Article

Targeting *Xist* with compounds that disrupt RNA structure and X inactivation

https://doi.org/10.1038/s41586-022-04537-z

Received: 27 November 2019

Accepted: 8 February 2022

Published online: 30 March 2022

Check for updates

Rodrigo Aguilar[1,2,5], Kerrie B. Spencer[3], Barry Kesner[1,2], Noreen F. Rizvi[3], Maulik D. Badmalia[4], Tyler Mrozowich[4], Jonathan D. Mortison[3], Carlos Rivera[1,2], Graham F. Smith[3], Julja Burchard[3], Peter J. Dandliker[3], Trushar R. Patel[4], Elliott B. Nickbarg[3] & Jeannie T. Lee[1,2]✉

Although more than 98% of the human genome is non-coding[1], nearly all of the drugs on the market target one of about 700 disease-related proteins. The historical reluctance to invest in non-coding RNA stems partly from requirements for drug targets to adopt a single stable conformation[2]. Most RNAs can adopt several conformations of similar stabilities. RNA structures also remain challenging to determine[3]. Nonetheless, an increasing number of diseases are now being attributed to non-coding RNA[4] and the ability to target them would vastly expand the chemical space for drug development. Here we devise a screening strategy and identify small molecules that bind the non-coding RNA prototype *Xist*[5]. The X1 compound has drug-like properties and binds specifically the RepA motif[6] of *Xist* in vitro and in vivo. Small-angle X-ray scattering analysis reveals that RepA can adopt multiple conformations but favours one structure in solution. X1 binding reduces the conformational space of RepA, displaces cognate interacting protein factors (PRC2 and SPEN), suppresses histone H3K27 trimethylation, and blocks initiation of X-chromosome inactivation. X1 inhibits cell differentiation and growth in a female-specific manner. Thus, RNA can be systematically targeted by drug-like compounds that disrupt RNA structure and epigenetic function.

The large non-coding transcriptome and its potential to encode disease variants make RNAs appealing targets for drug discovery. The search for RNA-binding small molecules has intensified in recent years[3,7,8]. Early examples did not have good drug-like properties, as they tend to be positively charged compounds with low selectivity, poor cellular uptake and high toxicity[3]. Recent examples such as ribocil (a flavin mononucleotide riboswitch inhibitor)[9] and evrysdi (a splice modulator for spinal muscular atrophy)[10] have been more promising. However, these examples were discovered in phenotypic screens and were only later determined to be RNA interactors. Scalable methods to screen RNA-targeting compounds are currently lacking. Although rational design by exploiting structural motifs is becoming more feasible[11], 3D structures for RNA are rarely available. To overcome conventional limitations, we devised an unbiased screen based on affinity-selection mass spectrometry. The automated ligand identification system[12] (ALIS) is based on a biophysical binding technology that identifies reversibly binding ligands with affinities stronger than 10 μM. ALIS includes a fast size-exclusion chromatography step to isolate the target-ligand complex away from unbound components and then uses liquid chromatography with mass spectrometry (LC–MS) to release, separate, and identify the bound ligand (Fig. 1b, bottom). Although ALIS has been used to identify protein hits, its potential for RNA drug discovery has been under-explored[13,14].

To test its applicability to RNA, we conducted a screen using *Xist*, the 17-kb transcript (Fig. 1a) that initiates X-chromosome inactivation (XCI) as it spreads along the inactive X chromosome[5,6] (Xi). RepA is a 431-nucleotide domain in *Xist* that comprises 8.5 units of a GC-rich motif responsible for gene silencing[6]. We synthesized mouse RepA RNA[15] and purified it under native conditions to retain secondary structures[16] (Extended Data Fig. 1). Using ALIS, we screened 50,000 compounds representing Merck's chemical diversity collection (Fig. 1b) and obtained one positive hit, X22, with a molecular weight of 382 Da. X22 exhibited reasonable pharmacological properties and 'drug-likeness'[17]. None of the 42 control RNAs[14] of varying lengths and GC content demonstrated binding to X22, including multiple other GC-rich RNAs (Supplementary Table 1). We carried out an expansion series around X22 for compounds with more than 70% structural similarity using Tanimoto scoring and identified 20 analogues (Fig. 1c; Supplementary Table 2). ALIS confirmed binding via competition studies. We rank-ordered their relative affinities for RepA in order to prioritize candidates. Among the 20 expanded compounds, 5 compounds did not bind appreciably (Fig. 1c, grey boxes) and were not pursued further. The remaining 15 demonstrated a range of binding affinities (Fig. 1c, Supplementary Table 2).

One compound, X1, has a unique indazole–benzimidazole scaffold and emerged as the highest affinity binder (dissociation constant (K_d) = 0.4 ± 0.3 μM (mean ± s.e.m.); Fig. 1d). Similar to X22, X1 displayed drug-likeness, with a molecular weight of 416 Da, 4 hydrogen bond donors, and 1 hydrogen bond acceptor. X1 has a low per cent effect of -11.7 at 40 μM in a thiazole orange displacement assay, indicating that

[1]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA. [2]Department of Genetics, The Blavatnik Institute, Harvard Medical School, Boston, MA, USA. [3]Merck & Co., Boston, MA, USA. [4]Alberta RNA Research and Training Institute, Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta, Canada. [5]Present address: Institute of Biomedical Sciences (ICB), Faculty of Medicine and Faculty of Life Sciences, Universidad Andres Bello, Santiago, Chile. ✉e-mail: lee@molbio.mgh.harvard.edu

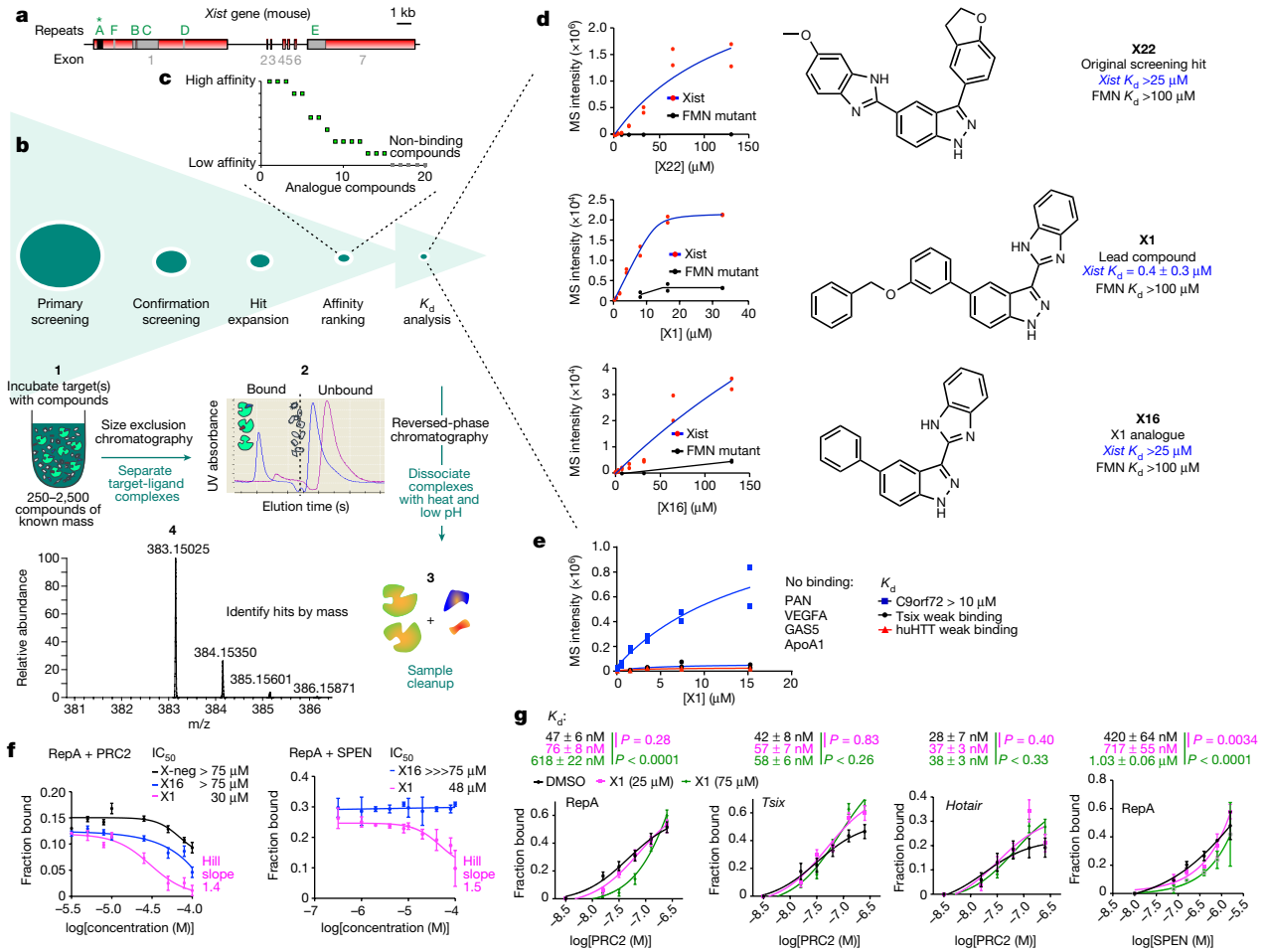


Fig. 1 | X1 binds RepA and weakens RNA binding to interacting protein partners in vitro. a, The mouse *Xist* gene with conserved repeat motifs A–F and exons 1–7. **b**, ALIS screening of 50,000 diverse small molecule compounds against RepA. (1) Compound mixture is equilibrated with target; (2) target–ligand complex is separated from unbound ligands by size-exclusion chromatography; (3) reverse-phased HPLC dissociates bound ligands and purifies sample; (4) mass spectrometry identifies ligand by mass. In the confirmation screening, a single hit (X22) is verified for target binding in ALIS. **c**, Hit expansion around the original hit, X22. Twenty analogues with at least 70% similarity were identified from the complete Merck collection, and affinity ranking was performed. The *y*-axis represents the relative binding affinity of each compound to RepA as determined by ALIS following the mass spectrometry step. **d**, The original hit X22, X1 from the expansion series, and X16 from synthetic analogues are shown with affinities for RepA. Negative control is the bacterial FMN Δ94–102 riboswitch[9]. Individual points are shown from two independent experiments. MS, mass spectrometry. **e**, X1 affinity for *PAN*, *VEGFA*, *GAS5*, *Apoa1*, *C9orf72*, *Tsix* and *huHTT* RNAs. Individual points are shown from two independent experiments. **f**, X1 weakens RepA–PRC2 and RepA–SPEN interactions. Densitometric analysis of EMSAs (Extended Data Fig. 2a) were used to determine IC_{50} and Hill coefficient. Increasing concentrations of indicated compounds (0, 5, 7.5, 10, 25, 50, 75 and 100 μM) were titrated against 0.5 nM RNA and 15.6 nM PRC2, or 0.1 nM RNA and 158 nM SPEN-RRM. Data are mean ± s.d.; $n = 3$ independent experiments. **g**, Quantification of RNA-EMSAs (Extended Data Fig. 2b) titrating PRC2 (0, 15.6, 31.2, 62.4, 124.9 and 250 nM) or SPEN-RRM (0, 79.2, 158, 396, 792 and 1,580 nM) against 25 or 75 μM X1. RNA concentration is 0.5 nM. Data are mean ± s.d.; *P*-values by one-way analysis of variance (ANOVA) with Dunnet's post-test comparing all conditions to control. $n = 3$ independent experiments.

it is not a nucleic acid intercalator. X1 did not appreciably bind other tested RNAs, including bacterial FMN Δ94–102 mutant riboswitch RNA[9] (K_d >>100 μM; Fig. 1d, black line) and various mammalian RNAs of similar size such as *PAN*, *VEGFA*, *GAS5*, *huHTT* and *APOA1* (Fig. 1e, Supplementary Table 1). The reverse complement of *Xist* (*Tsix*) and the CG-rich *C9orf72* (which binds a range of compounds nonspecifically[14]) also showed weak affinity for X1 (K_d >10 μM; Fig. 1e). Thus, X1 has high selectivity for RepA.

To determine whether X1 affects RepA function, we examined its interaction with cognate protein partners that facilitate *Xist* spreading and gene silencing. RepA binds polycomb repressive complex 2 (PRC2), an epigenetic complex that trimethylates histone H3 at lysine 27[15] (H3K27me3), and the RNA recognition motif (RRM) of SPEN (SPEN-RRM), a megadalton RNA-binding protein that interacts with histone deacetylases[18]. Using purified PRC2 (subunits EZH2, SUZ12, EED and RBAP48) and SPEN-RRM, we assessed whether X1 could disrupt RepA–PRC2 and RepA–SPEN interactions in electrophoretic mobility shift assays (EMSAs). In an X1 titration series (0–100 μM) against fixed concentrations of RepA RNA, PRC2 and SPEN-RRM, we observed progressively disruptive effects on RepA–PRC2 interactions, with a half-maximal inhibitory concentration (IC_{50}) of 30 μM; similarly, X1 also disrupted RepA–SPEN interactions, but with an IC_{50} of 48 μM (Fig. 1f, Extended Data Fig. 2a). These interactions have Hill coefficients of 1.4–1.5, indicating that X1 may exert cooperative binding, suggesting that more than one X1 molecule binds one RepA.

To perform structure–activity relationship studies, we synthesized analogues around X1. Among them, X16 lacks the benzyloxy group but still bound RepA. Consistent with its much lower affinity of 25 μM (Fig. 1d), X16 was also able to reduce RepA–PRC2 and RepA–SPEN interactions, but a higher compound concentration was required (IC_{50} > 75 μM) (Fig. 1f). We also tested a negative control compound (X-negative) that is known to bind influenza A virus RNA promoter[19]. Although X-negative bound similar-sized RNAs in other screenings[14],

it could not bind RepA and, accordingly, could not disrupt RepA–PRC2 interactions (Fig. 1f, Extended Data Fig. 2a).

To quantify disruptive effects, we measured K_d changes of RepA for PRC2 or SPEN in the presence of X1. Without X1, RepA bound PRC2 at 47 nM, whereas RepA bound SPEN with a tenfold lower affinity of 420 nM (Fig. 1g, Extended Data Fig. 2b). Addition of 25 μM X1 only modestly reduced the RepA–PRC2 affinity to 76 nM, and modestly reduced the RepA–SPEN affinity to 717 nM (Fig. 1g, Extended Data Fig. 2b). However, addition of 75 μM X1 caused a significant reduction of RepA affinity to 618 nM for PRC2 and 1.3 μM for SPEN. X1 had no significant effect on binding to other PRC2-interacting RNAs, including *Hotair* and *Tsix*, even at 75 μM X1 (Fig. 1g, Extended Data Fig. 2b–d). The effect of X1 effect was not affected by order of addition, as preincubating RepA with either X1 or PRC2 did not change the results (Extended Data Fig. 2e). Thus, X1 selectively inhibits interactions of RepA with its binding partners in vitro.

To assess the effects of X1 inside cells, we performed RNA immunoprecipitation (RIP) using anti-SUZ12 or anti-EZH2 antibodies to pull down PRC2-interacting RNAs and conducted quantitative PCR with reverse transcription (RT–qPCR) to quantify co-eluted RNAs. The female mouse embryonic stem (ES) cell model[20] *Tsix*$^{TST/+}$ (hereafter referred to as TST) initially carries two active X chromosomes (Xa) but recapitulates XCI when induced to differentiate into embryoid bodies for 4 days (Fig. 2a). We added 10 μM X1 on day 2, performed RIP on day 4, and observed that *Xist* recovery was reduced to 30–40% of control values in multiple biological replicates (Fig. 2b). By contrast, other PRC2 interactors[21] such as *Malat1*, *Gtl2*, *Htr-us* or *Nespas* were not affected by X1 (Fig. 2b, Extended Data Fig. 3a). Non-interacting RNAs such as *Gapdh*[21] were not immunoprecipitated, and X1 treatment had no effect on this outcome. Moreover, RIP using negative control IgG antibodies gave undetectable signals (Fig. 2b, c, Extended Data Fig. 3a). We also performed RIP using anti-SPEN antibodies and observed that X1 reduced RepA–SPEN interaction in vivo (Fig. 2b). Notably, X1 exerted a stronger inhibitory effect on SPEN than PRC2, consistent with the differential affinities of RepA for X1 (K_d = 400 nM), PRC2 (47 nM) and SPEN (420 nM). Because RepA binds PRC2 more avidly than it does SPEN or X1, RepA–PRC2 interactions are less inhibited by X1 than RepA–SPEN interactions. *Xist* RNA interacts with other protein partners through other domains, including CIZ1 and RBM15[22,23]. The m^{6}A RNA methyl-transferase RBM15 binds near RepA[22] and—notably—X1 did not disrupt RMB15–RepA interactions (Extended Data Fig. 3b). X1 also did not affect *Xist*–CIZ1 interactions (Fig. 2b), which occur via *Xist* exon 7 within repeat E[23]. Non-interacting RNAs such as *Malat1* and *Gapdh* were not pulled down by CIZ1 or RBM15 antibodies. Treating cells with X-negative also had no effect on *Xist*–PRC2 interactions (Fig. 2c). X1 can therefore selectively disrupt interactions of *Xist* with cognate proteins in vitro and inside cells.

To validate target engagement inside cells, we reasoned that if the effect of X1 was mediated by interaction with RepA, deleting RepA would eliminate the effect. We compared results of RIP in fibroblasts expressing either a wild-type *Xist* transgene[24] (X+P cells) or a version lacking RepA[25] (X–A cells). In 10 μM X1, SUZ12 RIP recovered significantly less *Xist* RNA from X+P cells relative to DMSO-treated cells (Fig. 2d; incomplete abrogation is consistent with there being other *Xist* domains (for example, repeat B) involved in recruiting PRC2[25]). The inhibitory effect of X1 was lost entirely when RepA was deleted (Fig. 2d), indicating that X1 indeed acts through RepA. To further validate target engagement, we synthesized a tritiated analogue of X1 (Fig. 2e, [^{3}H]X1) and performed RNA pulldown assays using biotinylated probes to capture *Xist* in TST cells expressing either a wild-type *Xist* allele (WT) or one lacking RepA[25] (X–A). When [^{3}H]X1 was added on day 3, pulldown assays on day 4 showed that radioactive X1 was recovered only from WT cells (Fig. 2e). Because X1 treatment had little effect on steady state *Xist* levels before or after pulldown (Extended Data Fig. 3c), the absence of [^{3}H]X1 counts in the X–A cells could not be attributed to absence of *Xist* RNA. The overall RepA dependence argues that X1 selectively binds RepA inside cells.

We tested whether the interaction of X1 with RepA affected XCI. To determine appropriate compound concentrations, we performed dose–response assays (0–20 μM). Female ES cells showed no growth retardation or inviability when treated with up to 10 μM X1 for 24 h (Extended Data Fig. 4a, b). We therefore performed phenotypic assays at 10 μM. When differentiated in the presence of 10 μM X1 for 24 h (day 3), female embryoid bodies remained healthy (Extended Data Fig. 4c). However, prolonged exposure to day 5 resulted in poor embryoid body outgrowth (Fig. 2f). Reducing X1 to 7.5 μM yielded the same result (Extended Data Fig. 4d). The less potent analogue X16 also suppressed embryoid body outgrowth at 10 μM, but to a lesser extent, whereas X-negative had no effect at all (Extended Data Fig. 4d, e), thereby correlating compound potency with growth suppression of embryoid bodies. The effect of X1 is *Xist*-dependent, as pre-XCI (day 0) female ES cells, as well as day 3–5 XY male and XO female cells, all grew normally when treated with X1 (Extended Data Figs. 4f, g, 5a). Thus, X1 exerts a phenotypic effect on embryoid body outgrowth. This effect is *Xist*-dependent and is observed only in XX female ES cells.

To determine whether female-specific effects resulted from aberrant XCI, we performed RNA fluorescence in situ hybridization (FISH). In both control and X1-treated cells, *Tsix* RNA was observed at the Xa[5] and *Xist* RNA (red) was properly upregulated at the Xi at day 3, indicating initiation of XCI (Extended Data Fig. 5b, 6). Immunostaining for EZH2 and the repressive histone mark, H3K27me3, showed proper PRC2 enrichment on the Xi in 40% cells and H3K27me3 enrichment in 70–80% at 24 h of X1 treatment, similar to those for controls (Extended Data Figs. 5b, 6). However, X1 exposure for 72 h to day 5 resulted in a strong effect on PRC2 recruitment. Whereas *Xist* foci continued to be observed in X1-treated cells, PRC2 and H3K27me3 enrichment was severely blunted ($P < 0.01$; Fig. 2g, h). Only 10.3% and 5% of treated cells showed enrichment, respectively, whereas more than 76.2% and more than 92% of control cells displayed enrichment. These data indicate that X1 requires 1–3 days to take effect, as is characteristic of some pharmacological agents. A three-way interaction between *Xist* RNA, PRC2 and PRC1 is required for spreading along the Xi[25]. Of note, X1 treatment also did not affect enrichment of PRC1 (the RING1B subunit) and associated H2AK119ub histone marks at day 3, but significantly compromised their enrichment at day 5 (Fig. 2g, h, Extended Data Figs. 5b, 6). Notably, in female ES cells lacking RepA[25] (ΔA), immunostaining showed that H3K27me3 enrichment on the Xi was severely blunted in untreated cells and remained low in X1-treated cells (Fig. 2i, Extended Data Fig. 6b), further affirming that X1 action depends on RepA. Unlike EPZ-6438 and PF-06821497 (two known EZH2 inhibitors), X1 is not a general PRC2 inhibitor, as bulk H3K27me3 in cells was unchanged (Extended Data Fig. 6c). Together, these findings demonstrate that X1 does not affect *Xist* expression but instead affects its ability to recruit interacting proteins. X1 thereby uncouples *Xist* expression from polycomb function in cells.

To obtain a high-resolution view and assess potential off-target effects, we performed allele-specific chromatin immunoprecipitation with sequencing (ChIP–seq) in hybrid ES cells to distinguish Xa from Xi. Owing to a *Tsix* mutation[20], TST cells invariably inactivate the X-chromosome of *Mus musculus* (X^{mus}) origin and maintain the X of *Mus castaneus* (X^{cas}) origin as Xa. Whereas DMSO-treated cells showed PRC2 and H3K27me3 enrichment on X^{mus}, X1-treated cells (10 μM) did not (Fig. 3a, b, Extended Data Fig. 7a, b). Specific X-linked genes including *Xist* displayed reduced H3K27me3 (Extended Data Fig. 7c, d). By contrast, genes that escape XCI and autosomal genes were not affected (Extended Data Fig. 7e, f). Metagene analysis showed that X1 treatment resulted in strongly inhibited H3K27me3 and PRC2 accumulation specifically on the Xi (in X^{mus}), whereas the Xa and autosomes were minimally affected (Fig. 3c, Extended Data Fig. 7g). Notably, although 30–48 μM (IC_{50}) of X1 was required to disrupt RepA-protein interactions in vitro (Fig. 1f), cellular effects were observed at 5 μM. Cells may concentrate X1 intracellularly, as quantification using radiolabelled

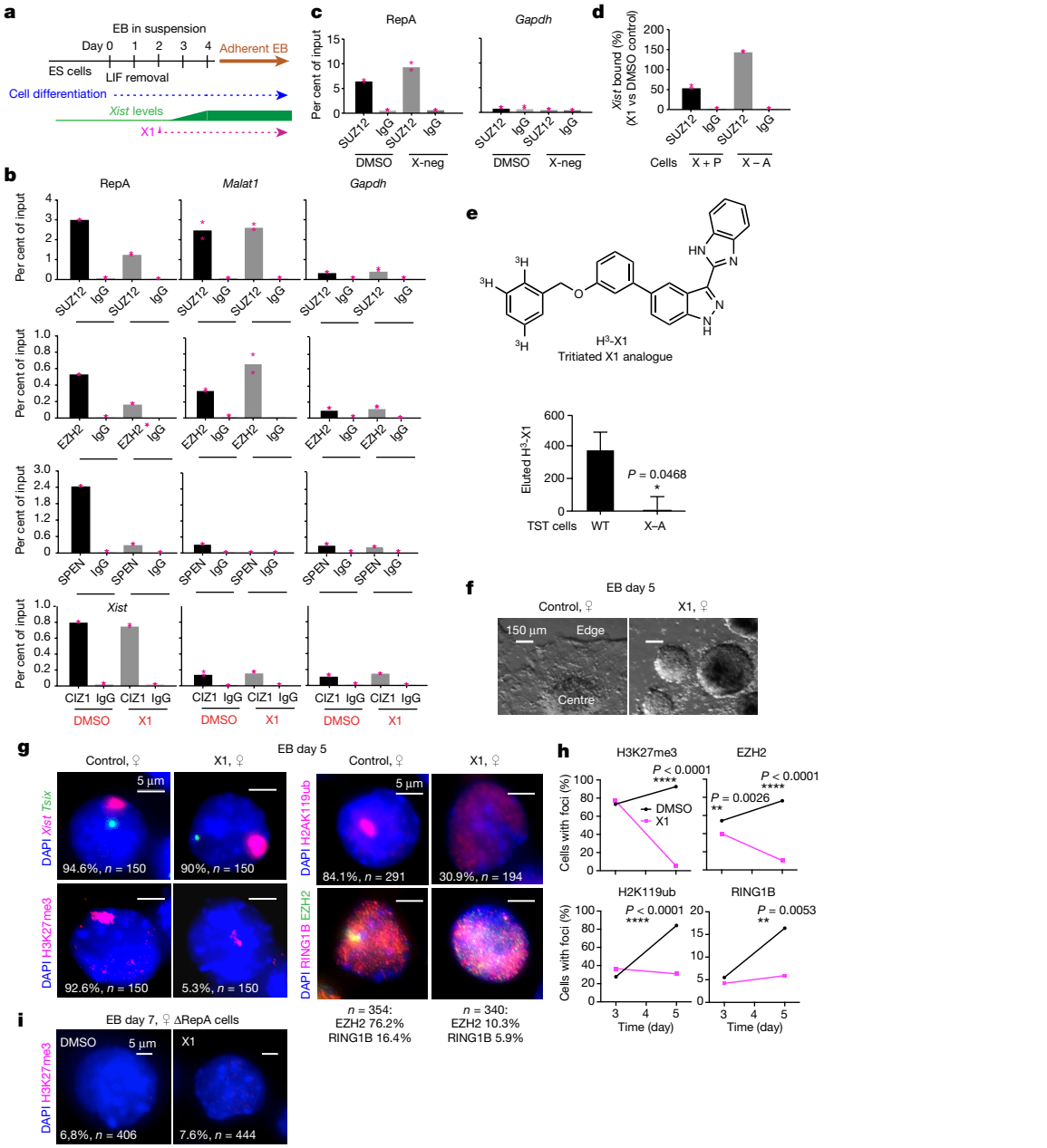


Fig. 2 | X1 weakens the *Xist* interaction with PRC2 and SPEN in a RepA-dependent manner inside cells. a, Schematic for analysis of XCI effects. X1 (10 µM) was added at day 2 and replenished daily with medium change. Embryoid bodies (EB) were grown in suspension, followed by adherent culture. **b**, **c**, RIP with RT–qPCR in day 4 female TST ES cells to evaluate *Xist* binding to SUZ12, EZH2, SPEN and CIZ1, as indicated, with or without X1 (**b**) or X-negative (X-neg) control compound (**c**). *Malat1* and *Gapdh* are negative control RNAs; IgG, control antibodies. Bars show mean; individual data points shown. $n = 3$ independent biological replicates, quantified in duplicate (representative graph shown). **d**, RIP with RT–qPCR in fibroblasts over-expressing transgenic *Xist* (X+P) or a RepA-deleted version (X–A). $n = 2$ independent biological replicates. **e**, Validation of target engagement using [^{3}H]X1. Data are mean ± s.d.; *P*-value by two-tailed Student's *t*-test; 3 biologically independent experiments. **f**, X1 inhibits growth of day 5 differentiating female TST cells. Scale bar, 150 µm. One representative field shown. **g**, *Xist* and *Tsix* RNA FISH and immunostaining for H3K27me3, H2AK119ub, EZH2 and RING1B in female TST embryoid bodies at day 5. One representative nucleus is shown. The percentage of cells with *Xist* foci is indicated. *n*, sample size. Scale bar, 5 µm. **h**, Enrichment of indicated protein epitopes across time. *P*-values by two-tailed χ^2 test from nuclei counted as shown in **g**. **i**, X1 inhibition depends on RepA, as shown by H3K27me3 immunostaining in day 7 female TST ES cells lacking RepA (ΔRepA). The percentage of cells with H3K27me3 enrichment is indicated. *n*, sample size from two biological replicates. Scale bar, 5 µm. See also Extended Data Fig. 4–6.

X1 (Fig. 2e) showed 180-fold higher concentrations inside cells (Supplementary Table 3). Thus, X1 shows strong on-target effects on PRC2 recruitment to the Xi.

Allele-specific RT–qPCR showed that X1 did not affect proper downregulation of *Tsix* (an *Xist* repressor[5]), upregulation of *Jpx* (an *Xist* activator[5]) or upregulation of *Xist* (Extended Data Fig. 8a); nor did it affect downregulation of pluripotency markers (for example, KLF4 and SOX2), suggesting normal ES cell differentiation. However, although there was also no effect on expression of Xa (in X^{cas}) genes, there was a major effect on silencing of Xi (in X^{mus}) genes between days 0–7 (Fig. 3d, pink line, Extended Data Fig. 8b). The effect was dose-dependent between 0–10 µM X1, with 10 µM resulting in full inhibition (Extended Data Fig. 8c). Notably, deleting RepA phenocopied X1 treatment (Fig. 3d, blue dots), further arguing that X1 effects are mediated through RepA.

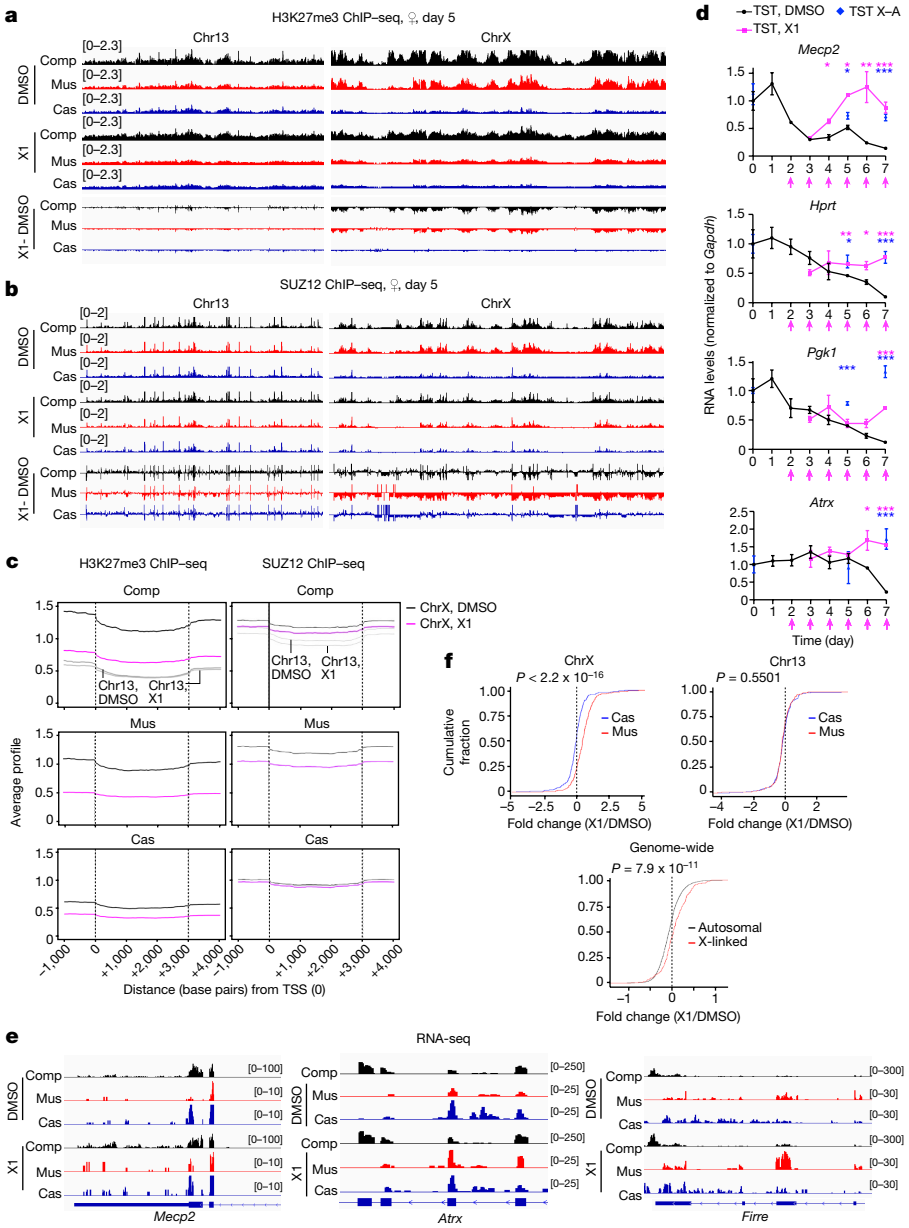


Fig. 3 | X1 treatment leads to Xi-specific loss of PRC2 and H3K27me3 enrichment and failure of XCI. a, **b**, Allele-specific H3K27me3 (**a**) and SUZ12 (**b**) ChIP–seq analyses of day 5 female embryoid bodies treated with 10 μM X1 or DMSO (control) for 72 h. Tracks for all reads (comp), *M. musculus* (Xi) and *M. castaneus* (Xa). X1–DMSO, subtraction of X1 reads from DMSO reads. Chromosomes 13 and X are shown, with a sliding window of 100 kb, step size 50 kb. Fragments per million mapped fragments (FPM) scale shown in brackets. **c**, Metagene analysis of the average Xi gene, Xa gene and Chr13 gene with or without X1 treatment. TSS, transcription start site. **d**, Time course of allele-specific RT–qPCR of indicated X-linked genes in DMSO- or X1-treated female embryoid bodies. TST X–A cells were analysed on days 5 and 7 phenocopy X1-treatment. *M. musculus* allele (Xi) shown. X1 was added on indicated days (pink arrows). *P*-values from two-tailed Student's *t*-test compared to DMSO control. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Data are mean ± s.d. of three biological replicates. **e**, RNA-seq analyses of day 5 DMSO- or X1-treated female embryoid bodies. Tracks shown for all reads (comp), *M. musculus* (Xi) and *M. castaneus* (Xa). FPM scale shown in brackets. **f**, Cumulative distribution plots of fold changes in gene expression in X1- versus DMSO-treated embryoid bodies. Genes with reads per kilobase of transcript, per million mapped reads (RPKM) >1 were included in the analysis. *P*-values by two-tailed Wilcoxon test.

The effects were also reversible, as Xi silencing and embryoid body growth resumed when X1 was withdrawn (Extended Data Fig. 8d, e). We conclude that X1 blocks PRC2 recruitment and Xi silencing in a RepA-dependent manner.

Allele-specific RNA-sequencing (RNA-seq) analysis verified the loss of Xi silencing. While control cells showed Xi silencing by day 5, X1-treated cells retained high-level Xi expression despite proper *Xist* upregulation (Fig. 3e, Extended Data Fig. 9a, b). Escapees and autosomal genes were not affected (Extended Data Fig. 9c, d). Cumulative distribution plots confirmed a significant shift in Xi expression relative to Xa expression in X1-treated cells (Fig. 3f, top left). Increased expression of Xi genes was also reflected in the right shift for X-linked genes relative to autosomal genes and to the genome (Fig. 3f, bottom). By contrast, there was no allele-specific effect on a representative autosome (Fig. 3f, top right). These data further support the on-target effect of X1 on Xi genes.

Although autosomes were unaffected in bulk, off-target changes could occur at individual genes. We therefore performed differential expression analysis and found that 197 autosomal genes showed significant changes, out of which 141 were upregulated and 56 were downregulated ($P < 0.05$). These changes could result directly from X1

treatment (true off-target effects) or indirectly from X-linked changes (secondary effects). To distinguish between them, we correlated the changes with H3K27me3 alterations, reasoning that direct effects would be accompanied by decreases in H3K27me3. Among the 197 affected autosomal genes, only 19 showed upregulation accompanied by H3K27me3 decrease (Extended Data Fig. 9e, quadrant II, Supplementary Table 4), and only 9 showed downregulation with increased H3K27me3 (quadrant IV). Quadrants I and III also contained only 14 and 1 genes, respectively. Therefore, autosomal effects were small relative to Xi effects (190 out of 335 X-linked genes expressed in ES cells) (Supplementary Table 5), and most may be secondary effects of changes in Xi.

The effect of X1 on RepA could occur either through steric blockade or a change in RNA conformation. Because X1 affects the interaction of RepA with both PRC2 and SPEN, we favoured the RNA conformation model. RNA can generally fold into many stable conformers, perhaps explaining why chemical probing of the RepA structure has yielded no consensus so far[26]. Here we used small-angle X-ray scattering (SAXS), which uses nanoscale density differences to determine size and shape distributions of biomolecules[27], in line with high-performance liquid chromatography (HPLC) to collect monodispersed samples without aggregation and degradation (Extended Data Fig. 10a). Quality analysis using PRIMUS showed an absence of upturned or downturned profile at low q regions, indicative that RNAs were free of aggregation or interparticulate interference (Extended Data Fig. 10b). The lower q region was used to generate the Guinier plot (Extended Data Fig. 10b, inset) and estimate the radius of gyration (R_g). Free RepA had an R_g of 83.11 ± 0.37 Å, and X1-treated RepA showed an R_g of 83.63±2.67 Å, suggesting similar overall dimensions and agreeing with HPLC data, in which RepA yielded similar elution profiles with or without X1 (Extended Data Fig. 10a). Dimensionless Kratky analysis showed a sharply bent profile (Extended Data Fig. 10c), implying that the RNA is folded in solution, similar to previous reports[28]. Pairwise distance distribution profile ($P(r)$) yielded real space R_g values for RepA versus RepA and X1 of 84.58 ± 0.17 and 85.66 ± 0.25 Å, respectively (Extended Data Fig. 10d), in close agreement with the Guinier analysis (Extended Data Fig. 10b, inset). The pairwise distance distribution profile also resulted in estimated maximum dimensions (D_{max}) of 263 and 277 Å for RepA and RepA with X1, respectively (Extended Data Fig. 10d), hinting at an underlying conformational change.

To determine the 3D structure of RepA in solution with and without X1, we used the ab initio modelling package DAMMIN. To reveal all possible conformations, 100 models were generated independently for each condition, followed by clustering of similar conformations using DAMCLUST. For RepA alone, although we identified 16 conformation clusters (C1–C16) (Fig. 4a, Extended Data Fig. 10e), C13 was dominant (Supplementary Video 1). These structures are distinct from previous models derived from population averaging in silico and/or chemical probing methods, which ranged from simple two-hairpin motifs[6] to complex structures involving inter-repeat duplexes[26]. An advantage of SAXS is the ability to account for conformational polydispersity[27]. The existence of 16 RepA conformers may explain why existing chemical studies have not resulted in a consensus structure for RepA[26]. Addition of X1 reduced RepA's conformational space to six, with one dominant form (C6′) that is distinct from all previous RepA structures (Fig. 4b, Supplementary Video 2). C6′ accounted for 78% of all possible conformations. Thus, SAXS analysis has enabled a 3D solution-based visualization of RepA and revealed that X1 reduces its conformational heterogeneity and induces formation of new structures. We propose that X1 acts on RepA by stabilizing a de novo conformation with a reduced affinity for cognate interactors, explaining the loss of XCI in X1-treated cells.

We have shown that non-coding RNA can be systematically targeted by a small molecule for phenotypic impact. X1 engages the RepA motif of *Xist*, disrupts its conformation, displaces interacting proteins (PRC2, PRC1 and SPEN) and abrogates XCI, without affecting *Xist* expression or localization (Fig. 4c). Medicinal chemistry studies of X1 could further increase its potency and specificity, potentially for Xi reactivation and treatment of X-linked diseases such as Rett syndrome[29]. Unlike compounds shown retrospectively to bind RNA[9,10], our hits arose not from phenotypic screens but from a direct RNA-binding screen. ALIS is agnostic to the mechanism of action and is potentially applicable to any RNA, including the vast majority for which there is no detailed structural information, despite most efforts being focused on sequence- or structure-based design[11,30]. The large size, multiple modular domains, likelihood of being mostly unstructured, and coexisting structural subpopulations of *Xist* present a unique challenge that was circumvented by ALIS. Total, there were fewer total hits for RepA than for typical protein targets using ALIS, possibly reflecting the original design of the diversity set to target proteins[13]. A goal of future research will be the development of curated sets with diverse RNA-binding properties to enhance identification of RNA-targeting drugs.

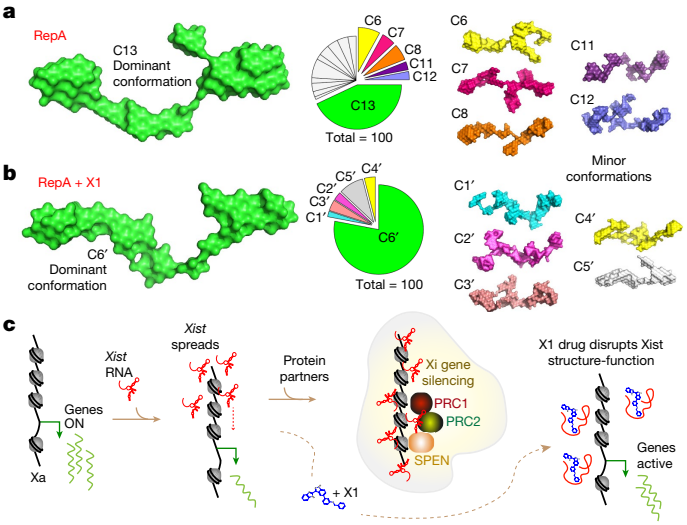


Fig. 4 | 3D structure of RepA with and without X1. a, Sixteen representative conformations of RepA (C1–C16). C13 (43% of species) is shown larger (left). The pie chart shows the relative abundance of structural clusters. See Extended Data Fig. 10 for all 16 conformations. **b**, Six representative conformations of RepA with X1 (C1′–C6′). C6′ (78% of species) is shown larger (left). The pie chart shows the relative abundance of structural clusters. **c**, Schematic showing XCI and the effects of adding X1.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04537-z.

1. Consortium, E. P. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
2. Santos, R. et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).
3. Warner, K. D. et al. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **17**, 547–558 (2018).
4. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–R110 (2015).
5. Lee, J. T. Epigenetic regulation by long noncoding RNAs. *Science* **338**, 1435–1439 (2012).
6. Wutz, A. et al. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat. Genet.* **30**, 167–174 (2002).
7. Rizvi, N. F. & Smith, G. F. RNA as a small molecule druggable target. *Bioorg. Med. Chem. Lett.* **27**, 5083–5088 (2017).
8. Disney, M. D. et al. Drugging the RNA world. *Cold Spring Harb. Perspect. Biol.* **10**, a034769 (2018).

9. Howe, J. A. et al. Selective small-molecule inhibition of an RNA structural element. *Nature* **526**, 672–677 (2015).
10. Palacino, J. et al. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.* **11**, 511–517 (2015).
11. Disney, M. D. et al. Inforna 2.0: a platform for the sequence-based design of small molecules targeting structured RNAs. *ACS Chem. Biol.* **11**, 1720–1728 (2016).
12. Allen Annis, D. et al. An affinity selection-mass spectrometry method for the identification of small molecule ligands from self-encoded combinatorial libraries—Discovery of a novel antagonist of E-coli dihydrofolate reductase. *Int. J. Mass Spectrom.* **238**, 77–83 (2004).
13. Rizvi, N. F. et al. Discovery of selective RNA-binding small molecules by affinity-selection mass spectrometry. *ACS Chem. Biol.* **13**, 820–831 (2018).
14. Rizvi, N. F. et al. Targeting RNA with small molecules: identification of selective, RNA-binding small molecules occupying drug-like chemical space. *SLAS Discov.* **25**, 384–396 (2020).
15. Cifuentes-Rojas, C. et al. Regulatory interactions between RNA and polycomb repressive complex 2. *Mol. Cell* **55**, 171–185 (2014).
16. Chillon, I. et al. Native purification and analysis of long RNAs. *Methods Enzymol.* **558**, 3–37 (2015).
17. Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **1**, 337–341 (2004).
18. Monfort, A. et al. Identification of Spen as a crucial factor for *Xist* function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.* **12**, 554–561 (2015).
19. Lee, M. K. et al. A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication. *Chem. Commun.* **50**, 368–370 (2014).
20. Ogawa, Y. et al. Intersection of the RNA interference and X-inactivation pathways. *Science* **320**, 1336–1341 (2008).
21. Zhao, J. et al. Genome-wide identification of polycomb-associated RNAs by RIP–seq. *Mol. Cell* **40**, 939–953 (2010).
22. Patil, D. P. et al. m^6A RNA methylation promotes *XIST*-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
23. Sunwoo, H. et al. Repeat E anchors *Xist* RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc. Natl Acad. Sci. USA* **114**, 10654–10659 (2017).
24. Jeon, Y. & Lee, J. T. YY1 tethers *Xist* RNA to the inactive X nucleation center. *Cell* **146**, 119–133 (2011).
25. Colognori, D. et al. *Xist* deletional analysis reveals an interdependency between *Xist* RNA and polycomb complexes for spreading along the inactive X. *Mol. Cell* **74**, 101–117 e110 (2019).
26. Liu, F. et al. Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat. Chem. Biol.* **13**, 282–289 (2017).
27. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570–2577 (2015).
28. Kim, D. N. et al. Zinc-finger protein CNBP alters the 3-D structure of lncRNA *Braveheart* in solution. *Nat. Commun.* **11**, 148 (2020).
29. Carrette, L. L. G. et al. A mixed modality approach towards Xi reactivation for Rett syndrome and other X-linked disorders. *Proc. Natl Acad. Sci. USA* **115**, 1715124115 (2017).
30. Stelzer, A. C. et al. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* **7**, 553–559 (2011).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

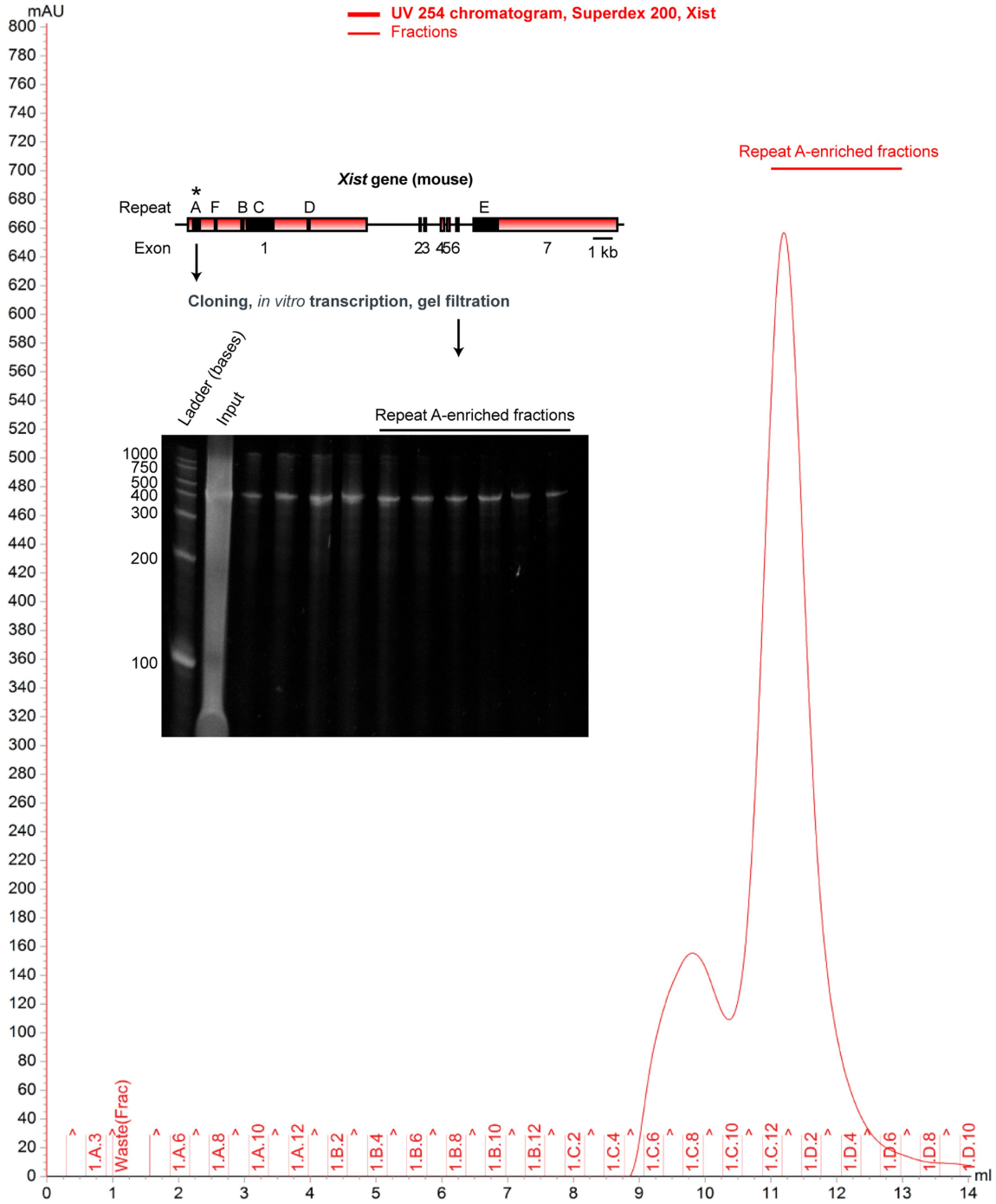
Sequence data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) with accession number GSE141683.

Acknowledgements We thank all members of the Lee and Patel laboratories, Charles Lesberg and other Merck team members for stimulating scientific discussions. We acknowledge the Merck MINt Award, a grant from the US National Institutes of Health (R01-HD097665), and funding from the Howard Hughes Medical Institute to J.T.L.; the Pew Charitable Trust Latin American Fellows Program to R.A. and C.R.; and the MGH Fund for Medical Discovery to R.A. M.D.B. and T.M. acknowledge MITACS and NSERC PGS-D fellowships respectively. T.R.P. is Canada Research Chair in RNA and Protein Biophysics. We thank DIAMOND Light Source B21 beamline staff for their help with data collection (BAG SM22113).

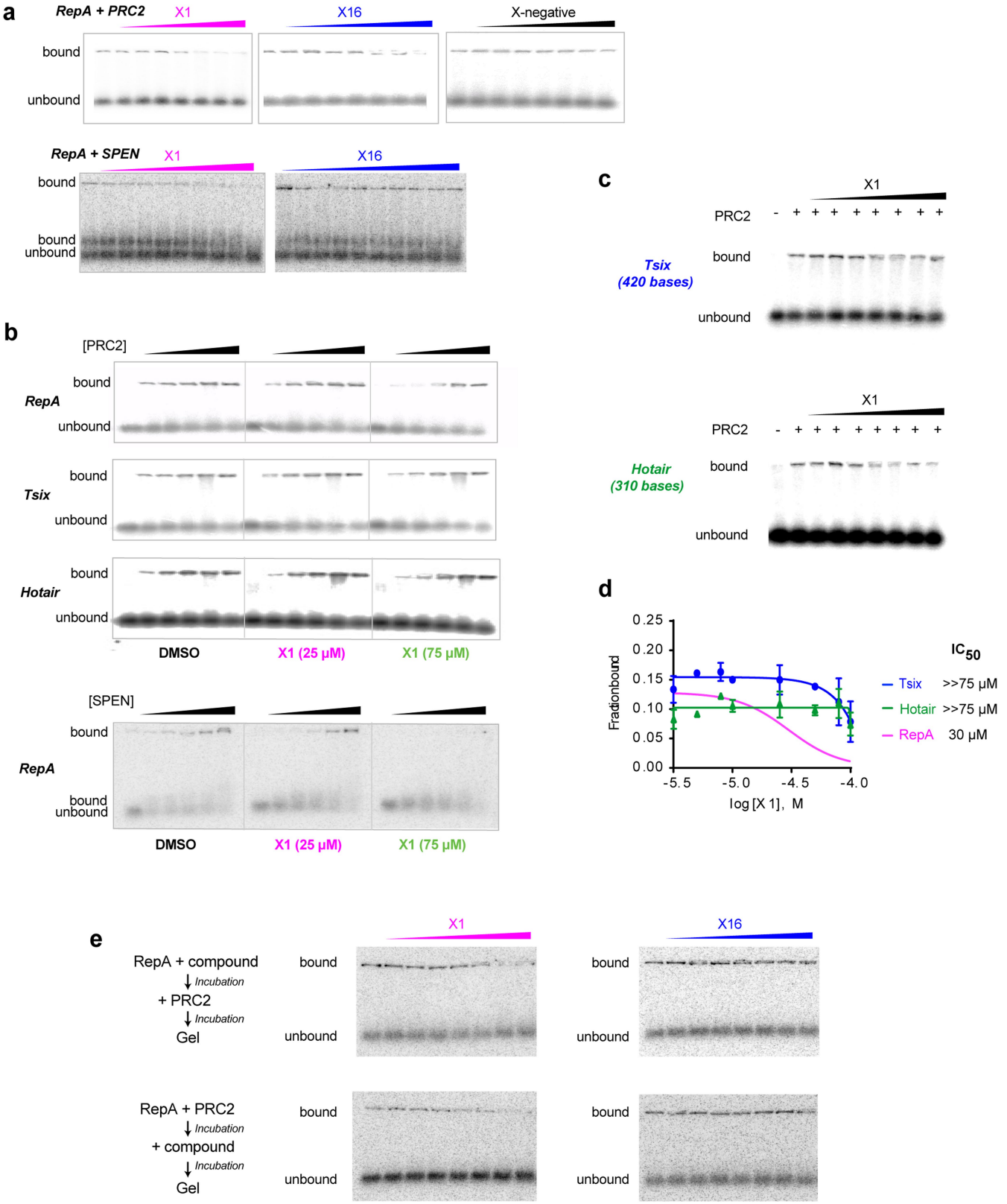
Author contributions R.A., K.S., B.K., N.R., T.M., M.D.B., J.M., G.S., J.B., P.J.D., T.R.P., E.N. and J.T.L. designed experiments and interpreted data. R.A. performed RNA purifications, biochemical assays, immunoFISH, ChIP–seq and RNA-seq. K.S. coordinated the small molecule synthesis and off-target activity evaluations. N.R. and E.N. performed the ALIS determinations. C.R. performed RNA purifications. B.K. conducted bioinformatics analyses. T.M., M.D.B. and T.R.P. performed and analysed SAXS data. R.A., K.S. and J.T.L. wrote the paper.

Competing interests K.B.P., N.F.R., J.D.M., G.F.S., J.B., P.J.D. and E.B.N. are current or former employees of Merck & Co. and may hold stock or other financial interests in Merck & Co. J.T.L. is a cofounder of Translate Bio and Fulcrum Therapeutics and is also a scientific advisor to Skyhawk Therapeutics.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04537-z.
Correspondence and requests for materials should be addressed to Jeannie T. Lee.
Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.

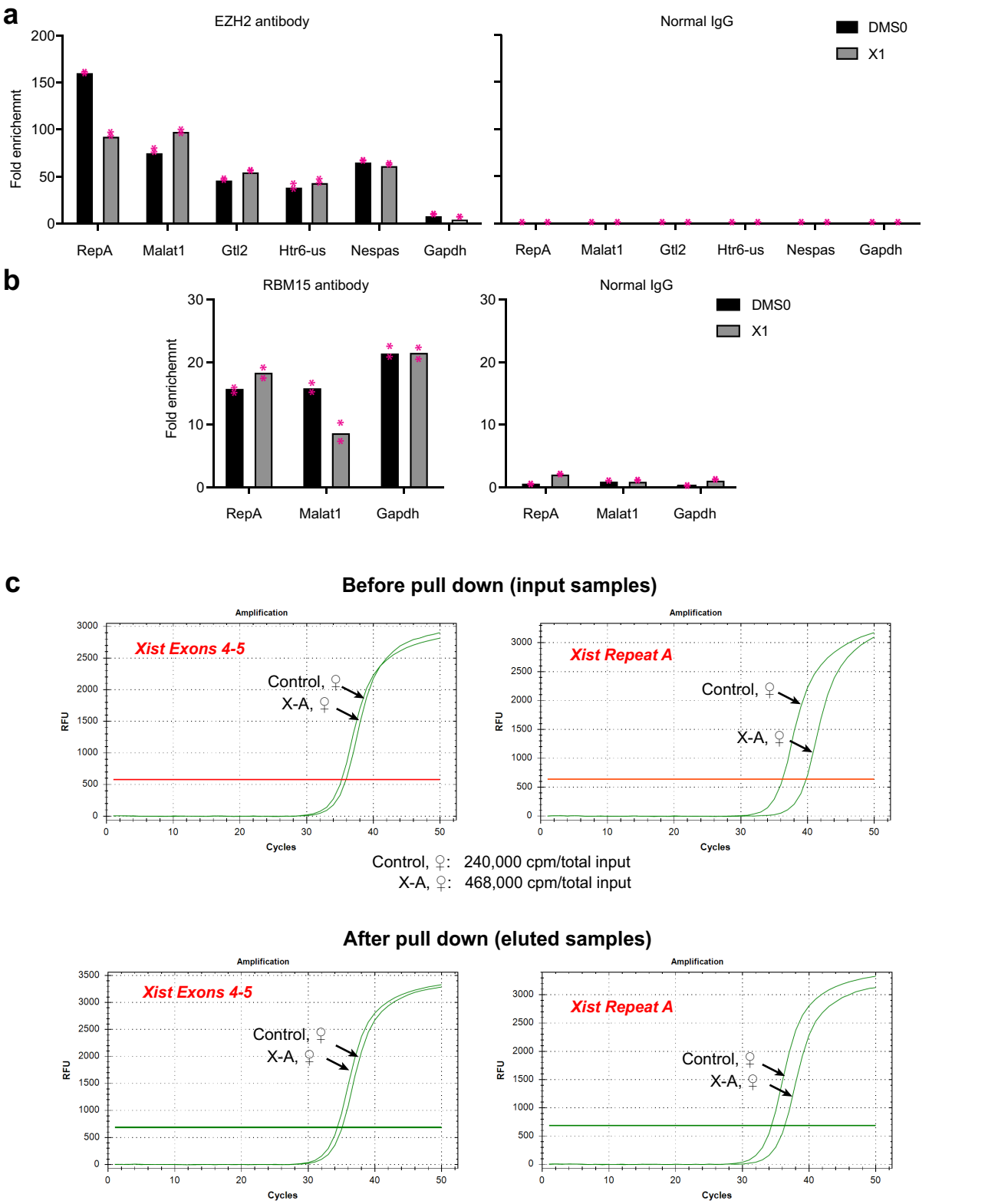


Extended Data Fig. 1 | Purification of Xist RepA RNA. A 431 Repeat A fragment of Xist RNA was *in vitro* transcribed and purified under native conditions by FPLC. A representative chromatogram is shown. To confirm size and stability of the sample just prior to ALIS, we visualized the RNA in a denaturing urea-PAGE.

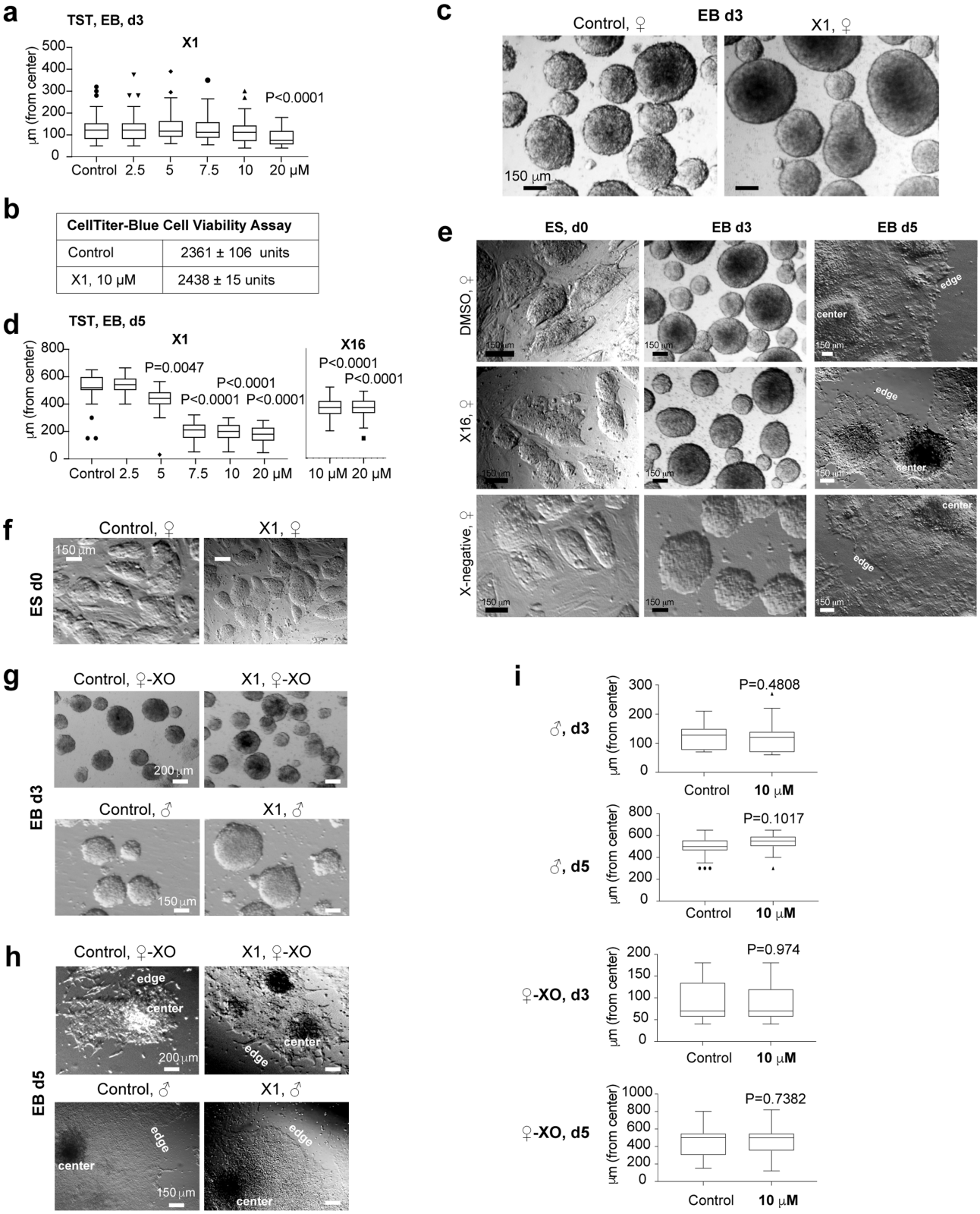


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | X1 inhibits interaction of Xist RepA with cognate interacting proteins *in vitro*. **a**, RNA EMSAs show that X1 weakens interaction between RepA and PRC2, and RepA and SPEN-RRM. Increasing concentrations of the compounds (0, 5, 7.5, 10, 25, 50, 75, 100 μM) were titrated against 0.5 nM RNA and 15.6 nM PRC2, or 0.1 nM RNA and 158 nM SPEN-RRM. Two replicates showed similar results. **b**, RNA EMSAs titrating PRC2 (0, 15.6, 31.2, 62.4, 124.9, 250 nM) or SPEN-RRM (0, 79.2, 158, 396, 792, 1580 nM) against a fixed concentration of X1 (25 or 75 μM) and 0.5 nM RepA, Tsix (reverse complement of RepA), or Hotair RNA—all of which are known PRC2 interactors. For SPEN, RNA was 0.1 nM. Two or more replicates showed similar results. **c**, Increasing concentrations of X1 (0, 5, 7.5, 10, 25, 50, 75, 100 μM) was titrated against 0.5 nM RNA (Tsix, Hotair) and 15.6 nM PRC2. One representative gel of two replicates is shown. **d**, Densitometric analysis to determine IC_{50}, which were too high to be measured for Tsix and Hotair. Data are represented as mean +/− SD. n = 2 independent experiments. RepA result from Fig. 1f is shown as reference. **e**, Order of addition does not affect X activity. Increasing concentrations of the compounds (0, 5, 7.5, 10, 25, 50, 75, 100 μM) was titrated against 0.5 nM RepA and 15.6 nM PRC2. One representative gel of two replicates is shown. **Top**, PRC2 was added to a RepA-molecule pre-incubated mix. **Bottom**, Molecule was added to a RepA-PRC2 pre-incubated mix.

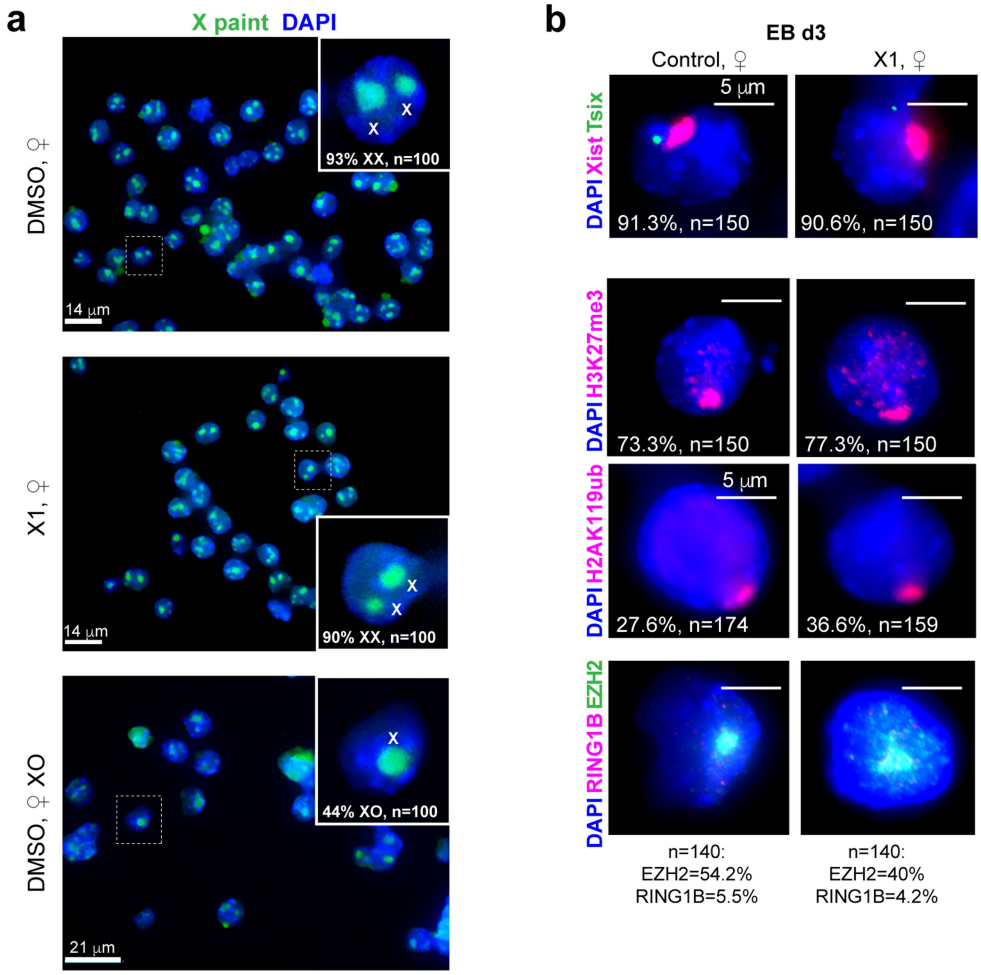


Extended Data Fig. 3 | X1 also inhibits interaction of Xist RepA with cognate interacting proteins *in vivo*. a-b, RIP-qPCR analysis in d4 female TST ES cells to evaluate Xist binding to EZH2 **(a)** and RBM15 **(b)** in 10 μM X1. IgG, negative control antibody. Other EZH2 interactors Malat1, Gtl2, Htr6-us and Nespas are shown. Gapdh, negative control RNA. Bars: mean. Individual data points included. n = 2 biologically independent experiments quantified in duplicate (Representative graph shown). **c, Top:** RT-qPCR confirms similar quantities of Xist RNA in control and X-A samples prior to Xist RNA pulldown. Xist exons 4-5 primers were used. **Bottom:** Similar quantities were also present following Xist RNA pulldown, thereby ruling out unequal Xist expression as a cause of unequal H^3 radioactive counts. X-A cells amplified poorly with RepA primers, consistent with deletion of RepA.

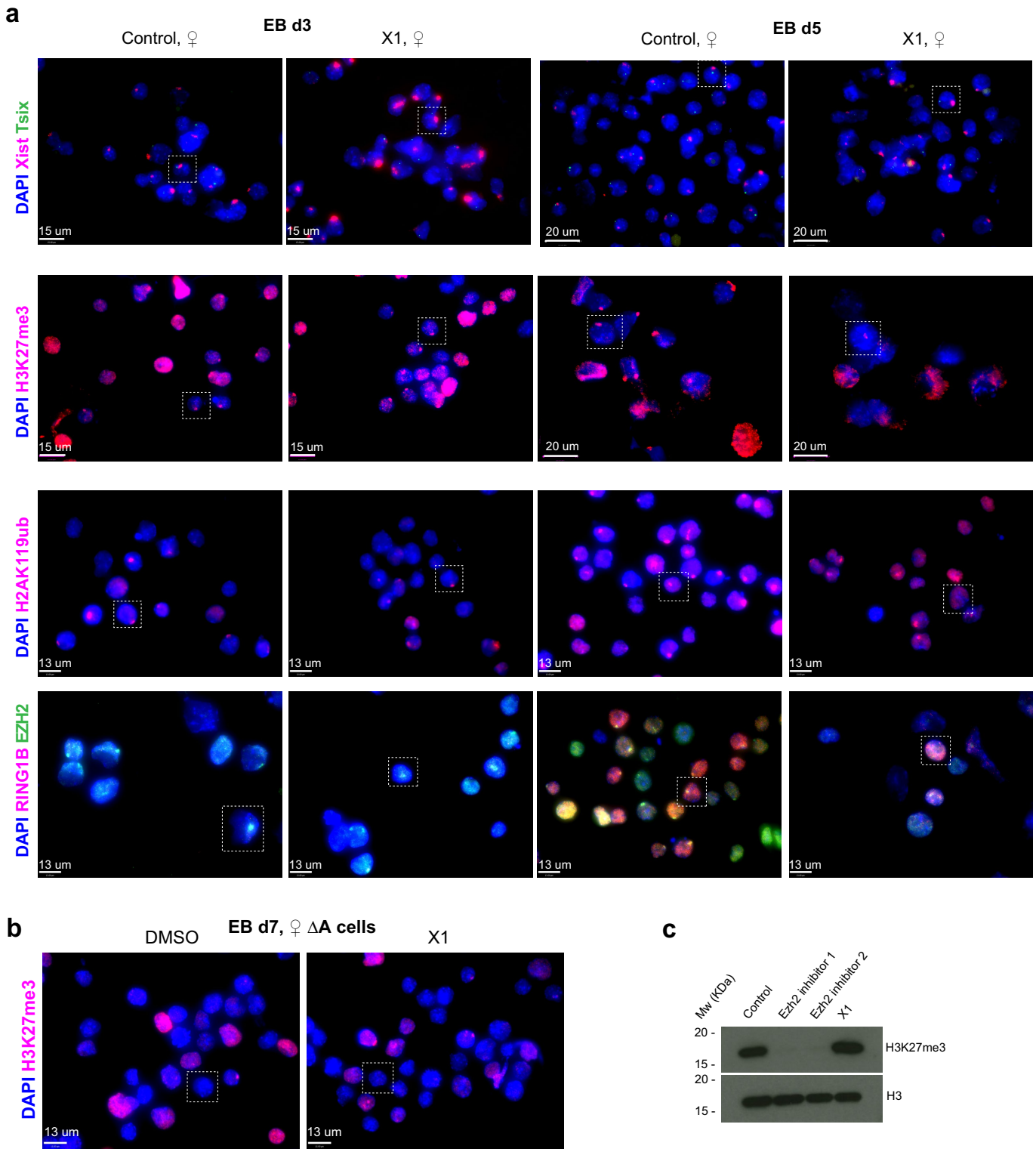


Extended Data Fig. 4 | See next page for caption.

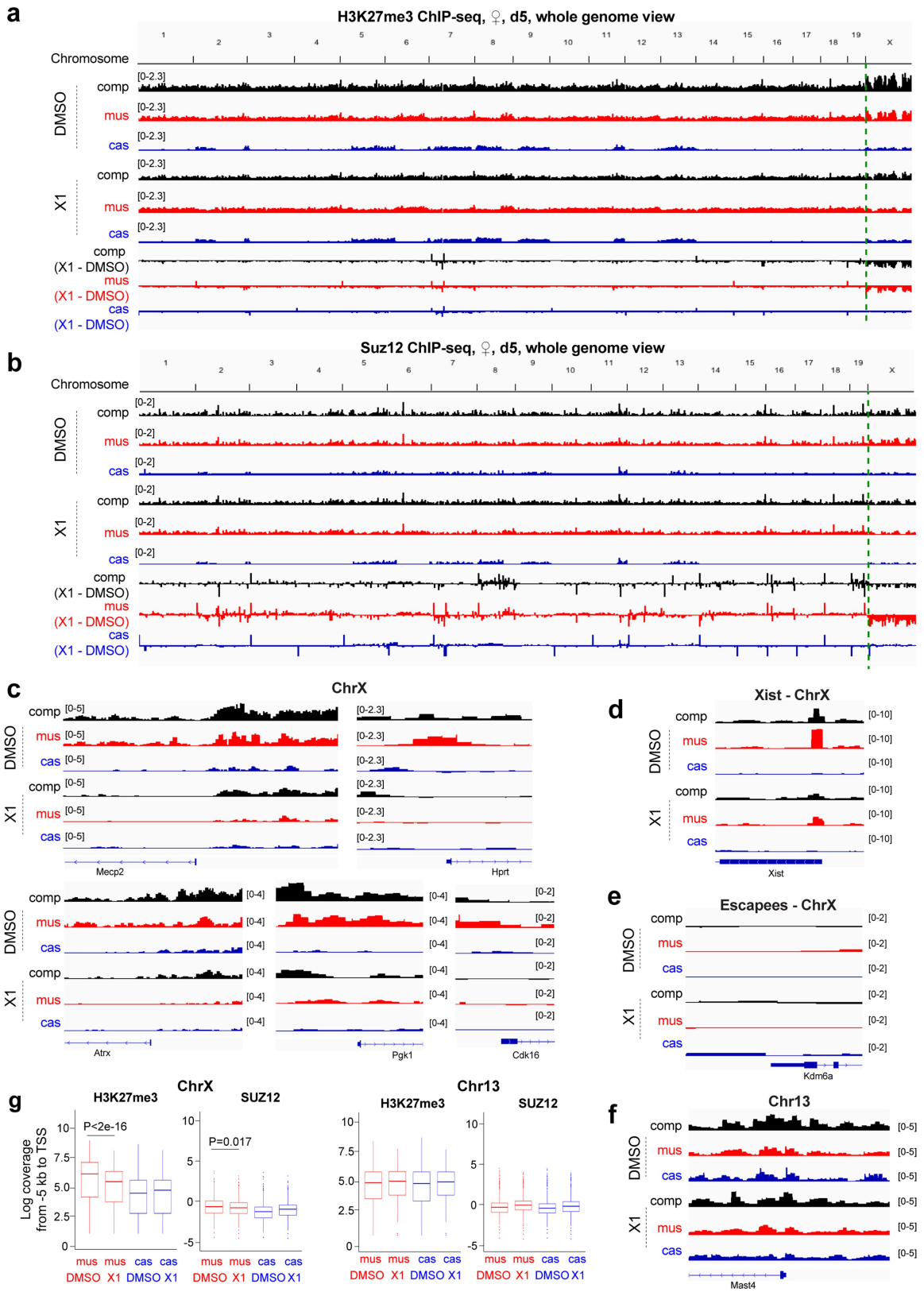
Extended Data Fig. 4 | X1 effects on EB outgrowth in ♀-TST-XX, ♀-XO, and ♂-XY EB cells. a, Growth of differentiating ♀-TST cells at day 3, or 24 h post-X1 treatment, up to 10 μM X1. Data are represented as Tukey box plots. Lower whisker: 25th percentile minus 1.5xInterquartile Range (IQR). Higher whisker: 75th percentile plus 1.5xIQR. Box range: 25th (bottom) to 75th (top) percentile. Line within box: median. Points beyond higher whisker are shown. *P*-values: one-way ANOVA with respect to control cells. n = 150 colonies combined from 3 independent experiments. **b**, Viability of d5 cells. n = 3 biologically independent experiments. **c**, No obvious effect on day 3 female EB growth after 24 h X1 treatment. **d**, Quantitation of EB outgrowth at day 5 (72 h post-drug application). The distance from EB center to edge of outgrowth was measured in 100 d3 or 30 d5 EBs combined from 3 independent experiments. Data presented as in panel (a). *P*-values: one-way ANOVA with respect to control cells. **e**, Weaker effect of X16 on ♀-TST EB outgrowth. No obvious effect of X-negative. One representative brightfield microscopy from 3 independent cultures is shown. Center of the EB and edge of outgrowth as marked. Scale as indicated. **f**, X1 had no effect on growth of pre-XCI (d0) female cells. **g,h**, X1 also did not inhibit ♀-TST-XO and ♂-XY ES cells at day 3 **(g)** or day 5 **(h)**. Neither cell line expresses Xist or undergoes XCI. One representative field is shown. Scale bar, 150 μm. **i**, Quantitation of EB outgrowth in XY male and XO female EBs at days 3 and 5. Distance from the EB center to the edge of outgrowth was measured. Day 3: n = 136, XO colonies; n = 112, XY colonies. Day 5: n = 40, XO colonies; n = 60, XY colonies). Data presented as in panel (a).



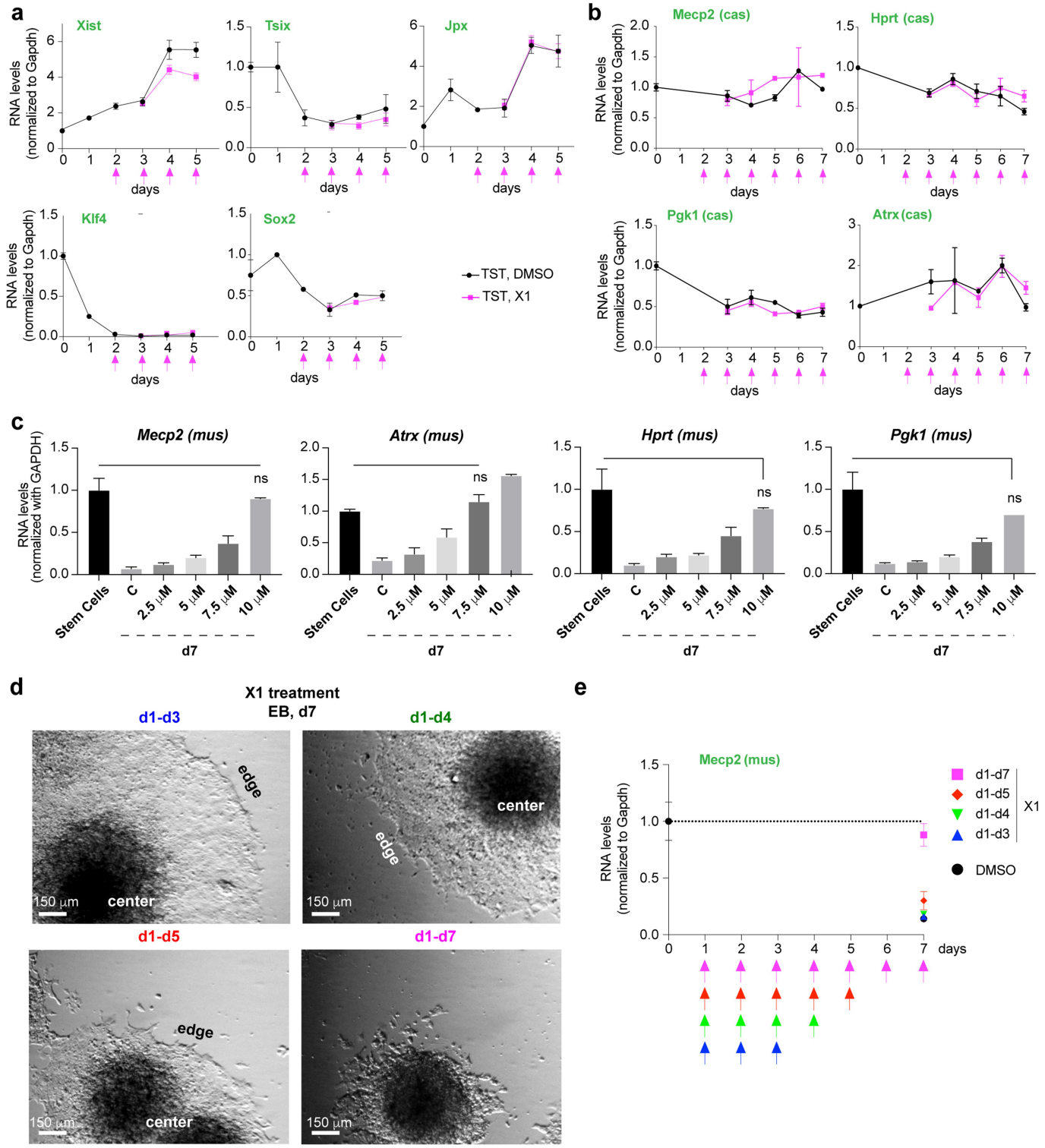
Extended Data Fig. 5 | Karyotype analysis of ES cells and RNA immunoFISH analysis of day 3 X1-treated cells. a, X-chromosome painting DNA FISH of DMSO- and X1-treated XX TST cells, and a DMSO-treated XO clone that spontaneously arose from the XX TST cells. Scale as shown. Inset: magnification of representative nucleus. %nuclei with indicated X chromosome number shown. n, sample size combining from 3 biologically independent experiments. **b**, Xist/Tsix RNA-FISH and immunostaing for H3K27me3, H2AK119ub, EZH2, and RING1B in ♀-TST EB at day 3. One representative nucleus is shown. %cells with Xist foci is indicated. n, sample size. Scale bar, 5 μm.



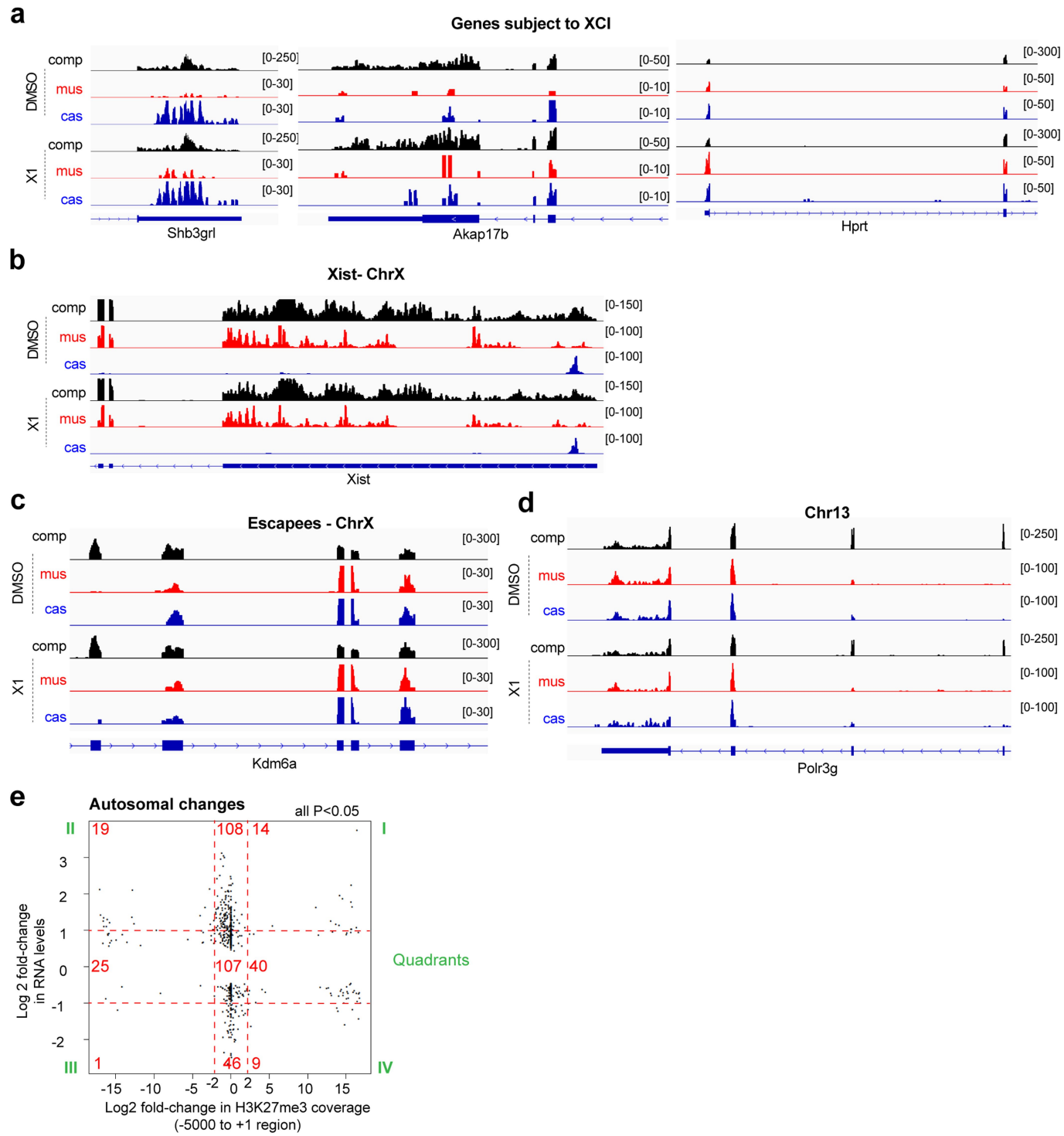
Extended Data Fig. 6 | Full fields for RNA immunoFISH experiments of Fig. 2 and Extended Data Fig. 5. a, Full fields for the RNA FISH and Immunofluorescence experiments, with boxed nuclei presented in Fig. 2g and Extended Data Fig. 5b. **b**, Full fields for H3K27me3 immunostaining of DMSO- or X1-treated ♀ TST-A cells, with boxed nuclei presented in Fig. 2h. %cells with foci on the Xi as indicated (sample size, n, from two biologically independent experiments combined). **c**, Western blot using H3K27me3 and total histone H3 antibodies. Total cell extracts were obtained from day 7 female EB cells after treating with 10 µM of various compounds from day 2. Compounds: EZH2 inhibitor 1 (EPZ-6438, MedChem Express), EZH2 inhibitor 2 (PF-06821497, Pfizer), or X1. One representative film of two replicates is shown.



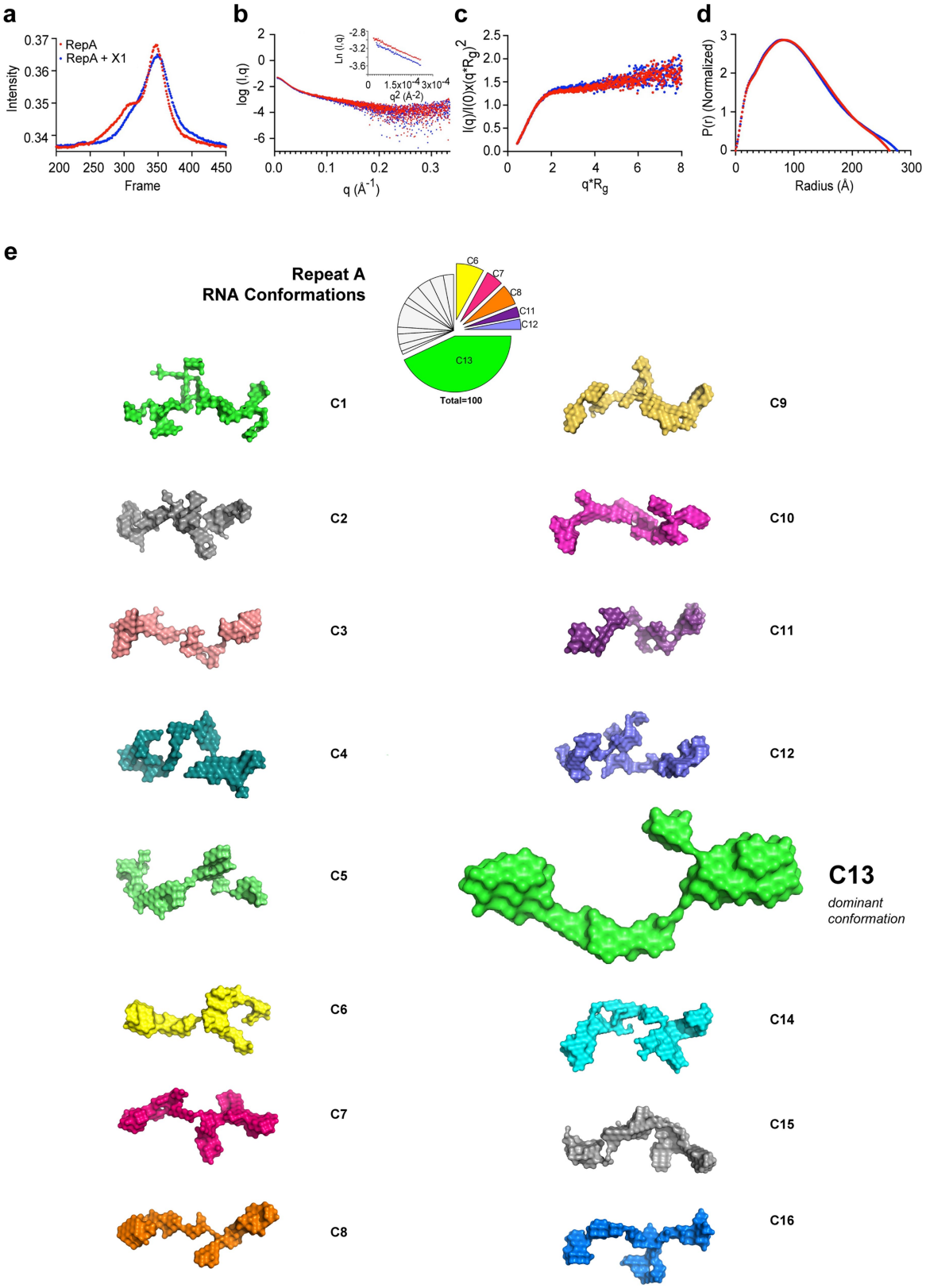
Extended Data Fig. 7 | Epigenomic analyses of PRC2 and H3K27me3 enrichment. a–b, Allele-specific H3K27me3 **(a)** and SUZ12 **(b)** ChIP-seq analyses of day 5 female EB treated with 10 μM X1 or DMSO (control) for 72 h. Tracks for all reads (composite, "comp"), mus (Xi), and cas (Xa). Dotted green lines separate ChrX. **c–f**, Zoom-ins for allele-specific H3K27me3 ChIP-seq analyses of day 5 female EB treated with 10 μM X1 or DMSO (control) for 72 h. Browser shots shown with sliding window 1 kb, step size 0.5 kb. Scale shown in brackets. **c**, X-linked genes subjected to XCI. **d**, the *Xist* gene. **e**, Escapees. **f**, Representative control autosomal gene on Chr13. **g**, Box plot of normalized read densities for the −5000 to +1 region of ChrX and Chr13 refSeq genes, parsed into mus and cas alleles. Lower whisker: 10th percentile. Higher whisker: 90th percentile. Box range: 25th (bottom) to 75th (top) percentile. Line within box: median. Points beyond whiskers are shown. *P*-values: two-tailed Wilcoxon test from data gathered from individual H3K27me3 and Suz12 ChIP-seq experiments.



Extended Data Fig. 8 | Analysis of gene expression and X1 reversibility. a, Time course RT-qPCR of indicated control genes in DMSO- or X1-treated female EB. X1 added on indicated days (pink arrows). Mean and S.D. shown for 3 biological replicates. **b**, Time course allele-specific RT-qPCR of indicated Xa genes in DMSO- or X1-treated female EB. X1 added on indicated days (pink arrows). Mean and S.D. shown for 3 biological replicates. **c**, Dose-response analysis in the range of 0–10 µM X1 compound. Allele-specific RT-qPCR of indicated X-linked genes in DMSO- or X1-treated female EB. Mus allele (Xi) shown. X1 was added on d2. *P*, two-tailed Student's *t*-test with respect to DMSO-treated TST control. Mean and S.D. shown for 2 replicates. At 10 µM X1, the Student's t-test reveal no significant difference between d7 cells and expression found in control ES cells. **d**–**e**, Female EB were grown from d1 in 10 µM X1 and the treatment was suspended on day 3, 4, 5, or maintained up to day 7. The growth morphology (**d**) and Mecp2 expression from the Xi is evaluated at d7 (**e**). One representative brightfield microscopy from 3 independent cultures is shown.



Extended Data Fig. 9 | Transcriptomic studies of on- and off-target effects. **a**–**d**, RNA-seq analyses of day 5 DMSO- or X1-treated female EB. Zoom-ins to representative X-linked genes subjected to XCI **(a)**, Xist **(b)**, escapee gene (**c**) and autosomal gene (**d**). Tracks for all reads (comp), mus reads (Xi), and cas reads (Xa). FPM scale shown in brackets. **e**, Differentially expressed autosomal genes (y axis) and their corresponding changes in H3K27me3 enrichment (x axis). Each dot represents a gene. Number of genes on each of the nine sections as shown. Comp tracks were sampled to the smallest library, then MultiTesting and IndependentFiltering DESeq2 filtering was performed reporting significance below 0.05 (Wald test) after Benjamini and Hochberg correction with the application of independent intensity filtering.



Extended Data Fig. 10 | Sixteen conformational clusters identified for native RepA RNA without X1 treatment. a, HPLC-SEC profile of the purified RepA RNA with or without X1 previous to SAXS data collection. **b**, PRIMUS analysis for initial data quality analysis. Inset: Guinier plot to determine the Radius of Gyration (R_g). **c**, Dimensionless Kratky analysis [q*Rg vs. I(q)/I(0) x(q*Rg)2] of samples. **d**, Pairwise distance distribution profile (P(r)) to estimate the real space dimensions of the molecule in Å. **e**, 16 clusters (C1–C16) of RepA are presented in their native state without X1. C13 is the dominant conformation. Pie-chart shows relative abundance of structural clusters. See also Fig. 4.

nature portfolio

Corresponding author(s): Jeannie T. Lee

Last updated by author(s): Nov 29, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Small-angle X-ray scattering (SAXS) - data collection step

The SAXS data was collected at the B21 BioSAXS beamline at the Diamond Light Source (Didcot, UK) using a high-performance liquid chromatography (HPLC-SAXS) as described previously. The SAXS data was imported in to ScÅtter (Rambo R. Diamond Light Source ScÅtter, a JAVA-based application for basic analysis of SAXS datasets. 2017; Didcot ver. 3.1R) where frames from peak region of each sample were selected and buffer subtracted.

Data analysis

Small-angle X-ray scattering (SAXS) - data analysis step

The selected frames were then merged using PRIMUS of ATSAS suite (version 3.0), and the merged files were used for all subsequent data analyses. First, Guinier analysis was performed to estimate the radius of gyration (RG) and evaluate homogeneity of the sample. Next, we performed dimensionless Kratky analysis to evaluate the folding extend of RepA, with and without the presence of X1. Finally, using GNOM of ATSAS suite (version 3.0) we generated pairwise distance distribution function (P(R)) to provide RG (real space) and maximum dimension (Dmax). The P(R) data was subsequently implemented in DAMMIN of ATSAS suite (version 3.0) to generate 100 dummy residue models. In order to identify the predominant scattering species, these models were averaged and clustered using DAMCLUST of ATSAS suite (version 3.0). All models are represented using PyMol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

Bioinformatic analyses (RNA-seq and ChIP-seq)

Adapters were removed from day 5 TST RNAseq library with software trim_galor/cutadapt (versions 0.4.3/1.7.1) with default parameters to remove adapters. Adapters were not removed from other libraries. Mapping performed using software Novalign (version 3.09) or STAR (version STAR_2.5.3a) against the mouse C57BL/6J (mm9) as well as variants M. musculus (mus) and M. castaneus (cas) 12, considering directional, first strand, for RNA-seq and unstranded reads for ChIP-seq. Salient parameters for Novalign are -i300 100 -t180 -F ILM1.8 -

rRandom -h180 180 -v180 and STAR are --outSAMprimaryFlag AllBestScore --outMultimapperOrder Random --outSAMtype BAM Unsorted --outFilterMultimapNmax 100 --outFilterMismatchNmax 6 --outSAMattributes All –sjdbFileChrStartEnd. Bigwig files were generated 12, and read in the Integrative Genomics Viewer.
DEseq2 (version 1.24.0 or greater) is used to quantitate RNA seq counts over genes using software Seq Monk (version 1.44.0 or greater). We limited our analyses to highly expressed genes (RPKM > 1 in the DMSO control) and for cumulative distribution we set a liberal DEseq2 p-values (p<1) to see the general trend. Plots were attained converting counts to log2 fold change in awk and then importing to R where ggplot2 (version 3.2.0 or greater), stat_ecdf() and Wilcox.test were used to generate statistics. For scatterplots comparing ChIP to RNA seq results, comp tracks were sampled to the smallest library, then MultiTesting and IndependentFiltering DESeq2 filtering is performed reporting significance below 0.05 after Benjamini and Hochberg correction with the application of independent intensity filtering, reporting shrank log2 values.
Metagene analysis of d5 tst cells of H3K27me3 and Suz12 marks is performed using the software ceas (version 1.0.2 under python 2.7) on FPM scaled coverage (wiggle format) generated over active genes plus or minus 3000 nt. Highly expressed genes are used to avoid conflating repressed genes on Xa with those on Xi, and are defined as having FPKM greater than 1 in d5 tst cells strand specific RNA sequencing experiments. We used several equal methods to determine FPKM, including software packages SeqMonk and Homer (version 4.10). When we use the software package Homer we performed the following operations: tags directories are created from uniquely aligned reads to mm9 using makeTagDirectory tags with parameters -flip -sspe -removeSpikes 10000 8 -tbp 1 -genome mm9 -checkGC. Note effective removal of PCRd dups using tbp. The function AnalyzeRepeats is used to generate raw counts over genes. FPKM values are derived scaling to gene length per 1000 and per million. Obtaining FPKM from SeqMonk is derived first using the RNAseq quantification pipeline to generate raw counts. Then, quantitate existing probes using read count quantification choosing correct for total read count using per million reads and correct for probe length.ragments derived from uniquely aligned paired ends without PCR dups removed.
Additional ChIP analysis for comparing to RNAseq, is performed in Seq Monk using feature probe creation 5000 nt upstream of genes, read count quantification over these probes. For allelic analysis these values are scaled to the number of read per million in their associated comp library prior to log2 transformation and filtering out reads with very small counts. For non-allelic analysis we did similar analysis however, after scaling the comp track, we used a script to subtract input replacing any values less than 0 with a pseudo count of 1e-5 to prevent divide by zero errors. We then derived log2 fold change of treated over control. The software R ggplot2 library is used to create plots. Additionally, stat_compare_means is used to define p-values.
In addition, to obtain allelic skew of treated vs non-treated we used the metric (mus/(cas+mus))14. This has the advantage that it is invariant to experiment size and therefore can be used to understand the relative enrichment of one allele over the other in a non-bias manner. We defined active genes as genes having FPKM > 0.5. In SeqMonk, we created groups that combined, for each replicate of RNA seq, each allelic type (cas and mus). Then, using the RNA-seq pipeline, we quantified raw probe counts within genes (merged transcripts isoforms) for both mus and the newly created groups. Then we obtained mus/(cas+mus) by using relative quantification, dividing each mus by its respective cas and mus group count. We obtained only values associated with chromosome X by creating a positional filter. Using the resulting X chromosome probes, we create an annotation report with exact overlapping genes. Finally, to compare the drug response in a way that is invariant to experiment size, using this report, we compared X1 treated to non-treated and counted genes that had a 10 percent increase in mus/(cas+mus) over non-treated. All performed under mac OS 10.14.6 using a late 2013 Macbook Pro.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) repository with accession number GSE141683

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical analysis was used to determine the sample size. For each experiment, we choose a sample size based on similar publications. The determined sample size was adequate as the differences between experimental groups was significant and reproducible. Examples of publications used as example for sample size determination as follows: ALIS: https://doi.org/10.1177/2472555218796656 EMSA: https://doi.org/10.1016/j.molcel.2014.05.009 Micrographies, RT-qPCR: https://doi.org/10.1016/j.cell.2010.09.049 Micrographies, RIP, RT-qPCR: https://doi.org/10.1016/j.cell.2013.05.028 Sequencing: https://doi.org/10.1101/gr.133751.111 SAXS:

Data exclusions	No data were excluded from the analysis.
Replication	Independent in vitro experiments (ALIS, EMSAs, SAXS) were performed using different batches/aliquots of purified proteins, RNAs or molecules for each replicate whenever possible. Each replicate (mixing of components and chromatography or gel separation) was performed at different times . Independent studies that required cell culture such as micrographies (RNA-FISH, DNA-FISH, immunostaining), and RT-qPCR, RIP and ChIP experiments were performed starting from different passages each time. Each biological replicate was performed at different days. All attempts at replication were successful.
Randomization	Randomization is not applicable to studies, as they compare a treatment with a drug vs. a control in cell lines.
Blinding	Experiments comparing multiple molecules were performed blindly. Experiments were cells were treated with one drug vs. its vehicle were not performed blindly.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Immunofluorescence: H3K27me3 antibodies were from Active Motif (#39155, 1:500, ON), H2AK119ub from Cell Signaling (#8240T, rabbit, 1:500), Ezh2 from Beckton Dickinson (#612666, mouse, 1:500), and Ring1B from Cell Signaling (#5694S, rabbit, 1:500). Anti-rabbit-Alexa fluor 555 and anti-mouse-Alexa fluor 488 antibodies were from Thermo Fisher (1:500, 2 h). RNA Immunoprecipitation: Nuclear lysate was incubated with 5 ug of antibodies or IgG immobilized on Dynabeads Protein A (Thermo Fisher), overnight at 4°C. Antibodies used rabbit anti Suz12 (#39057, Active Motif, CA), mouse anti Ezh2 clone AC22 (#39875, Active Motif), rabbit anti Ciz1 (#PA5-27625, Thermo Fisher), rabbit anti SPEN (#ab72266, Abcam), rabbit anti RBM15 (#ab70549, Abcam), normal rabbit IgG (#2729, Cell Signaling, MA), normal mouse IgG (#12-371, MilliporeSigma). Western blotting: anti H3K27me3 (Active Motif, #39535, 1:5,000) or anti histone H3 antibody (Abcam, #ab1791, 1:5,000). ChIP-seq: 5 ug of Rabbit anti Histone H3K27me3 antibody (#39155, Active Motif), rabbit anti Suz12 (#39057, Active Motif, CA) or normal rabbit IgG
Validation	Antibodies were used according to manufacturer recommendations and the references provided by them: - H3K27me3, Active Motif #39155 (ChIP-seq, IF) https://www.activemotif.com/catalog/details/39155 - H3K27me3, Active Motif #39535 (WB) https://www.activemotif.com/catalog/details/39535.html - H2AK119ub, Cell Signaling #8240T (IF) https://www.cellsignal.com/products/primary-antibodies/ubiquityl-histone-h2a-lys119-d27c4-xp-rabbit-mab/8240?site-search-type=Products&N=4294956287&Ntt=8240t&fromPage=plp&_requestid=371647 - Ezh2, Beckton Dickinson #612666 (IF) https://www.bdbiosciences.com/en-us/products/reagents/microscopy-imaging-reagents/immunofluorescence-reagents/purified-mouse-anti-ezh2.612666 - Ring1B, Cell Signaling #5694S (IF) https://www.cellsignal.com/products/primary-antibodies/ring1b-d22f2-xp-rabbit-mab/5694 - Suz12, Active Motif #39057 (ChIP-seq) https://www.activemotif.com/catalog/details/39357 (RNA-IP) https://doi.org/10.1371/journal.ppat.1003366 - Ezh2 clone AC22, Active Motif #39875 (RIP) https://doi.org/10.1016/j.celrep.2019.08.028 - Ciz1, Thermo Fisher #PA5-27625 (IP) https://www.thermofisher.com/antibody/product/CIZ1-Antibody-Polyclonal/PA5-27625 - SPEN, Abcam #ab72266 (IP) https://www.abcam.com/SPEN-antibody-ab72266.html - RBM15, Abcam #ab70549 (IP) https://www.abcam.com/rbm15ott-antibody-ab70549.html - Histone H3 antibody, Abcam #ab1791 (WB) https://www.abcam.com/histone-h3-antibody-nuclear-marker-and-chip-grade-ab1791.html

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	All cell lines used in this study, except Sf9 cells, were generated by the Lee Lab. The Sf9 insect cells were kindly donated by Dr. Robert Kingston (MGH, Boston)
Authentication	Xist (X+P cells) or Xist lacking the Repeat A domain (X-A cells) were generated as described in Jeon, 2011. Female mouse embryonic stem cells TsixTST/+ (named TST cells) were generated as described in Ogawa, 2008. Female embryonic TST cells where RepA was deleted using CRISPR-Cas9 technology (TST, X-A cells) were generated as described in Sunwoo, 2017. All references are included in the manuscript. Our Sf9 cells were cultured as recommended by ATCC with modifications introduced by Dr. Robert Kington's laboratory (MGH, Boston) (https://doi.org/10.1016/S1097-2765(00)80315-9).
Mycoplasma contamination	Cell lines are free of mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141683
Files in database submission	Raw files: ChiP d5 Control-1 Input : RAb3-1_TGACCA_L004_R1_001.fastq.gz, RAb3-1_TGACCA_L004_R2_001.fastq.gzz ChiP d5 Control-1 H3K27me3: RAb3-2_GCCAAT_L004_R1_001.fastq.gz, RAb3-2_GCCAAT_L004_R2_001.fastq.gz ChiP d5 Control-2 Input : C_I_CKDL190140968-1a-16_H55VGBBXX_L8_1.fq.gz, C_I_CKDL190140968-1a-16_H55VGBBXX_L8_2.fq.gz ChiP d5 Control-2 Suz12: C_S_CKDL190140968-1a-18_H55VGBBXX_L8_1.fq.gz, C_S_CKDL190140968-1a-18_H55VGBBXX_L8_2.fq.gz ChiP d5 Treatment-1 Input: RAb3-3_CAGATC_L004_R1_001.fastq.gz, RAb3-3_CAGATC_L004_R2_001.fastq.gz ChiP d5 Treatment-1 H3K27me3: RAb3-4_CTTGTA_L004_R1_001.fastq.gz, RAb3-4_CTTGTA_L004_R2_001.fastq.gz ChiP d5 Treatment-2 Input: X1_I_CKDL190140968-1a-20_H55VGBBXX_L8_1.fq.gz, X1_I_CKDL190140968-1a-20_H55VGBBXX_L8_2.fq.gz ChiP d5 Treatment-2 Suz12: X1_S_CKDL190140968-1a-23_H55VGBBXX_L8_1.fq.gz, X1_S_CKDL190140968-1a-23_H55VGBBXX_L8_2.fq.gz Bigwig and Bed files: ChIP H3K27me3 over input: GSE141683_RAb3-2.div.RAb3-1.cas.wig.bw ChIP H3K27me3 over input: GSE141683_RAb3-2.div.RAb3-1.comp.wig.bw ChIP H3K27me3 over input: GSE141683_RAb3-2.div.RAb3-1.mus.wig.bw ChIP H3K27me3 over input - treatment: GSE141683_RAb3-4.div.RAb3-3.cas.wig.bw ChIP H3K27me3 over input - treatment: GSE141683_RAb3-4.div.RAb3-3.comp.wig.bw ChIP H3K27me3 over input - treatment: GSE141683_RAb3-4.div.RAb3-3.mus.wig.bw GSE141683_RAb4.AnalyzeRepeats.fpkm_gt_1.genes_pm3000.chr13.comp.bed.gz GSE141683_RAb4.AnalyzeRepeats.fpkm_gt_1.genes_pm3000.chrX.comp.bed.gz
Genome browser session (e.g. UCSC)	No longer applicable

Methodology

Replicates	For each Control and Treatment conditions, 2 ChIPseq were performed. Each ChIP-seq contained an input and one antibody. Antibodies used were anti Suz12 (part of the PRC2 complex that generated H3K27me3), and anti H3K27me3. H3K27me3 and Suz12 agree.
Sequencing depth	All sequencing was performed aiming for 60 million reads, paired end, with average library sizes of 300-350 pb confirmed by Bioanalyzer. Number of reads uniquely aligned: ChiP d5 Control-1 Input : 37234932 ChiP d5 Control-1 H3K27me3: 61732763 ChiP d5 Control-2 Input : 41226162 ChiP d5 Control-2 Suz12: 38600154

ChiP d5 Treatment-1 Input: 21506283
ChiP d5 Treatment-1 H3K27me3: 62456446
ChiP d5 Treatment-2 Input: 33301306
ChiP d5 Treatment-2 Suz12: 42945400

Antibodies

Rabbit anti Histone H3K27me3 antibody (#39155, Active Motif), rabbit anti Suz12 (#39057, Active Motif, CA)

Peak calling parameters

Peak calling was not required for the analyses performed in this study.

Data quality

No data quality on peak calling applies to this study. Peak calling was not required for the analyses performed in this study.

Software

Adapters were removed from day 5 TST RNAseq library with software trim_galor/cutadapt (versions 0.4.3/1.7.1) with default parameters to remove adapters. Adapters were not removed from other libraries. Mapping performed using software Novalign (version 3.09) or STAR (version STAR_2.5.3a) against the mouse C57BL/6J (mm9) as well as variants M. musculus (mus) and M. castaneus (cas) 11, considering directional, first strand, for RNA-seq and unstranded reads for ChIP-seq. Salient parameters for Novalign are -i300 100 -t180 -F ILM1.8 -rRandom -h180 180 -v180 and STAR are --outSAMprimaryFlag AllBestScore --outMultimapperOrder Random --outSAMtype BAM Unsorted --outFilterMultimapNmax 100 --outFilterMismatchNmax 6 --outSAMattributes All –sjdbFileChrStartEnd. Bigwig files were generated 11, and read in the Integrative Genomics Viewer.
DEseq2 (version 1.24.0 or greater) is used to quantitate RNA seq counts over genes using software Seq Monk (version 1.44.0 or greater). We limited our analyses to highly expressed genes (RPKM > 1 in the DMSO control) and for cumulative distribution we set a liberal DEseq2 p-values ($p<1$) to see the general trend. Plots were attained converting counts to log2 fold change in awk and then importing to R where ggplot2 (version 3.2.0 or greater), stat_ecdf() and Wilcox.test were used to generate statistics. For scatterplots comparing CHiP to RNA seq results, comp tracks were sampled to the smallest library, then MultiTesting and IndependentFiltering DESeq2 filtering is performed reporting significance below 0.05 after Benjamimi and Hochberg correction with the application of independent intensity filtering, reporting shrank log2 values.
Metagene analysis of d5 tst cells of H3K27me3 marks is performed using the software ceas (version 1.0.2 under python 2.7) on FPM scaled coverage (wiggle format) generated over active genes plus or minus 3000 nt. Active genes are used to avoid conflating repressed genes on Xa with those on Xi. Active genes are defined as having FPKM greater than 1 in d5 tst cells strand specific RNA sequencing experiments. The software package homer (version 4.10) is used to determine FPKM. Tags directories are created from uniquely aligned reads to mm9 using makeTagDirectory tags with parameters -flip -sspe -removeSpikes 10000 8 -tbp 1 -genome mm9 -checkGC. Note effective removal of PCRd dups using tbp. The function AnalyzeRepeats is used to generate raw counts over genes. FPKM values are derived scaling to gene length per 1000 and per million fragments derived from uniquely aligned paired ends without PCR dups removed.
Additional CHiP analysis for comparing to RNAseq, is performed in Seq Monk using feature probe creation 5000 nt upstream of genes, read count quantification over these probes. For allelic analysis these values are scaled to the number of read per million in their associated comp library prior to log2 transformation and filtering out reads with very small counts. For non-allelic analysis we did similar analysis however, after scaling the comp track, we used a script to subtract input replacing any values less than 0 with a psudo count of 1e-5 to prevent divide by zero errors. We then derived log2 fold change of treated over control. The software R ggplot2 library is used to create plots. Additionally, stat_compare_means is used to define p-values. All performed under mac OS 10.14.6 using a late 2013 Macbook Pro.