Identifying new regulators of evolutionarily conserved differential chromosome X dosage compensation

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 gene groups (ploidy-sensitive chrX [PSX] and ploidy-insensitive chrX [PIX] genes) on chromosome X, with distinct sensitivity to genome ploidy and showing distinct but evolutionarily-conserved (across tissues in both primates and mouse) dosage compensations (X-over-Autosome dosage Ratio, or XAR). We also demonstrated that the trends of XAR-PIX and XAR-PSX in cancers were divergent, with XAR-PIX down-regulated and XAR-PSX up-regulated, compared to corresponding normal tissues, which was supported at both the RNA and protein level across tissue types. We further

explored potential XAR regulators systematically and identified and experimentally validated two new independent regulators, mild *TP53* (verified in human) and potent *ATRX* (verified in both human and mouse), expanding the current dosage compensation machinery with only the MOF complex in mammals. Lastly, we presented interesting findings of XAR-PIX and XAR-PSX dynamics during stem cell differentiation and embryo preimplantation, as well as in female-predominant lupus patients. For example, we showed the opposite trends of XARs during stem cell differentiation and in cancer. Collectively, we provided insights to further unravel the mystery of X dosage compensation during evolution.

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). 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 chromosome X 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 chromosome X (Deng, et al. 2013). 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 and colleagues 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 chromosome X (Larsson, et al. 2019). However, these factors were unknown and the hypothesis of enhancer factors in X upregulation was not verified.

X dosage plays an important role in cancer initiation, because excessive X upregulation by somatic X reactivation under chronic stress leads to cancer (Yang, et al. 2020). Although X upregulation has been contributed to the loss of Barr body (the inactivated X chromosome) in breast and ovarian cancers (Pageau, et al. 2007), 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) (Davoli, et al. 2013). 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 noncoding RNA (lncRNA) genes, to uncover the elusive mechanism of dosage compensation in humans. However, these possibilities have not been explored yet.

71 In this work, we identified two distinct groups of genes (PIX and PSX) on chromosome X and
72 characterized their dosage compensations in about 20,000 RNA sequencing (RNA-Seq)
73 samples from TCGA and the Genotype-Tissue Expression (GTEx) project (GTEx-Consortium
74 2013) and thousands of proteomic samples from Clinical Proteomic Tumor Analysis
75 Consortium (CPTAC), providing a systemic view of both XAR-PIX and XAR-PSX in human. We
76 found that the differential dosage between PIX and PSX genes were evolutionarily conserved.
77 We also identified and experimentally validated *TP53* as a new specific regulator of XAR-PIX
78 and *ATRX* as a new regulator of both XAR-PIX and XAR-PSX. We further explored the dynamics
79 of both XAR-PIX and XAR-PSX during stem cell differentiation and embryo preimplantation,
80 and in female-predominant autoimmune diseases. Our work provided insights into
81 understanding of physiological and pathological roles of the evolutionarily conserved
82 differential X chromosome dosage compensation in humans.
83

84 # Results

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

86

87 We are interested in which genes on chrX require less dosage compensation, since only a
88 subset of chrX genes were affected after *KAT8* knockdown (Deng, et al. 2013). To this end, we
89 examined chrX gene sensitivities to genome ploidy, by correlating protein-coding gene
90 expression levels with genome ploidies of cancer samples from TCGA (Taylor, et al. 2018),
91 after filtering out non-coding genes and genes with extremely low expression in normal
92 samples from the Genotype-Tissue Expression (GTEx) project (GTEx-Consortium 2013) (n=615
93 chrX genes and n=16,622 autosome genes remained, **Table S1**). We identified two groups of
94 chrX genes (**Fig. 1A,** termed ploidy-sensitive chrX [PSX] genes, and ploidy-insensitive chrX [PIX]

genes, see **Methods**) showing distinct expression distribution. Interestingly, PSX genes (n=286, **Table S1**) had highly similar expression pattern to autosome genes (**Fig. 1B**), while PIX genes (n=329, **Table S1**) had much lower expression levels, with a median of about 2/3 of PSX genes expression (**Fig. 1C**). Of note, no preference for genes escaping X inactivation (Dunford, et al. 2017) was observed in PSX or PIX.

We found three PIX-specific gene clusters on chrX, compared to PSX genes (**Fig. 1D**). Interestingly, the PIX cluster located at around 100Mb (hg38) contained eight out of nine Transcription Elongation Factor A Protein-Like (TCEAL) genes, encoding nuclear phosphoproteins (Pillutla, et al. 1999). Although most genes from this family were not intensively studied, most reports suggested that they acted as tumor suppressors, especially for *TCEAL7* (Lafferty-Whyte, et al. 2010; Yue, et al. 2019). For these tumor suppressors, lower sensitivity to dosage may protect both males and females from cancer equally. Interestingly, the chrX region around 100Mb showed the lowest expression level after zygote genome activation (ZGA) at E4 during embryo implantation (Petropoulos, et al. 2016).

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 2kb to 8kb, while lengths of PSX genes peaked at around 4kb (**Fig. 1E**). This trend was similar when including introns (**Fig. S1A**). Although GC content of mRNAs is critical for their storage and decay (Courel, et al. 2019), we found similar GC content between PIX and PSX genes, although slightly higher in PSX genes (**Fig. S1B**). Moreover, we

observed lower conservation of PIX genes, which had lower expression, than PSX genes (**Fig. 1F**), which resembled the faster-X effect in non-coding regions on chrX (Meisel and Connallon 2013).

RNA half-lives (one type of post-transcriptional regulation) were reported to be longer for X-linked than autosomal genes (Deng, et al. 2013). Inspired by this, we examined the mRNA half-lives of PIX and PSX genes by reanalyzing data from a previous study (Tani, et al. 2012). We found that PIX (4.9h) tended to have shorter mRNA half-lives than PSX genes (6.6h) (**Fig. 1G**), consistent with lower expression levels of PIX than PSX genes.

Evolutionally conserved differential dosage compensation between PIX and PSX genes

Based on these two groups of chrX genes (PIX and PSX), we calculated two dosage ratios of chromosome X over autosome (XAR, see **Methods**) to evaluate their dosage compensation, termed XAR-PIX and XAR-PSX, respectively. As expected, XAR-PSX dosages were around one across GTEx tissues (**Fig. 2A**), but mostly above one across TCGA cancer types (**Fig. 2B**). In contrast, XAR-PIX dosages were much lower than XAR-PSX and fluctuating around 0.6 both across GTEx tissues (**Fig. 2C**) and across TCGA cancer types (**Fig. 2D**). Furthermore, the difference between XAR-PIX and XAR-PSX were also observed in non-human primates (**Fig. 2E**) and mouse (**Fig. 2F**) 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

143 types, we observed that XAR-PSX dosages were significantly increased (**Fig. 2G**), while XAR-
144 PIX showed the opposite trend (**Fig. 2H**), compared to normal tissues from GTEx. Moreover,
145 most magnitudes of changes (or effect sizes) of both XAR-PSX and XAR-PIX were moderate to
146 large, as indicated by the Cohen's D statistics. Of note, most tissues or cancer types showed
147 no XAR difference between males and females, especially for XAR-PIX (**Fig. S2A-D**), which
148 suggests that the possibility of XAR-PSX upregulation due to reactivation of the inactivated X
149 in females can be excluded. Accordingly, when investigating proteomic data generated by
150 mass spectrometry, the opposite trends between XAR-PIX and XAR-PSX were also observed,
151 comparing primary tumors with normal tissue samples in CPTAC **(Fig. 2I-J)**.
152
153 **Potential regulators of X dosage compensation were enriched in histone modification**
154
155 The MSL complex was known to regulate XAR through H4K16 acetylation in both human and
156 mouse (Deng, et al. 2013). We found that only about 8.8% (29/329) and 10.5% (30/286) of
157 PIX and PSX genes showed expression changes of more than 1.2-fold after knockdown of the
158 key MSL component *KAT8* in mouse (Deng, et al. 2013). To identify other new regulators of
159 XAR, we examined TCGA mutation spectrums that may be associated with XAR in each cancer
160 type. We reasoned that mutated XAR regulators should result in differential XAR compared
161 with wildtype regulators. In total, we obtained 209 and 515 potential regulators (FDR < 0.5)
162 for XAR-PIX and XAR-PSX in male tumors, respectively (**Fig. 3A, Fig. S3A**). For the top 200
163 regulators, gene ontology (GO) analysis was performed and similar GO terms were found for
164 XAR-PIX and XAR-PSX regulators, mainly about post-translational modifications, including
165 histone methylation, peptidyl-lysine methylation, and protein methylation (**Fig. 3B-C**). These
166 results indicates the impact on dosage compensation in tumor by histone modifications,

167 which were known to influence the state and structure of chromatin and thus regulate
168 chromatin accessibility and gene expression (Millan-Zambrano, et al. 2022). Among the
169 potentially new regulators of XAR, we found that *TP53* and *ATRX* were promising regulators
170 for XAR-PIX in both males (**Fig. 3A**) and females (**Fig. S3B**) and both genes are involved in
171 genome-wide histone modifications (Allison and Milner 2004; Dyer, et al. 2017); therefore,
172 we selected these two genes for further study.
173
174 **TP53 is a new regulator of PIX-specific dosage compensation**
175
176 *TP53* mutations are the most frequent events across all cancers (Hainaut and Pfeifer 2016),
177 and more than 91% of tumors with *TP53* mutations show functional loss of both *TP53* copies
178 (Donehower, et al. 2019). Interestingly, *TP53* mutations contributed to the sex discrepancy of
179 cancer incidences by differentially regulating X-linked genes between two sexes (Haupt, et al.
180 2019). However, it is unknown whether *TP53* regulates X dosage. *TP53* was reported to
181 interact with key MSL components, *KAT8* and *MSL2* (Kruse and Gu 2009; Li, et al. 2009). We
182 found three and four p53 binding sites (Nguyen, et al. 2018) in *KAT8* (one site in promoter)
183 and *MSL2* (three sites in promoter), respectively, which indicates direct regulation of *KAT8*
184 and *MSL2* transcription by p53. Of note, *TP53* mRNA itself is a critical regulator of cellular
185 stress (Haronikova, et al. 2019). Meanwhile, we found that *TP53* mutations were associated
186 with lower XAR-PIX (**Fig. 3A**). However, we observed that *TP53* mRNA was universally up-
187 regulated in cancer samples, compared with non-cancerous samples, across cancer types (**Fig.
188 S3C**). These results were consistent with a previous study finding that the deadenylase
189 poly(A)-specific ribonuclease (PARN) destabilized *TP53* mRNA while p53 increase activated

190 PARN (Devany, et al. 2013). There exists a negative feedback loop between *TP53* mRNA and
191 p53 protein levels.
192
193 To validate our finding, we constructed *TP53*-knockout (*TP53*-KO) cells from two *TP53*-
194 wildtype (*TP53*-WT) cell lines, one male (HT1080) and one female (MCF7, containing three
195 active X and no inactive X chromosomes (Sirchia, et al. 2005)), by using CRISPR technologies
196 (**Fig. 3D**, see **Methods**), followed by RNA-seq. We found that the expression distributions of
197 PSX and autosome genes were almost unchanged (**Fig. S3D** and **Fig. S3E**). In contrast, we
198 observed XAR-PIX dosage increase after *TP53*-KO in both sexes (**Fig. 3E** and **Fig. 3F**). We
199 further demonstrated that *TP53* may regulate XAR-PIX by repressing KAT8, since *TP53*-KO
200 enhanced *KAT8* expression in both sexes (**Fig. S3F** and **Fig. S3G**). *TP53*-KO had a larger effect
201 in female cells (8% vs. 6% increase of XAR-PIX) than male cells, which may due to MCF7 cells
202 containing three active X chromosomes (Sirchia, et al. 2005) and *TP53* affected *MSL1* in
203 female cells only.
204
205 ***ATRX* is a new positive regulator of both PIX and PSX dosage compensations**
206
207 *ATRX* is an X-linked chromatin remodeler that functions in genome stabilization and
208 heterochromatin formation (Dyer, et al. 2017; Teng, et al. 2021), regulating transcription in
209 both activation and repression manners (Levy, et al. 2015; Voon, et al. 2015). We noted that
210 *ATRX* mutation was associated with lower XAR-PIX in tumor samples (**Fig. 3A**). Accordingly,
211 we observed significant positive correlations between *ATRX* mRNA expression and XAR-PIX
212 across 33 tumors (**Fig. S4A**), but weak correlations with XAR-PSX (**Fig. S4B**). To further confirm
213 the association between ATRX and XAR, we investigated XAR before and after *ATRX* knockout

214 in another two studies. Interestingly, *ATRX* knockout reduced both XAR-PIX and XAR-PSX in
215 mouse (**Fig. 4A-C**), suggesting that *ATRX* might help to maintain dosage compensation.
216 Moreover, XAR-PIX was more affected than XAR-PSX after *ATRX* knockout in GBM cell line (**Fig.**
217 **4D-F**), consistent with XAR-PIX being more correlated to *ATRX* expression and PIX genes under
218 more stringent dosage control.
219
220 Next, we examined whether X-linked genes are more sensitive to the change of *ATRX*
221 expression. For autosomal and X-linked genes co-expressed with *ATRX*, we compared their
222 coefficient estimates of *ATRX* expression by linear regression models. We reasoned that larger
223 absolute values of ATRX coefficients indicate higher impact on gene expression. As a result,
224 *ATRX* regression coefficients were significantly higher in X-linked genes than in autosomal
225 genes across a variety of tumors, at both mRNA (**Fig. 4G**) and protein level (**Fig. 4H**),
226 suggesting that *ATRX* might be more closely related to X-linked genes. Besides, *ATRX* loss
227 could lead to changes in chromatin accessibility in liver cancer cell line (Liang, et al. 2020) and
228 we noted that more X-linked regions became less accessible, in agreement with the
229 observation of reduced XAR after *ATRX* knockout. Taken together, our results indicated that
230 *ATRX* is a new regulator of X dosage compensation, especially for XAR-PIX.
231
232 **Potential lncRNA regulators of XAR**
233
234 LncRNAs regulate gene expression at multiple levels, including chromatin remodeling,
235 transcription and post-transcription regulation (Statello, et al. 2021). One of the best-known
236 examples of gene expression regulation by lncRNA is X inactive-specific transcript (*XIST*)-
237 mediated X inactivation in females (Loda, et al. 2022). Considering the important roles of

238 lncRNAs in gene expression regulation, we speculated that there were other lncRNAs related
239 to dosage compensation to be explored. Therefore, we correlated lncRNA with XAR in each
240 cancer in male, and obtained 2,689 and 31 lncRNAs significantly related to XAR-PIX and XAR-
241 PSX, respectively. Next, we focused on the co-expressed genes and RNA-protein interactions
242 of these lncRNAs, which might provide clues for how these lncRNAs affect dosage
243 compensation in tumors. As a results, co-expressed genes of these lncRNAs were most
244 enriched in histone modification (**Fig. 3G**), consistent with the results in potential regulators
245 of XAR (**Fig. 3B-C**), suggesting that lncRNAs might involve in dosage compensation by
246 regulating genes related to histone modification. In addition, genes functioning in nuclear
247 transport, chromatin organization, and mRNA metabolic processes were enriched. In terms
248 of predicted interactions with XAR-associated lncRNAs, proteins involved in RNA processing
249 and stability as well as translational regulation were enriched (**Table S2**).
250
251 **XAR and stemness were closely related**
252
253 Increased stemness is an important feature or cancer cells (Malta, et al. 2018). We thus
254 examined the relationship between XAR and stemness. We observed mostly positive
255 correlations with cancer stemness for XAR-PSX (**Fig. 5A**), but mostly negative correlations for
256 XAR-PIX (**Fig. 5B**), across cancer types. We further examined XAR in embryonic stem cells (ESC)
257 and induced pluripotent stem cells (iPSC) using gene expression data from the Progenitor Cell
258 Biology Consortium (PCBC) (Salomonis, et al. 2016) . As expected, we found no difference
259 between ESC and iPSC, for both XAR-PIX and XAR-PSX (**Fig. 5C** and **Fig. 5D**), probably because
260 ESC and iPSC had similar expression pattern (Bock, et al. 2011). However, we found much

261 lower XAR-PIX and XAR-PSX in male than in female iPSC (**Fig. 5E-F**), but the signals of sex
262 difference in ESC were weak.
263
264 We further examined XAR dynamics during iPSC differentiation culturing. Interestingly, we
265 found that XAR-PSX was reduced along iPSC culturing time (**Fig. 5G-H**), but XAR-PIX showed
266 the opposite trend (**Fig. 5I-J**), in both human and mouse (Barry, et al. 2017). These patterns
267 were consistent with overall reduction of XAR-PIX and elevation of XAR-PSX across cancer
268 types above. Therefore, our results demonstrated that chromosome X genes dosages were
269 converging during stem cell differentiation but diverging in cancer cells (**Fig. 5K**).
270
271 **XAR dynamics during human embryo implantation**
272
273 We further explored the XAR dynamics during embryo implantation, when female embryo
274 cells undergo both chromosome X inactivation and X upregulation by dosage compensation,
275 while male embryo cells undergo X upregulation. We reanalyzed single-cell (n=1,529 cells)
276 RNA sequencing data of human preimplantation embryos ranging from day 3 to 7
277 (Petropoulos, et al. 2016). Consistent with the original report, for XAR-PSX, both male and
278 female cells peaked at day 4 when ZGA happened (**Fig. 6A**). Median XAR-PSX in females
279 reached near 2.0 but only 1.5 in males. However, XAR-PIX dosages were much lower than
280 those from adult tissues (GTEx) and were similar between day 3 and day 4, after which XAR-
281 PIX was decreased to an extremely low level around 0.1 (**Fig. 6B**). Furthermore, higher XAR-
282 PSX in female than in male embryo cells was most significant at day 6, especially for both
283 primitive endoderm (PE) and trophectoderm (TE) (**Fig. 6C**) cells.
284

285 **Insufficient X gene dosage in autoimmune patients**

286

287 Besides cancer, we were also interested in whether XAR was also dysregulated in female-
288 predominant (female-male ratio is about 9:1) systemic lupus erythematosus (SLE), which is a
289 prototypical autoimmune disease characterized by systemic organ damages and abundant
290 autoantibodies, such as anti-Ro (Hung, et al. 2015). However, its pathogenesis and extreme
291 sex bias were still incompletely understood. By reanalyzing a previous dataset from SLE
292 patients and healthy controls (Hung, et al. 2015), we found that XAR-PIX, but not XAR-PSX,
293 were specifically down-regulated in SLE patients compared with healthy controls (**Fig. 6D** and
294 **Fig. S5A**), which means insufficient dosage of X genes in SLE patients. Moreover, SLE patients
295 with anti-Ro autoantibodies showed lower XAR-PIX dosage than those without anti-Ro
296 autoantibodies (**Fig. 6E** and **Fig. S5B**). Similarly, SLE patients with more active interferon
297 response (indicated by the Interferon Signature Metric, or ISM) showed lower XAR-PIX dosage
298 (**Fig. 6F** and **Fig. S5C**). These findings may provide insights to understand SLE pathogenesis.

299

300 Discussions

301 Here, we identified two distinct groups of genes on chromosome X, whose differential dosage
302 compensations across tissues were observed at both RNA- and protein-level and were
303 evolutionarily conserved in primates and in mouse. We explored the potential XAR regulators
304 (both protein-coding and lncRNA) and identified XAR-PIX-specific regulator *TP53* and general
305 XAR regulator *ATRX*, and experimentally verified them. We further investigated XAR-PIX and
306 XAR-PSX associations with stemness of cancer cells and embryonic stem cells and with sex-
307 biased autoimmune disease. Our findings of distinct characteristics of XAR-PIX and XAR-PSX
308 may help resolve controversies described in our previous study (Xiong, et al. 2010).

309 Collectively, integrating large-scale sequencing data from both cancer and normal samples
310 greatly enhanced our ability to investigate both pathological and physiological roles of X
311 dosage compensation.
312
313 The frequency of transcriptional bursts and possible trans-acting enhancer-binding regulators
314 were proposed to be the driving force behind X upregulation in mouse (Larsson, et al. 2019).
315 Interestingly, p53 can bind pervasively to genome-wide enhancers (Younger and Rinn 2017)
316 and regulate the transcriptional burst of target genes (Friedrich, et al. 2019), further
317 supporting our finding of *TP53* regulating XAR. Moreover, the crosstalk between *TP53* and
318 nonsense-mediated decay (NMD) pathway may contribute to XAR regulation by *TP53*
319 (Gewandter, et al. 2011; Deng, et al. 2013). Therefore, *TP53* may regulate XAR by various
320 mechanisms. We expect XAR regulation by *TP53* would be conserved among mammals, due
321 to ultra-conservation of *TP53* during evolution (Lu, et al. 2009). However, further efforts are
322 needed to support our hypothesis.
323
324 XAR in most cancers showed no difference between sexes, suggesting negligible contribution
325 to XAR changes in cancer by reactivation of the inactivated X in females. Since *TP53* mutation-
326 induced aneuploidies were not specific to chromosome X, they were unlikely to drive both
327 XAR-PSX and XAR-PIX to change in such a coordinated way across cancers, as observed in this
328 study. At least, XAR-PIX should be less affected by aneuploidies, by PIX definition.
329
330 Several limitations of this study represent directions for future research. First, it is unknown
331 how distinct dosages of PIX and PSX genes were maintained and why did they show divergent
332 trends in cancers compared to normal tissues. Second, despite the high cost and long period

333 of CRISPR-based *TP53*-KO experiments, it would further solidify our conclusion if CRISPR-
334 mediated *TP53*-KO can be established in more cell lines of different lineages. Third, additional
335 XAR regulators remain to be identified by other ways, because this analysis was limited by
336 mutation frequencies in cancer patients. Genome-wide screening experiments in the future
337 probably will identify more XAR regulators. Until then, the problem of abnormal XAR-PIX and
338 XAR-PSX in cancer may be solved.
339
340

341 # Materials and Methods

342 **Datasets**

343 Gene expression levels (transcripts per million, or TPM) and related clinical information of
344 RNA-seq samples from The Cancer Genome Atlas project (TCGA,
345 https://portal.gdc.cancer.gov) and the Genotype-Tissue Expression project (GTEx,
346 https://www.gtexportal.org), processed by the same computational pipeline, were obtained
347 from the UCSC Xena database (https://xenabrowser.net) (Goldman, et al. 2020). For GTEx,
348 tissues with less than 30 samples were excluded. Gene expression levels of cell lines
349 sequenced by the Progenitor Cell Biology Consortium (PCBC) were downloaded from Synapse
350 (https://www.synapse.org/#!Synapse:syn1773109/wiki/54962). Gene expression levels of
351 ESCs during differentiation culturing for both human and mouse were downloaded from
352 (Barry, et al. 2017). Somatic mutations for TCGA tumor samples were obtained the GDC Portal
353 (https://portal.gdc.cancer.gov). Genome ploidy information for TCGA samples were
354 downloaded from (Taylor, et al. 2018). Cancer stemness of TCGA samples were obtained from
355 (Malta, et al. 2018). Processed single-cell RNA sequencing data for embryonic stem cells
356 during preimplantation were downloaded from (Petropoulos, et al. 2016). Gene expression

357 levels of SLE RNA-seq data were obtained from (Hung, et al. 2015). Half-life data of mRNAs
358 were obtained from (Tani, et al. 2012). Our RNA-seq data for *TP53*-WT and *TP53*-KO cell lines
359 (MCF7 and HT1080) from this study were deposited in the SRA database (PRJNA777384).
360 Tissue-specific gene expression levels (Read counts) for non-human primates were
361 downloaded from NHPRTR (http://nhprtr.org/) (Peng, et al. 2015). Gene expression levels
362 (Fragments Per Kilobase of transcript per Million mapped reads, or FPKM) for the standard
363 mouse strain, C57BL/6JJcl were collected from (Li, et al. 2017). Quantitative proteomic
364 expression levels (TMT ratios) for tumors were obtained from CPTAC
365 (https://pdc.cancer.gov/pdc/), covering data from 11 studies (PDC000120, PDC000116,
366 PDC000110, PDC000127, PDC000204, PDC000221, PDC000153, PDC000234, PDC000270,
367 PDC000125 and PDC000198). RNA-seq data for mouse ATRX knockout cell lines were
368 collected from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/),
369 including datasets GSE178113 (Qin, et al. 2022) and GSE107878 (Deneault, et al. 2018).
370 LncRNA expression profiles for GTEx and TCGA samples were downloaded from RefLnc
371 (http://reflnc.gao-lab.org/) (Jiang, et al. 2019).
372

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

374 For both chrX and autosomes, only protein-coding genes were considered and those with
375 median expression levels TPM<=1 in GTEx samples were further excluded. For the remaining
376 chrX genes, Pearson correlations (R) between gene expression and genome ploidy (Taylor, et
377 al. 2018) were calculated in each cancer (except CHOL, DLBC, UCS, KICH, and BRCA) separately
378 from TCGA, using only male samples to avoid the confounding of possible reactivation of the
379 inactivated X in female samples. Then, the median of absolute R (absR) was calculated for
380 each gene and the average of these absR medians (MR) was determined for these genes.

381 Lastly, chrX genes with absR medians less or higher than MR were defined as PIX and PSX
382 genes, respectively. Remaining PIX, PSX, and autosome genes were used for XAR-PIX and XAR-
383 PSX calculation.
384
385 **Comparison of genomic features for PIX and PSX genes**
386 Gene lengths were calculated for each gene either including or excluding introns, based on
387 GENCODE v25 (Harrow, et al. 2012). PhastCons100way scores were downloaded from the
388 UCSC Genome Browser (https://hgdownload.soe.ucsc.edu/downloads.html) and gene
389 conservation were calculated by averaging phastCons100way scores within exonic regions.
390 GC content for each gene were calculated from its mRNA sequence as the percentage of G
391 and C bases.
392
393 **Calculation of XAR-PSX and XAR-PIX**
394 First, median expression levels of PSX, PIX, and autosomal genes were calculated. Then, XAR-
395 PSX or XAR-PIX were defined as the ratio of PSX or PIX median over autosomal median. The
396 magnitude (Cohen's D) of XAR difference between TCGA and GTEx was estimated by the
397 function cohens_d from the R package rstatix (v0.6). For non-human species, genes were
398 classified according to their homologous gene categories (PSX, PIX or autosomal genes) in
399 humans, and then XAR-PSX and XAR-PIX were calculated as above, based on their processed
400 gene expression levels (TPM). For proteomic expression levels, TMT ratios were used to
401 calculate XAR, while XAR-PSX and XAR-PIX were compared between primary tumors and solid
402 tissue normal samples.
403
404 **Association of XAR with somatic mutations**

405 Damaging mutations were defined as Missense_Mutation, Nonsense_Mutation,
406 Nonstop_Mutation, Splice_Region, Splice_Site, Translation_Start_Site, Frame_Shift_Del,
407 Frame_Shift_Ins, In_Frame_Ins, or In_Frame_Del. For each gene, samples were grouped by
408 whether having damaging mutations in each cancer for each gender. Genes with less than 20
409 damaging mutations were not considered. XAR between these two groups were compared
410 using two-sided *t*-test and *P*-values were corrected using false discovery rate (FDR).
411
412 **Gene ontology analysis**
413 GO enrichment analyses were performed using R package clusterProfiler (v4.2.2). Enriched
414 biological processes were identified by right-tailed Fisher's exact test, and Benjamini-
415 Hochberg correction was used for multiple comparisons. For potential protein-coding
416 regulators of XAR, genes expressed in at least in one tumor were used as the background gene
417 set. For lncRNA analysis, neighboring protein-coding genes of XAR-associated lncRNAs were
418 used as background gene set.
419
420 **ATRX regression coefficients comparison between autosomal and X-linked genes**
421 For autosomal and X-linked genes, linear regression models with ATRX expression were fit at
422 the mRNA or proteomic level in each tumor, in which *ATRX* coefficient estimates were
423 collected and converted to absolute form for comparison between the two gene groups.
424
425 **Correlations for lncRNAs and XAR in tumors**
426 For 77,900 lncRNAs collected, those with median TPM greater than 0 in at least 1 tumor were
427 selected. Next, linear regression models were fit to lncRNAs expression and XAR in each tumor,
428 with lncRNAs correlated to XAR-PIX or XAR-PSX in at least 3 male tumors extracted for further

429 study. Neighboring co-expressing protein-coding genes of these lncRNAs were identified by
430 Pearson correlation coefficient (FDR<0.05) and subjected to GO analysis. RNA-protein
431 interactions of XAR-associated lncRNAs were collected from LncSEA2.0
432 (http://bio.liclab.net/LncSEA/index.php).
433
434 **Data visualization**
435 The R package ggpubr (v.04, based on ggplot2) was used for basic data visualizations in this
436 study. The heatmaps of stemness correlation with XAR were visualized by R package
437 ComplexHeatmap (v2.2). Genomic distribution of PIX and PSX genes were visualized by R
438 package circlize (v0.4). Correlation heatmaps between XAR predictions and observations
439 were visualized by R package ggcorrplot2 (v0.1). Cumulative distribution of expression levels
440 were visualized by R package ggplot2 (v3.3).
441
442 **Cell culture and CRISPR-Cas9 knockout**
443 MCF7 and HT1080 cells were obtained from the American Type Culture Collection (ATCC). All
444 cells were cultured in high-glucose DMEM containing 10% fetal bovine serum (FBS) and 1%
445 penicillin/ streptomycin and maintained at 37 °C with 5% CO_2.
446
447 The sgRNAs of *TP53* were designed on the CHOPCHOP website (https://chopchop.cbu.uib.no)
448 as described before (Montague, et al. 2014). sgRNAs with highly predicted efficiency and
449 lower off-target were chosen for in vitro transcription. The sgRNA sequence targeted *TP53*
450 were as follows: sgRNA-1, 5'-TATCTGAGCAGCGCTCATGG-3' and sgRNA-2, 5'-
451 GGTGAGGCTCCCCTTTCTTG-3'. The sgRNA oligos were cloned into the pDR274 vector and in
452 vitro transcribed using the MEGAshortscript T7 transcription kit (Thermo Fisher), then

453 purifified using the MEGAclear kit (Thermo Fisher). The sgRNAs were then dissolved in RNase-
454 free water and quantified using NanoDrop 1000 (Thermo Fisher).
455
456 To generate *TP53* knockout cells, electroporation of Cas9 RNP to the cells was performed as
457 previously reported (Kim, et al. 2014). Specifically, 10 µg TrueCut Cas9 Protein v2 (Thermo)
458 and 2 µg each of the in vitro transcribed sgRNAs were incubated together at 37°C for 20 min,
459 and 5×10^5 MCF7 or HT1080 cells (70%-90% confluent) were harvested and centrifuged at 90×
460 g for 10 min. Then the Cas9 RNP was electroporated to the cell pellet by Lonza 4D-
461 Nucleofector (Lonza) under a pre-optimized program.
462
463 **Western Blotting**
464 Cells were harvested and lysed with RIPA buffer (Beyotime). Cell extracts of equal total
465 protein (20 µg each) were separated by SDS-PAGE and transferred to Nylon membrane
466 (Merck Millipore). The membranes were blocked with 5% nonfat milk in TBST washing buffer
467 (10 mM Tris pH 8.0, 150 mM NaCl with 0.5% Tween-20) for 1 h at room temperature, then
468 incubated at 4°C overnight with anti-*TP53* (1:1000; sc-126,Santa Cruz) and anti-GAPDH
469 (1:5000; P30008M, Abmart) primary antibodies. After washing, the membranes were
470 incubated with fluorescence-conjugated secondary antibodies and detected with Odyssey
471 Imagers.
472
473 **RNA extraction and real-time quantitative PCR**
474 Total RNA was extracted from cultured cells using RNAiso Plus (Takara) and reverse
475 transcribed into cDNA by the PrimeScript™ RT reagent Kit (Takara) , following the
476 manufacturer's protocol. Then real-time quantitative PCR was performed using TB Green®

477 Premix Ex Taq™ II (Takara) on the Real-Time PCR Thermal Cycler qTOWER3 (Analytik Jena).
478 Primers used in this study (forward/reverse; 5' to 3') were as follows:
479 *KAT8*: TCACTCGCAACCAAAAGCG/GATCGCCTCATGCTCCTTCT
480 *MSL1*: CCCATCACCGTTACCATTACG/GGAACAGCCAAGACTGAAGTTT
481 *MSL2*: GTAGCCACTGACTTATGTTCCAC/GCTGCAAATTAGGGCAACAGAC
482 GAPDH: GGAGCGAGATCCCTCCAAAAT/GGCTGTTGTCATACTTCTCATGG
483 Data analysis was performed using the 2(-ΔΔCT) method, and mRNA expression levels were
484 normalized to GAPDH.
485
486 **RNA sequencing of MCF7 and HT1080 cell lines**
487 Total RNA was extracted using TransZol Up Plus RNA Kit (Cat#ER501-01, Trans) following the
488 manufacturer's instructions and checked for RNA integrity by an Agilent Bioanalyzer 2100
489 (Agilent technologies, Santa Clara, CA, US). Qualified total RNA was further purified by
490 RNAClean XP Kit (Cat#A63987, Beckman Coulter, Inc. Kraemer Boulevard Brea, CA,USA) and
491 RNase-Free DNase Set (Cat#79254, QIAGEN, GmBH, Germany). Sequencing libraries were
492 constructed following steps from the manufacturer's instructions (VAHTS Universal V6 RNA-
493 seq Library Prep Kit for Illumina®, Vazyme, NR604-02), followed by paired-end (2x 150bp)
494 RNA-seq by Illumina HiSeq 2500. Cleaned reads were aligned to the human genome (hg38)
495 by Kallisto (v0.46, default parameters) (Bray, et al. 2016) to obtain gene expression levels
496 (TPM), based on gene annotations from GENCODE v25 (Harrow, et al. 2012). Reads were also
497 mapped to hg38 using HISAT2 (v2.0.4, default parameters), and alignments of *TP53*-WT
498 samples were visualized by IGV (v2.5.2) for manual inspection to confirm TP53 status.
499
500 **Data availability**

501 The RNA-seq data generated in this study have been submitted to the NCBI BioProject
502 database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA777384.
503

504 ## Declaration of interests

505 The authors declare no competing interests.

506 ## Acknowledgements

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

510 ## Author contributions

511 MBG and YYX conceived the project. MBG and ZWF collected the data, performed the analysis,
512 and interpreted the results. MBG wrote the manuscript with input from ZWF. ZWF and BHC
513 performed the experimental validation. ZSY was involved in project discussion. YYX was
514 involved in critical discussions and reviewed the manuscript. All authors approved the
515 manuscript.
516

517 ## References

518 Allison SJ, Milner J. 2004. Remodelling chromatin on a global scale: a novel protective function
519 of p53. Carcinogenesis 25:1551-1557.
520 Barry C, Schmitz MT, Jiang P, Schwartz MP, Duffin BM, Swanson S, Bacher R, Bolin JM, Elwell
521 AL, McIntosh BE, et al. 2017. Species-specific developmental timing is maintained by
522 pluripotent stem cells ex utero. Dev Biol 423:101-110.

523 Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW,
524 Oakley DH, et al. 2011. Reference Maps of human ES and iPS cell variation enable high-
525 throughput characterization of pluripotent cell lines. Cell 144:439-452.
526 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq
527 quantification. Nat Biotechnol 34:525-527.
528 Courel M, Clement Y, Bossevain C, Foretek D, Vidal Cruchez O, Yi Z, Benard M, Benassy MN,
529 Kress M, Vindry C, et al. 2019. GC content shapes mRNA storage and decay in human cells.
530 Elife 8.
531 Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, Park PJ, Elledge SJ. 2013. Cumulative
532 haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer
533 genome. Cell 155:948-962.
534 Deneault E, White SH, Rodrigues DC, Ross PJ, Faheem M, Zaslavsky K, Wang Z, Alexandrova R,
535 Pellecchia G, Wei W, et al. 2018. Complete Disruption of Autism-Susceptibility Genes by Gene
536 Editing Predominantly Reduces Functional Connectivity of Isogenic Human Neurons. Stem
537 Cell Reports 11:1211-1225.
538 Deng X, Berletch JB, Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Disteche CM. 2013.
539 Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life,
540 and MOF-mediated H4K16 acetylation. Dev Cell 25:55-68.
541 Devany E, Zhang X, Park JY, Tian B, Kleiman FE. 2013. Positive and negative feedback loops in
542 the p53 and mRNA 3' processing pathways. Proc Natl Acad Sci U S A 110:3351-3356.
543 Disteche CM. 2012. Dosage compensation of the sex chromosomes. Annu Rev Genet 46:537-
544 560.

545 Donehower LA, Soussi T, Korkut A, Liu Y, Schultz A, Cardenas M, Li X, Babur O, Hsu TK,
546 Lichtarge O, et al. 2019. Integrated Analysis of TP53 Gene and Pathway Alterations in The
547 Cancer Genome Atlas. Cell Rep 28:3010.
548 Dunford A, Weinstock DM, Savova V, Schumacher SE, Cleary JP, Yoda A, Sullivan TJ, Hess JM,
549 Gimelbrant AA, Beroukhim R, et al. 2017. Tumor-suppressor genes that escape from X-
550 inactivation contribute to cancer sex bias. Nat Genet 49:10-16.
551 Dyer MA, Qadeer ZA, Valle-Garcia D, Bernstein E. 2017. ATRX and DAXX: Mechanisms and
552 Mutations. Cold Spring Harb Perspect Med 7.
553 Friedrich D, Friedel L, Finzel A, Herrmann A, Preibisch S, Loewer A. 2019. Stochastic
554 transcription in the p53-mediated response to DNA damage is modulated by burst frequency.
555 Mol Syst Biol 15:e9068.
556 Gewandter JS, Bambara RA, O'Reilly MA. 2011. The RNA surveillance protein SMG1 activates
557 p53 in response to DNA double-strand breaks but not exogenously oxidized mRNA. Cell Cycle
558 10:2561-2567.
559 Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D,
560 Brooks AN, et al. 2020. Visualizing and interpreting cancer genomics data via the Xena
561 platform. Nat Biotechnol 38:675-678.
562 GTEx-Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580-
563 585.
564 Hainaut P, Pfeifer GP. 2016. Somatic TP53 Mutations in the Era of Genome Sequencing. Cold
565 Spring Harb Perspect Med 6.
566 Haronikova L, Olivares-Illana V, Wang L, Karakostis K, Chen S, Fahraeus R. 2019. The p53
567 mRNA: an integral part of the cellular stress response. Nucleic Acids Res 47:3257-3271.

568 Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D,
569 Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The
570 ENCODE Project. Genome Res 22:1760-1774.
571 Haupt S, Caramia F, Herschtal A, Soussi T, Lozano G, Chen H, Liang H, Speed TP, Haupt Y. 2019.
572 Identification of cancer sex-disparity in the functional integrity of p53 and its X chromosome
573 network. Nat Commun 10:5385.
574 Hung T, Pratt GA, Sundararaman B, Townsend MJ, Chaivorapol C, Bhangale T, Graham RR,
575 Ortmann W, Criswell LA, Yeo GW, et al. 2015. The Ro60 autoantigen binds endogenous
576 retroelements and regulates inflammatory gene expression. Science 350:455-459.
577 Jiang S, Cheng SJ, Ren LC, Wang Q, Kang YJ, Ding Y, Hou M, Yang XX, Lin Y, Liang N, et al. 2019.
578 An expanded landscape of human long noncoding RNA. Nucleic Acids Res 47:7842-7856.
579 Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schutz F, Daish T, Grutzner F,
580 Kaessmann H. 2012. Mechanisms and evolutionary patterns of mammalian and avian dosage
581 compensation. PLoS Biol 10:e1001328.
582 Kim S, Kim D, Cho SW, Kim J, Kim JS. 2014. Highly efficient RNA-guided genome editing in
583 human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res 24:1012-1019.
584 Kruse JP, Gu W. 2009. MSL2 promotes Mdm2-independent cytoplasmic localization of p53. J
585 Biol Chem 284:3250-3263.
586 Lafferty-Whyte K, Bilsland A, Hoare SF, Burns S, Zaffaroni N, Cairney CJ, Keith WN. 2010.
587 TCEAL7 inhibition of c-Myc activity in alternative lengthening of telomeres regulates hTERT
588 expression. Neoplasia 12:405-414.
589 Larsson AJM, Coucoravas C, Sandberg R, Reinius B. 2019. X-chromosome upregulation is
590 driven by increased burst frequency. Nat Struct Mol Biol 26:963-969.

591 Levy MA, Kernohan KD, Jiang Y, Berube NG. 2015. ATRX promotes gene expression by
592 facilitating transcriptional elongation through guanine-rich coding regions. Hum Mol Genet
593 24:1824-1835.
594 Li B, Qing T, Zhu J, Wen Z, Yu Y, Fukumura R, Zheng Y, Gondo Y, Shi L. 2017. A Comprehensive
595 Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. Sci Rep 7:4200.
596 Li X, Wu L, Corsa CA, Kunkel S, Dou Y. 2009. Two mammalian MOF complexes regulate
597 transcription activation by distinct mechanisms. Mol Cell 36:290-301.
598 Liang J, Liu H, Li G, Qian J, Gao R, Zhou Y, Wang X. 2020. Global changes in chromatin
599 accessibility and transcription following ATRX inactivation in human cancer cells. FEBS Lett
600 594:67-78.
601 Loda A, Collombet S, Heard E. 2022. Gene regulation in time and space during X-chromosome
602 inactivation. Nat Rev Mol Cell Biol 23:231-249.
603 Lu WJ, Amatruda JF, Abrams JM. 2009. p53 ancestry: gazing through an evolutionary lens. Nat
604 Rev Cancer 9:758-762.
605 Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B,
606 Huelsken J, Omberg L, Gevaert O, et al. 2018. Machine Learning Identifies Stemness Features
607 Associated with Oncogenic Dedifferentiation. Cell 173:338-354 e315.
608 Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data. Trends Genet
609 29:537-544.
610 Millan-Zambrano G, Burton A, Bannister AJ, Schneider R. 2022. Histone post-translational
611 modifications - cause and consequence of genome function. Nat Rev Genet.
612 Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. 2014. CHOPCHOP: a CRISPR/Cas9
613 and TALEN web tool for genome editing. Nucleic Acids Res 42:W401-407.

614 Nguyen TT, Grimm SA, Bushel PR, Li J, Li Y, Bennett BD, Lavender CA, Ward JM, Fargo DC,
615 Anderson CW, et al. 2018. Revealing a human p53 universe. Nucleic Acids Res 46:8153-8167.
616 Pageau GJ, Hall LL, Ganesan S, Livingston DM, Lawrence JB. 2007. The disappearing Barr body
617 in breast and ovarian cancers. Nat Rev Cancer 7:628-633.
618 Peng X, Thierry-Mieg J, Thierry-Mieg D, Nishida A, Pipes L, Bozinoski M, Thomas MJ, Kelly S,
619 Weiss JM, Raveendran M, et al. 2015. Tissue-specific transcriptome sequencing analysis
620 expands the non-human primate reference transcriptome resource (NHPRTR). Nucleic Acids
621 Res 43:D737-742.
622 Petropoulos S, Edsgard D, Reinius B, Deng Q, Panula SP, Codeluppi S, Plaza Reyes A, Linnarsson
623 S, Sandberg R, Lanner F. 2016. Single-Cell RNA-Seq Reveals Lineage and X Chromosome
624 Dynamics in Human Preimplantation Embryos. Cell 165:1012-1026.
625 Pillutla RC, Shimamoto A, Furuichi Y, Shatkin AJ. 1999. Genomic structure and chromosomal
626 localization of TCEAL1, a human gene encoding the nuclear phosphoprotein p21/SIIR.
627 Genomics 56:217-220.
628 Qin T, Mullan B, Ravindran R, Messinger D, Siada R, Cummings JR, Harris M, Muruganand A,
629 Pyaram K, Miklja Z, et al. 2022. ATRX loss in glioma results in dysregulation of cell-cycle phase
630 transition and ATM inhibitor radio-sensitization. Cell Rep 38:110216.
631 Salomonis N, Dexheimer PJ, Omberg L, Schroll R, Bush S, Huo J, Schriml L, Ho Sui S, Keddache
632 M, Mayhew C, et al. 2016. Integrated Genomic Analysis of Diverse Induced Pluripotent Stem
633 Cells from the Progenitor Cell Biology Consortium. Stem Cell Reports 7:110-125.
634 Sirchia SM, Ramoscelli L, Grati FR, Barbera F, Coradini D, Rossella F, Porta G, Lesma E, Ruggeri
635 A, Radice P, et al. 2005. Loss of the inactive X chromosome and replication of the active X in
636 BRCA1-defective and wild-type breast cancer cells. Cancer Res 65:2139-2146.

637 Statello L, Guo CJ, Chen LL, Huarte M. 2021. Gene regulation by long non-coding RNAs and its
638 biological functions. Nat Rev Mol Cell Biol 22:96-118.
639 Tani H, Mizutani R, Salam KA, Tano K, Ijiri K, Wakamatsu A, Isogai T, Suzuki Y, Akimitsu N. 2012.
640 Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding
641 transcripts in mammals. Genome Res 22:947-956.
642 Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, Schumacher SE, Wang C, Hu H, Liu J, et
643 al. 2018. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. Cancer
644 Cell 33:676-689 e673.
645 Teng YC, Sundaresan A, O'Hara R, Gant VU, Li M, Martire S, Warshaw JN, Basu A, Banaszynski
646 LA. 2021. ATRX promotes heterochromatin formation to protect cells from G-quadruplex
647 DNA-mediated stress. Nat Commun 12:3887.
648 Vallot C, Ouimette JF, Rougeulle C. 2016. Establishment of X chromosome inactivation and
649 epigenomic features of the inactive X depend on cellular contexts. Bioessays 38:869-880.
650 Voon HP, Hughes JR, Rode C, De La Rosa-Velazquez IA, Jenuwein T, Feil R, Higgs DR, Gibbons
651 RJ. 2015. ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci
652 and Imprinted Genes. Cell Rep 11:405-418.
653 Xiong Y, Chen X, Chen Z, Wang X, Shi S, Wang X, Zhang J, He X. 2010. RNA sequencing shows
654 no dosage compensation of the active X-chromosome. Nat Genet 42:1043-1047.
655 Yang L, Yildirim E, Kirby JE, Press W, Lee JT. 2020. Widespread organ tolerance to Xist loss and
656 X reactivation except under chronic stress in the gut. Proc Natl Acad Sci U S A 117:4262-4272.
657 Younger ST, Rinn JL. 2017. p53 regulates enhancer accessibility and activity in response to
658 DNA damage. Nucleic Acids Res 45:9889-9900.

659 Yue X, Lan F, Xia T. 2019. Hypoxic Glioma Cell-Secreted Exosomal miR-301a Activates
660 Wnt/beta-catenin Signaling and Promotes Radiation Resistance by Targeting TCEAL7. Mol
661 Ther 27:1939-1949.
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682

Figures and Legends

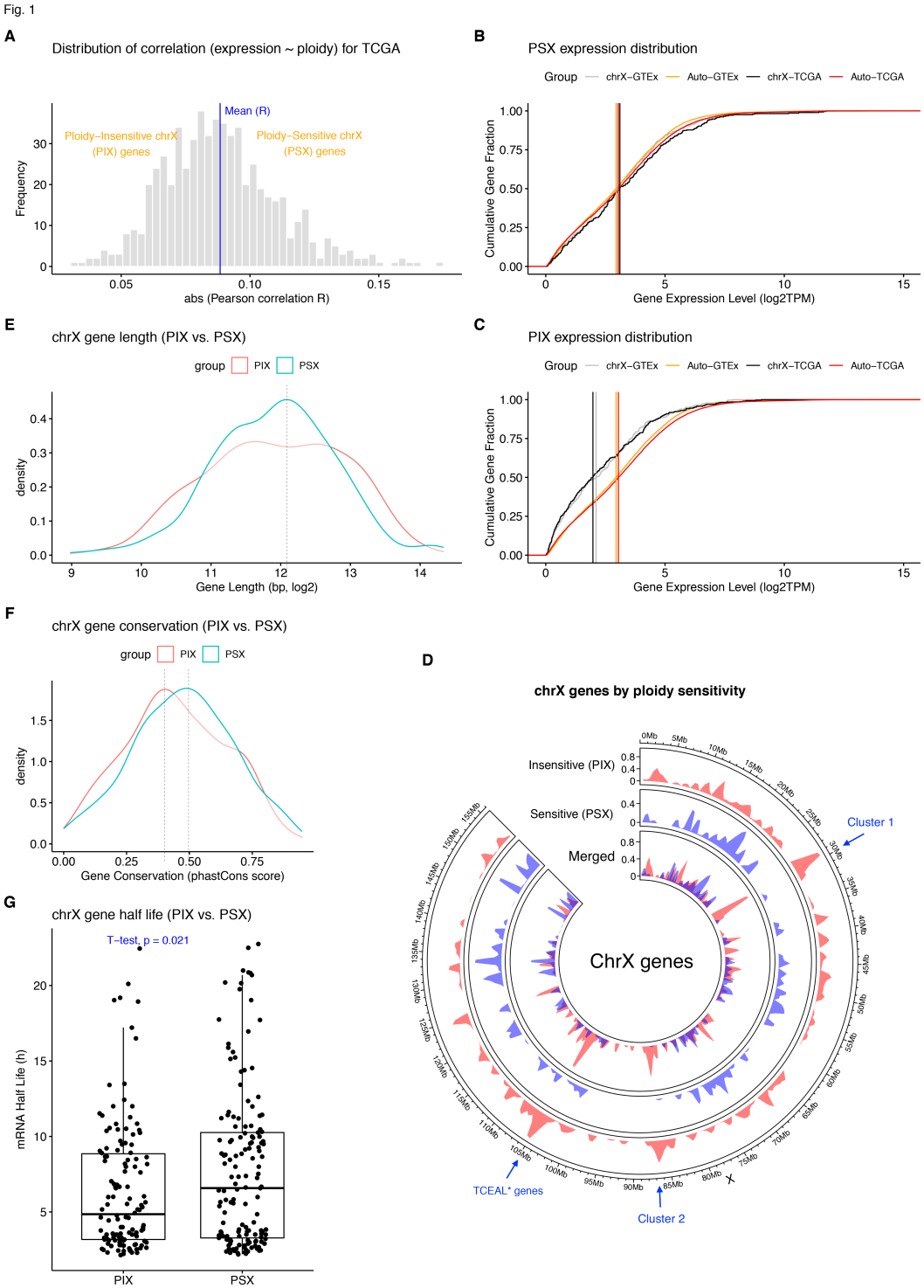


Fig. 1, Identification and characterization of PSX and PIX genes on chromosome X. (**A**) Classification chrX genes based on absolute values of Pearson correlations (x-axis) compared to the average absolute correlation value (blue vertical line). (**B-C**) Cumulative distribution of

688 expression levels of PSX (**B**) and PIX (**C**) genes on chrX, compared to autosomal genes, using
689 samples from the TCGA project or the GTEx project. A total of 100 samples were randomly
690 selected and their median expression levels for genes were shown for visualization. Vertical
691 lines indicate medians. (**D**) PIX and PSX genes distribution on chrX. PIX enriched regions were
692 marked by arrows. (**E-G**) Comparing gene length (**E**), gene sequence conservation (**F**), and
693 mRNA half-life (two-sided *t*-test, **G**) between PIX and PSX genes. PIX: ploidy-insensitive chrX
694 genes, PSX: ploidy-sensitive chrX genes.
695

Fig. 2

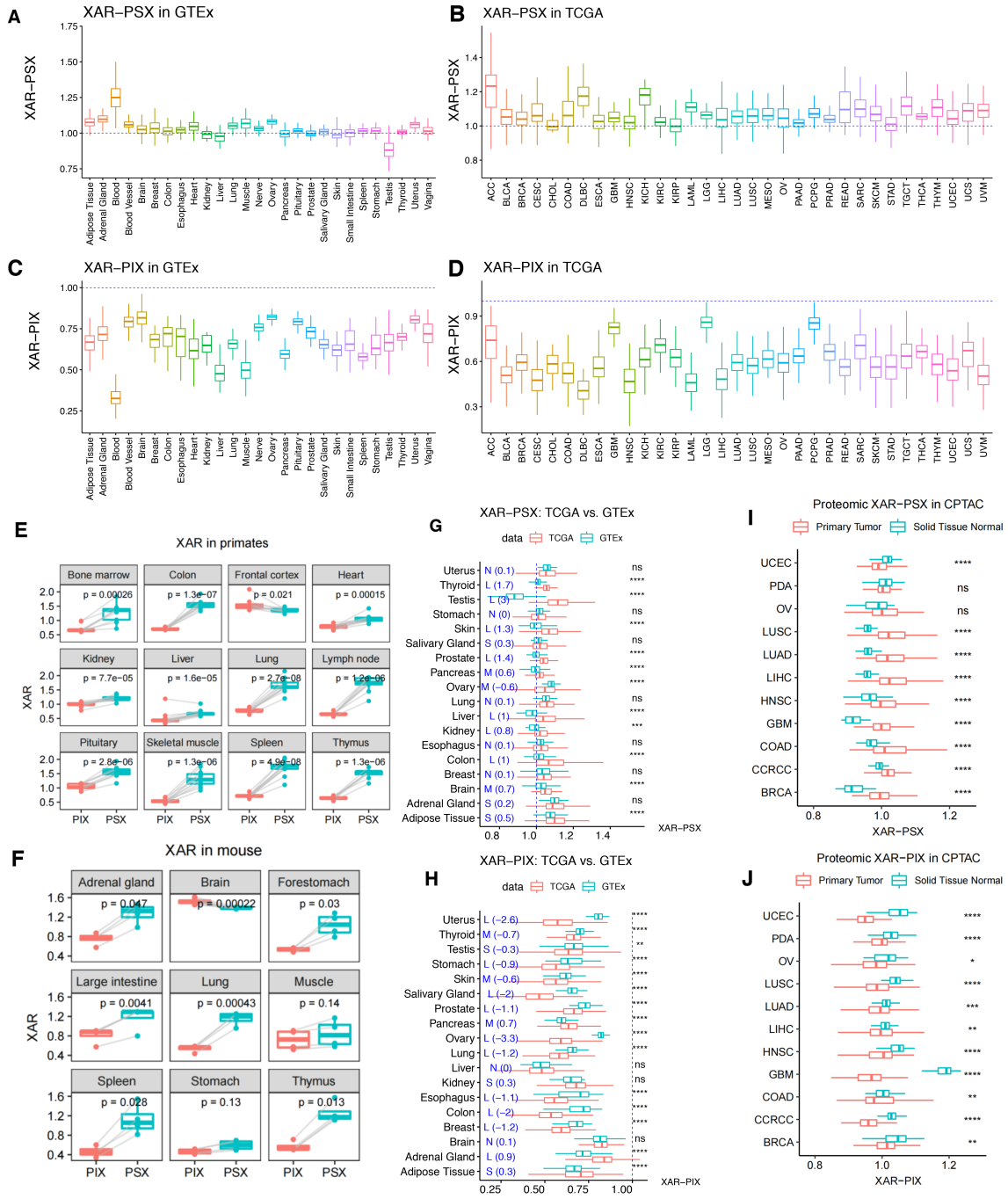


Fig. 2, Evolutionarily conserved differential dosages of PSX and PIX genes across normal tissues and cancer types. (**A**) XAR-PSX across 27 normal tissues from GTEx were close to one, except for blood and testis. (**B**) XAR-PSX across 33 cancer types from TCGA. (**C**) XAR-PIX across GTEx normal tissues. (**D**) XAR-PIX across TCGA cancer types. (**E**) Evolutionarily conserved differential dosages between PIX and PSX genes in primates across 12 tissues. (**F**)

702 Evolutionarily conserved differential dosages between PIX and PSX genes in mouse across 9
703 tissues. (**G**) XAR-PSX were higher in cancers than in matched normal tissues. Effect sizes were
704 indicated by Cohen's D statistics in blue (negligible[N]<0.2, 0.2<=small[S]<0.5,
705 0.5<=moderate[M]<0.8, large[L]>=0.8). (**H**) XAR-PIX were lower in cancers than in matched
706 normal tissues. Effect sizes were indicated by Cohen's D statistics in blue. (**I**) Protein-level
707 XAR-PSX were higher in tumors than normal tissues in most cancer types. (**J**) Protein-level
708 XAR-PIX were lower in tumors than normal tissues of all investigated cancer types.
709 (**E-F**) Two-sided paired t-test. (**G-J**) Two-sided Wilcoxon test *P*-value significance: ns >0.05, *
710 <= 0.05, ** <=0.01, *** <= 0.001, **** <= 0.0001. Matching of tumor types to normal tissues
711 were as follows: ACC, Adrenal Gland; BRCA, Breast; COAD, Colon; ESCA, Esophagus; GBM,
712 Brain; HNSC, Salivary Gland; KICH, Kidney; KIRC, Kidney; KIRP, Kidney; LGG, Brain; LIHC, Liver;
713 LUAD, Lung; LUSC, Lung; OV, Ovary; PAAD, Pancreas; PRAD, Prostate; PCPG, Adrenal Gland;
714 READ, Colon; SARC, Adipose Tissue; SKCM, Skin; STAD, Stomach; TGCT, Testis; THCA, Thyroid;
715 UCS, Uterus; UCEC, Uterus.
716 XAR: X-over-autosome dosage ratio, ACC: Adrenocortical carcinoma, BLCA: Bladder urothelial
717 Carcinoma, BRCA: Breast invasive carcinoma, CESC: Cervical squamous cell carcinoma and
718 endocervical adenocarcinoma, CHOL: Cholangiocarcinoma, COAD: Colon adenocarcinoma,
719 CCRCC: Clear cell renal cell carcinoma, DLBC: Lymphoid neoplasm diffuse large B-cell
720 lymphoma, ESCA: Esophageal carcinoma, GBM: Glioblastoma multiforme, HNSC: Head and
721 neck squamous cell carcinoma, KICH: Kidney chromophobe, KIRC: Kidney renal clear cell
722 carcinoma, KIRP: Kidney renal papillary cell carcinoma, LAML: Acute myeloid leukemia, LGG:
723 Brain lower grade glioma, LIHC: Liver hepatocellular carcinoma, LUAD: Lung adenocarcinoma,
724 LUSC: Lung squamous cell carcinoma, MESO: Mesothelioma, OV: Ovarian serous
725 cystadenocarcinoma, PAAD: Pancreatic adenocarcinoma, PCPG: Pheochromocytoma and

726 paraganglioma, PDA: Pancreatic ductal adenocarcinoma, PRAD: Prostate adenocarcinoma,
727 READ: Rectum adenocarcinoma, SARC: Sarcoma, SKCM: Skin Cutaneous melanoma, STAD:
728 Stomach adenocarcinoma, TGCT: Testicular germ cell tumors, THCA: Thyroid carcinoma,
729 THYM: Thymoma, UCEC: Uterine corpus endometrial carcinoma, UCS: Uterine
730 carcinosarcoma, UVM: Uveal melanoma.
731
732
733
734
735
736

Fig. 3

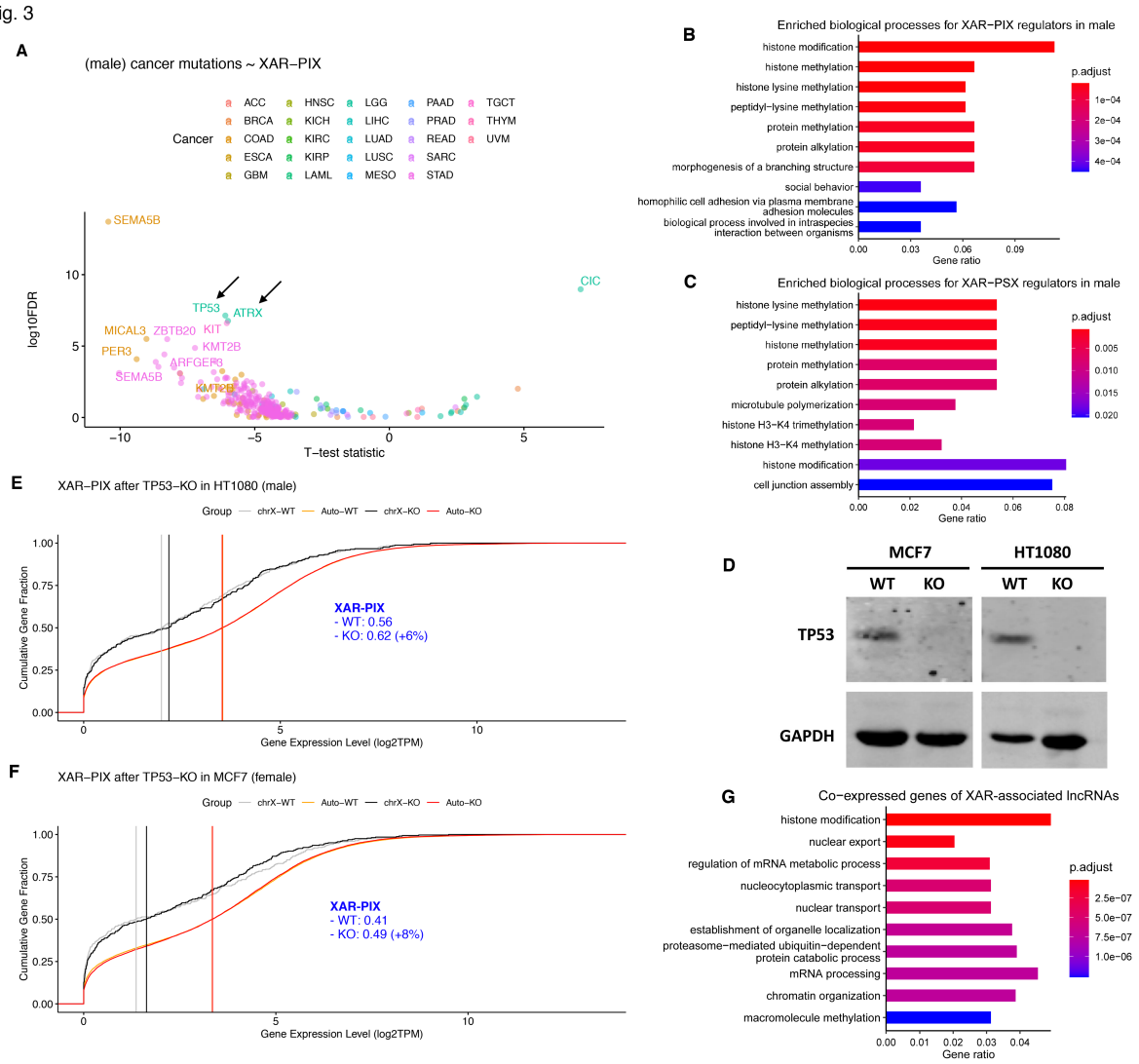


Fig. 3, Characterizing potential regulators of XAR, with experimental validation of *TP53* regulation of XAR-PIX. (**A**) Significance of XAR-PIX difference (y-axis, two-sided *t*-test) between wildtype and mutated genes across cancers in male. *TP53* and *ATRX* were marked by arrows. For better visualization, only genes with FDR<1 were shown as points and those with FDR<1e-4 labeled. (**B**-**C**) Top 10 enriched biological processes in GO analysis for top 200 potential regulators of XAR-PIX (**B**) and XAR-PSX (**C**). (**D**) Western blot results for CRISPR-based *TP53*-KO in MCF7 (female) and HT1080 (male) cells. (**E-F**) Cumulative distribution of PIX and autosomal gene expression levels for HT1080 (**E**) and MCF7 (**F**), before and after *TP53*-KO.

746 Vertical lines indicate median expression used to calculate XAR-PIX dosages. (**G**) Top 10

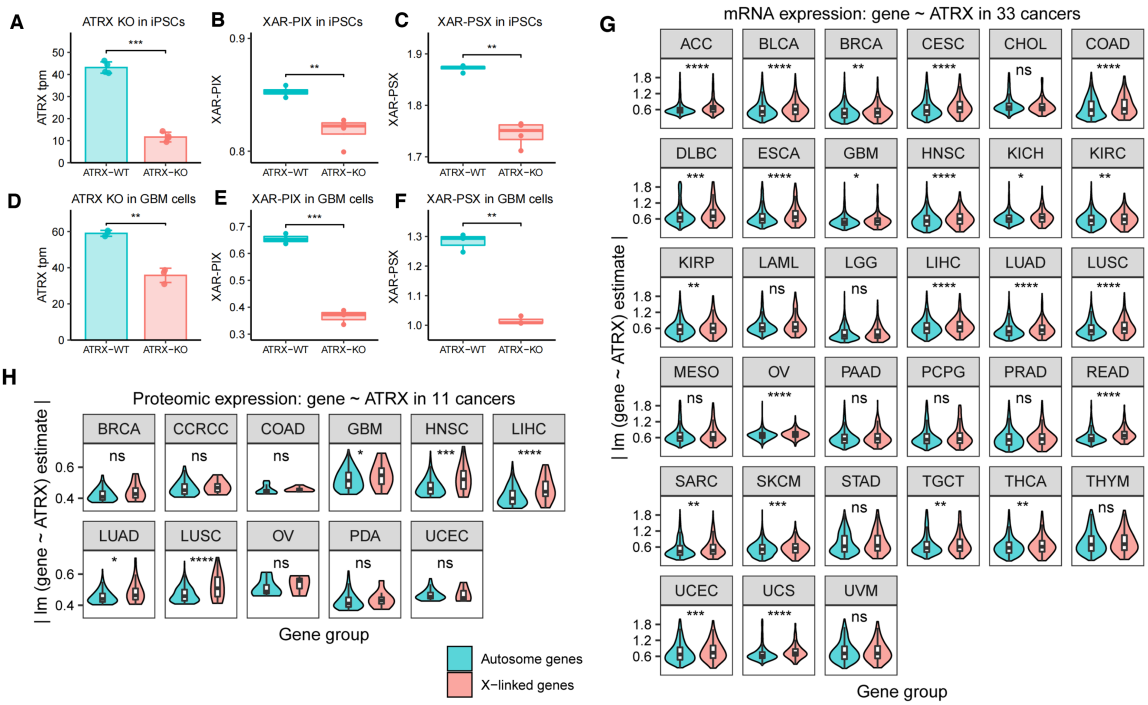
747 enriched biological processes for co-expressed genes of XAR-associated lnRNAs.

748

749

750

Fig. 4



751 **Fig. 4, *ATRX* regulates dosages of both PIX and PSX genes**. (**A-C**) *ATRX* knockout (KO) in mouse

752 induced pluripotent stem cells (iPSCs) significantly reduced both XAR-PIX (**B**) and XAR-PSX (**C**).

753 (**D-F**) *ATRX* knockout (KO) in human GBM cells significantly reduced both XAR-PIX (**E**) and XAR-

754 PSX (**F**). (**G-H**) For autosomal and X-linked genes, linear regression models (each gene

755 expression ~ *ATRX* expression) were fit to obtain regression coefficients of *ATRX* in each tumor,

756 using mRNA (**G**) or proteomic (**H**) expression data. Absolute values of *ATRX* regression

757 coefficients were compared between the two gene groups. Larger absolute values of *ATRX*

758 coefficients indicate higher impact on gene expression.

759 (**A-F**) Two-sided paired t-test. ** *P*<=0.01, *** <= 0.001. (**G-H**) Two-sided Wilcoxon test. * *P*<=

760 0.05, ** <=0.01, *** <= 0.001, **** <= 0.0001.

761

762

763

764

765

766

767

768

Fig. 5

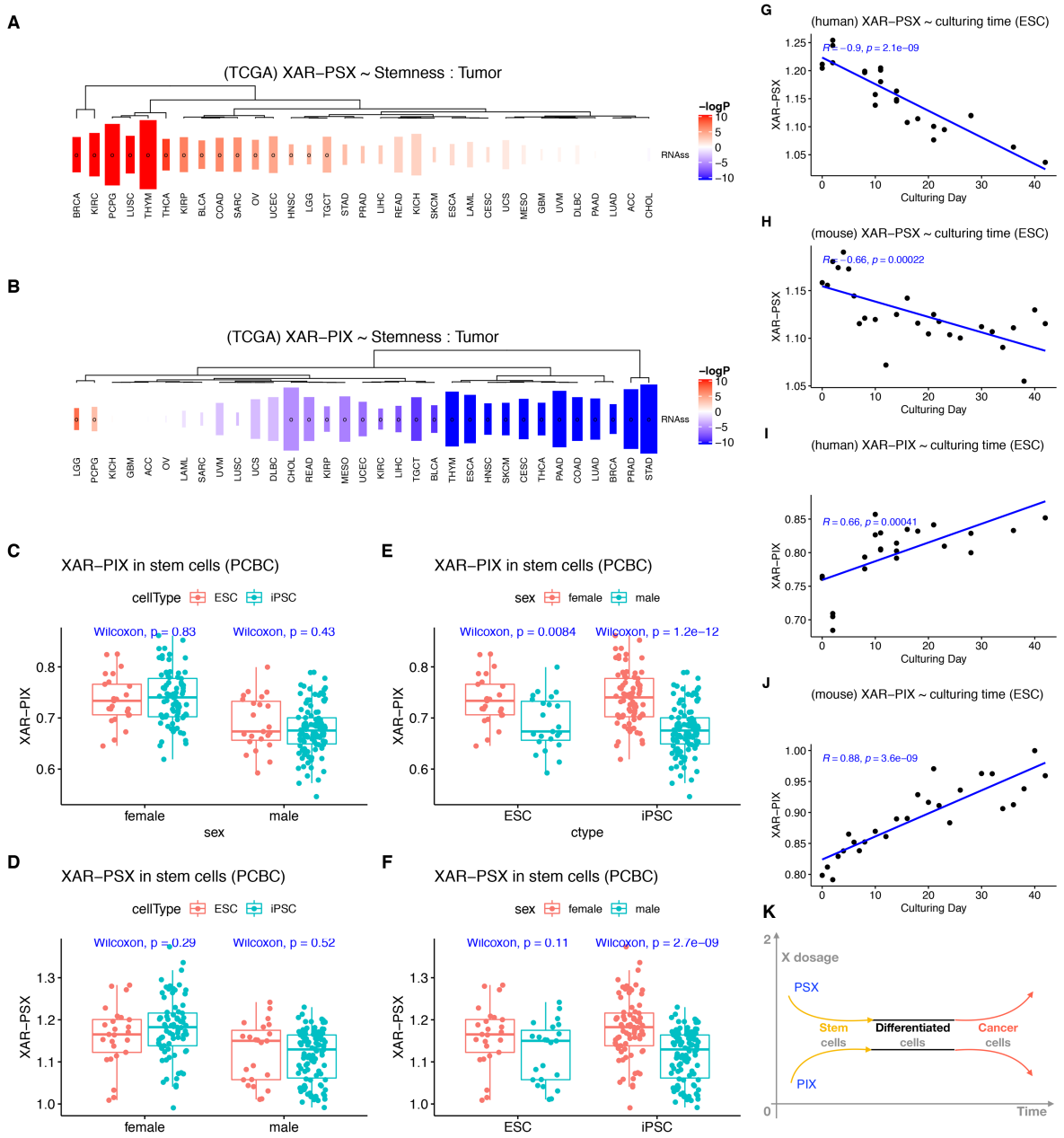


Fig. 5, XAR-PSX and XAR-PIX were inversely associated with stemness. (**A-B**) Cancer stemness (RNAss) was positively correlated with XAR-PSX (**A**) and negatively with XAR-PIX (**B**). colors indicate -log10 P-values, and sizes of rectangles represent Pearson correlation. Correlations with FDR<0.1 were marked by circles. (**C-D**) Comparison of XAR-PIX (**C**) and XAR-PSX (**D**) between ESCs and iPSCs from females or males. (**E-F**) Comparison of XAR-PIX (**E**) and XAR-PSX (**F**) between sexes from ESCs or iPSCs. (**G-H**) Strong negative correlation of XAR-PSX

776 with differentiation culturing time in both human (**G**) and mouse (**H**). (**I-J**). Strong positive

777 correlation of XAR-PSX with differentiation culturing time in both human (**I**) and mouse (**J**). (**K**)

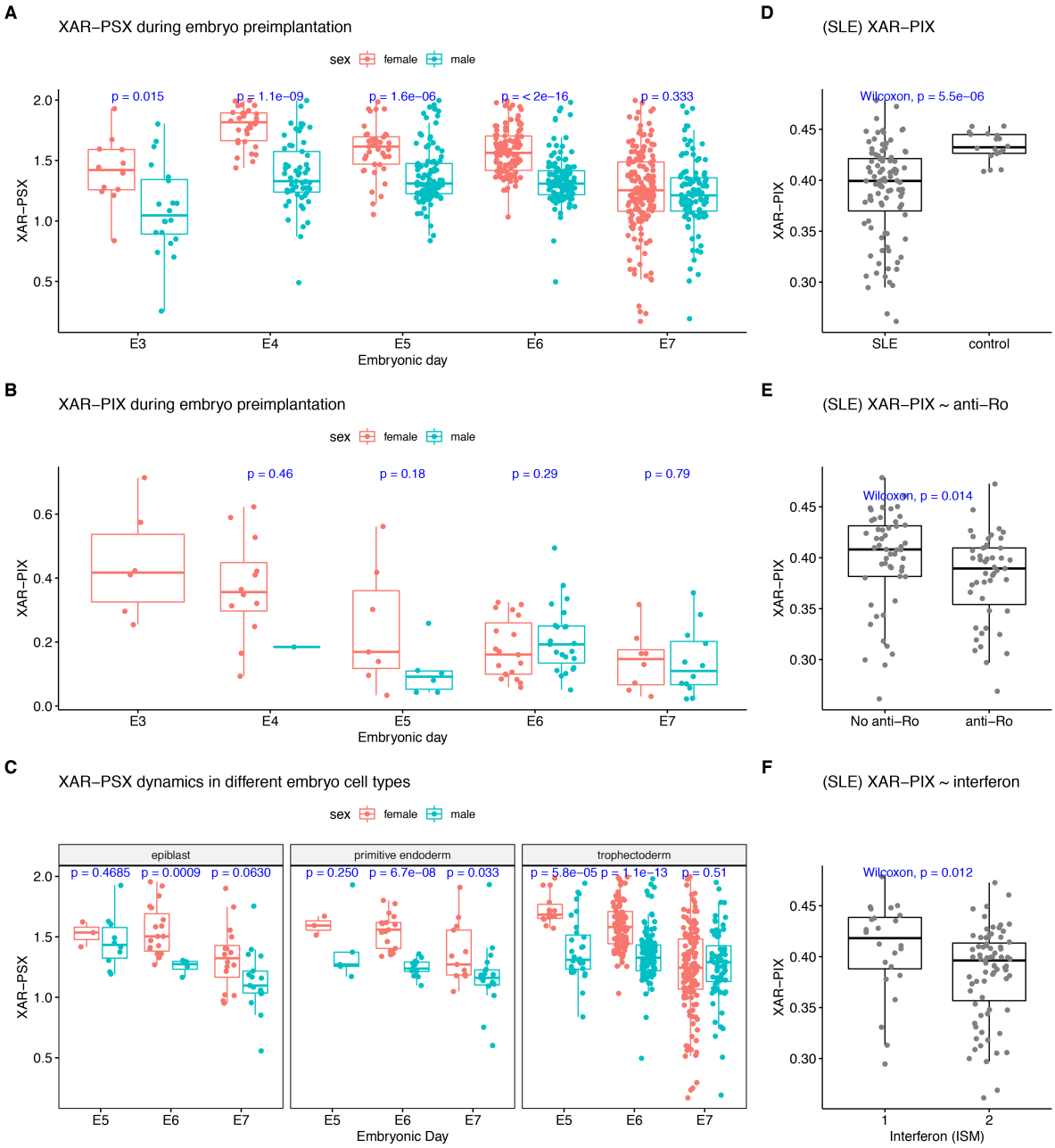
778 Major findings of this study were converging X dosages during stem cell differentiation but

779 diverging X dosages in cancer. (**C-F**) Two-sided Wilcoxon test.

780

781

Fig. 6



782 **Fig. 6, XAR-PSX and XAR-PIX dynamics during embryo preimplantation and in autoimmune**
783 **disease.** (**A-B**) XAR-PSX (**A**) and XAR-PIX (**B**) changed differently in embryonic stem cells
784 (grouped by sex) during implantation from day 3 (E3) to day 7 (E7). (**C**) Comparison of XAR-
785 PSX between sexes from E5 to E7, across three cell types, epiblast (EPI), primitive endoderm
786 (PE), and trophectoderm (TE). (**D-F**) XAR-PIX were down-regulated in SLE patients compared
787 to healthy controls (**D**), lower in patients with anti-Ro autoantibodies than those without anti-
788 Ro (**E**), and also lower in patients with more active interferon response (ISM) than those with
789 less ISM (**F**). (**A-F**) Two-sided Wilcoxon test.
790
791