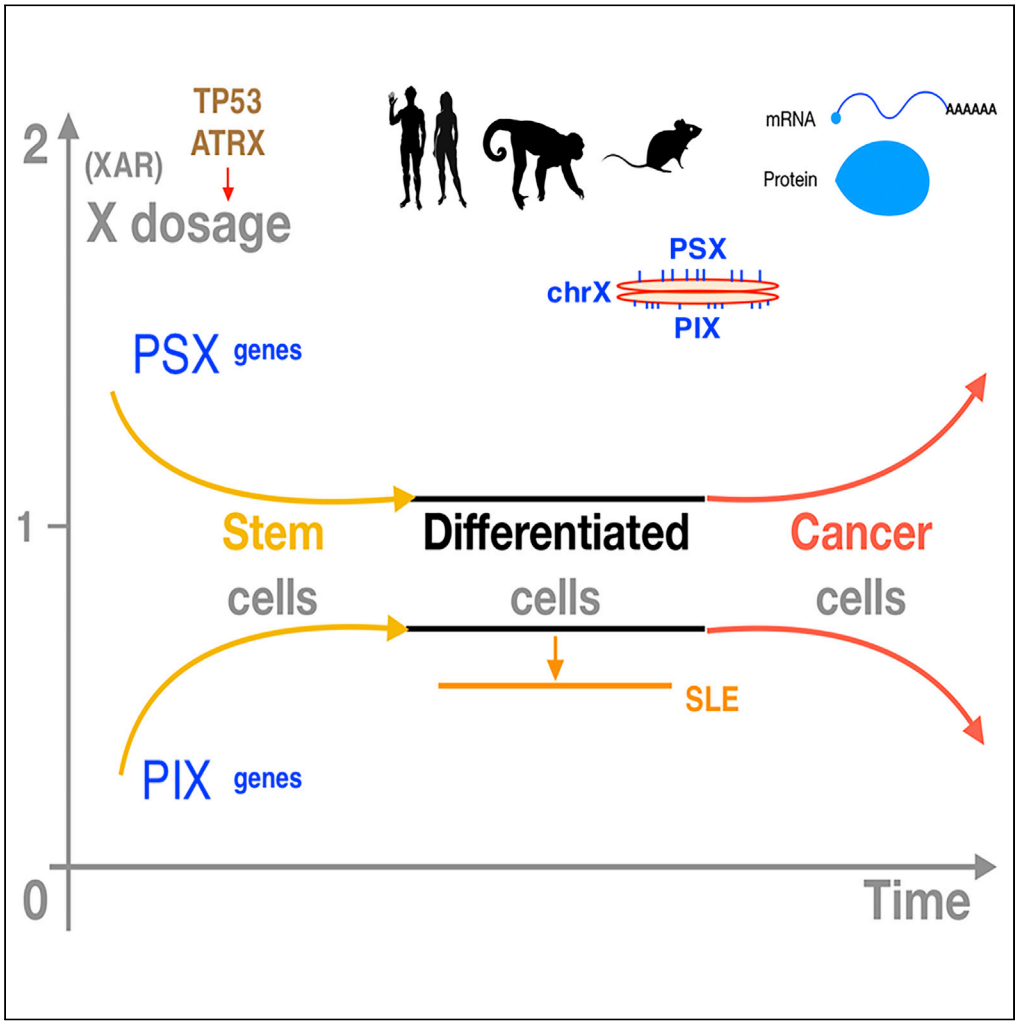
iScience

CellPress
OPEN ACCESS

Article

Distinct dosage compensations of ploidy-sensitive and -insensitive X chromosome genes during development and in diseases



Mengbiao Guo, Zhengwen Fang, Bohong Chen, Zhou Songyang, Yuanyan Xiong

xyyan@mail.sysu.edu.cn

Highlights

Conserved differential dosages of PIX and PSX genes at both mRNA and protein level

PIX and PSX dosages showed distinct trends in stem cells and in cancer cells

PIX dosage was lower in SLE autoimmune patients than in healthy individuals

TP53 and ATRX were identified and validated as two XAR regulators

Guo et al., iScience 26, 105997
February 17, 2023
https://doi.org/10.1016/j.isci.2023.105997

iScience

CellPress
OPEN ACCESS

Article

Distinct dosage compensations of ploidy-sensitive and -insensitive X chromosome genes during development and in diseases

Mengbiao Guo,[1,2] Zhengwen Fang,[1,2] Bohong Chen,[1,2] Zhou Songyang,[1] and Yuanyan Xiong[1,3,*]

SUMMARY

The active X chromosome in mammals is upregulated to balance its dosage to autosomes during evolution. However, it is elusive why the known dosage compensation machinery showed uneven and small influence on X genes. Here, based on >20,000 transcriptomes, we identified two X gene groups (ploidy-sensitive [PSX] and ploidy-insensitive [PIX]), showing distinct but evolutionarily conserved dosage compensations (termed XAR). We demonstrated that XAR-PIX was downregulated whereas XAR-PSX upregulated at both RNA and protein levels across cancer types, in contrast with their trends during stem cell differentiation. XAR-PIX, but not XAR-PSX, was lower and correlated with autoantibodies and inflammation in patients of lupus, suggesting that insufficient dosage of PIX genes contribute to lupus pathogenesis. We further identified and experimentally validated two XAR regulators, *TP53* and *ATRX*. Collectively, we provided insights into X dosage compensation in mammals and demonstrated different regulation of PSX and PIX and their pathophysiological roles in human diseases.

INTRODUCTION

In mammals, females have two X chromosomes whereas males have only one and females evolved to silence one copy of X to match the single X copy in males, a phenomenon called X chromosome inactivation (XCI)[1] with exceptions of escaping genes.[2] Initiated early in female embryo development, XCI is essentially stable for life after complete.[3] To balance the dosage between the single active X chromosome and other autosomes with two active copies, both sexes evolved to upregulate the active X chromosome,[1,4,5] although with controversies.[6,7]

Meanwhile, some possible mechanisms underlying X upregulation in mammals were revealed, mainly through epigenetic regulation, including transcription elongation and initiation enhancement and RNA stability improvement.[5] Expectedly, *KAT8* (lysine acetyltransferase 8, also known as *MOF* or *MYST1*), *MSL1* (male-specific lethal 1 homolog), and *MSL2*, whose homologs are responsible for X upregulation in Drosophila, were found to behave similarly in mammals and regulate histone acetylation to enhance transcription preferably on X chromosome.[5] However, it is intriguing that *KAT8* affected X genes unevenly and only a small percentage of X genes showed considerable effect sizes.

Recently, using published single-cell RNA sequencing datasets, Larsson et al. reported that the frequency of transcriptional bursts was the driving force behind X upregulation in mouse, and they proposed that *trans*-acting enhancer-binding regulators may be the underlying factor behind the higher transcriptional burst frequency of X chromosome.[4] However, these factors were unknown and the hypothesis of enhancer factors in X upregulation was not verified.

TP53 is a highly conserved gene important for various stress responses.[8] It was reported to interact with key MSL components, *KAT8* and *MSL2*.[9,10] The p53 binding sites[11] in *KAT8* (one site in promoter) and *MSL2* (three sites in promoter) indicate direct regulation of *KAT8* and *MSL2* transcription by p53. ATRX, an X-linked telomere-associating protein studied by our team,[12,13] is a chromatin remodeler that functions in genome stabilization and heterochromatin formation.[14,15] ATRX works with XIST and PRC2 to remodel chromatins[16] and ATRX loss can affect X inactivation in embryonic trophoblast.[17] Importantly, it has been

[1]Key Laboratory of Gene Engineering of the Ministry of Education, Institute of Healthy Aging Research, School of Life Sciences, Sun Yat-sen University, Guangzhou 510006, China

[2]These authors contributed equally

[3]Lead contact

*Correspondence: xyyan@mail.sysu.edu.cn

https://doi.org/10.1016/j.isci.2023.105997

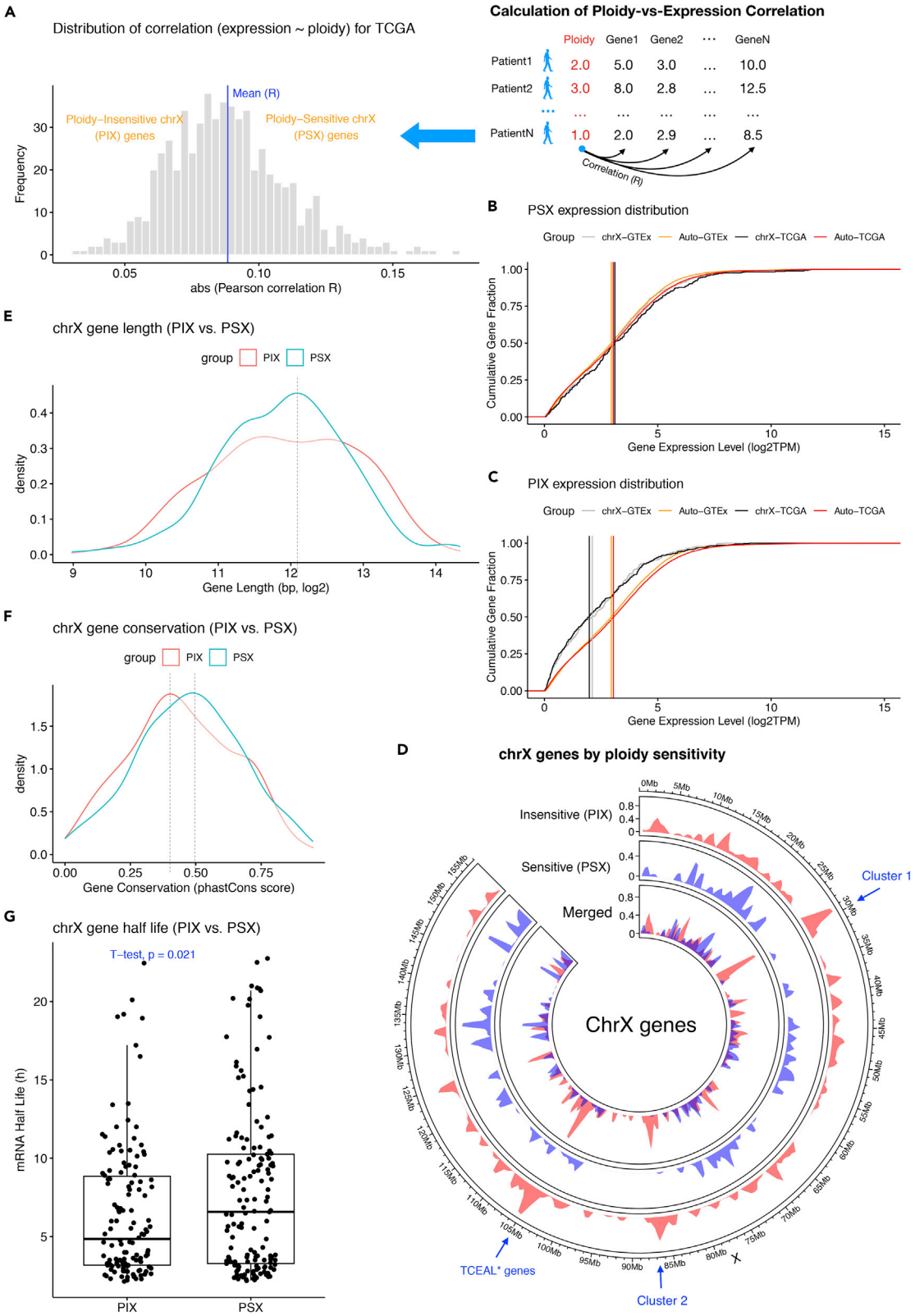

A
Distribution of correlation (expression ~ ploidy) for TCGA
Mean (R)
Ploidy–Insensitive chrX (PIX) genes
Ploidy–Sensitive chrX (PSX) genes
Frequency
abs (Pearson correlation R)
Calculation of Ploidy-vs-Expression Correlation
Ploidy
Gene1
Gene2
GeneN
Patient1
Patient2
PatientN
Correlation (R)
B
PSX expression distribution
Group
chrX–GTEx
Auto–GTEx
chrX–TCGA
Auto–TCGA
Cumulative Gene Fraction
Gene Expression Level (log2TPM)
C
PIX expression distribution
E
chrX gene length (PIX vs. PSX)
group
PIX
PSX
density
Gene Length (bp, log2)
F
chrX gene conservation (PIX vs. PSX)
Gene Conservation (phastCons score)
G
chrX gene half life (PIX vs. PSX)
T–test, p = 0.021
mRNA Half Life (h)
D
chrX genes by ploidy sensitivity
Insensitive (PIX)
Sensitive (PSX)
Merged
ChrX genes
Cluster 1
Cluster 2
TCEAL* genes

Figure 1. Identification and characterization of PSX and PIX genes on X chromosome

(A) Classification X chromosome genes based on absolute values of Pearson correlations (xaxis) compared to the average absolute correlation value (blue vertical line). The illustration of calculation of ploidy-versus-expression correlation was shown on the right.

(B and C) Cumulative distribution of expression levels of PSX (B) and PIX (C) genes on X chromosome, compared to autosomal genes, using samples from the TCGA project or the GTEx project. A total of 100 samples were randomly selected and their median expression levels for genes were shown for visualization. Vertical lines indicate medians.

(D) PIX and PSX genes distribution on X chromosome. PIX enriched regions were marked by arrows.

(E–G) Comparing gene length (E), gene sequence conservation (F), and mRNA half-life (two-sided *t*-test, G; data are represented as median +/− IQR) between PIX and PSX genes. PIX: ploidy-insensitive X chromosome genes, PSX: ploidy-sensitive X chromosome genes. See also Figure S1.

reported that TP53 and ATRX (also known as RAD54) physically interact with each other[18] and work together in the same pathway.[19–21] Both genes are involved in genome-wide histone modifications,[15,22] thus capable of regulating a large number of genes. We therefore hypothesized that TP53 and ATRX probably regulate X dosage compensation.

Of interest, *TP53* mutations contributed to the sex discrepancy of cancer incidences by differentially regulating X-linked genes between two sexes.[23] Notably, *TP53* mutations are the most frequent events across all cancers,[24] and more than 91% of tumors with *TP53* mutations show functional loss of both *TP53* copies.[25] p53 is also implicated in immune regulation of autoimmune diseases.[26] On the other hand, X-linked genes have been implicated in many diseases, including cancer[23] and autoimmune diseases like systemic lupus erythematosus (SLE).[27] Abnormal X dosage has also been reported in both cancer and SLE.[28,29] For example, X upregulation has been contributed to the loss of Barr body (the inactivated X chromosome) in breast and ovarian cancers. X dosage also plays an important role in cancer initiation, because excessive X upregulation by somatic X reactivation under chronic stress leads to cancer.[30] However, the dynamics of X-over-autosome dosage ratios (XAR) remain to be studied systematically across cancers, which accumulates a large number of high-throughput sequencing datasets, for example, The Cancer Genome Atlas (TCGA).[31] More importantly, large numbers of cancer mutations and genome-wide expression profiles from TCGA have the potential to reveal new regulators of XAR, including both protein-coding and long non-coding RNA (lncRNA) genes, to uncover the elusive mechanism of dosage compensation in humans. However, these possibilities have not been explored yet.

In this work, we identified two distinct groups of genes (PIX and PSX) on X chromosome and characterized their dosage compensations in about 20,000 RNA sequencing (RNA-Seq) samples from TCGA and the Geno-type-Tissue Expression (GTEx) project,[32] and thousands of proteomic samples from Clinical Proteomic Tumor Analysis Consortium (CPTAC). We provided a systemic view of both XAR-PIX and XAR-PSX in human. We showed that the differential dosage between PIX and PSX genes were evolutionarily conserved. Of interest, XAR-PIX and XAR-PSX were affected differently in sex-biased diseases and during stem cell differentiation which is closely linked with XCI.[33] Lastly, we identified and experimentally validated two new XAR regulators, *TP53* and *ATRX*. In summary, our work contributed to better understanding of X dosage compensation under different conditions.

RESULTS

Two groups of X chromosome genes with distinct sensitivities to genome ploidy

We are interested in which genes on X chromosome require less dosage compensation, because only a subset of X chromosome genes were affected after *KAT8* knockdown.[5] To this end, we examined X chromosome gene sensitivities to genome ploidy, defined by the absolute Pearson's correlation coefficients between protein-coding gene expression levels and genome ploidies (data downloaded directly from a previous study[34]) of cancer samples from TCGA. After quality control, 615 X chromosome genes (Table S1, see STAR Methods) were used for the expression-versus-ploidy correlation analysis. Based on the approximately normal distribution of these correlations, we identified two groups of X chromosome genes, termed ploidy-sensitive X chromosome (PSX) genes, and ploidy-insensitive X chromosome (PIX) genes (see STAR Methods), separated by the mean of the correlation distribution (Figure 1A). As expected, PSX and PIX genes showed distinct expression distribution. PSX genes (n = 286, Table S1) had highly similar expression pattern to autosome genes (Figure 1B), whereas PIX genes (n = 329, Table S1) had much lower expression levels, with a median of about 2/3 of PSX genes expression (Figure 1C). Of note, no preference for genes escaping X inactivation was observed in PSX or PIX. Of interest, we found three PIX-specific gene clusters on X chromosome, compared to PSX genes (Figure 1D). The PIX cluster located at around 100 Mb

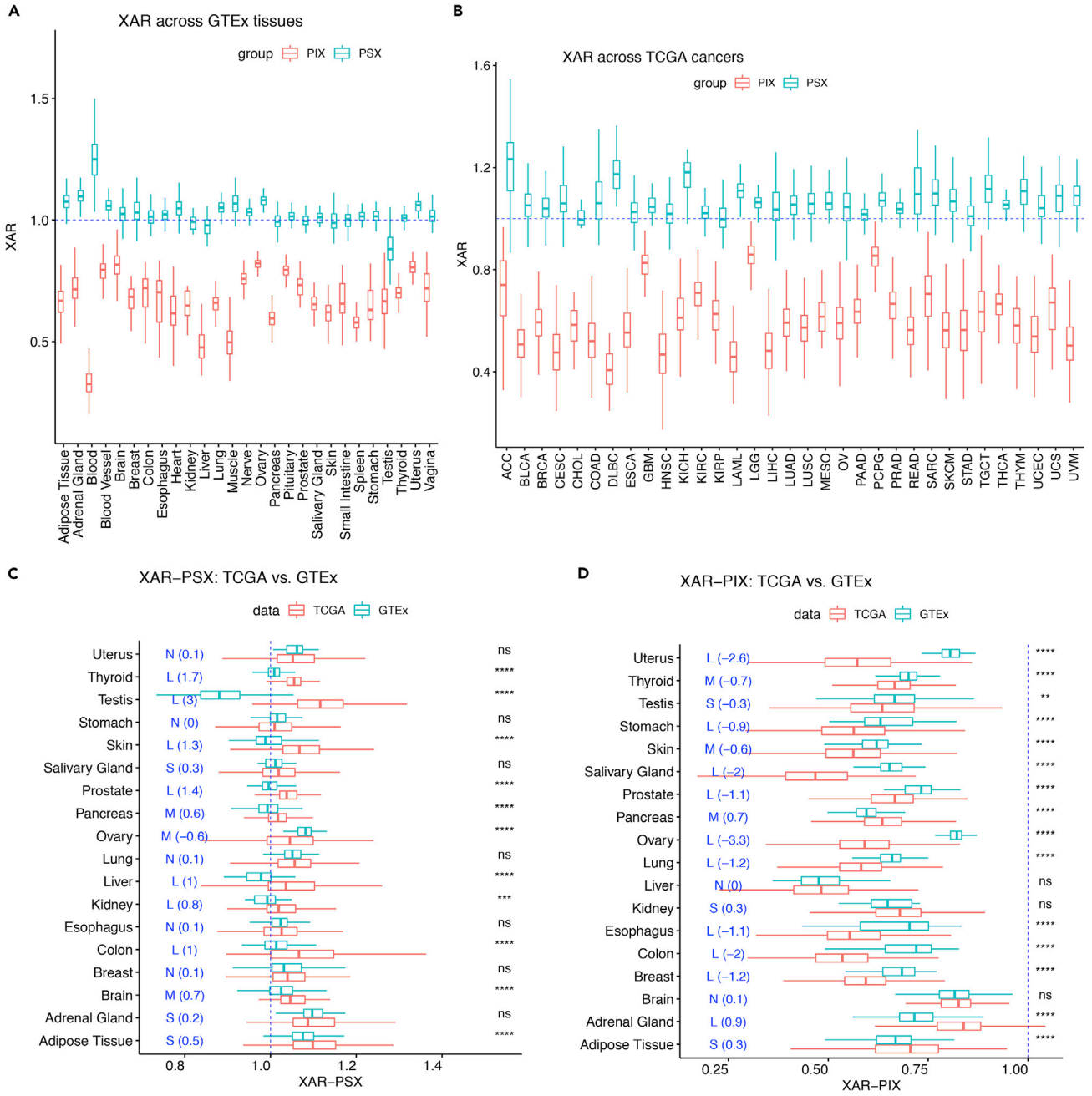


Figure 2. Differential dosages of PSX and PIX genes across normal tissues and cancer types

(A) XAR-PSX and XAR-PIX across 27 normal tissues from GTEx were close to one, except for blood and testis.

(B) XAR-PSX and XAR-PIX across 33 cancer types from TCGA.

(C) XAR-PSX was higher in cancers than in matched normal tissues. Effect sizes were indicated by Cohen'SD statistics in blue (negligible[N]<0.2, 0.2<=small [S]<0.5, 0.5<=moderate[M]<0.8, large[L]≥0.8).

(D) XAR-PIX were lower in cancers than in matched normal tissues. Effect sizes were indicated by Cohen'SD statistics in blue.

(C and D) Two-sided Wilcoxon test p-value significance: ns >0.05, * ≤ 0.05, ** ≤0.01, *** ≤ 0.001, **** ≤ 0.0001. (A–D) Data are represented as median +/− IQR. See also Figures S2 and S3. Matching of tumor types to normal tissues were as follows: ACC, Adrenal Gland; BRCA, Breast; COAD, Colon; ESCA, Esophagus; GBM, Brain; HNSC, Salivary Gland; KICH, Kidney; KIRC, Kidney; KIRP, Kidney; LGG, Brain; LIHC, Liver; LUAD, Lung; LUSC, Lung; OV, Ovary; PAAD, Pancreas; PRAD, Prostate; PCPG, Adrenal Gland; READ, Colon; SARC, Adipose Tissue; SKCM, Skin; STAD, Stomach; TGCT, Testis; THCA, Thyroid; UCS, Uterus; UCEC, Uterus. XAR: X-over-autosome dosage ratio, ACC: Adrenocortical carcinoma, BLCA: Bladder urothelial Carcinoma, BRCA: Breast invasive carcinoma, CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma, CHOL: Cholangiocarcinoma, COAD: Colon adenocarcinoma, CCRCC: Clear cell renal cell carcinoma, DLBC: Lymphoid neoplasm diffuse large B-cell lymphoma, ESCA: Esophageal carcinoma,

Figure 2. *Continued*
GBM: Glioblastoma multiforme, HNSC: Head and neck squamous cell carcinoma, KICH: Kidney chromophobe, KIRC: Kidney renal clear cell carcinoma, KIRP: Kidney renal papillary cell carcinoma, LAML: Acute myeloid leukemia, LGG: Brain lower grade glioma, LIHC: Liver hepatocellular carcinoma, LUAD: Lung adenocarcinoma, LUSC: Lung squamous cell carcinoma, MESO: Mesothelioma, OV: Ovarian serous cystadenocarcinoma, PAAD: Pancreatic adenocarcinoma, PCPG: Pheochromocytoma and paraganglioma, PDA: Pancreatic ductal adenocarcinoma, PRAD: Prostate adenocarcinoma, READ: Rectum adenocarcinoma, SARC: Sarcoma, SKCM: Skin Cutaneous melanoma, STAD: Stomach adenocarcinoma, TGCT: Testicular germ cell tumors, THCA: Thyroid carcinoma, THYM: Thymoma, UCEC: Uterine corpus endometrial carcinoma, UCS: Uterine carcinosarcoma, UVM: Uveal melanoma.

(hg38) contained eight out of nine Transcription Elongation Factor A Protein-Like (TCEAL) genes, encoding nuclear phosphoproteins.

Different features and regulation of PIX and PSX genes

The distinct expression pattern between PIX and PSX genes prompted us to explore their differential genomic features. First, we found that the lengths (excluding introns) of PIX genes spanned a large range from 2 kb to 8 kb, whereas lengths of PSX genes peaked at around 4 kb (Figure 1E). This trend was similar when including introns (Figure S1A). Although GC content of mRNAs is critical for their storage and decay,[35] we found similar GC content between PIX and PSX genes (Figure S1B). Moreover, we observed lower conservation of PIX genes than PSX genes (Figure 1F), which resembled the faster-X effect in non-coding regions on X chromosome.[36] RNA half-lives were reported to be longer for X-linked than autosomal genes.[5] Inspired by this, we examined the mRNA half-lives of PIX and PSX genes and found that PIX (4.9h) tended to have shorter mRNA half-lives than PSX genes (6.6h) (Figure 1G), consistent with lower expression levels of PIX than PSX genes.

Evolutionally conserved differential dosage compensations between PIX and PSX genes

Based on these two groups of X chromosome genes (PIX and PSX), we calculated two dosage ratios of X chromosome overautosome (XAR, see STAR Methods) to evaluate their dosage compensation, termed XAR-PIX and XAR-PSX, respectively. As expected, XAR-PSX dosages were around one across GTEx tissues (Figure 2A), but mostly above one across TCGA cancer types (Figure 2B). In contrast, XAR-PIX dosages were much lower than XAR-PSX and fluctuated around 0.6 both across GTEx tissues and across TCGA cancer types. Furthermore, the difference between XAR-PIX and XAR-PSX was also observed in non-human primates (Figure S2A) and mouse (Figure S2B) across tissues, except for the brain. It suggests that PIX and PSX gene classification and their dosage compensation are evolutionally conserved.

Variations in TCGA cancers were generally larger than GTEx tissues for both XAR-PSX and XAR-PIX, reflecting abnormal dosage regulation across cancers. Surprisingly, across TCGA cancer types, we observed that XAR-PSX dosages were significantly increased (Figure 2C), whereas XAR-PIX showed the opposite trend (Figure 2D), compared to normal tissues from GTEx. Moreover, most magnitudes of changes (or effect sizes) of both XAR-PSX and XAR-PIX were moderate to large, as indicated by the Cohen's d statistics. Of note, most tissues or cancer types showed no XAR difference between males and females, especially for XAR-PIX (Figures S3A–S3D), which suggests that the possibility of XAR-PSX upregulation because of reactivation of the inactivated X in females[37] can be excluded. Accordingly, when investigating proteomic data generated by mass spectrometry, the opposite trend between XAR-PIX and XAR-PSX was also observed when we compared primary tumors with normal tissue samples in CPTAC (Figures S2C and S2D).

XAR-PSX and XAR-PIX were differentially associated with stemness

It is well-known that XAR of stem cells changes dramatically during epigenetic reprogramming of early embryo development. On the other hand, increased stemness is an important feature or cancer cells.[38] We thus examined the relationship between XAR and stemness in cancer. We observed mostly positive correlations with cancer stemness for XAR-PSX, but mostly negative correlations for XAR-PIX, across cancer types (Figure 3A). We further examined XAR in embryonic stem cells (ESC) and induced pluripotent stem cells (iPSC) using gene expression data from the Progenitor Cell Biology Consortium (PCBC). As expected, we found no difference between ESC and iPSC, for both XAR-PIX and XAR-PSX (Figure 3B), probably because ESC and iPSC had similar expression pattern.[39] However, we found much lower XAR-PIX and XAR-PSX in male than in female iPSC (Figure 3C), but the signals of sex difference in ESC were weak.

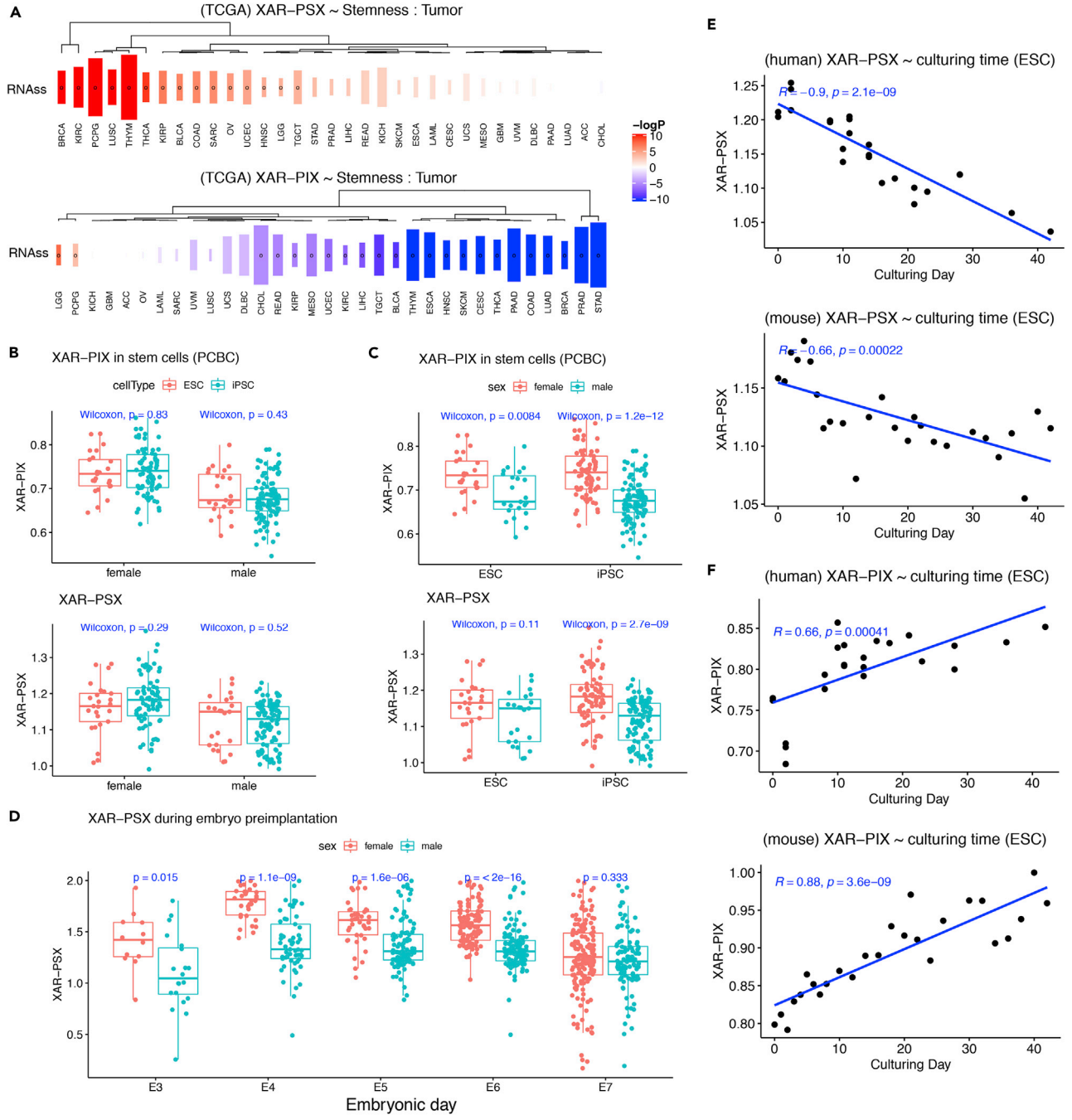


Figure 3. XAR-PSX and XAR-PIX were inversely associated with stemness

(A) Cancer stemness (RNAss) was positively correlated with XAR-PSX (top) and negatively with XAR-PIX (bottom). Colors indicate-log10 p-values, and sizes of rectangles represent Pearson correlation. Correlations with FDR<0.1 were marked by circles.

(B) Comparison of XAR-PIX (top) and XAR-PSX (bottom) between ESCs and iPSCs from females or males.

(C) Comparison of XAR-PIX (top) and XAR-PSX (bottom) between sexes from ESCs or iPSCs.

(D) XAR-PSX dynamics of embryonic stem cells (grouped by sex) during preimplantation from day 3 (E3) to 7 (E7).

(E) Strong negative correlation of XAR-PSX with differentiation culturing time in both human (top) and mouse (bottom).

(F) Strong positive correlation of XAR-PIX with differentiation culturing time in both human (top) and mouse (bottom).

(B–D) Two-sided Wilcoxon test. Data are represented as median +/– IQR.

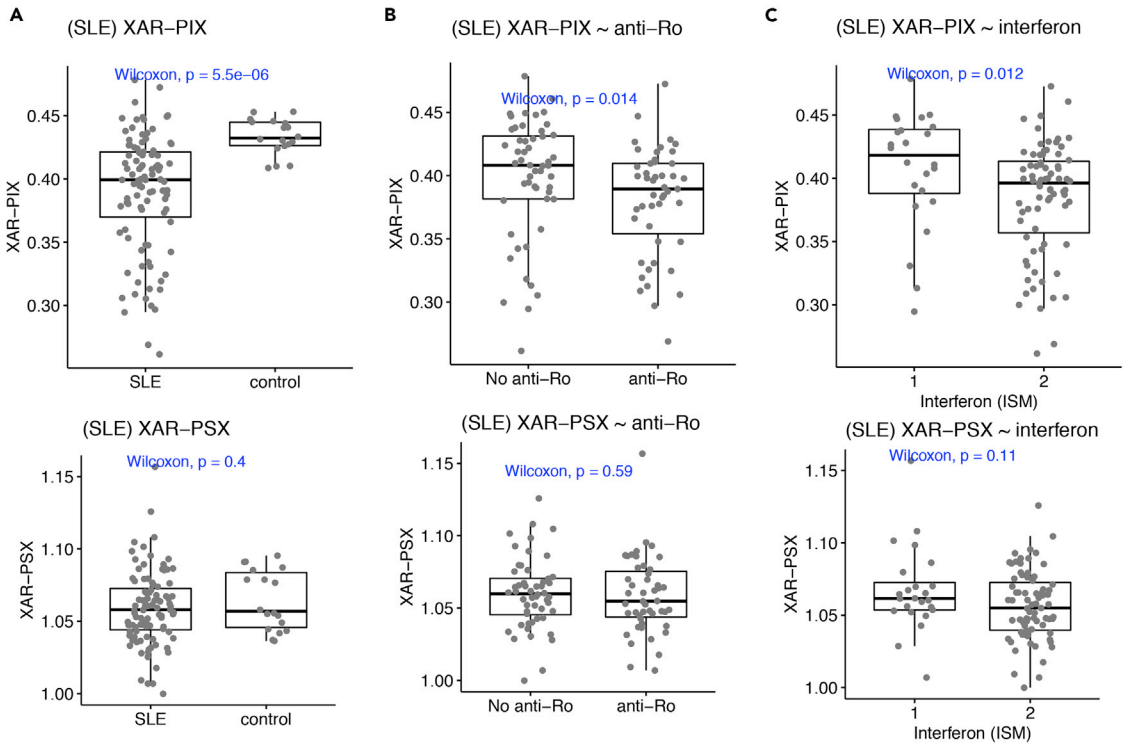


Figure 4. XAR in autoimmune disease

(A) XAR-PIX (top) and XAR-PSX (bottom) in SLE patients compared to healthy controls.

(B) XAR-PIX (top) and XAR-PSX (bottom) in SLE patients with anti-Ro autoantibodies compared to those without anti-Ro.

(C) XAR-PIX (top) and XAR-PSX (bottom) in SLE patients with more active interferon response (ISM) compared to those with less ISM.

(A–C) Two-sided Wilcoxon test.Data are represented as median +/− IQR.

We noticed that XAR-PSX was between 1.1 and 1.2 in both ESCs and iPSCs, although ESCs and some iPSCs have two active X chromosomes.[40] This may be related to early embryonic development, during which XAR-PSX has peaked (as expected, close to 1.9 in female and 1.4 in male) at embryo day 4 (E4) and decreased dramatically to be equal in both sexes (around 1.2) (Figure 3D), which was consistent with our observations of XAR-PSX in ESC and iPSC cell lines.

We further examined XAR dynamics during iPSC differentiation culturing. Of interest, we found that XAR-PSX was reduced along iPSC culturing time (Figure 3E), but XAR-PIX showed the opposite trend (Figure 3F), in both human and mouse. These patterns were consistent with overall reduction of XAR-PIX and elevation of XAR-PSX across cancer types above.

Insufficient PIX gene dosage in autoimmune patients

Both altered X-inactivation and X-linked risk genes were associated with SLE, we thus investigated which type of XAR (XAR-PIX or XAR-PSX) was dysregulated in SLE. By reanalyzing a previous dataset from SLE patients and healthy controls,[41] we found that XAR-PIX, but not XAR-PSX, were specifically down-regulated in SLE patients compared with healthy controls (Figure 4A), which means insufficient dosage of X genes in SLE patients. Moreover, SLE patients with anti-Ro autoantibodies showed lower XAR-PIX dosage than those without anti-Ro autoantibodies (Figure 4B). Similarly, SLE patients with more active interferon response (indicated by the Interferon Signature Metric, or ISM) showed lower XAR-PIX dosage (Figure 4C). These findings may provide insights to understand SLE pathogenesis.

Potential lncRNA regulators of XAR

LncRNAs regulate gene expression at multiple levels, including chromatin remodeling, transcription and post-transcription regulation.[42] One of the best-known examples of gene expression regulation by lncRNA is X inactive-specific transcript (*XIST*)-mediated X inactivation in females.[43] Considering the important roles

of lncRNAs in gene expression regulation, we speculated that there were other lncRNAs related to dosage compensation to be explored. Therefore, we correlated lncRNA with XAR in each cancer in male, and obtained 2,689 and 31 lncRNAs significantly related to XAR-PIX and XAR-PSX, respectively. Next, we focused on the co-expressed genes and RNA-protein interactions of these lncRNAs, which might provide clues for how these lncRNAs affect dosage compensation in tumors. As a result, co-expressed genes of these lncRNAs were most enriched in histone modification (Figure S4A), suggesting that lncRNAs might involve in dosage compensation by regulating genes related to histone modification. In addition, genes functioning in nuclear transport, chromatin organization, and mRNA metabolic processes were enriched. In terms of predicted interactions with XAR-associated lncRNAs, proteins involved in RNA processing and stability as well as translational regulation were enriched (Table S2).

TP53 and ATRX are new regulators of X dosage compensation

We found that only about 8.8% (29/329) and 10.5% (30/286) of PIX and PSX genes showed expression changes of more than 1.2-fold after knockdown of the key MSL component *KAT8* in mouse.[5] To identify new regulators of XAR, we examined TCGA mutation spectra that may be associated with XAR in each cancer type. We reasoned that mutated XAR regulators should result in differential XAR compared with wild-type regulators. In total, we obtained 209 and 515 potential regulators (FDR <0.5) for XAR-PIX and XAR-PSX in male tumors, respectively (Figures 5A and S4B). For the top 200 regulators, gene ontology (GO) analysis was performed and similar GO terms were found for XAR-PIX and XAR-PSX regulators, mainly about post-translational modifications, including histone methylation, peptidyl-lysine methylation, and protein methylation (Figures S4C and S4D).

Among the potentially new regulators of XAR, we found that *TP53* and *ATRX* were promising regulators for XAR-PIX in both males and females (Figures 5A and S4E), based on previous literature mentioned above. They are both genome-wide chromatin remodelers, capable of regulating a large number of genes. We chose them for further investigation also for the following two reasons. We found that the top 200 XAR-PIX dependent genes formed a PPI network mostly centered on TP53 and ATRX (Figure S5). TP53 was also a center node in the PPI network (Figure S4F) with RNA-binding proteins (RBPs) associated with XAR-associated lncRNAs.

Next, we constructed *TP53*-knockout (*TP53*-KO) cells from two *TP53*-wildtype (*TP53*-WT) cell lines, one male (HT1080) and one female (MCF7, containing three active X and no inactive X chromosomes[44]), by using CRISPR technologies (Figure S4G, see STAR Methods), followed by RNA-seq. We observed XAR-PIX increase after *TP53*-KO in both sexes (Figures 5B and 5C), in contrast to the unchanged XAR-PSX (Figures 5D and 5E). We further demonstrated that *TP53* may partly regulate XAR-PIX by repressing KAT8, because *TP53*-KO enhanced *KAT8* expression in both sexes (Figure 5F). *TP53*-KO had a larger effect in female cells (8% vs. 6% increase of XAR-PIX) than male cells, which may because of MCF7 cells containing three active X chromosomes and *TP53* affected *MSL1* in female cells only.

Finally, we examined the role of ATRX in XAR. We investigated XAR before and after *ATRX* knockout in two studies. Of interest, ATRX-KO reduced both XAR-PIX and XAR-PSX in mouse iPSCs (Figure 5G) and GBM cells (Figure 5H). Similarly, we examined MSL expression changes after ATRX-KO, we observed consistent down-regulation of both MSL1 and MSL2 in either GBM or iPSC cells in mice, although KAT8 was found up-regulated in mouse iPSC cells whereas it was unchanged in GBM cells (Figure S6A). To further support *ATRX* as a new regulator of XAR, we examined whether X-linked genes are more sensitive to the change of *ATRX* expression. For autosomal and X-linked genes co-expressed with *ATRX*, we compared their coefficient estimates of *ATRX* expression by linear regression models. We reasoned that larger absolute values of ATRX coefficients indicate higher impact on gene expression. As a result, *ATRX* regression coefficients were significantly higher in X-linked genes than in autosomal genes across a variety of tumors, at both mRNA (Figure S6B) and protein level (Figure S6C).

DISCUSSIONS

Here, we identified and comprehensively characterized two distinct groups of genes on the X chromosome. Integrating large-scale sequencing data from both cancer and normal samples greatly enhanced our ability to investigate both pathological and physiological roles of XAR. Our findings may help resolve XAR controversies described in our previous study.[6] We summarized our findings in Figure 6. Our results showed that dosages of PIX and PSX genes were converging during stem cell differentiation but diverging in cancer

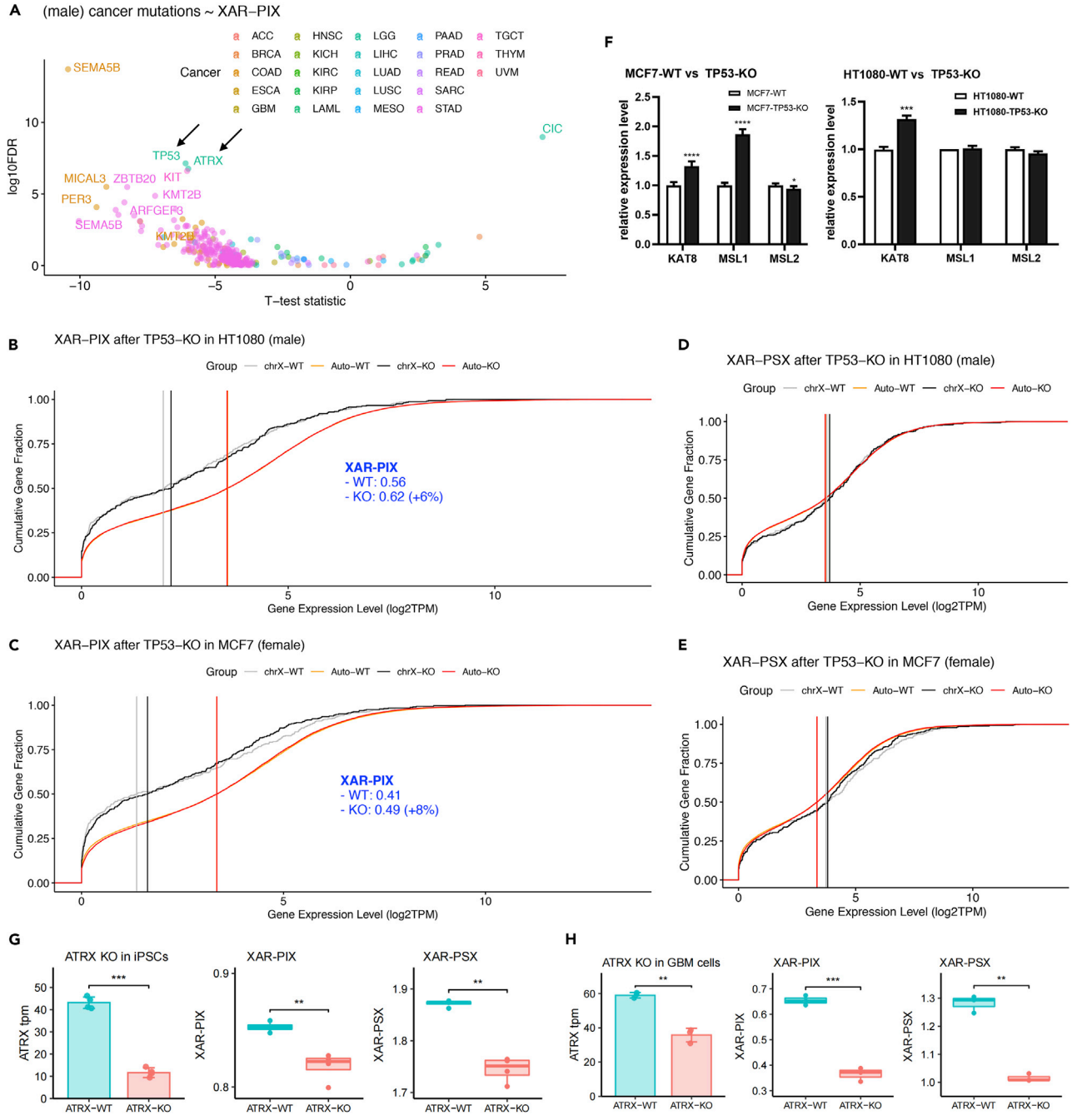


Figure 5. *TP53* and *ATRX* were two regulators of XAR

(A) Significance of XAR-PIX difference (yaxis, two-sided *t*-test) between wildtype and mutated genes across cancers in male. *TP53* and *ATRX* were marked by arrows. For better visualization, only genes with FDR<1 were shown as points and those with FDR<1e-4 labeled.

(B and C) Cumulative distribution of PIX and autosomal gene expression levels for HT1080 (B) and MCF7 (C), before and after *TP53*-KO. Vertical lines indicate median expression used to calculate XAR-PIX dosages.

(D and E) Cumulative distribution of PSX and autosomal gene expression levels for HT1080 (D) and MCF7 (E) before and after *TP53*-KO. Vertical lines indicate median expression used to calculate XAR-PSX dosages.

(F) Gene expression changes of MSL complex subunits *KAT8*, *MSL1*, and *MSL2* in MCF7 (left) and HT1080 (right) cells, before and after *TP53*-KO. Data are represented as mean ±SD.

(G) *ATRX* knockout (KO) in mouse induced pluripotent stem cells (iPSCs) significantly reduced both XAR-PIX (middle) and XAR-PSX (right).

(H) *ATRX* knockout (KO) in human GBM cells significantly reduced both XAR-PIX (middle) and XAR-PSX (right).

(G and H) Data are represented as mean ± SD (barplot) or median +/− IQR (boxplot).

(F–H) Two-sided *t*-test, *P<= 0.05, ** ≤0.01, *** ≤ 0.001, **** ≤ 0.0001. See also Figures S4–S6.

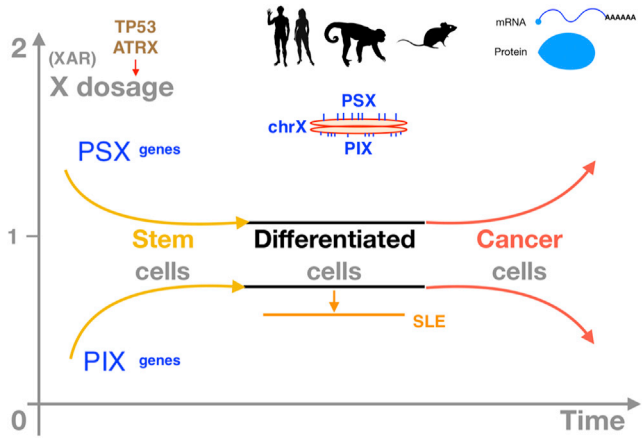


Figure 6. Major findings of this study

Differential dosages of two groups of X chromosome genes (PIX and PSX) were observed in human and in animals at both mRNA and protein levels. PIX and PSX dosages were converging during stem cell differentiation but diverging in cancer. PIX dosage was lower in blood cells of SLE autoimmune patients than in healthy individuals. TP53 and ATRX were identified and validated as two XAR regulators. XAR: X-over-autosome dosage ratio, PIX: ploidy-insensitive X genes, PSX: ploidy-sensitive X genes.

cells. However, in SLE patients, only dosages of PIX genes, but not PSX genes, were down-regulated compared to healthy individuals. Finally, TP53 and ATRX were identified and validated as two XAR regulators.

Although most genes from the TCEAL family, located within one of PIX gene clusters, were not intensively studied, most reports suggested that they acted as tumor suppressors, especially for *TCEAL7*.[45,46] For these tumor suppressors, lower sensitivity to dosage may protect both males and females from cancer equally. Of interest, the chrX region around 100 Mb showed the lowest expression level after zygote genome activation (ZGA) at E4 during embryo implantation.[47]

Of note, *TP53* mRNA itself is a critical regulator of cellular stress.[48] A previous study found that the deadenylase poly(A)-specific ribonuclease (PARN) destabilized *TP53* mRNA whereas p53 increase activated PARN,[49] indicating a negative feedback loop between *TP53* mRNA and p53 protein levels. We observed that *TP53* mRNA was universally up-regulated in cancer samples, compared with non-cancerous samples, across cancer types (Figure S7). Moreover, XAR-PIX was lower in tumors than non-cancerous samples across cancer types. These results were consistent with observations that p53 mutations (or higher *TP53* mRNA) were associated with lower XAR-PIX in cancer and *TP53*-KO resulted in XAR-PIX increase.

The frequency of transcriptional bursts and possible *trans*-acting enhancer-binding regulators were proposed to be the driving force behind X upregulation in mouse.[4] Of interest, p53 can bind pervasively to genome-wide enhancers[50] and regulate the transcriptional burst of target genes,[51]further supporting our finding of *TP53* regulating XAR. Moreover, the crosstalk between *TP53* and nonsense-mediated decay (NMD) pathway may contribute to XAR regulation by *TP53*.[5,52] Therefore, *TP53* may regulate XAR by various mechanisms. We expect XAR regulation by *TP53* would be conserved among mammals, because of ultra-conservation of *TP53* during evolution.[8] However, further efforts are needed to support our hypothesis.

XAR in most cancers showed no difference between sexes, suggesting negligible contribution to XAR changes in cancer by reactivation of the inactivated X in females. Because *TP53* mutation-induced aneuploidies were not specific to X chromosome, they were unlikely to drive both XAR-PSX and XAR-PIX to change in such a coordinated way across cancers, as observed in this study. At least, XAR-PIX should be less affected by aneuploidies, by PIX definition.

Limitations of the study

Several limitations of this study represent directions for future research. First, it is unknown how distinct dosages of PIX and PSX genes were maintained and why did they show divergent trends in cancers compared to normal tissues. Second, despite the high cost and long period of CRISPR-based *TP53*-KO experiments, it would further solidify our conclusion if CRISPR-mediated *TP53*-KO can be established in more cell lines of different lineages. Third, additional XAR regulators remain to be identified by other ways, because this analysis was limited by mutation frequencies in cancer patients. Genome-wide screening experiments in the future probably will identify more XAR regulators. Until then, the problem of abnormal XAR-PIX and XAR-PSX in cancer may be solved. Lastly, it would be very interesting to further explore abnormal XAR for both PIX and PSX in more diseases, including other types of autoimmune diseases like rheumatoid arthritis and inflammatory neurodegenerative diseases.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
- METHOD DETAILS
  - Data collection and preprocessing
  - CRISPR-Cas9 knockout
  - Western blotting
  - RNA extraction and real-time quantitative PCR
  - RNA sequencing of MCF7 and HT1080 cell lines
  - Defining PIX, PSX, and autosomal genes for dosage compensation estimation
  - Comparison of genomic features for PIX and PSX genes
  - Calculation of XAR-PSX and XAR-PIX
  - Association of XAR with somatic mutations
  - Gene ontology analysis
  - ATRX regression coefficients comparison between autosomal and X-linked genes
  - Correlations for lncRNAs and XAR in tumors
  - Data visualization
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.105997.

ACKNOWLEDGMENTS

The work was supported by National Natural Science Foundation of China (NSFC) (Grant number 92249303) and Guangdong Basic and Applied Basic Research Foundation (2021A1515110972). We would like to thank Prof. Wanling Yang for providing helpful writing advice. The authors would like to thank TCGA, GTEx, PCBC, NHPRTR, and other projects for making their research data publicly available.

AUTHOR CONTRIBUTIONS

M.B.G. and Y.Y.X. conceived the project. M.B.G. and Z.W.F. collected the data, performed the analysis, and interpreted the results. M.B.G. wrote the manuscript with input from Z.W.F. Z.W.F. and B.H.C. performed the experimental validation. Z.S.Y. was involved in project discussion. Y.Y.X. was involved in critical discussions and reviewed the manuscript. All authors approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: July 27, 2022
Revised: December 12, 2022
Accepted: January 12, 2023
Published: January 18, 2023

REFERENCES

1. Disteche, C.M. (2012). Dosage compensation of the sex chromosomes. Annu. Rev. Genet. *46*, 537–560. https://doi.org/10.1146/annurev-genet-110711-155454.

2. Dunford, A., Weinstock, D.M., Savova, V., Schumacher, S.E., Cleary, J.P., Yoda, A., Sullivan, T.J., Hess, J.M., Gimelbrant, A.A., Beroukhim, R., et al. (2017). Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat. Genet. *49*, 10–16. https://doi.org/10.1038/ng.3726.

3. Vallot, C., Ouimette, J.F., and Rougeulle, C. (2016). Establishment of X chromosome inactivation and epigenomic features of the inactive X depend on cellular contexts. Bioessays *38*, 869–880. https://doi.org/10.1002/bies.201600121.

4. Larsson, A.J.M., Coucoravas, C., Sandberg, R., and Reinius, B. (2019). X-chromosome upregulation is driven by increased burst frequency. Nat. Struct. Mol. Biol. *26*, 963–969. https://doi.org/10.1038/s41594-019-0306-y.

5. Deng, X., Berletch, J.B., Ma, W., Nguyen, D.K., Hiatt, J.B., Noble, W.S., Shendure, J., and Disteche, C.M. (2013). Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. Dev. Cell *25*, 55–68. https://doi.org/10.1016/j.devcel.2013.01.028.

6. 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. Nat. Genet. *42*, 1043–1047. https://doi.org/10.1038/ng.711.

7. Julien, P., Brawand, D., Soumillon, M., Necsulea, A., Liechti, A., Schütz, F., Daish, T., Grützner, F., and Kaessmann, H. (2012). Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. PLoS Biol. *10*, e1001328. https://doi.org/10.1371/journal.pbio.1001328.

8. Lu, W.J., Amatruda, J.F., and Abrams, J.M. (2009). p53 ancestry: gazing through an evolutionary lens. Nat. Rev. Cancer *9*, 758–762. https://doi.org/10.1038/nrc2732.

9. Kruse, J.P., and Gu, W. (2009). MSL2 promotes Mdm2-independent cytoplasmic localization of p53. J. Biol. Chem. *284*, 3250–3263. https://doi.org/10.1074/jbc.M805658200.

10. Li, X., Wu, L., Corsa, C.A.S., Kunkel, S., and Dou, Y. (2009). Two mammalian MOF complexes regulate transcription activation by distinct mechanisms. Mol. Cell *36*, 290–301. https://doi.org/10.1016/j.molcel.2009.07.031.

11. Nguyen, T.A.T., Grimm, S.A., Bushel, P.R., Li, J., Li, Y., Bennett, B.D., Lavender, C.A., Ward, J.M., Fargo, D.C., Anderson, C.W., et al. (2018). Revealing a human p53 universe. Nucleic Acids Res. *46*, 8153–8167. https://doi.org/10.1093/nar/gky720.

12. He, Q., Kim, H., Huang, R., Lu, W., Tang, M., Shi, F., Yang, D., Zhang, X., Huang, J., Liu, D., and Songyang, Z. (2015). The daxx/atrx complex protects tandem repetitive elements during DNA hypomethylation by promoting H3K9 trimethylation. Cell Stem Cell *17*, 273–286. https://doi.org/10.1016/j.stem.2015.07.022.

13. Hu, Y., Shi, G., Zhang, L., Li, F., Jiang, Y., Jiang, S., Ma, W., Zhao, Y., Songyang, Z., and Huang, J. (2016). Switch telomerase to ALT mechanism by inducing telomeric DNA damages and dysfunction of ATRX and DAXX. Sci. Rep. *6*, 32280. https://doi.org/10.1038/srep32280.

14. Teng, Y.C., Sundaresan, A., O'Hara, R., Gant, V.U., Li, M., Martire, S., Warshaw, J.N., Basu, A., and Banaszynski, L.A. (2021). ATRX promotes heterochromatin formation to protect cells from G-quadruplex DNA-mediated stress. Nat. Commun. *12*, 3887. https://doi.org/10.1038/s41467-021-24206-5.

15. Dyer, M.A., Qadeer, Z.A., Valle-Garcia, D., and Bernstein, E. (2017). ATRX and DAXX: mechanisms and mutations. Cold Spring Harb.Perspect.Med. *7*, a026567. https://doi.org/10.1101/cshperspect.a026567.

16. Ren, W., Medeiros, N., Warneford-Thomson, R., Wulfridge, P., Yan, Q., Bian, J., Sidoli, S., Garcia, B.A., Skordalakes, E., Joyce, E., et al. (2020). Disruption of ATRX-RNA interactions uncovers roles in ATRX localization and PRC2 function. Nat. Commun. *11*, 2219. https://doi.org/10.1038/s41467-020-15902-9.

17. Garrick, D., Sharpe, J.A., Arkell, R., Dobbie, L., Smith, A.J.H., Wood, W.G., Higgs, D.R., and Gibbons, R.J. (2006). Loss of Atrx affects trophoblast development and the pattern of X-inactivation in extraembryonic tissues. PLoS Genet. *2*, e58. https://doi.org/10.1371/journal.pgen.0020058.

18. Linke, S.P., Sengupta, S., Khabie, N., Jeffries, B.A., Buchhop, S., Miska, S., Henning, W., Pedeux, R., Wang, X.W., Hofseth, L.J., et al. (2003). p53 interacts with hRAD51 and hRAD54, and directly modulates homologous recombination. Cancer Res. *63*, 2596–2605.

19. Oppel, F., Tao, T., Shi, H., Ross, K.N., Zimmerman, M.W., He, S., Tong, G., Aster, J.C., and Look, A.T. (2019). Loss of atrx cooperates with p53-deficiency to promote the development of sarcomas and other malignancies. PLoS Genet. *15*, e1008039. https://doi.org/10.1371/journal.pgen.1008039.

20. Seah, C., Levy, M.A., Jiang, Y., Mokhtarzada, S., Higgs, D.R., Gibbons, R.J., and Bérubé, N.G. (2008). Neuronal death resulting from targeted disruption of the Snf2 protein ATRX is mediated by p53. J. Neurosci. *28*, 12570–12580. https://doi.org/10.1523/JNEUROSCI.4048-08.2008.

21. Gulve, N., Su, C., Deng, Z., Soldan, S.S., Vladimirova, O., Wickramasinghe, J., Zheng, H., Kossenkov, A.V., and Lieberman, P.M. (2022). DAXX-ATRX regulation of p53 chromatin binding and DNA damage response. Nat. Commun. *13*, 5033. https://doi.org/10.1038/s41467-022-32680-8.

22. Allison, S.J., and Milner, J. (2004). Remodelling chromatin on a global scale: a novel protective function of p53. Carcinogenesis *25*, 1551–1557. https://doi.org/10.1093/carcin/bgh212.

23. Haupt, S., Caramia, F., Herschtal, A., Soussi, T., Lozano, G., Chen, H., Liang, H., Speed, T.P., and Haupt, Y. (2019). Identification of cancer sex-disparity in the functional integrity of p53 and its X chromosome network. Nat. Commun. *10*, 5385. https://doi.org/10.1038/s41467-019-13266-3.

24. Hainaut, P., and Pfeifer, G.P. (2016). Somatic TP53 mutations in the era of genome sequencing. Cold Spring Harb.Perspect.Med. *6*, a026179. https://doi.org/10.1101/cshperspect.a026179.

25. Donehower, L.A., Soussi, T., Korkut, A., Liu, Y., Schultz, A., Cardenas, M., Li, X., Babur, O., Hsu, T.K., Lichtarge, O., et al. (2019). Integrated analysis of TP53 gene and pathway alterations in the cancer genome Atlas. Cell Rep. *28*, 3010. https://doi.org/10.1016/j.celrep.2019.08.061.

26. Muñoz-Fontela, C., Mandinova, A., Aaronson, S.A., and Lee, S.W. (2016). Emerging roles of p53 and other tumour-suppressor genes in immune regulation. Nat. Rev. Immunol. *16*, 741–750. https://doi.org/10.1038/nri.2016.99.

27. Hewagama, A., Gorelik, G., Patel, D., Liyanarachchi, P., McCune, W.J., Somers, E., Gonzalez-Rivera, T., Michigan Lupus Cohort, Strickland, F., and Richardson, B. (2013). Overexpression of X-linked genes in T cells

from women with lupus. J. Autoimmun. *41*, 60–71. https://doi.org/10.1016/j.jaut.2012.12.006.

28. Pageau, G.J., Hall, L.L., Ganesan, S., Livingston, D.M., and Lawrence, J.B. (2007). The disappearing Barr body in breast and ovarian cancers. Nat. Rev. Cancer *7*, 628–633. https://doi.org/10.1038/nrc2172.

29. Syrett, C.M., Paneru, B., Sandoval-Heglund, D., Wang, J., Banerjee, S., Sindhava, V., Behrens, E.M., Atchison, M., and Anguera, M.C. (2019). Altered X-chromosome inactivation in T cells may promote sex-biased autoimmune diseases. JCI Insight *4*, e126751. https://doi.org/10.1172/jci.insight.126751.

30. Yang, L., Yildirim, E., Kirby, J.E., Press, W., and Lee, J.T. (2020). Widespread organ tolerance to Xist loss and X reactivation except under chronic stress in the gut. Proc. Natl. Acad. Sci. USA *117*, 4262–4272. https://doi.org/10.1073/pnas.1917203117.

31. Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell *155*, 948–962. https://doi.org/10.1016/j.cell.2013.10.011.

32. GTEx Consortium (2013). The genotype-tissue expression (GTEx) project. Nat. Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.

33. Hall, L.L., Byron, M., Butler, J., Becker, K.A., Nelson, A., Amit, M., Itskovitz-Eldor, J., Stein, J., Stein, G., Ware, C., and Lawrence, J.B. (2008). X-inactivation reveals epigenetic anomalies in most hESC but identifies sublines that initiate as expected. J. Cell. Physiol. *216*, 445–452. https://doi.org/10.1002/jcp.21411.

34. Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., Liu, J., et al. (2018). Genomic and functional approaches to understanding cancer aneuploidy. Cancer Cell *33*, 676–689.e3. https://doi.org/10.1016/j.ccell.2018.03.007.

35. Courel, M., Clément, Y., Bossevain, C., Foretek, D., Vidal Cruchez, O., Yi, Z., Bénard, M., Benassy, M.N., Kress, M., Vindry, C., et al. (2019). GC content shapes mRNA storage and decay in human cells. Elife *8*, e49708. https://doi.org/10.7554/eLife.49708.

36. Meisel, R.P., and Connallon, T. (2013). The faster-X effect: integrating theory and data. Trends Genet. *29*, 537–544. https://doi.org/10.1016/j.tig.2013.05.009.

37. Chaligné, R., Popova, T., Mendoza-Parra, M.A., Saleem, M.A.M., Gentien, D., Ban, K., Piolot, T., Leroy, O., Mariani, O., Gronemeyer, H., et al. (2015). The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. Genome Res. *25*, 488–503. https://doi.org/10.1101/gr.185926.114.

38. Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell *173*, 338–354.e15. https://doi.org/10.1016/j.cell.2018.03.034.

39. Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439–452. https://doi.org/10.1016/j.cell.2010.12.032.

40. Bar, S., Seaton, L.R., Weissbein, U., Eldar-Geva, T., and Benvenisty, N. (2019). Global characterization of X chromosome inactivation in human pluripotent stem cells. Cell Rep. *27*, 20–29.e3. https://doi.org/10.1016/j.celrep.2019.03.019.

41. Hung, T., Pratt, G.A., Sundararaman, B., Townsend, M.J., Chaivorapol, C., Bhangale, T., Graham, R.R., Ortmann, W., Criswell, L.A., Yeo, G.W., and Behrens, T.W. (2015). The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. Science *350*, 455–459. https://doi.org/10.1126/science.aac7442.

42. Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat. Rev. Mol. Cell Biol. *22*, 96–118. https://doi.org/10.1038/s41580-020-00315-9.

43. Loda, A., Collombet, S., and Heard, E. (2022). Gene regulation in time and space during X-chromosome inactivation. Nat. Rev. Mol. Cell Biol. *23*, 231–249. https://doi.org/10.1038/s41580-021-00438-7.

44. Sirchia, S.M., Ramoscelli, L., Grati, F.R., Barbera, F., Coradini, D., Rossella, F., Porta, G., Lesma, E., Ruggeri, A., Radice, P., et al. (2005). Loss of the inactive X chromosome and replication of the active X in BRCA1-defective and wild-type breast cancer cells. Cancer Res. *65*, 2139–2146. https://doi.org/10.1158/0008-5472.CAN-04-3465.

45. Yue, X., Lan, F., and Xia, T. (2019). Hypoxic glioma cell-secreted exosomal miR-301a activates wnt/beta-catenin signaling and promotes radiation resistance by targeting TCEAL7. Mol. Ther. *27*, 1939–1949. https://doi.org/10.1016/j.ymthe.2019.07.011.

46. Lafferty-Whyte, K., Bilsland, A., Hoare, S.F., Burns, S., Zaffaroni, N., Cairney, C.J., and Keith, W.N. (2010). TCEAL7 inhibition of c-Myc activity in alternative lengthening of telomeres regulates hTERT expression. Neoplasia *12*, 405–414. https://doi.org/10.1593/neo.10180.

47. Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell *165*, 1012–1026. https://doi.org/10.1016/j.cell.2016.03.023.

48. Haronikova, L., Olivares-Illana, V., Wang, L., Karakostis, K., Chen, S., and Fåhraeus, R. (2019). The p53 mRNA: an integral part of the cellular stress response. Nucleic Acids Res. *47*, 3257–3271. https://doi.org/10.1093/nar/gkz124.

49. Devany, E., Zhang, X., Park, J.Y., Tian, B., and Kleiman, F.E. (2013). Positive and negative feedback loops in the p53 and mRNA 3′ processing pathways. Proc. Natl. Acad. Sci. USA *110*, 3351–3356. https://doi.org/10.1073/pnas.1212533110.

50. Younger, S.T., and Rinn, J.L. (2017). p53 regulates enhancer accessibility and activity in response to DNA damage. Nucleic Acids Res. *45*, 9889–9900. https://doi.org/10.1093/nar/gkx577.

51. Friedrich, D., Friedel, L., Finzel, A., Herrmann, A., Preibisch, S., and Loewer, A. (2019). Stochastic transcription in the p53-mediated response to DNA damage is modulated by burst frequency. Mol. Syst. Biol. *15*, e9068. https://doi.org/10.15252/msb.20199068.

52. Gewandter, J.S., Bambara, R.A., and O'Reilly, M.A. (2011). The RNA surveillance protein SMG1 activates p53 in response to DNA double-strand breaks but not exogenously oxidized mRNA. Cell Cycle *10*, 2561–2567. https://doi.org/10.4161/cc.10.15.16347.

53. Salomonis, N., Dexheimer, P.J., Omberg, L., Schroll, R., Bush, S., Huo, J., Schriml, L., Ho Sui, S., Keddache, M., Mayhew, C., et al. (2016). Integrated genomic analysis of diverse induced pluripotent stem cells from the progenitor cell Biology Consortium. Stem Cell Rep. *7*, 110–125. https://doi.org/10.1016/j.stemcr.2016.05.006.

54. Barry, C., Schmitz, M.T., Jiang, P., Schwartz, M.P., Duffin, B.M., Swanson, S., Bacher, R., Bolin, J.M., Elwell, A.L., McIntosh, B.E., et al. (2017). Species-specific developmental timing is maintained by pluripotent stem cells ex utero. Dev. Biol. *423*, 101–110. https://doi.org/10.1016/j.ydbio.2017.02.002.

55. Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res. *22*, 947–956. https://doi.org/10.1101/gr.130559.111.

56. Peng, X., Thierry-Mieg, J., Thierry-Mieg, D., Nishida, A., Pipes, L., Bozinoski, M., Thomas, M.J., Kelly, S., Weiss, J.M., Raveendran, M., et al. (2015). Tissue-specific transcriptome sequencing analysis expands the non-human primate reference transcriptome resource (NHPRTR). Nucleic Acids Res. *43*, D737–D742. https://doi.org/10.1093/nar/gku1110.

57. Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y., and Shi, L. (2017). A comprehensive mouse transcriptomic BodyMap across 17 tissues by RNA-seq. Sci. Rep. *7*, 4200. https://doi.org/10.1038/s41598-017-04520-z.

58. Qin, T., Mullan, B., Ravindran, R., Messinger, D., Siada, R., Cummings, J.R., Harris, M., Muruganand, A., Pyaram, K., Miklja, Z., et al.

(2022). ATRX loss in glioma results in dysregulation of cell-cycle phase transition and ATM inhibitor radio-sensitization. Cell Rep. *38*, 110216. https://doi.org/10.1016/j.celrep.2021.110216.

59. Deneault, E., White, S.H., Rodrigues, D.C., Ross, P.J., Faheem, M., Zaslavsky, K., Wang, Z., Alexandrova, R., Pellecchia, G., Wei, W., et al. (2018). Complete disruption of autism-susceptibility genes by gene editing predominantly reduces functional connectivity of isogenic human neurons. Stem Cell Rep. *11*, 1211–1225. https://doi.org/10.1016/j.stemcr.2018.10.003.

60. Jiang, S., Cheng, S.J., Ren, L.C., Wang, Q., Kang, Y.J., Ding, Y., Hou, M., Yang, X.X., Lin, Y., Liang, N., and Gao, G. (2019). An expanded landscape of human long noncoding RNA. Nucleic Acids Res. *47*, 7842–7856. https://doi.org/10.1093/nar/gkz621.

61. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678. https://doi.org/10.1038/s41587-020-0546-8.

62. Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. *42*, W401–W407. https://doi.org/10.1093/nar/gku410.

63. Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. *24*, 1012–1019. https://doi.org/10.1101/gr.171322.113.

64. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.

65. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for the ENCODE Project. Genome Res. *22*, 1760–1774. https://doi.org/10.1101/gr.135350.111.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Human anti-TP53	N/A	Cat# sc-126, Santa Cruz; RRID:AB_628082
Human anti-GAPDH	N/A	Cat# P30008M, Abmart
Deposited data		
Gene expression levels and related clinical information of RNA-seq samples from The Cancer Genome Atlas project (TCGA, https://portal.gdc.cancer.gov) and the Genotype-Tissue Expression project (GTEx, https://www.gtexportal.org)	UCSC Xena database. Goldman, et al. 2020	(https://xenabrowser.net)
Gene expression levels of cell lines sequenced by the Progenitor Cell Biology Consortium (PCBC)	Synapse. Salomonis et al.[53]	https://www.synapse.org/#!Synapse:syn1773109/wiki/54962
Gene expression levels of ESCs during differentiation culturing for both human and mouse	Barry et al.[54]	GSE90053
Somatic mutations for TCGA tumor samples were obtained the GDC Portal	N/A	https://portal.gdc.cancer.gov
Genome ploidy information for TCGA samples	Taylor et al.[34]	PMID:29622463
Cancer stemness of TCGA samples	Malta et al.[38]	PMID:29625051
Processed single-cell RNA sequencing data for embryonic stem cells during preimplantation	Petropoulos et al.[47]	PMID: 27062923
Gene expression levels of SLE RNA-seq data	Hung et al.[41]	GSE72509
Half-life data of mRNAs	Tani et al.[55]	PMID:22369889
RNA-seq data for TP53-WT and TP53-KO cell lines (MCF7 and HT1080)	This study	SRA: PRJNA777384
Calculated XAR values for TCGA and GTEx	This study	http://bioinfo-sysu.com:3838/sample-apps/panxar
Tissue-specific gene expression levels (Read counts) for non-human primates	Peng, et al.[56]	NHPRTR http://nhprtr.org
Gene expression levels for the standard mouse strain, C57BL/6JJcl	Li et al.[57]	PMID:28646208
Quantitative proteomic expression levels (TMT ratios) for BRCA	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000120
Quantitative proteomic expression levels (TMT ratios) for COAD	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000116
Quantitative proteomic expression levels (TMT ratios) for OV	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000110
Quantitative proteomic expression levels (TMT ratios) for CCRCC	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000127
Quantitative proteomic expression levels (TMT ratios) for GBM	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000204
Quantitative proteomic expression levels (TMT ratios) for HNSCC	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000221
Quantitative proteomic expression levels (TMT ratios) for LUAD	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000153
Quantitative proteomic expression levels (TMT ratios) for LSCC	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000234

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Quantitative proteomic expression levels (TMT ratios) for PDA	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000270
Quantitative proteomic expression levels (TMT ratios) for UCEC	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000125
Quantitative proteomic expression levels (TMT ratios) for HCC	CPTAC (https://pdc.cancer.gov/pdc/)	PDC000198
RNA-seq data for mouse ATRX knockout cell lines	Qin et al.[58], Deneault et al.[59]	GSE178113
RNA-seq data for mouse ATRX knockout cell lines	Qin et al.[58], Deneault et al.[59]	GSE107878
LncRNA expression profiles for GTEx and TCGA samples	Jiang et al.[60]	RefLnc (http://reflnc.gao-lab.org/)
Experimental models: Cell lines		
MCF7	AmericanType Culture Collection (ATCC)	N/A
HT1080	AmericanType Culture Collection (ATCC)	N/A
Oligonucleotides		
RT-PCR primer KAT8: TCACTCGCAACCAAAAGC G/GATCGCCTCATGCTCCTTCT	Thermo Fisher	N/A
RT-PCR primer MSL1: CCCATCACCGTTACCAT TACG/GGAACAGCCAAGACTGAAGTTT	Thermo Fisher	N/A
RT-PCR primer MSL2: GTAGCCACTGACTTATG TTCCAC/GCTGCAAATTAGGGCAACAGAC	Thermo Fisher	N/A
RT-PCR primer GAPDH: GGAGCGAGATCCCT CCAAAAT/GGCTGTTGTCATACTTCTCATGG	Thermo Fisher	N/A
sgRNA sequence targeted TP53: sgRNA-1, 5′-TATCTGAGCAGCGCTCATGG-3′ and sgRNA-2, 5′-GGTGAGGCTCCCCTTTCTTG-3′.	Thermo Fisher	N/A
Software and algorithms		
R v3.6.3	https://www.r-project.org/	https://cran.r-project.org/
ComplexHeatmap v2.2.0	Gu et al., 2016	https://jokergoo.github.io/ComplexHeatmap-reference/book/
ggpubr v0.4.0	N/A	https://rpkgs.datanovia.com/ggpubr/
circlize v0.4.0	N/A	https://cran.r-project.org/web/packages/circlize/index.html
ggplot2 v3.3.0	N/A	https://ggplot2.tidyverse.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Yuanyan Xiong (xyyan@mail.sysu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. RNA-seq sequencing data generated by this study were deposited in SRA with the accession number PRJNA777384. Calculated XAR values for TCGA and GTEx samples can be interactively

explored at http://bioinfo-sysu.com:3838/sample-apps/panxar. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

MCF7 and HT1080 cells were obtained from the AmericanType Culture Collection (ATCC). All cells were cultured in high-glucose DMEM containing 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin and maintained at 37°C with 5% CO_2.

METHOD DETAILS

Data collection and preprocessing

Gene expression levels (transcripts per million, or TPM) and related clinical information of RNA-seq samples from The Cancer Genome Atlas project (TCGA, https://portal.gdc.cancer.gov) and the Genotype-Tissue Expression project (GTEx, https://www.gtexportal.org), processed by the same computational pipeline, were obtained from the UCSC Xena database (https://xenabrowser.net).[61] For GTEx, tissues with less than 30 samples were excluded. Gene expression levels of cell lines sequenced by the Progenitor Cell Biology Consortium (PCBC)[53] were downloaded from Synapse (https://www.synapse.org/#!Synapse:syn1773109/wiki/54962). Gene expression levels of ESCs during differentiation culturing for both human and mouse were downloaded from.[54] Somatic mutations for TCGA tumor samples were obtained the GDC Portal (https://portal.gdc.cancer.gov). Genome ploidy information for TCGA samples were downloaded from.[34] Cancer stemness of TCGA samples were obtained from.[38] Processed single-cell RNA sequencing data for embryonic stem cells during preimplantation were downloaded from.[47] Gene expression levels of SLE RNA-seq data were obtained from.[41] Half-life data of mRNAs were obtained from.[55] Our RNA-seq data for *TP53*-WT and *TP53*-KO cell lines (MCF7 and HT1080) from this study were deposited in the SRA database (PRJNA777384 [https://dataview.ncbi.nlm.nih.gov/object/PRJNA777384?reviewer=vtlrbhl2kos1m8fe4lij3ul1uk]). Tissue-specific gene expression levels (Read counts) for non-human primates were downloaded from NHPRTR (http://nhprtr.org/).[56] Gene expression levels (Fragments Per Kilobase of transcript per Million mapped reads, or FPKM) for the standard mouse strain, C57BL/6JJcl were collected from.[57] Quantitative proteomic expression levels (TMT ratios) for tumors were obtained from CPTAC (https://pdc.cancer.gov/pdc/), covering data from 11 studies (PDC000120, PDC000116, PDC000110, PDC000127, PDC000204, PDC000221, PDC000153, PDC000234, PDC000270, PDC000125 and PDC000198). RNA-seq data for mouse ATRX knockout cell lines were collected from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/), including datasets GSE178113[58] and GSE107878.[59] LncRNA expression profiles for GTEx and TCGA samples were downloaded from RefLnc (http://reflnc.gao-lab.org/).[60]

CRISPR-Cas9 knockout

The sgRNAs of *TP53* were designed on the CHOPCHOP website (https://chopchop.cbu.uib.no) as described before.[62] sgRNAs with highly predicted efficiency and lower off-target were chosen for *in vitro* transcription. The sgRNA sequence targeted *TP53* were as follows: sgRNA-1, 5′-TATCTGAG CAGCGCTCATGG-3′ and sgRNA-2, 5′-GGTGAGGCTCCCCTTTCTTG-3′. The sgRNA oligos were cloned into the pDR274 vector and *in vitro* transcribed using the MEGAshortscript T7 transcription kit (Thermo Fisher), then purified using the MEGAclear kit (Thermo Fisher). The sgRNAs were then dissolved in RNase-free water and quantified using NanoDrop 1000 (Thermo Fisher).

To generate *TP53* knockout cells, electroporation of Cas9 RNP to the cells was performed as previously reported.[63] Specifically, 10 μg TrueCut Cas9 Protein v2 (Thermo) and 2 μg each of the *in vitro* transcribed sgRNAs were incubated together at 37°C for 20 min, and 5×10^5 MCF7 or HT1080 cells (70%-90% confluent) were harvested and centrifuged at 90× g for 10 min. Then the Cas9 RNP was electroporated to the cell pellet by Lonza 4D-Nucleofector (Lonza) under a pre-optimized program.

Western blotting

Cells were harvested and lysed with RIPA buffer (Beyotime). Cell extracts of equal total protein (20 μg each) were separated by SDS-PAGE and transferred to Nylon membrane (Merck Millipore). The membranes were blocked with 5% nonfat milk in TBST washing buffer (10 mM Tris pH 8.0, 150 mM NaCl with 0.5% Tween-20)

for 1hat room temperature, then incubated at 4°C overnight with anti-TP53 (1:1000; sc-126,Santa Cruz) and anti-GAPDH (1:5000; P30008M, Abmart) primary antibodies. After washing, the membranes were incubated with fluorescence-conjugated secondary antibodies and detected with Odyssey Imagers.

RNA extraction and real-time quantitative PCR

Total RNA was extracted from cultured cells using RNAiso Plus (Takara) and reverse transcribed into cDNA by the PrimeScript™ RT reagent Kit (Takara), following the manufacturer's protocol. Then real-time quantitative PCR was performed using TB Green® Premix Ex Taq™ II (Takara) on the Real-Time PCR Thermal Cycler qTOWER3 (Analytik Jena). Primers used in this study (forward/reverse; 5' to 3') were as follows:

KAT8: TCACTCGCAACCAAAAGCG/GATCGCCTCATGCTCCTTCT

MSL1: CCCATCACCGTTACCATTACG/GGAACAGCCAAGACTGAAGTTT

MSL2: GTAGCCACTGACTTATGTTCCAC/GCTGCAAATTAGGGCAACAGAC

GAPDH: GGAGCGAGATCCCTCCAAAAT/GGCTGTTGTCATACTTCTCATGG

Data analysis was performed using the 2(-ΔΔCT) method, and mRNA expression levels were normalized to GAPDH.

RNA sequencing of MCF7 and HT1080 cell lines

Total RNA was extracted using TransZol Up Plus RNA Kit (Cat#ER501-01, Trans) following the manufacturer's instructions and checked for RNA integrity by an Agilent Bioanalyzer 2100 (Agilent technologies, Santa Clara, CA, US). Qualified total RNA was further purified by RNAClean XP Kit (Cat#A63987, Beckman Coulter, Inc. Kraemer Boulevard Brea, CA,USA) and RNase-Free DNase Set (Cat#79254, QIAGEN, GmBH, Germany). Sequencing libraries were constructed following steps from the manufacturer's instructions (VAHTS Universal V6 RNA-seq Library Prep Kit for Illumina®, Vazyme, NR604-02), followed by paired-end (2x 150bp) RNA-seq by Illumina HiSeq 2500. Cleaned reads were aligned to the human genome (hg38) by Kallisto (v0.46, default parameters)[64] to obtain gene expression levels (TPM), based on gene annotations from GENCODE v25.[65] Reads were also mapped to hg38 using HISAT2 (v2.0.4, default parameters), and alignments of *TP53*-WT samples were visualized by IGV (v2.5.2) for manual inspection to confirm TP53 status.

Defining PIX, PSX, and autosomal genes for dosage compensation estimation

For both X chromosome and autosomes, only protein-coding genes were considered and those with median expression levels TPM<=1 in GTEx samples were further excluded. After filtering, a total of 615 X chromosome genes and 16,622 autosome genes remained. For the remaining X chromosome genes, Pearson correlations (R) between gene expression and genome ploidy[34] were calculated in each cancer (except CHOL, DLBC, UCS, KICH, and BRCA) separately from TCGA, using only male samples to avoid the confounding of possible reactivation of the inactivated X in female samples. Then, the median of absolute R (absR) was calculated for each gene and the average of these absR medians (MR) was determined for these genes. Lastly, X chromosome genes with absR medians less or higher than MR were defined as PIX and PSX genes, respectively. Remaining PIX, PSX, and autosome genes were used for XAR-PIX and XAR-PSX calculation.

Comparison of genomic features for PIX and PSX genes

Gene lengths were calculated for each gene either including or excluding introns, based on GENCODE v25.[65] PhastCons100way scores were downloaded from the UCSC Genome Browser (https://hgdownload.soe.ucsc.edu/downloads.html) and gene conservation were calculated by averaging phastCons100way scores within exonic regions. GC content for each gene were calculated from its mRNA sequence as the percentage of G and C bases.

Calculation of XAR-PSX and XAR-PIX

First, median expression levels of PSX, PIX, and autosomal genes were calculated. Then, XAR-PSX or XAR-PIX were defined as the ratio of PSX or PIX median over autosomal median, instead of the mean.[6] The

magnitude (Cohen's D) of XAR difference between TCGA and GTEx was estimated by the function cohens_d from the R package rstatix (v0.6). For non-human species, genes were classified according to their homologous gene categories (PSX, PIX or autosomal genes) in humans, and then XAR-PSX and XAR-PIX were calculated as above, based on their processed gene expression levels (TPM). For proteomic expression levels, TMT ratios were used to calculate XAR, while XAR-PSX and XAR-PIX were compared between primary tumors and solid tissue normal samples.

Association of XAR with somatic mutations

Damaging mutations were defined as Missense_Mutation, Nonsense_Mutation, Nonstop_Mutation, Splice_Region, Splice_Site, Translation_Start_Site, Frame_Shift_Del, Frame_Shift_Ins, In_Frame_Ins, or In_Frame_Del. For each gene, samples were grouped by whether having damaging mutations in each cancer for each gender. Genes with less than 20 damaging mutations were not considered. XAR between these two groups were compared using two-sided *t*-test and *P*-values were corrected using false discovery rate (FDR).

Gene ontology analysis

GO enrichment analyses were performed using R package clusterProfiler (v4.2.2). Enriched biological processes were identified by right-tailed Fisher's exact test, and Benjamini-Hochberg correction was used for multiple comparisons. For potential protein-coding regulators of XAR, genes expressed in at least in one tumor were used as the background gene set. For lncRNA analysis, neighboring protein-coding genes of XAR-associated lncRNAs were used as background gene set.

ATRX regression coefficients comparison between autosomal and X-linked genes

For autosomal and X-linked genes, linear regression models with ATRX expression were fit at the mRNA or proteomic level in each tumor, in which*ATRX* coefficient estimates were collected and converted to absolute form for comparison between the two gene groups.

Correlations for lncRNAs and XAR in tumors

For 77,900 lncRNAs collected, those with median TPM greater than 0 in at least 1 tumor were selected. Next, linear regression models were fit to lncRNAs expression and XAR in each tumor, with lncRNAs correlated to XAR-PIX or XAR-PSX in at least 3 male tumors extracted for further study. Neighboring co-expressing protein-coding genes of these lncRNAs were identified by Pearson correlation coefficient (FDR<0.05) and subjected to GO analysis. RNA-protein interactions of XAR-associated lncRNAs were collected from LncSEA2.0 (http://bio.liclab.net/LncSEA/index.php).

Data visualization

The R package ggpubr (v.04, based on ggplot2) was used for basic data visualizations in this study. The heatmaps of stemness correlation with XAR were visualized by R package ComplexHeatmap (v2.2). Genomic distribution of PIX and PSX genes were visualized by R package circlize (v0.4). Cumulative distribution of expression levels were visualized by R package ggplot2 (v3.3). Calculated XAR values for TCGA and GTEx samples can be interactively explored at http://bioinfo-sysu.com:3838/sample-apps/panxar.

QUANTIFICATION AND STATISTICAL ANALYSIS

Two-sided Wilcoxon test or Student's test were used for statistical analysis in R (https://www.r-project.org). Data are presented as median ± interquartile range (IQR) or mean ± standard deviation (SD). The level of significance was indicated by the p value. In all figures where applicable, levels of statistical significance were indicated as: *P<= 0.05, ** <=0.01, *** <= 0.001, **** <= 0.0001.