticed that many were clustered along the X chromosome: 33 of 65 genes with $\Delta E_X>0$ in LCLs reside in 13 clusters of expressed genes, and 21 of 45 NPX genes with $\Delta E_X>0$ in fibroblasts reside in 8 clusters (**Table S15**). Most such clusters contain at least one gene with an NPY partner in humans or other eutherian mammals, which led us to speculate that dosage-insensitive genes were influenced by the expression strategies of their dosage-sensitive neighbors. To test this, we examined ΔE_X values for neighbors of NPX-NPY pairs in humans or other eutherian mammals with $\Delta E_X>0$. These neighbors had significantly higher ΔE_X values than genes that lack NPY-partnered neighbors (**Fig. 5D, Table S16**). Conversely, unpartnered NPX genes with $\Delta E_X>0$ were more likely to have a NPY-partnered neighbor than would be expected by chance (**Fig. S11**). Like their NPY-partnered neighbors, most of these adjacent genes were biallelically expressed (**Table S16**).

We considered whether inherent properties of the adjacent genes contributed to their expression from all copies of Chr X ($\Delta E_X>0$), or whether this is explained solely by their neighbors. To address this question, we examined the adjacent genes' sensitivity to increases or decreases in gene dosage. The genes separated into two groups, the first being as dosage-sensitive (**Fig. 5E**) and broadly-expressed throughout the body (**Fig. 5F**) as genes with human NPY partners, suggesting that their inherent properties contributed to $\Delta E_X>0$. Indeed, two of these genes – *SMC1A* and *MED14* – are essential for cell viability. Given these results, we were surprised that these genes had not retained NPY homologs and surmised that they utilized other evolutionary mechanisms to preserve gene dosage. In at least one case – for *MED14* in rat – this

was accomplished by retrotransposition of the NPX gene back to NPY (*41*). Complete sequencing of the sex chromosomes of additional mammals may clarify this question.

The second group of adjacent genes with $\Delta E_X>0$ showed little evidence of dosage-sensitivity (in quadrant 4 of **Fig. 5E**), consistent with them being epigenetic hitchhikers. We suspect that they lost their Y copies but were not under significant evolutionary pressure to be attenuated as their precise dosage is inconsequential. Their NPX-NPY neighbors may have provided a euchromatic environment that allowed them to have $\Delta E_X>0$.

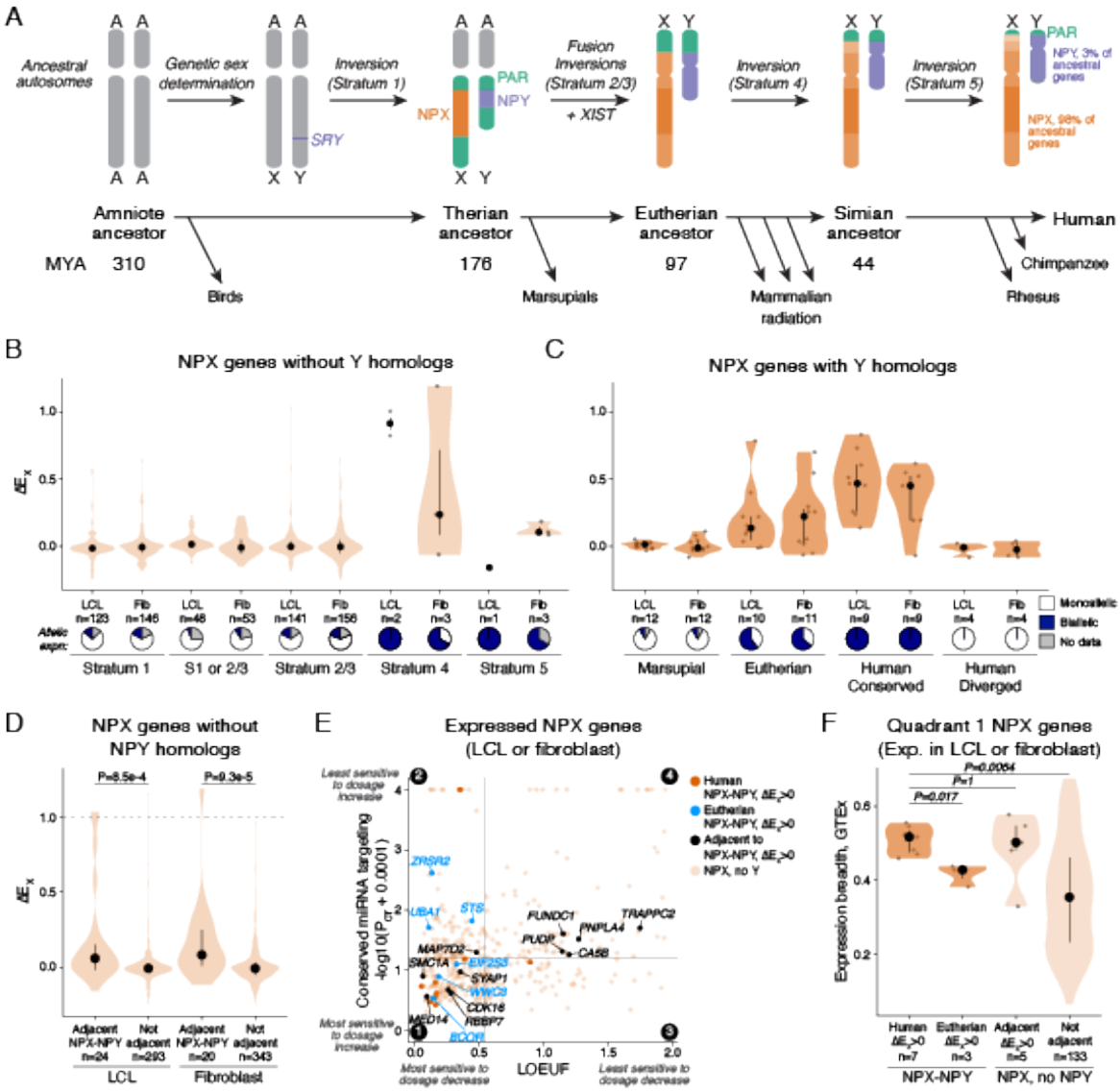

Figure 5. The impact of NPY homology on NPX gene regulation. (A) Timeline of human sex chromosome evolution. Dates in millions of years ago (MYA). **(B-D)** Violin plots, with median (dot) and interquartile ranges (whiskers) indicated, of ΔE_X for NPX genes without NPY homologs, by evolutionary stratum **(B)**, NPX genes with NPY homologs (see text for details of four "diverged" NPY homologs), or that are adjacent to NPX-NPY genes (**C**), and genes adjacent to NPX-NPY pairs in humans or eutherian mammals **(D)**. **(E)** Dosage sensitivity (see **Fig. 1B**) of NPX genes expressed in LCLs or fibroblasts, with eutherian NPY homologs and adjacent genes indicated. **(F)** For genes in quadrant 1, expression breadth is similar for NPX-NPY and adjacent genes. P-values, Wilcoxon rank sum test.

A mosaic of dosage compensation strategies on the X chromosome

In contrast to what is seen in *D. melanogaster*, *C. elegans,* and *M. musculus*, we find abundant evidence of multiple dosage compensation strategies across the human X chromosome – strategies rooted in the properties of individual genes (**Fig. 6**). Conventional thinking holds that X-chromosome gene expression in human males and females is equalized through monoallelic expression – a product of *XIST*-mediated XCI. Indeed, we find that monoallelic expression is commonplace among NPX genes following evolutionary loss of NPY homologs. Surprisingly, however, for at least 29 monoallelically expressed NPX genes, the expressed allele is transcriptionally upregulated with increasing numbers of X chromosomes, such that $\Delta E_X>0$ (**Fig. 4B**). These NPX genes do not "escape" XCI, yet they are expressed at higher levels in 46,XX females than in 46,XY males. The mechanism and biological impact of this monoallelic upregulation with increasing numbers of X chromosomes must now be explored. This transcriptional modulation of a single allele (on the "active" X chromosome) brings to mind the

transcriptional activation observed in *D. melanogaster* males and proposed as an intermediate step in the evolution of XCI in mammals (*24, 42*).

Other dosage compensation strategies are employed by X-chromosome genes with Y-chromosome homologs. We find that PAR genes, which maintain crossing-over in XY males, retain nearly equal, but slightly attenuated, expression on heterochromatinized X chromosomes. By contrast, NPX genes whose NPY homologs were preserved by natural selection due to their extraordinary dosage sensitivity (*18, 43*) are partially attenuated on heterochromatinized X chromosomes. We find that this strategy relies on continued functional interchangeability of NPX and NPY homologs in somatic tissues. When human NPY genes acquire testis-specific functions not interchangeable with the broader somatic roles of their NPX partners, those NPX genes become fully attenuated and monoallelically expressed (**Fig. 5C**). These PAR and NPX-NPY dosage scenarios, in which a functionally interchangeable Y Chr homolog is maintained, distinguish the human sex chromosomes from those of *D. melanogaster*, where the Y chromosome is not expressed somatically; from *C. elegans*, which has no Y chromosome; and from *M. Musculus*, with few somatically-expressed Y Chr genes.

We began this study by investigating metrics of natural selection against under- or over-expression of individual genes, aiming to differentiate among the dosage sensitivities of genes on the human X Chr. Integrating all of our current findings, we see that the most dosage sensitive genes – the NPX genes with functionally interchangeable NPY partners – did not adopt the commonplace dosage compensation strategy of monoallelic expression and $\Delta E_X=0$, but instead display biallelic expression and the modest levels of attenuation documented here. We suggest that, with regard to the preservation of both male and female fitness in the face of sex

chromosome evolution, intense and diverging selective pressures promoted a mosaic of dosage compensation strategies tailored to the properties of broad classes of genes.

Our model of gene-by-gene mosaicism based on properties of individual genes has its limitations. We see that dosage-sensitive NPX genes influence the regulation of their neighbors on the X Chr, and even a history of having a dosage-sensitive, NPY-partnered neighbor is impactful.

Implications of ΔE_X and ΔE_Y for understanding human health and disease

The numbers of X or Y chromosomes differ between euploid males (46,XY) and females (46,XX), and more strikingly in individuals with Turner syndrome (45,X), Klinefelter syndrome (47,XXY), and other sex chromosome aneuploidies. We propose that phenotypes associated with sex chromosome copy number are caused by genes whose expression levels change with chromosome copy number, and whose precise dosage affects fitness or viability. Further, we propose that this holds in both euploid and aneuploid individuals. From our analysis of gene expression across the spectrum of sex chromosome aneuploidy, we derived the metrics ΔE_X and ΔE_Y, which identify genes whose expression changes with chromosome copy number. Although PAR genes display the highest ΔE_X and ΔE_Y values, none except *SHOX* shows evidence of dosage sensitivity. By contrast, NPX-NPY pairs, also with high ΔE_X and ΔE_Y values, are exquisitely dosage-sensitive and represent the most promising candidates to contribute to phenotypes related to X Chr or Y Chr copy number – including the 99% in utero lethality of monosomy X. These NPX-NPY pair genes are expressed throughout the body and are involved in basic cellular functions including transcription, translation, and protein stability (*18*). Other

NPX genes that lost their NPY partners but have $\Delta E_X>0$ and high dosage sensitivity may also contribute to these phenotypes.

This work also allows us to better understand the features of Chr X, Y, and 21 that may account for the milder phenotypes associated with an extra X or Y compared to trisomy 21. As with the sex chromosomes, it is thought that dosage-sensitive genes on Chr 21 are responsible for the Down syndrome phenotype (*44*). Although we observed similar numbers of genes with $\Delta E>0$ on Chr X and Chr 21, attenuation of almost all genes on heterochromatinized X chromosomes resulted in smaller average ΔE values compared to Chr 21. While average ΔE values were comparable for Chr Y and Chr 21, the smaller number of genes on Chr Y may account for the milder phenotypes associated with supernumerary Y chromosomes. We conclude that dosage compensation on Chr X – in all of its forms – permits the viability of individuals with up to three additional chromosomes (e.g., 49,XXXXY).

Comparisons among sex chromosome aneuploidies have revealed specific phenotypes associated with the copy number of X or Y chromosomes. For example, 47,XXY and 47,XYY males display a higher prevalence of autism spectrum disorder (ASD) traits than 46,XY males, which display a higher prevalence of ASD traits than 46,XX females (*45-47*). Interestingly, the addition of an extra Y chromosome leads to an even higher prevalence of ASD traits than an extra X chromosome. This suggests that both sex chromosomes may influence certain phenotypes, but that one has a stronger effect than the other. We suspect that NPX-NPY pairs may play a role in these phenotypes, with the higher prevalence in 47,XYY reflecting the fact that ΔE_Y values for NPY genes are greater than the corresponding ΔE_X values of their NPX homologs. Other neurodevelopmental phenotypes, such as verbal and language weaknesses, are more prevalent among individuals with extra X chromosomes (*48*), implicating dosage-sensitive

NPX genes with $\Delta E_X>0$ that have no NPY partners or are functionally differentiated from their NPY partners.

Ultimately, the derivation of ΔE_X required the wide range of chromosome copy number observed among individuals with sex chromosome aneuploidy. Once measured, ΔE_X will have great value for understanding how X-linked genes contribute to differences in health and disease among euploid (46,XX and 46,XY) individuals. In addition, the quantitative, gene-by-gene reframing of dosage compensation provided by our ΔE_X metric provides an opportunity to revisit many aspects of X chromosome biology, including its epigenetic regulation and evolution.

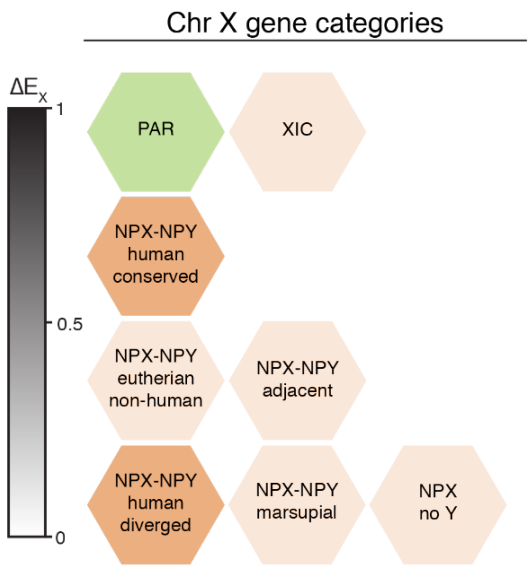

Figure 6. Model of mosaic dosage compensation strategies on the human X chromosome. Hexagons represent categories of X chromosome genes we analyzed in this study, organized vertically by ΔE_X. Colors distinguish between structural categories: green, PAR genes, dark orange, NPX-NPY pair in humans; light orange, NPX without NPY pair. XIC, lncRNA genes in the X inactivation center.

References

1. A. S. Mukherjee, W. Beermann, Synthesis of ribonucleic acid by the X-chromosomes of Drosophila melanogaster and the problem of dosage compensation. *Nature* **207**, 785-786 (1965).
2. B. J. Meyer, L. P. Casson, Caenorhabditis elegans compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. *Cell* **47**, 871-881 (1986).
3. C. J. Brown *et al*., A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38-44 (1991).
4. C. B. Bridges, Non-Disjunction as Proof of the Chromosome Theory of Heredity (Concluded). *Genetics* **1**, 107-163 (1916).
5. W. J. Welshons, L. B. Russell, The Y-Chromosome as the Bearer of Male Determining Factors in the Mouse. *Proc Natl Acad Sci U S A* **45**, 560-566 (1959).
6. V. Nigon, Les modalités de la reproduction et le déterminisme du sexe chez quelques Nématodes libres. *Ann. Sci. Nat. Zool. Ser.* **11**, 1-32 (1949).
7. E. B. Hook, D. Warburton, Turner syndrome revisited: review of new data supports the hypothesis that all viable 45,X cases are cryptic mosaics with a rescue cell line, implying an origin by mitotic loss. *Hum Genet* **133**, 417-424 (2014).
8. E. B. Hook, D. Warburton, The distribution of chromosomal genotypes associated with Turner's syndrome: livebirth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism. *Hum Genet* **64**, 24-27 (1983).

9. C. E. Ford, K. W. Jones, P. E. Polani, J. C. Dealmeida, J. H. Briggs, A Sex-Chromosome Anomaly in a Case of Gonadal Dysgenesis (Turners Syndrome). *Lancet* **1**, 711-713 (1959).

10. H. H. Turner, A syndrome of infantilism, congenital webbed neck, and cubitus valgus. *Endocrinology* **23**, 566-574 (1938).

11. B. P. Balaton, A. M. Cotton, C. J. Brown, Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biol Sex Differ* **6**, (2015).

12. T. H. Morgan, No crossing over in the male of drosophila of genes in the second and third pairs of chromosomes. *Biol Bull-Us* **26**, 195-204 (1914).

13. H. J. Muller, A gene for the fourth chromosome of drosophila. *J Exp Zool* **17**, 325-336 (1914).

14. A. B. Carvalho, B. A. Dobo, M. D. Vibranovski, A. G. Clark, Identification of five new genes on the Y chromosome of Drosophila melanogaster. *Proc Natl Acad Sci U S A* **98**, 13225-13230 (2001).

15. H. J. Cooke, W. R. A. Brown, G. A. Rappold, Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature* **317**, 687-692 (1985).

16. J. B. Berletch *et al*., Escape from X Inactivation Varies in Mouse Tissues. *Plos Genet* **11**, e1005079 (2015).

17. A. K. Godfrey *et al*., Quantitative analysis of Y-Chromosome gene expression across 36 human tissues. *Genome Res*, (2020).

18. D. W. Bellott *et al*., Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494-499 (2014).

19. Y. Q. S. Soh *et al*., Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* **159**, 800-813 (2014).

20. K. J. Karczewski *et al*., The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).

21. R. C. Friedman, K. K. Farh, C. B. Burge, D. P. Bartel, Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92-105 (2009).

22. D. P. Bartel, MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233 (2009).

23. B. T. Lahn, D. C. Page, Four evolutionary strata on the human X chromosome. *Science* **286**, 964-967 (1999).

24. S. Ohno. (Springer Berlin Heidelberg, Berlin, Heidelberg, 1967), vol. 1, pp. 1-201.

25. D. Freije, C. Helms, M. S. Watson, H. Donis-Keller, Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* **258**, 1784-1787 (1992).

26. H. Skaletsky *et al*., The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825-837 (2003).

27. M. T. Ross *et al*., The DNA sequence of the human X chromosome. *Nature* **434**, 325-337 (2005).

28. E. Rao *et al*., Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat Genet* **16**, 54-63 (1997).

29. M. Clement-Jones *et al*., The short stature homeobox gene SHOX is involved in skeletal abnormalities in Turner syndrome. *Hum Mol Genet* **9**, 695-702 (2000).

30. T. Ogata, N. Matsuo, Sex chromosome aberrations and stature: deduction of the principal factors involved in the determination of adult height. *Hum Genet* **91**, 551-562 (1993).

31. G. Consortium *et al*., Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).

32. A. Tsherniak *et al*., Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e516 (2017).

33. A. Megarbane *et al*., The 50th anniversary of the discovery of trisomy 21: the past, present, and future of research and treatment of Down syndrome. *Genet Med* **11**, 611-616 (2009).

34. T. Tukiainen *et al*., Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244-248 (2017).

35. G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, N. Brockdorff, Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-137 (1996).

36. H. Karner *et al*., Functional Conservation of LncRNA JPX Despite Sequence and Structural Divergence. *J Mol Biol* **432**, 283-300 (2020).

37. D. Tian, S. Sun, J. T. Lee, The long noncoding RNA, Jpx, is a molecular switch for X-chromosome inactivation. *Cell* **143**, 390-403 (2010).

38. A. M. Cotton *et al*., Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol* **14**, R122 (2013).

39. M. Garieri *et al*., Extensive cellular heterogeneity of X inactivation revealed by single-cell allele-specific expression in human fibroblasts. *Proc Natl Acad Sci U S A* **115**, 13015-13020 (2018).

40. D. W. Bellott, D. C. Page, Dosage-sensitive functions in embryonic development drove the survival of genes on sex-specific chromosomes in snakes, birds, and mammals. *Genome Res*, (2021).

41. J. W. Prokop *et al*., Analysis of Sry duplications on the Rattus norvegicus Y-chromosome. *BMC Genomics* **14**, 792 (2013).

42. K. Jegalian, D. C. Page, A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature* **394**, 776-780 (1998).

43. S. Naqvi, D. W. Bellott, K. S. Lin, D. C. Page, Conserved microRNA targeting reveals preexisting gene dosage sensitivities that shaped amniote sex chromosome evolution. *Genome Res* **28**, 474-483 (2018).

44. T. Makino, A. McLysaght, Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci U S A* **107**, 9270-9274 (2010).

45. J. L. Ross *et al*., Behavioral and Social Phenotypes in Boys With 47,XYY Syndrome or 47,XXY Klinefelter Syndrome. *PEDIATRICS* **129**, 769-778 (2012).

46. N. R. Tartaglia *et al*., Autism Spectrum Disorder in Males with Sex Chromosome Aneuploidy: XXY/Klinefelter Syndrome, XYY, and XXYY. *J Dev Behav Pediatr* **38**, 197-207 (2017).

47. S. van Rijn, A review of neurocognitive functioning and risk for psychopathology in sex chromosome trisomy (47,XXY, 47,XXX, 47, XYY). *Curr Opin Psychiatry* **32**, 79-84 (2019).

48. T. Green, S. Flash, A. L. Reiss, Sex differences in psychiatric disorders: what we can learn from sex chromosome aneuploidies. *Neuropsychopharmacology* **44**, 9-21 (2019).

49. J. L. Mueller *et al*., Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* **45**, 1083-1087 (2013).

50. E. K. Jackson *et al*., Large palindromes on the primate X Chromosome are preserved by natural selection. *bioRxiv*, (2021).

51. M. Vangipuram, D. Ting, S. Kim, R. Diaz, B. Schüle, Skin Punch Biopsy Explant Culture for Derivation of Primary Human Fibroblasts. *JoVE (Journal of Visualized Experiments)*, e3779-e3779 (2013).

52. S. Naqvi *et al*., Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**, 249-+ (2019).

53. K. D. Pruitt *et al*., The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res* **19**, 1316-1323 (2009).

54. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* **34**, 525-527 (2016).

55. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).

56. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

57. M. E. Ritchie *et al*., limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

58. A. Ciccodicola *et al*., Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudoautosomal region. *Hum Mol Genet* **9**, 395-401 (2000).

59. W. A. Bickmore, H. J. Cooke, Evolution of homologous sequences on the human X and Y chromosomes, outside of the meiotic pairing segment. *Nucleic Acids Res* **15**, 6261-6271 (1987).

60. F. Rouyer *et al*., A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* **319**, 291-295 (1986).

Acknowledgements:

We thank members of the Page laboratory, especially Lukas Chmatal, for helpful comments on the manuscript, Sahin Naqvi for advice on RNA-sequencing and analysis, Jorge Adarme and Susan Tocio for laboratory support, and the Whitehead Institute Genome Technology Core facility for library preparation and sequencing.

Funding:

National Institutes of Health grant F32HD091966 (AKSR)

National Institutes of Health grant U01HG0007587 (DCP and MM)

National Institutes of Health grant K23HD092588 (SMD)

Schmidt Science Fellows (AFG)

Howard Hughes Medical Institute (DCP)

National Human Genome Research Institute Intramural Research Program (MM)

NIH/NCATS Colorato CTSA grant UL1 TR002535 (NRT)

Contents are the authors' sole responsibility and do not necessarily represent official NIH views.

Philanthropic gifts from:

Brit and Alexander d'Arbeloff

Arthur W. and Carol Tobin Brill

Matthew Brill

Charles Ellis

Author contributions:

Conceptualization: AKSR and DCP

Data curation: AKSR and HS

Formal analysis: AKSR, AKG, HS, DWB, AG, LVB, and JFH

Funding acquisition: AKSR, AFG, SMD, NRT, MM, and DCP

Investigation: AKSR, SP, LB, AD, and EP

Methodology: AKSR, AKG, HS, and AFG

Project administration: AKSR, JFH, LB, AB, PK, NB, PCL, CK, and SMD

Resources: AB, PK, NB, PCL, CK, SMD, NRT, CSS, and MM

Software: AKSR, AKG, HS, AFG, and LVB

Supervision: AKSR, JFH, NRT, CSS, MM, and DCP

Validation: AKG, HS, DWB, and LVB

Visualization: AKSR and AKG

Writing - original draft preparation: AKSR and DCP

Writing - review and editing: AKSR, AKG, DWB, AFG, LVB, JFH, and DCP

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All RNA-sequencing data has been deposited to dbGaP, accession number phs002481.v1.p1. Code and processed data for recreating our analysis is available at https://github.com/aksanroman/dosage_compensation_strategies.git. Cell lines are available upon request.