# Comprehensively characterizing evolutionarily conserved differential dosage compensations of the X chromosome during development and in diseases

Mengbiao Guo[1,#], Zhengwen Fang[1,#], Bohong Chen[1,#], Zhou Songyang[1], Yuanyan Xiong[1,*]

[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.

[#] These authors contributed equally.

[*] Corresponding author: Yuanyan Xiong, xyyan@mail.sysu.edu.cn, Telephone: +86-20-39943531, Fax: +86-20-39943778

## Abstract

The active X chromosome in mammals is upregulated to balance its dosage to autosomes during evolution. However, it is elusive why the currently known dosage compensation machinery showed uneven and small influence on X genes, necessitating systemic investigation of X dosage in different angles and identification of new dosage regulators. Here, based on >20,000 transcriptomes, we identified two X gene groups (genome ploidy-sensitive [PSX] and ploidy-insensitive [PIX] genes), showing distinct but evolutionarily-conserved (in both primates and mouse) dosage compensations (termed X-over-Autosome dosage Ratio, or XAR). We then explored XAR in diseases and in stem cells, where XAR is potentially important. We demonstrated that XAR-PIX was downregulated while XAR-PSX upregulated across cancer types at both RNA and protein levels. In contrast, XAR-PIX was upregulated while XAR-PSX downregulated during stem cell differentiation. Interestingly,

XAR-PIX, but not XAR-PSX, was significantly lower and associated with autoantibodies and inflammation in lupus patients, suggesting that insufficient dosage of PIX genes may contribute to lupus pathogenesis. We further identified and experimentally validated two new XAR regulators, *TP53* and *ATRX*. Collectively, we provided insights to further unravel the mystery of X dosage compensation in mammals and its pathophysiological roles in human diseases.

## Introduction

In mammals, females have two X chromosomes while 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) (Disteche 2012) with exceptions of escaping genes (Dunford, et al. 2017). Initiated early in female embryo development, XCI is essentially stable for life after complete (Vallot, et al. 2016). 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 (Disteche 2012; Deng, et al. 2013; Larsson, et al. 2019), although with controversies (Xiong, et al. 2010; Julien, et al. 2012).

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 (Deng, et al. 2013). 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 (Deng, et al. 2013). However, it is intriguing that

48 *KAT8* affected X genes unevenly and only a small percentage of X genes showed
49 considerable effect sizes.
50
51 Recently, using published single-cell RNA sequencing datasets, Larsson and colleagues
52 reported that the frequency of transcriptional bursts was the driving force behind X
53 upregulation in mouse, and they proposed that trans-acting enhancer-binding regulators
54 may be the underlying factor behind the higher transcriptional burst frequency of X
55 chromosome (Larsson, et al. 2019). However, these factors were unknown and the
56 hypothesis of enhancer factors in X upregulation was not verified.
57
58 *TP53* is a highly-conserved gene important for various stress responses (Lu, et al. 2009). It
59 was reported to interact with key MSL components, *KAT8* and *MSL2* (Kruse and Gu 2009; Li,
60 et al. 2009). The p53 binding sites (Nguyen, et al. 2018) in *KAT8* (one site in promoter) and
61 *MSL2* (three sites in promoter) indicate direct regulation of *KAT8* and *MSL2* transcription by
62 p53. ATRX, an X-linked telomere-associating protein studied by our team (He, et al. 2015; Hu,
63 et al. 2016), is a chromatin remodeler that functions in genome stabilization and
64 heterochromatin formation (Dyer, et al. 2017; Teng, et al. 2021). ATRX works with XIST and
65 PRC2 to remodel chromatins (Ren, et al. 2020) and ATRX loss can affect X inactivation in
66 embryonic trophoblast (Garrick, et al. 2006). Importantly, it has been reported that TP53
67 and ATRX (also known as RAD54) physically interact with each other (Linke, et al. 2003) and
68 work together in the same pathway (Seah, et al. 2008; Oppel, et al. 2019; Gulve, et al. 2022).
69 Both genes are involved in genome-wide histone modifications (Allison and Milner 2004;
70 Dyer, et al. 2017), thus capable of regulating a large number of genes. We therefore
71 hypothesized that TP53 and ATRX probably regulate X dosage compensation.

72

73 Interestingly, *TP53* mutations contributed to the sex discrepancy of cancer incidences by
74 differentially regulating X-linked genes between two sexes (Haupt, et al. 2019). Notably,
75 *TP53* mutations are the most frequent events across all cancers (Hainaut and Pfeifer 2016),
76 and more than 91% of tumors with *TP53* mutations show functional loss of both *TP53* copies
77 (Donehower, et al. 2019). p53 is also implicated in immune regulation of autoimmune
78 diseases (Munoz-Fontela, et al. 2016). On the other hand, X-linked genes have been
79 implicated in many diseases, including cancer (Haupt, et al. 2019) and autoimmune diseases
80 like systemic lupus erythematosus (SLE) (Hewagama, et al. 2013). Abnormal X dosage has
81 also been reported in both cancer and SLE (Pageau, et al. 2007; Syrett, et al. 2019). For
82 example, X upregulation has been contributed to the loss of Barr body (the inactivated X
83 chromosome) in breast and ovarian cancers. X dosage also plays an important role in cancer
84 initiation, because excessive X upregulation by somatic X reactivation under chronic stress
85 leads to cancer (Yang, et al. 2020). However, the dynamics of X-over-autosome dosage
86 ratios (XAR) remain to be studied systematically across cancers, which accumulates a large
87 number of high-throughput sequencing datasets, for example, The Cancer Genome Atlas
88 (TCGA) (Davoli, et al. 2013). More importantly, large numbers of cancer mutations and
89 genome-wide expression profiles from TCGA have the potential to reveal new regulators of
90 XAR, including both protein-coding and long noncoding RNA (lncRNA) genes, to uncover the
91 elusive mechanism of dosage compensation in humans. However, these possibilities have
92 not been explored yet.

93

94 In this work, we identified two distinct groups of genes (PIX and PSX) on X chromosome and
95 characterized their dosage compensations in about 20,000 RNA sequencing (RNA-Seq)

samples from TCGA and the Genotype-Tissue Expression (GTEx) project (GTEx-Consortium 2013), 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. Interestingly, XAR-PIX and XAR-PSX were affected differently in sex-biased diseases and during stem cell differentiation which is closely linked with XCI (Hall, et al. 2008). 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, since only a subset of X chromosome genes were affected after *KAT8* knockdown (Deng, et al. 2013). 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 (n=615 X chromosome genes and n=16,622 autosome genes remained, **Table S1**) with genome ploidies of cancer samples from TCGA. 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 **Methods**), separated by the mean of the correlation distribution (**Fig. 1A**). As expected, PSX and PIX genes showed distinct expression distribution. PSX genes (n=286,

120 **Table S1**) had highly similar expression pattern to autosome genes (**Fig. 1B**), while PIX genes
121 (n=329, **Table S1**) had much lower expression levels, with a median of about 2/3 of PSX
122 genes expression (**Fig. 1C**). Of note, no preference for genes escaping X inactivation was
123 observed in PSX or PIX. Interestingly, we found three PIX-specific gene clusters on X
124 chromosome, compared to PSX genes (**Fig. 1D**). The PIX cluster located at around 100Mb
125 (hg38) contained eight out of nine Transcription Elongation Factor A Protein-Like (TCEAL)
126 genes, encoding nuclear phosphoproteins.
127
128 **Different features and regulation of PIX and PSX genes**
129
130 The distinct expression pattern between PIX and PSX genes prompted us to explore their
131 differential genomic features. First, we found that the lengths (excluding introns) of PIX
132 genes spanned a large range from 2kb to 8kb, while lengths of PSX genes peaked at around
133 4kb (**Fig. 1E**). This trend was similar when including introns (**Fig. S1A**). Although GC content
134 of mRNAs is critical for their storage and decay (Courel, et al. 2019), we found similar GC
135 content between PIX and PSX genes (**Fig. S1B**). Moreover, we observed lower conservation
136 of PIX genes than PSX genes (**Fig. 1F**), which resembled the faster-X effect in non-coding
137 regions on X chromosome (Meisel and Connallon 2013). RNA half-lives were reported to be
138 longer for X-linked than autosomal genes (Deng, et al. 2013). Inspired by this, we examined
139 the mRNA half-lives of PIX and PSX genes and found that PIX (4.9h) tended to have shorter
140 mRNA half-lives than PSX genes (6.6h) (**Fig. 1G**), consistent with lower expression levels of
141 PIX than PSX genes.
142
143 **Evolutionally conserved differential dosage compensation between PIX and PSX genes**

144

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

154

155 Variations in TCGA cancers were generally larger than GTEx tissues for both XAR-PSX and
156 XAR-PIX, reflecting abnormal dosage regulation across cancers. Surprisingly, across TCGA
157 cancer types, we observed that XAR-PSX dosages were significantly increased (**Fig. 2C**), while
158 XAR-PIX showed the opposite trend (**Fig. 2D**), compared to normal tissues from GTEx.
159 Moreover, most magnitudes of changes (or effect sizes) of both XAR-PSX and XAR-PIX were
160 moderate to large, as indicated by the Cohen's D statistics. Of note, most tissues or cancer
161 types showed no XAR difference between males and females, especially for XAR-PIX (**Fig.**
162 **S3A-D**), which suggests that the possibility of XAR-PSX upregulation due to reactivation of
163 the inactivated X in females (Chaligne, et al. 2015) can be excluded. Accordingly, when
164 investigating proteomic data generated by mass spectrometry, the opposite trend between
165 XAR-PIX and XAR-PSX was also observed when we compared primary tumors with normal
166 tissue samples in CPTAC **(Fig. S2C-D)**.

167

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

169

170 It is well-known that XAR of stem cells changes dramatically during epigenetic
171 reprogramming of early embryo development. On the other hand, increased stemness is an
172 important feature or cancer cells (Malta, et al. 2018). We thus examined the relationship
173 between XAR and stemness in cancer. We observed mostly positive correlations with cancer
174 stemness for XAR-PSX, but mostly negative correlations for XAR-PIX, across cancer types (**Fig.**
175 **3A**). We further examined XAR in embryonic stem cells (ESC) and induced pluripotent stem
176 cells (iPSC) using gene expression data from the Progenitor Cell Biology Consortium (PCBC).
177 As expected, we found no difference between ESC and iPSC, for both XAR-PIX and XAR-PSX
178 (**Fig. 3B**), probably because ESC and iPSC had similar expression pattern (Bock, et al. 2011).
179 However, we found much lower XAR-PIX and XAR-PSX in male than in female iPSC (**Fig. 3C**),
180 but the signals of sex difference in ESC were weak.

181

182 We noticed that XAR-PSX was between 1.1 to 1.2 in both ESCs and iPSCs, although ESCs and
183 some iPSCs have two active X chromosomes (Bar, et al. 2019). This may be related to early
184 embryonic development, during which XAR-PSX has peaked (as expected, close to 1.9 in
185 female and 1.4 in male) at embryo day 4 (E4) and decreased dramatically to be equal in both
186 sexes (around 1.2) (**Fig. 3D**), which was consistent with our observations of XAR-PSX in ESC
187 and iPSC cell lines.

188

189 We further examined XAR dynamics during iPSC differentiation culturing. Interestingly, we
190 found that XAR-PSX was reduced along iPSC culturing time (**Fig. 3E**), but XAR-PIX showed the

191 opposite trend (**Fig. 3F**), in both human and mouse. These patterns were consistent with
192 overall reduction of XAR-PIX and elevation of XAR-PSX across cancer types above.
193
194 **Insufficient PIX gene dosage in autoimmune patients**
195
196 Both altered X-inactivation and X-linked risk genes were associated with SLE, we thus
197 investigated which type of XAR (XAR-PIX or XAR-PSX) was dysregulated in SLE. By
198 reanalyzing a previous dataset from SLE patients and healthy controls (Hung, et al. 2015),
199 we found that XAR-PIX, but not XAR-PSX, were specifically down-regulated in SLE patients
200 compared with healthy controls (**Fig. 4A**), which means insufficient dosage of X genes in SLE
201 patients. Moreover, SLE patients with anti-Ro autoantibodies showed lower XAR-PIX dosage
202 than those without anti-Ro autoantibodies (**Fig. 4B**). Similarly, SLE patients with more active
203 interferon response (indicated by the Interferon Signature Metric, or ISM) showed lower
204 XAR-PIX dosage (**Fig. 4C**). These findings may provide insights to understand SLE
205 pathogenesis.
206
207 **TP53 and ATRX are new regulators of X dosage compensation**
208
209 We found that only about 8.8% (29/329) and 10.5% (30/286) of PIX and PSX genes showed
210 expression changes of more than 1.2-fold after knockdown of the key MSL component *KAT8*
211 in mouse (Deng, et al. 2013). To identify new regulators of XAR, we examined TCGA
212 mutation spectrums that may be associated with XAR in each cancer type. We reasoned
213 that mutated XAR regulators should result in differential XAR compared with wildtype
214 regulators. In total, we obtained 209 and 515 potential regulators (FDR < 0.5) for XAR-PIX

215 and XAR-PSX in male tumors, respectively (**Fig. 5A, Fig. S4A**). For the top 200 regulators,
216 gene ontology (GO) analysis was performed and similar GO terms were found for XAR-PIX
217 and XAR-PSX regulators, mainly about post-translational modifications, including histone
218 methylation, peptidyl-lysine methylation, and protein methylation (**Fig. S4B-C**).
219
220 Among the potentially new regulators of XAR, we found that *TP53* and *ATRX* were promising
221 regulators for XAR-PIX in both males and females (**Fig. 5A, Fig. S4D**), based on previous
222 literatures mentioned above. They are both genome-wide chromatin remodelers, capable of
223 regulating a large number of genes. We chose them for further investigation also for the
224 following two reasons. We found that the top 200 XAR-PIX dependent genes formed a PPI
225 network mostly centered on TP53 and ATRX (**Fig. S5**). TP53 was also a center node in the PPI
226 network (**Fig. S4E**) with RNA-binding proteins (RBPs) associated with XAR-associated
227 lncRNAs (**Supplementary Results, Fig. S4F, Table S2**).
228
229 Next, we constructed *TP53*-knockout (*TP53*-KO) cells from two *TP53*-wildtype (*TP53*-WT) cell
230 lines, one male (HT1080) and one female (MCF7, containing three active X and no inactive X
231 chromosomes (Sirchia, et al. 2005)), by using CRISPR technologies (**Fig. S4G**, see **Methods**),
232 followed by RNA-seq. We observed XAR-PIX increase after *TP53*-KO in both sexes (**Fig. 5B**
233 and **Fig. 5C**), in contrast to the unchanged XAR-PSX (**Fig. 5D** and **Fig. 5E**). We further
234 demonstrated that *TP53* may partly regulate XAR-PIX by repressing KAT8, since *TP53*-KO
235 enhanced *KAT8* expression in both sexes (**Fig. 5F**). *TP53*-KO had a larger effect in female cells
236 (8% vs. 6% increase of XAR-PIX) than male cells, which may due to MCF7 cells containing
237 three active X chromosomes and *TP53* affected *MSL1* in female cells only.
238

239 Finally, we examined the role of ATRX in XAR. We investigated XAR before and after *ATRX*
240 knockout in two studies. Interestingly, ATRX-KO reduced both XAR-PIX and XAR-PSX in
241 mouse iPSCs (**Fig. 5G**) and GBM cells (**Fig. 5H**). Similarly, we examined MSL expression
242 changes after ATRX-KO, we observed consistent down-regulation of both MSL1 and MSL2 in
243 either GBM or iPSC cells in mice, although KAT8 was found up-regulated in mouse iPSC cells
244 while it was unchanged in GBM cells (**Fig. S6A**). To further support *ATRX* as a new regulator
245 of XAR, we examined whether X-linked genes are more sensitive to the change of *ATRX*
246 expression. For autosomal and X-linked genes co-expressed with *ATRX,* we compared their
247 coefficient estimates of *ATRX* expression by linear regression models. We reasoned that
248 larger absolute values of ATRX coefficients indicate higher impact on gene expression. As a
249 result, *ATRX* regression coefficients were significantly higher in X-linked genes than in
250 autosomal genes across a variety of tumors, at both mRNA (**Fig. S6B**) and protein level (**Fig.**
251 **S6C**).
252

253 ## Discussions

254 Here, we identified and comprehensively characterized two distinct groups of genes on the
255 X chromosome. Integrating large-scale sequencing data from both cancer and normal
256 samples greatly enhanced our ability to investigate both pathological and physiological roles
257 of XAR. Our findings may help resolve XAR controversies described in our previous study
258 (Xiong, et al. 2010). We summarized our findings in **Figure 6**. Our results showed that
259 dosages of PIX and PSX genes were converging during stem cell differentiation but diverging
260 in cancer cells. However, in SLE patients, only dosages of PIX genes, but not PSX genes, were
261 down-regulated compared to healthy individuals. Finally, TP53 and ATRX were identified
262 and validated as two XAR regulators.

263

264 Although most genes from the TCEAL family, located within one of PIX gene clusters, were
265 not intensively studied, most reports suggested that they acted as tumor suppressors,
266 especially for *TCEAL7* (Lafferty-Whyte, et al. 2010; Yue, et al. 2019). For these tumor
267 suppressors, lower sensitivity to dosage may protect both males and females from cancer
268 equally. Interestingly, the chrX region around 100Mb showed the lowest expression level
269 after zygote genome activation (ZGA) at E4 during embryo implantation (Petropoulos, et al.
270 2016).

271

272 Of note, *TP53* mRNA itself is a critical regulator of cellular stress (Haronikova, et al. 2019). A
273 previous study found that the deadenylase poly(A)-specific ribonuclease (PARN) destabilized
274 *TP53* mRNA while p53 increase activated PARN (Devany, et al. 2013), indicating a negative
275 feedback loop between *TP53* mRNA and p53 protein levels. We observed that *TP53* mRNA
276 was universally up-regulated in cancer samples, compared with non-cancerous samples,
277 across cancer types (**Fig. S7**). Moreover, XAR-PIX was lower in tumors than non-cancerous
278 samples across cancer types. These results were consistent with observations that p53
279 mutations (or higher *TP53* mRNA) were associated with lower XAR-PIX in cancer and *TP53*-
280 KO resulted in XAR-PIX increase.

281

282 The frequency of transcriptional bursts and possible trans-acting enhancer-binding
283 regulators were proposed to be the driving force behind X upregulation in mouse (Larsson,
284 et al. 2019). Interestingly, p53 can bind pervasively to genome-wide enhancers (Younger
285 and Rinn 2017) and regulate the transcriptional burst of target genes (Friedrich, et al. 2019),
286 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* (Gewandter, et al. 2011; Deng, et al. 2013). Therefore, *TP53* may regulate XAR by various mechanisms. We expect XAR regulation by *TP53* would be conserved among mammals, due to ultra-conservation of *TP53* during evolution (Lu, et al. 2009). 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. Since *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.

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.

## Materials and Methods

311 **Datasets**

312 Gene expression levels (transcripts per million, or TPM) and related clinical information of
313 RNA-seq samples from The Cancer Genome Atlas project (TCGA,
314 https://portal.gdc.cancer.gov) and the Genotype-Tissue Expression project (GTEx,
315 https://www.gtexportal.org), processed by the same computational pipeline, were obtained
316 from the UCSC Xena database (https://xenabrowser.net) (Goldman, et al. 2020). For GTEx,
317 tissues with less than 30 samples were excluded. Gene expression levels of cell lines
318 sequenced by the Progenitor Cell Biology Consortium (PCBC) (Salomonis, et al. 2016) were
319 downloaded from Synapse (https://www.synapse.org/#!Synapse:syn1773109/wiki/54962).
320 Gene expression levels of ESCs during differentiation culturing for both human and mouse
321 were downloaded from (Barry, et al. 2017). Somatic mutations for TCGA tumor samples
322 were obtained the GDC Portal (https://portal.gdc.cancer.gov). Genome ploidy information
323 for TCGA samples were downloaded from (Taylor, et al. 2018). Cancer stemness of TCGA
324 samples were obtained from (Malta, et al. 2018). Processed single-cell RNA sequencing data
325 for embryonic stem cells during preimplantation were downloaded from (Petropoulos, et al.
326 2016). Gene expression levels of SLE RNA-seq data were obtained from (Hung, et al. 2015).
327 Half-life data of mRNAs were obtained from (Tani, et al. 2012). Our RNA-seq data for *TP53*-
328 WT and *TP53*-KO cell lines (MCF7 and HT1080) from this study were deposited in the SRA
329 database (PRJNA777384). Tissue-specific gene expression levels (Read counts) for non-
330 human primates were downloaded from NHPRTR (http://nhprtr.org/) (Peng, et al. 2015).
331 Gene expression levels (Fragments Per Kilobase of transcript per Million mapped reads, or
332 FPKM) for the standard mouse strain, C57BL/6JJcl were collected from (Li, et al. 2017).
333 Quantitative proteomic expression levels (TMT ratios) for tumors were obtained from
334 CPTAC (https://pdc.cancer.gov/pdc/), covering data from 11 studies (PDC000120,

335 PDC000116, PDC000110, PDC000127, PDC000204, PDC000221, PDC000153, PDC000234,
336 PDC000270, PDC000125 and PDC000198). RNA-seq data for mouse ATRX knockout cell lines
337 were collected from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/),
338 including datasets GSE178113 (Qin, et al. 2022) and GSE107878 (Deneault, et al. 2018).
339 LncRNA expression profiles for GTEx and TCGA samples were downloaded from RefLnc
340 (http://reflnc.gao-lab.org/) (Jiang, et al. 2019).
341

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

343 For both X chromosome and autosomes, only protein-coding genes were considered and
344 those with median expression levels TPM<=1 in GTEx samples were further excluded. For
345 the remaining X chromosome genes, Pearson correlations (R) between gene expression and
346 genome ploidy (Taylor, et al. 2018) were calculated in each cancer (except CHOL, DLBC, UCS,
347 KICH, and BRCA) separately from TCGA, using only male samples to avoid the confounding
348 of possible reactivation of the inactivated X in female samples. Then, the median of absolute
349 R (absR) was calculated for each gene and the average of these absR medians (MR) was
350 determined for these genes. Lastly, X chromosome genes with absR medians less or higher
351 than MR were defined as PIX and PSX genes, respectively. Remaining PIX, PSX, and
352 autosome genes were used for XAR-PIX and XAR-PSX calculation.
353

354 **Comparison of genomic features for PIX and PSX genes**

355 Gene lengths were calculated for each gene either including or excluding introns, based on
356 GENCODE v25 (Harrow, et al. 2012). PhastCons100way scores were downloaded from the
357 UCSC Genome Browser (https://hgdownload.soe.ucsc.edu/downloads.html) and gene
358 conservation were calculated by averaging phastCons100way scores within exonic regions.

359 GC content for each gene were calculated from its mRNA sequence as the percentage of G
360 and C bases.
361
362 **Calculation of XAR-PSX and XAR-PIX**
363 First, median expression levels of PSX, PIX, and autosomal genes were calculated. Then,
364 XAR-PSX or XAR-PIX were defined as the ratio of PSX or PIX median over autosomal median,
365 instead of the mean (Xiong, et al. 2010). The magnitude (Cohen's D) of XAR difference
366 between TCGA and GTEx was estimated by the function cohens_d from the R package
367 rstatix (v0.6). For non-human species, genes were classified according to their homologous
368 gene categories (PSX, PIX or autosomal genes) in humans, and then XAR-PSX and XAR-PIX
369 were calculated as above, based on their processed gene expression levels (TPM). For
370 proteomic expression levels, TMT ratios were used to calculate XAR, while XAR-PSX and
371 XAR-PIX were compared between primary tumors and solid tissue normal samples.
372
373 **Association of XAR with somatic mutations**
374 Damaging mutations were defined as Missense_Mutation, Nonsense_Mutation,
375 Nonstop_Mutation, Splice_Region, Splice_Site, Translation_Start_Site, Frame_Shift_Del,
376 Frame_Shift_Ins, In_Frame_Ins, or In_Frame_Del. For each gene, samples were grouped by
377 whether having damaging mutations in each cancer for each gender. Genes with less than
378 20 damaging mutations were not considered. XAR between these two groups were
379 compared using two-sided *t*-test and *P*-values were corrected using false discovery rate
380 (FDR).
381
382 **Gene ontology analysis**

383 GO enrichment analyses were performed using R package clusterProfiler (v4.2.2). Enriched
384 biological processes were identified by right-tailed Fisher's exact test, and Benjamini-
385 Hochberg correction was used for multiple comparisons. For potential protein-coding
386 regulators of XAR, genes expressed in at least in one tumor were used as the background
387 gene set. For lncRNA analysis, neighboring protein-coding genes of XAR-associated lncRNAs
388 were used as background gene set.
389
390 **ATRX regression coefficients comparison between autosomal and X-linked genes**
391 For autosomal and X-linked genes, linear regression models with ATRX expression were fit at
392 the mRNA or proteomic level in each tumor, in which *ATRX* coefficient estimates were
393 collected and converted to absolute form for comparison between the two gene groups.
394
395 **Data visualization**
396 The R package ggpubr (v.04, based on ggplot2) was used for basic data visualizations in this
397 study. The heatmaps of stemness correlation with XAR were visualized by R package
398 ComplexHeatmap (v2.2). Genomic distribution of PIX and PSX genes were visualized by R
399 package circlize (v0.4). Correlation heatmaps between XAR predictions and observations
400 were visualized by R package ggcorrplot2 (v0.1). Cumulative distribution of expression levels
401 were visualized by R package ggplot2 (v3.3).
402
403 **Cell culture and CRISPR-Cas9 knockout**
404 MCF7 and HT1080 cells were obtained from the American Type Culture Collection (ATCC).
405 All cells were cultured in high-glucose DMEM containing 10% fetal bovine serum (FBS) and 1%
406 penicillin/ streptomycin and maintained at 37 °C with 5% $CO_2$.

407

408 The sgRNAs of *TP53* were designed on the CHOPCHOP website
409 (https://chopchop.cbu.uib.no) as described before (Montague, et al. 2014). sgRNAs with
410 highly predicted efficiency and lower off-target were chosen for in vitro transcription. The
411 sgRNA sequence targeted *TP53* were as follows: sgRNA-1, 5'-TATCTGAGCAGCGCTCATGG-3'
412 and sgRNA-2, 5'-GGTGAGGCTCCCCTTTCTTG-3'. The sgRNA oligos were cloned into the
413 pDR274 vector and in vitro transcribed using the MEGAshortscript T7 transcription kit
414 (Thermo Fisher), then purifified using the MEGAclear kit (Thermo Fisher). The sgRNAs were
415 then dissolved in RNase-free water and quantified using NanoDrop 1000 (Thermo Fisher).

416

417 To generate *TP53* knockout cells, electroporation of Cas9 RNP to the cells was performed as
418 previously reported (Kim, et al. 2014). Specifically, 10 µg TrueCut Cas9 Protein v2 (Thermo)
419 and 2 µg each of the in vitro transcribed sgRNAs were incubated together at 37°C for 20 min,
420 and $5\times10^5$ MCF7 or HT1080 cells (70%-90% confluent) were harvested and centrifuged at
421 90× g for 10 min. Then the Cas9 RNP was electroporated to the cell pellet by Lonza 4D-
422 Nucleofector (Lonza) under a pre-optimized program.

423

424 **Western Blotting**

425 Cells were harvested and lysed with RIPA buffer (Beyotime). Cell extracts of equal total
426 protein (20 µg each) were separated by SDS-PAGE and transferred to Nylon membrane
427 (Merck Millipore). The membranes were blocked with 5% nonfat milk in TBST washing
428 buffer (10 mM Tris pH 8.0, 150 mM NaCl with 0.5% Tween-20) for 1 h at room temperature,
429 then incubated at 4°C overnight with anti-*TP53* (1:1000; sc-126,Santa Cruz) and anti-GAPDH
430 (1:5000; P30008M, Abmart) primary antibodies. After washing, the membranes were

431 incubated with fluorescence-conjugated secondary antibodies and detected with Odyssey
432 Imagers.
433
434 **RNA extraction and real-time quantitative PCR**
435 Total RNA was extracted from cultured cells using RNAiso Plus (Takara) and reverse
436 transcribed into cDNA by the PrimeScript™ RT reagent Kit (Takara) , following the
437 manufacturer's protocol. Then real-time quantitative PCR was performed using TB Green®
438 Premix Ex Taq™ II (Takara) on the Real-Time PCR Thermal Cycler qTOWER3 (Analytik Jena).
439 Primers used in this study (forward/reverse; 5' to 3') were as follows:
440 *KAT8*: TCACTCGCAACCAAAAGCG/GATCGCCTCATGCTCCTTCT
441 *MSL1*: CCCATCACCGTTACCATTACG/GGAACAGCCAAGACTGAAGTTT
442 *MSL2*: GTAGCCACTGACTTATGTTCCAC/GCTGCAAATTAGGGCAACAGAC
443 GAPDH: GGAGCGAGATCCCTCCAAAAT/GGCTGTTGTCATACTTCTCATGG
444 Data analysis was performed using the 2(-ΔΔCT) method, and mRNA expression levels were
445 normalized to GAPDH.
446
447 **RNA sequencing of MCF7 and HT1080 cell lines**
448 Total RNA was extracted using TransZol Up Plus RNA Kit (Cat#ER501-01, Trans) following the
449 manufacturer's instructions and checked for RNA integrity by an Agilent Bioanalyzer 2100
450 (Agilent technologies, Santa Clara, CA, US). Qualified total RNA was further purified by
451 RNAClean XP Kit (Cat#A63987, Beckman Coulter, Inc. Kraemer Boulevard Brea, CA,USA) and
452 RNase-Free DNase Set (Cat#79254, QIAGEN, GmBH, Germany). Sequencing libraries were
453 constructed following steps from the manufacturer's instructions (VAHTS Universal V6 RNA-
454 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) (Bray, et al. 2016) to obtain gene expression levels (TPM), based on gene annotations from GENCODE v25 (Harrow, et al. 2012). 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.

## Data availability

The RNA-seq data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA777384.

## Declaration of interests

The authors declare no competing interests.

## Acknowledgements

The work was supported by National Natural Science Foundation of China (NSFC) (Grant number 31571350, U1611265, and 31871323) and Guangdong Basic and Applied Basic Research Foundation (2021A1515110972). The authors would like to thank TCGA, GTEx, PCBC, NHPRTR, and other projects for making their research data publicly available.

## Author contributions

MBG and YYX conceived the project. MBG and ZWF collected the data, performed the analysis, and interpreted the results. MBG wrote the manuscript with input from ZWF. ZWF and BHC performed the experimental validation. ZSY was involved in project discussion. YYX

was involved in critical discussions and reviewed the manuscript. All authors approved the manuscript.

## References


Allison SJ, Milner J. 2004. Remodelling chromatin on a global scale: a novel protective function of p53. Carcinogenesis 25:1551-1557.

Bar S, Seaton LR, Weissbein U, Eldar-Geva T, Benvenisty N. 2019. Global Characterization of X Chromosome Inactivation in Human Pluripotent Stem Cells. Cell Rep 27:20-29 e23.

Barry C, Schmitz MT, Jiang P, Schwartz MP, Duffin BM, Swanson S, Bacher R, Bolin JM, Elwell AL, McIntosh BE, et al. 2017. Species-specific developmental timing is maintained by pluripotent stem cells ex utero. Dev Biol 423:101-110.

Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH, et al. 2011. Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell 144:439-452.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525-527.

Chaligne R, Popova T, Mendoza-Parra MA, Saleem MA, 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.

Courel M, Clement Y, Bossevain C, Foretek D, Vidal Cruchez O, Yi Z, Benard M, Benassy MN, Kress M, Vindry C, et al. 2019. GC content shapes mRNA storage and decay in human cells. Elife 8.

Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, Park PJ, Elledge SJ. 2013. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell 155:948-962.

Deneault E, White SH, Rodrigues DC, Ross PJ, 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 Reports 11:1211-1225.

Deng X, Berletch JB, Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Disteche CM. 2013. Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. Dev Cell 25:55-68.

Devany E, Zhang X, Park JY, Tian B, Kleiman FE. 2013. Positive and negative feedback loops in the p53 and mRNA 3' processing pathways. Proc Natl Acad Sci U S A 110:3351-3356.

Disteche CM. 2012. Dosage compensation of the sex chromosomes. Annu Rev Genet 46:537-560.

Donehower LA, Soussi T, Korkut A, Liu Y, Schultz A, Cardenas M, Li X, Babur O, Hsu TK, Lichtarge O, et al. 2019. Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. Cell Rep 28:3010.

Dunford A, Weinstock DM, Savova V, Schumacher SE, Cleary JP, Yoda A, Sullivan TJ, Hess JM, Gimelbrant AA, Beroukhim R, et al. 2017. Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat Genet 49:10-16.

519 Dyer MA, Qadeer ZA, Valle-Garcia D, Bernstein E. 2017. ATRX and DAXX: Mechanisms and
520 Mutations. Cold Spring Harb Perspect Med 7.
521 Friedrich D, Friedel L, Finzel A, Herrmann A, Preibisch S, Loewer A. 2019. Stochastic
522 transcription in the p53-mediated response to DNA damage is modulated by burst
523 frequency. Mol Syst Biol 15:e9068.
524 Garrick D, Sharpe JA, Arkell R, Dobbie L, Smith AJ, Wood WG, Higgs DR, Gibbons RJ. 2006.
525 Loss of Atrx affects trophoblast development and the pattern of X-inactivation in
526 extraembryonic tissues. PLoS Genet 2:e58.
527 Gewandter JS, Bambara RA, O'Reilly MA. 2011. The RNA surveillance protein SMG1 activates
528 p53 in response to DNA double-strand breaks but not exogenously oxidized mRNA. Cell
529 Cycle 10:2561-2567.
530 Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers
531 D, Brooks AN, et al. 2020. Visualizing and interpreting cancer genomics data via the Xena
532 platform. Nat Biotechnol 38:675-678.
533 GTEx-Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580-
534 585.
535 Gulve N, Su C, Deng Z, Soldan SS, Vladimirova O, Wickramasinghe J, Zheng H, Kossenkov AV,
536 Lieberman PM. 2022. DAXX-ATRX regulation of p53 chromatin binding and DNA damage
537 response. Nat Commun 13:5033.
538 Hainaut P, Pfeifer GP. 2016. Somatic TP53 Mutations in the Era of Genome Sequencing. Cold
539 Spring Harb Perspect Med 6.
540 Hall LL, Byron M, Butler J, Becker KA, Nelson A, Amit M, Itskovitz-Eldor J, Stein J, Stein G,
541 Ware C, et al. 2008. X-inactivation reveals epigenetic anomalies in most hESC but identifies
542 sublines that initiate as expected. J Cell Physiol 216:445-452.
543 Haronikova L, Olivares-Illana V, Wang L, Karakostis K, Chen S, Fahraeus R. 2019. The p53
544 mRNA: an integral part of the cellular stress response. Nucleic Acids Res 47:3257-3271.
545 Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D,
546 Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The
547 ENCODE Project. Genome Res 22:1760-1774.
548 Haupt S, Caramia F, Herschtal A, Soussi T, Lozano G, Chen H, Liang H, Speed TP, Haupt Y.
549 2019. Identification of cancer sex-disparity in the functional integrity of p53 and its X
550 chromosome network. Nat Commun 10:5385.
551 He Q, Kim H, Huang R, Lu W, Tang M, Shi F, Yang D, Zhang X, Huang J, Liu D, et al. 2015. The
552 Daxx/Atrx Complex Protects Tandem Repetitive Elements during DNA Hypomethylation by
553 Promoting H3K9 Trimethylation. Cell Stem Cell 17:273-286.
554 Hewagama A, Gorelik G, Patel D, Liyanarachchi P, McCune WJ, Somers E, Gonzalez-Rivera T,
555 Michigan Lupus C, Strickland F, Richardson B. 2013. Overexpression of X-linked genes in T
556 cells from women with lupus. J Autoimmun 41:60-71.
557 Hu Y, Shi G, Zhang L, Li F, Jiang Y, Jiang S, Ma W, Zhao Y, Songyang Z, Huang J. 2016. Switch
558 telomerase to ALT mechanism by inducing telomeric DNA damages and dysfunction of ATRX
559 and DAXX. Sci Rep 6:32280.
560 Hung T, Pratt GA, Sundararaman B, Townsend MJ, Chaivorapol C, Bhangale T, Graham RR,
561 Ortmann W, Criswell LA, Yeo GW, et al. 2015. The Ro60 autoantigen binds endogenous
562 retroelements and regulates inflammatory gene expression. Science 350:455-459.
563 Jiang S, Cheng SJ, Ren LC, Wang Q, Kang YJ, Ding Y, Hou M, Yang XX, Lin Y, Liang N, et al.
564 2019. An expanded landscape of human long noncoding RNA. Nucleic Acids Res 47:7842-
565 7856.

566 Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schutz F, Daish T, Grutzner F,
567 Kaessmann H. 2012. Mechanisms and evolutionary patterns of mammalian and avian
568 dosage compensation. PLoS Biol 10:e1001328.
569 Kim S, Kim D, Cho SW, Kim J, Kim JS. 2014. Highly efficient RNA-guided genome editing in
570 human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res 24:1012-1019.
571 Kruse JP, Gu W. 2009. MSL2 promotes Mdm2-independent cytoplasmic localization of p53. J
572 Biol Chem 284:3250-3263.
573 Lafferty-Whyte K, Bilsland A, Hoare SF, Burns S, Zaffaroni N, Cairney CJ, Keith WN. 2010.
574 TCEAL7 inhibition of c-Myc activity in alternative lengthening of telomeres regulates hTERT
575 expression. Neoplasia 12:405-414.
576 Larsson AJM, Coucoravas C, Sandberg R, Reinius B. 2019. X-chromosome upregulation is
577 driven by increased burst frequency. Nat Struct Mol Biol 26:963-969.
578 Li B, Qing T, Zhu J, Wen Z, Yu Y, Fukumura R, Zheng Y, Gondo Y, Shi L. 2017. A
579 Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. Sci Rep
580 7:4200.
581 Li X, Wu L, Corsa CA, Kunkel S, Dou Y. 2009. Two mammalian MOF complexes regulate
582 transcription activation by distinct mechanisms. Mol Cell 36:290-301.
583 Linke SP, Sengupta S, Khabie N, Jeffries BA, Buchhop S, Miska S, Henning W, Pedeux R, Wang
584 XW, Hofseth LJ, et al. 2003. p53 interacts with hRAD51 and hRAD54, and directly modulates
585 homologous recombination. Cancer Res 63:2596-2605.
586 Lu WJ, Amatruda JF, Abrams JM. 2009. p53 ancestry: gazing through an evolutionary lens.
587 Nat Rev Cancer 9:758-762.
588 Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B,
589 Huelsken J, Omberg L, Gevaert O, et al. 2018. Machine Learning Identifies Stemness
590 Features Associated with Oncogenic Dedifferentiation. Cell 173:338-354 e315.
591 Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data. Trends Genet
592 29:537-544.
593 Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. 2014. CHOPCHOP: a CRISPR/Cas9
594 and TALEN web tool for genome editing. Nucleic Acids Res 42:W401-407.
595 Munoz-Fontela C, Mandinova A, Aaronson SA, Lee SW. 2016. Emerging roles of p53 and
596 other tumour-suppressor genes in immune regulation. Nat Rev Immunol 16:741-750.
597 Nguyen TT, Grimm SA, Bushel PR, Li J, Li Y, Bennett BD, Lavender CA, Ward JM, Fargo DC,
598 Anderson CW, et al. 2018. Revealing a human p53 universe. Nucleic Acids Res 46:8153-8167.
599 Oppel F, Tao T, Shi H, Ross KN, Zimmerman MW, He S, Tong G, Aster JC, Look AT. 2019. Loss
600 of atrx cooperates with p53-deficiency to promote the development of sarcomas and other
601 malignancies. PLoS Genet 15:e1008039.
602 Pageau GJ, Hall LL, Ganesan S, Livingston DM, Lawrence JB. 2007. The disappearing Barr
603 body in breast and ovarian cancers. Nat Rev Cancer 7:628-633.
604 Peng X, Thierry-Mieg J, Thierry-Mieg D, Nishida A, Pipes L, Bozinoski M, Thomas MJ, Kelly S,
605 Weiss JM, Raveendran M, et al. 2015. Tissue-specific transcriptome sequencing analysis
606 expands the non-human primate reference transcriptome resource (NHPRTR). Nucleic Acids
607 Res 43:D737-742.
608 Petropoulos S, Edsgard D, Reinius B, Deng Q, Panula SP, Codeluppi S, Plaza Reyes A,
609 Linnarsson S, Sandberg R, Lanner F. 2016. Single-Cell RNA-Seq Reveals Lineage and X
610 Chromosome Dynamics in Human Preimplantation Embryos. Cell 165:1012-1026.

611 Qin T, Mullan B, Ravindran R, Messinger D, Siada R, Cummings JR, Harris M, Muruganand A,
612 Pyaram K, Miklja Z, et al. 2022. ATRX loss in glioma results in dysregulation of cell-cycle
613 phase transition and ATM inhibitor radio-sensitization. Cell Rep 38:110216.
614 Ren W, Medeiros N, Warneford-Thomson R, Wulfridge P, Yan Q, Bian J, Sidoli S, Garcia BA,
615 Skordalakes E, Joyce E, et al. 2020. Disruption of ATRX-RNA interactions uncovers roles in
616 ATRX localization and PRC2 function. Nat Commun 11:2219.
617 Salomonis N, Dexheimer PJ, Omberg L, Schroll R, Bush S, Huo J, Schriml L, Ho Sui S,
618 Keddache M, Mayhew C, et al. 2016. Integrated Genomic Analysis of Diverse Induced
619 Pluripotent Stem Cells from the Progenitor Cell Biology Consortium. Stem Cell Reports
620 7:110-125.
621 Seah C, Levy MA, Jiang Y, Mokhtarzada S, Higgs DR, Gibbons RJ, Berube NG. 2008. Neuronal
622 death resulting from targeted disruption of the Snf2 protein ATRX is mediated by p53. J
623 Neurosci 28:12570-12580.
624 Sirchia SM, Ramoscelli L, Grati FR, Barbera F, Coradini D, Rossella F, Porta G, Lesma E,
625 Ruggeri A, Radice P, et al. 2005. Loss of the inactive X chromosome and replication of the
626 active X in BRCA1-defective and wild-type breast cancer cells. Cancer Res 65:2139-2146.
627 Syrett CM, Paneru B, Sandoval-Heglund D, Wang J, Banerjee S, Sindhava V, Behrens EM,
628 Atchison M, Anguera MC. 2019. Altered X-chromosome inactivation in T cells may promote
629 sex-biased autoimmune diseases. JCI Insight 4.
630 Tani H, Mizutani R, Salam KA, Tano K, Ijiri K, Wakamatsu A, Isogai T, Suzuki Y, Akimitsu N.
631 2012. Genome-wide determination of RNA stability reveals hundreds of short-lived
632 noncoding transcripts in mammals. Genome Res 22:947-956.
633 Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, Schumacher SE, Wang C, Hu H, Liu J, et
634 al. 2018. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. Cancer
635 Cell 33:676-689 e673.
636 Teng YC, Sundaresan A, O'Hara R, Gant VU, Li M, Martire S, Warshaw JN, Basu A,
637 Banaszynski LA. 2021. ATRX promotes heterochromatin formation to protect cells from G-
638 quadruplex DNA-mediated stress. Nat Commun 12:3887.
639 Vallot C, Ouimette JF, Rougeulle C. 2016. Establishment of X chromosome inactivation and
640 epigenomic features of the inactive X depend on cellular contexts. Bioessays 38:869-880.
641 Xiong Y, Chen X, Chen Z, Wang X, Shi S, Wang X, Zhang J, He X. 2010. RNA sequencing shows
642 no dosage compensation of the active X-chromosome. Nat Genet 42:1043-1047.
643 Yang L, Yildirim E, Kirby JE, Press W, Lee JT. 2020. Widespread organ tolerance to Xist loss
644 and X reactivation except under chronic stress in the gut. Proc Natl Acad Sci U S A 117:4262-
645 4272.
646 Younger ST, Rinn JL. 2017. p53 regulates enhancer accessibility and activity in response to
647 DNA damage. Nucleic Acids Res 45:9889-9900.
648 Yue X, Lan F, Xia T. 2019. Hypoxic Glioma Cell-Secreted Exosomal miR-301a Activates
649 Wnt/beta-catenin Signaling and Promotes Radiation Resistance by Targeting TCEAL7. Mol
650 Ther 27:1939-1949.
651

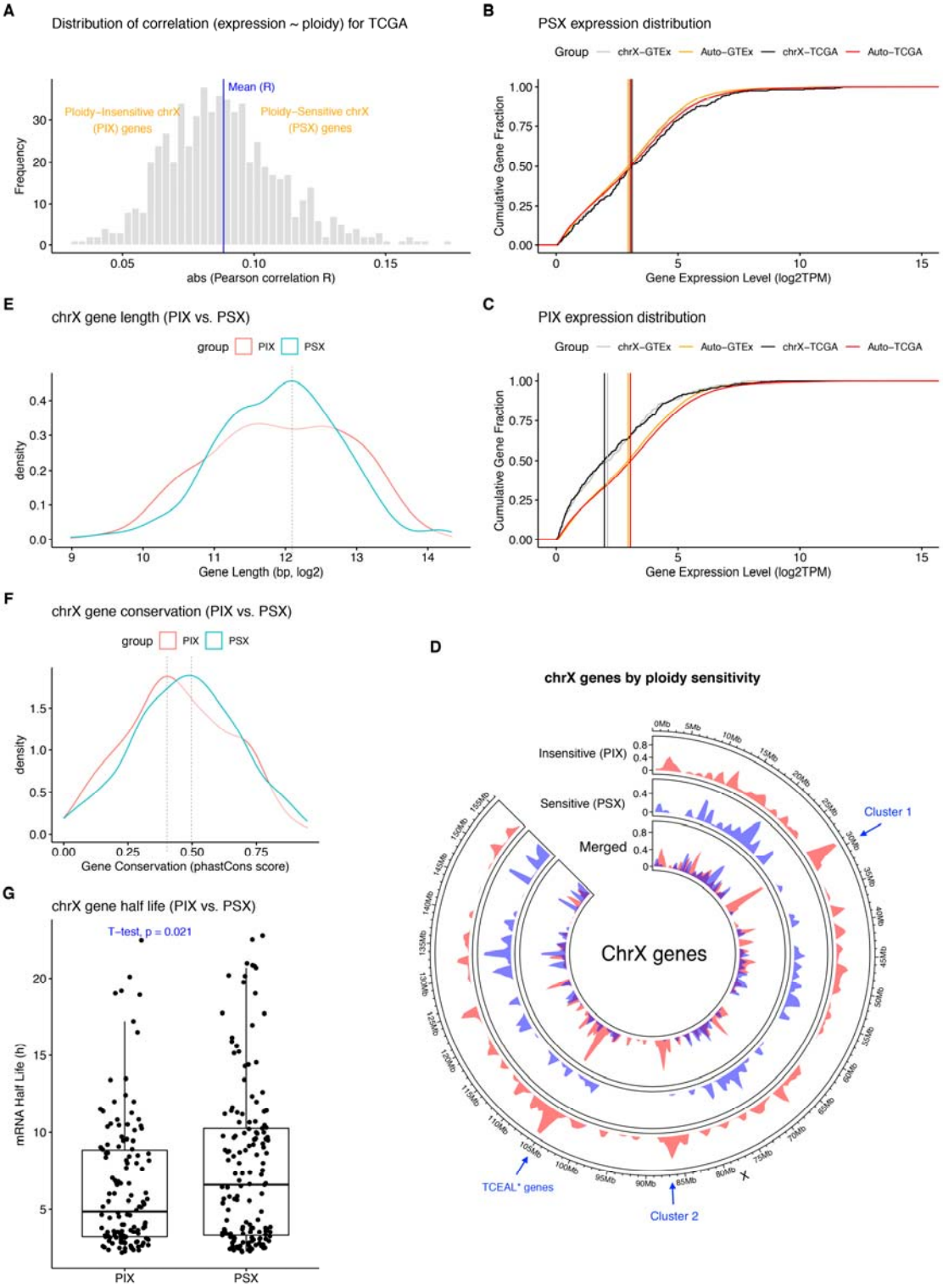
652

653

654

## Figures and Legends

Fig. 1


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)
B
PSX expression distribution
Group chrX−GTEx Auto−GTEx chrX−TCGA Auto−TCGA
Cumulative Gene Fraction
Gene Expression Level (log2TPM)
E
chrX gene length (PIX vs. PSX)
group PIX PSX
density
Gene Length (bp, log2)
C
PIX expression distribution
Group chrX−GTEx Auto−GTEx chrX−TCGA Auto−TCGA
Cumulative Gene Fraction
Gene Expression Level (log2TPM)
F
chrX gene conservation (PIX vs. PSX)
group PIX PSX
density
Gene Conservation (phastCons score)
D
chrX genes by ploidy sensitivity
Insensitive (PIX)
Sensitive (PSX)
Merged
ChrX genes
Cluster 1
TCEAL* genes
Cluster 2
G
chrX gene half life (PIX vs. PSX)
T−test, p = 0.021
mRNA Half Life (h)
PIX
PSX

664

665 **Fig. 1, Identification and characterization of PSX and PIX genes on X chromosome.** (**A**)
666 Classification X chromosome genes based on absolute values of Pearson correlations (x-axis)
667 compared to the average absolute correlation value (blue vertical line). (**B-C**) Cumulative
668 distribution of expression levels of PSX (**B**) and PIX (**C**) genes on X chromosome, compared
669 to autosomal genes, using samples from the TCGA project or the GTEx project. A total of 100
670 samples were randomly selected and their median expression levels for genes were shown
671 for visualization. Vertical lines indicate medians. (**D**) PIX and PSX genes distribution on X
672 chromosome. PIX enriched regions were marked by arrows. (**E-G**) Comparing gene length (**E**),
673 gene sequence conservation (**F**), and mRNA half-life (two-sided *t*-test, **G**) between PIX and
674 PSX genes. PIX: ploidy-insensitive X chromosome genes, PSX: ploidy-sensitive X chromosome
675 genes.
676
677
678
679
680
681
682
683
684
685
686
687
688

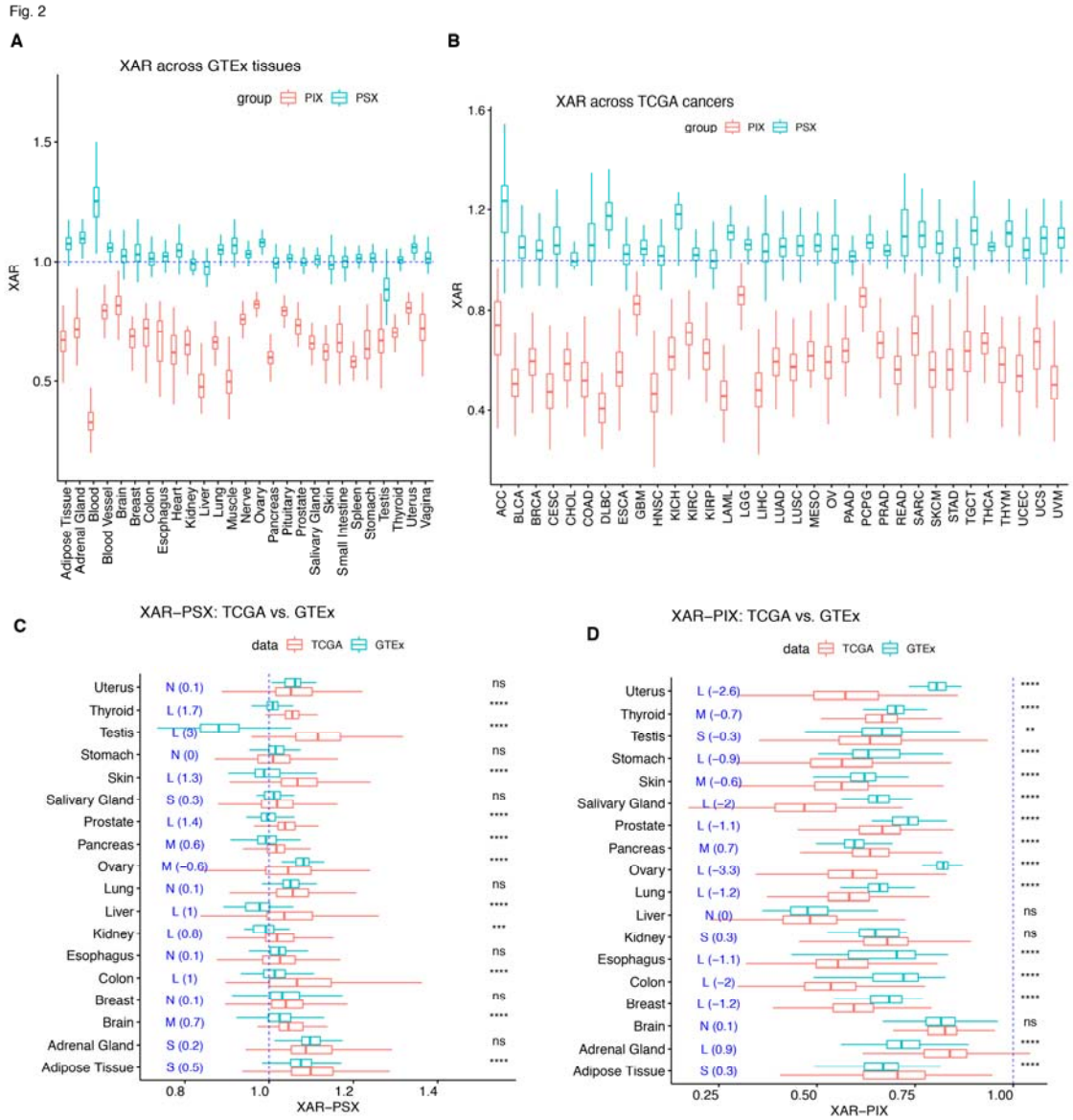



**Fig. 2, Differential dosages of PSX and PIX genes across normal tissues and cancer types.** (**A**) XAR-PSX and XAR-PIX across 27 normal tissues from GTEx were close to one, except for blood and testis. (**B**) XAR-PSX and XAR-PIX across 33 cancer types from TCGA. (**C**) XAR-PSX were higher in cancers than in matched normal tissues. Effect sizes were indicated by Cohen's D statistics in blue (negligible[N]<0.2, 0.2<=small[S]<0.5, 0.5<=moderate[M]<0.8,

695 large[L]>=0.8). (**D**) XAR-PIX were lower in cancers than in matched normal tissues. Effect
696 sizes were indicated by Cohen's D statistics in blue.
697 (**C-D**) Two-sided Wilcoxon test *P*-value significance: ns >0.05, * <= 0.05, ** <=0.01, *** <=
698 0.001, **** <= 0.0001. Matching of tumor types to normal tissues were as follows: ACC,
699 Adrenal Gland; BRCA, Breast; COAD, Colon; ESCA, Esophagus; GBM, Brain; HNSC, Salivary
700 Gland; KICH, Kidney; KIRC, Kidney; KIRP, Kidney; LGG, Brain; LIHC, Liver; LUAD, Lung; LUSC,
701 Lung; OV, Ovary; PAAD, Pancreas; PRAD, Prostate; PCPG, Adrenal Gland; READ, Colon; SARC,
702 Adipose Tissue; SKCM, Skin; STAD, Stomach; TGCT, Testis; THCA, Thyroid; UCS, Uterus; UCEC,
703 Uterus.
704 XAR: X-over-autosome dosage ratio, ACC: Adrenocortical carcinoma, BLCA: Bladder
705 urothelial Carcinoma, BRCA: Breast invasive carcinoma, CESC: Cervical squamous cell
706 carcinoma and endocervical adenocarcinoma, CHOL: Cholangiocarcinoma, COAD: Colon
707 adenocarcinoma, CCRCC: Clear cell renal cell carcinoma, DLBC: Lymphoid neoplasm diffuse
708 large B-cell lymphoma, ESCA: Esophageal carcinoma, GBM: Glioblastoma multiforme, HNSC:
709 Head and neck squamous cell carcinoma, KICH: Kidney chromophobe, KIRC: Kidney renal
710 clear cell carcinoma, KIRP: Kidney renal papillary cell carcinoma, LAML: Acute myeloid
711 leukemia, LGG: Brain lower grade glioma, LIHC: Liver hepatocellular carcinoma, LUAD: Lung
712 adenocarcinoma, LUSC: Lung squamous cell carcinoma, MESO: Mesothelioma, OV: Ovarian
713 serous cystadenocarcinoma, PAAD: Pancreatic adenocarcinoma, PCPG: Pheochromocytoma
714 and paraganglioma, PDA: Pancreatic ductal adenocarcinoma, PRAD: Prostate
715 adenocarcinoma, READ: Rectum adenocarcinoma, SARC: Sarcoma, SKCM: Skin Cutaneous
716 melanoma, STAD: Stomach adenocarcinoma, TGCT: Testicular germ cell tumors, THCA:
717 Thyroid carcinoma, THYM: Thymoma, UCEC: Uterine corpus endometrial carcinoma, UCS:
718 Uterine carcinosarcoma, UVM: Uveal melanoma.

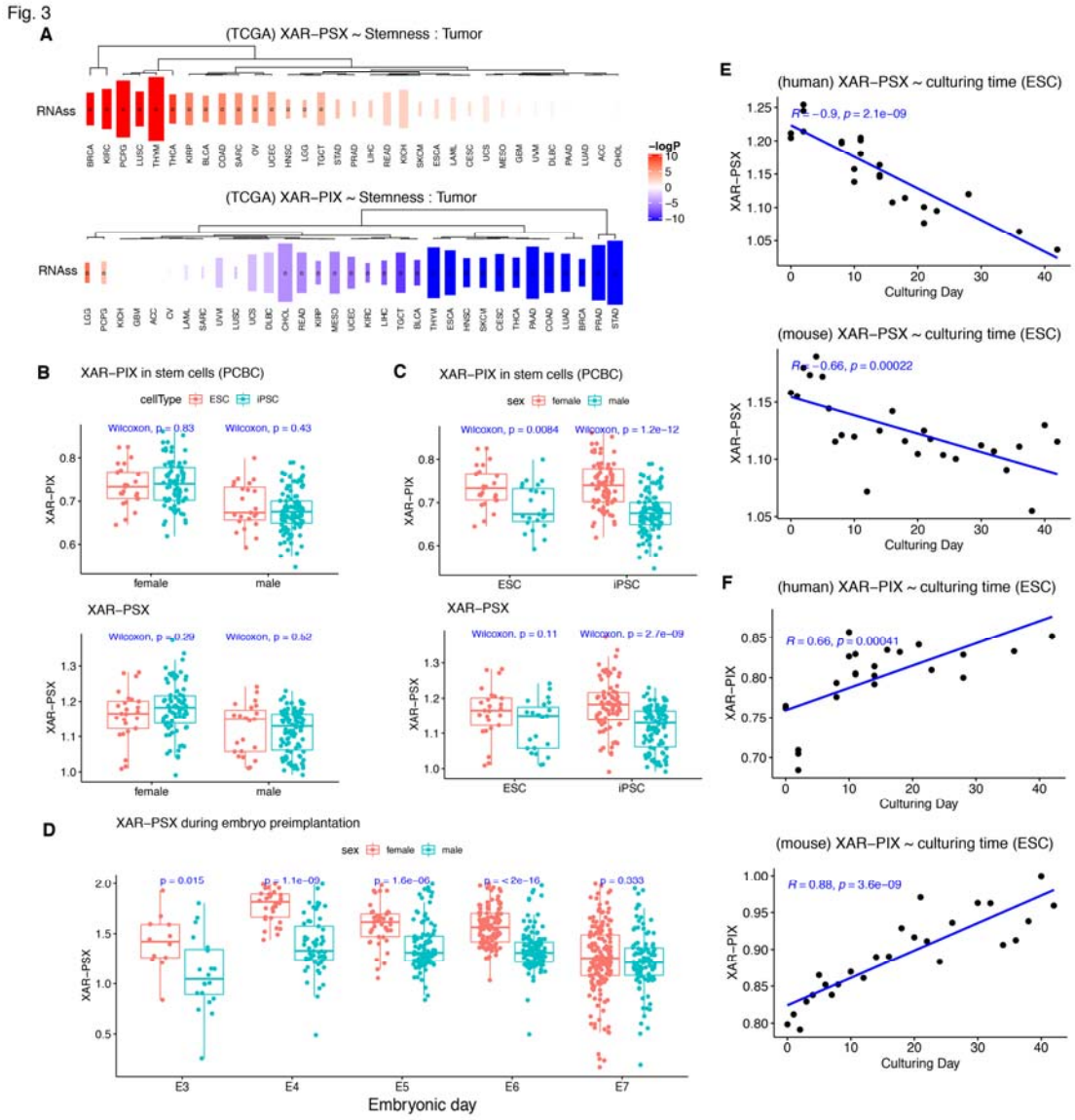
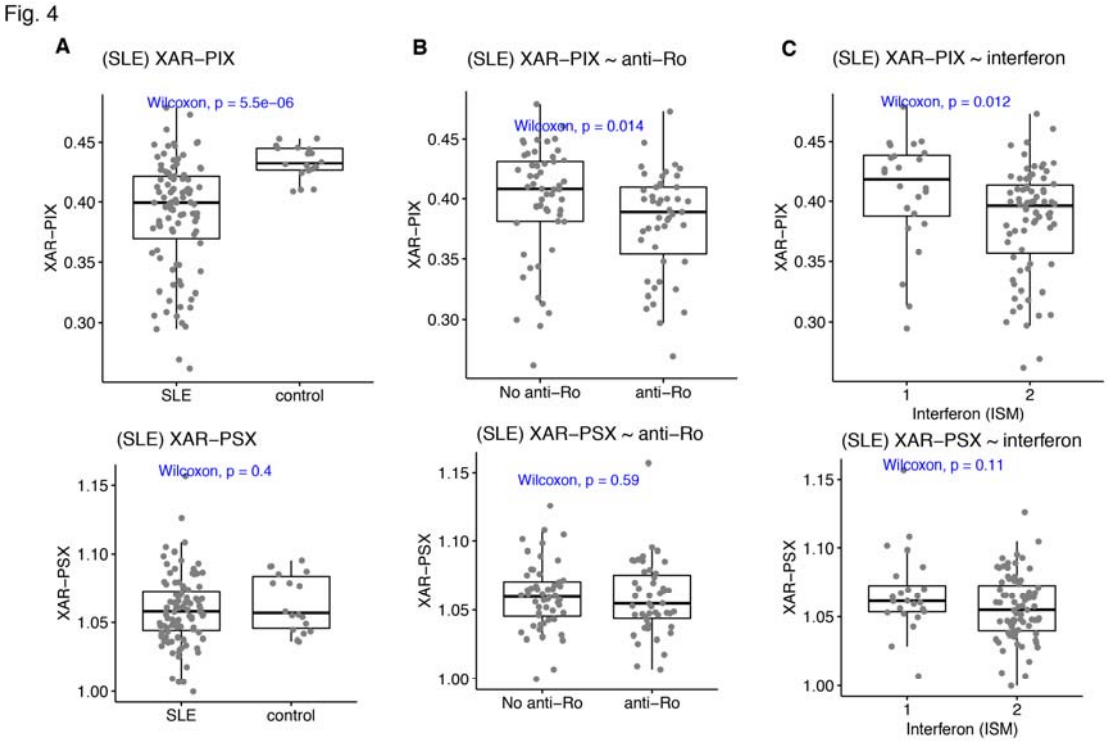



**Fig. 3, XAR-PSX and XAR-PIX were inversely associated with stemness.** (**A**) Cancer stemness (RNAss) was positively correlated with XAR-PSX (top) and negatively with XAR-PIX (bottom). colors indicate -log10 P-values, and sizes of rectangles represent Pearson correlation. Correlations with FDR<0.1 were marked by circles. (**B**) Comparison of XAR-PIX (top) and

726 XAR-PSX (bottom) between ESCs and iPSCs from females or males. (**C**) Comparison of XAR-
727 PIX (top) and XAR-PSX (bottom) between sexes from ESCs or iPSCs. (**D**) XAR-PSX dynamics of
728 embryonic stem cells (grouped by sex) during preimplantation from day 3 (E3) to 7 (E7). (**E**)
729 Strong negative correlation of XAR-PSX with differentiation culturing time in both human
730 (top) and mouse (bottom). (**F**) Strong positive correlation of XAR-PIX with differentiation
731 culturing time in both human (top) and mouse (bottom). (**B-D**) Two-sided Wilcoxon test.
732



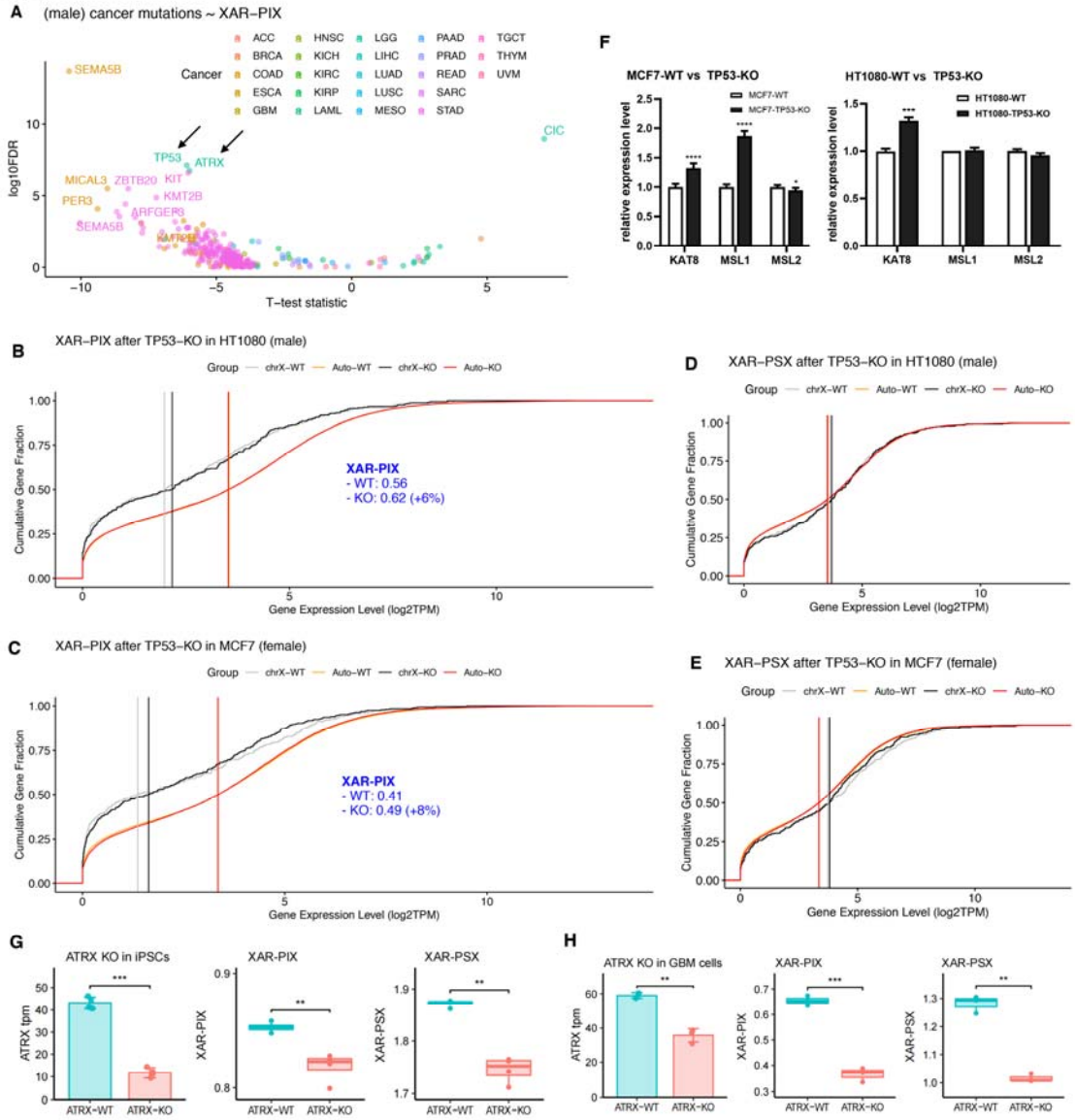

733

734 **Fig. 4, XAR in autoimmune disease.** (**A**) XAR-PIX (top) and XAR-PSX (bottom) in SLE patients
735 compared to healthy controls. (**B**) XAR-PIX (top) and XAR-PSX (bottom) in SLE patients with
736 anti-Ro autoantibodies compared to those without anti-Ro. (**C**) XAR-PIX (top) and XAR-PSX

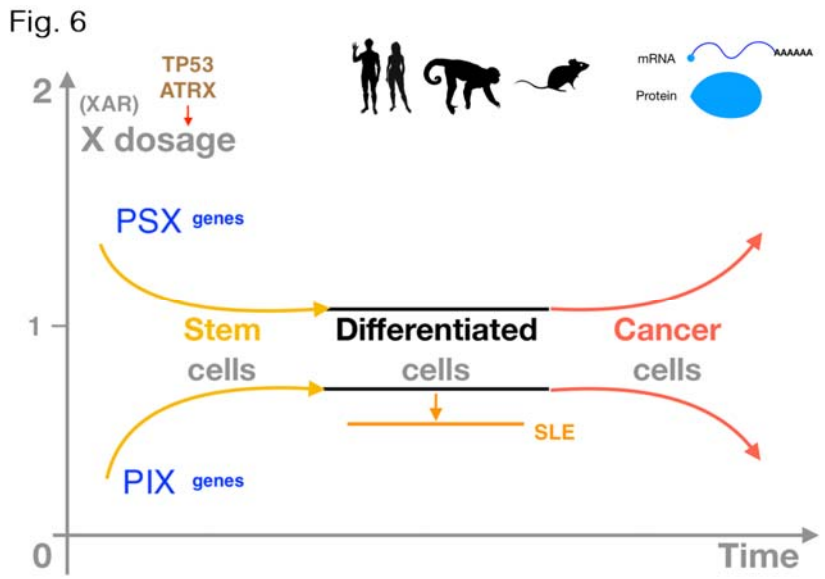
(bottom) in SLE patients with more active interferon response (ISM) compared to those with less ISM.

(**A-C**) Two-sided Wilcoxon test.

Fig. 5



743 **Fig. 5, *TP53* and *ATRX* were two regulators of XAR.** (**A**) Significance of XAR-PIX difference (y-
744 axis, two-sided *t*-test) between wildtype and mutated genes across cancers in male. *TP53*
745 and *ATRX* were marked by arrows. For better visualization, only genes with FDR<1 were
746 shown as points and those with FDR<1e-4 labeled. (**B-C**) Cumulative distribution of PIX and
747 autosomal gene expression levels for HT1080 (**B**) and MCF7 (**C**), before and after *TP53*-KO.
748 Vertical lines indicate median expression used to calculate XAR-PIX dosages. (**D-E**)
749 Cumulative distribution of PSX and autosomal gene expression levels for HT1080 (**D**) and
750 MCF7 (**E**), before and after *TP53*-KO. Vertical lines indicate median expression used to
751 calculate XAR-PSX dosages. (**F**) Gene expression changes of MSL complex subunits *KAT8*,
752 *MSL1*, and *MSL2* in MCF7 (left) and HT1080 (right) cells, before and after *TP53*-KO. (**G**) *ATRX*
753 knockout (KO) in mouse induced pluripotent stem cells (iPSCs) significantly reduced both
754 XAR-PIX (middle) and XAR-PSX (right). (**H**) *ATRX* knockout (KO) in human GBM cells
755 significantly reduced both XAR-PIX (middle) and XAR-PSX (right). (**F-H**) Two-sided *t*-test, *
756 *P*<= 0.05, ** <=0.01, *** <= 0.001, **** <= 0.0001.
757




758

759 **Fig. 6**, **Major findings of this study**. Differential dosages of two groups of X chromosome
760 genes (PIX and PSX) were observed in human and animals at both mRNA and protein levels.
761 X dosages were converging during stem cell differentiation but diverging in cancer. X dosage
762 compensation was lower in blood cells of SLE autoimmune patients than in healthy
763 individuals. TP53 and ATRX were identified and validated as two XAR regulators. XAR: X-
764 over-autosome dosage ratio, PIX: ploidy-insensitive X genes, PSX: ploidy-sensitive X genes.

765