Article

A protein assembly mediates *Xist* localization and gene silencing

Amy Pandya-Jones[1], Yolanda Markaki[1,2], Jacques Serizay[1,3,11], Tsotne Chitiashvili[1,2], Walter R. Mancia Leon[1,12], Andrey Damianov[4], Constantinos Chronis[1,13], Bernadett Papp[1,14], Chun-Kan Chen[5,15], Robin McKee[1], Xiao-Jun Wang[4], Anthony Chau[1], Shan Sabri[1], Heinrich Leonhardt[2], Sika Zheng[4,16], Mitchell Guttman[5], Douglas. L. Black[4,6,7,8,9 ✉] & Kathrin Plath[1,6,7,8,9,10 ✉]

Nuclear compartments have diverse roles in regulating gene expression, yet the molecular forces and components that drive compartment formation remain largely unclear[1]. The long non-coding RNA *Xist* establishes an intra-chromosomal compartment by localizing at a high concentration in a territory spatially close to its transcription locus[2] and binding diverse proteins[3–5] to achieve X-chromosome inactivation (XCI)[6,7]. The XCI process therefore serves as a paradigm for understanding how RNA-mediated recruitment of various proteins induces a functional compartment. The properties of the inactive X (Xi)-compartment are known to change over time, because after initial *Xist* spreading and transcriptional shutoff a state is reached in which gene silencing remains stable even if *Xist* is turned off[8]. Here we show that the *Xist* RNA-binding proteins PTBP1[9], MATR3[10], TDP-43[11] and CELF1[12] assemble on the multivalent E-repeat element of *Xist*[7] and, via self-aggregation and heterotypic protein–protein interactions, form a condensate[1] in the Xi. This condensate is required for gene silencing and for the anchoring of *Xist* to the Xi territory, and can be sustained in the absence of *Xist*. Notably, these E-repeat-binding proteins become essential coincident with transition to the *Xist*-independent XCI phase[8], indicating that the condensate seeded by the E-repeat underlies the developmental switch from *Xist*-dependence to *Xist*-independence. Taken together, our data show that *Xist* forms the Xi compartment by seeding a heteromeric condensate that consists of ubiquitous RNA-binding proteins, revealing an unanticipated mechanism for heritable gene silencing.

Although many *Xist*-interacting proteins have a defined function during the initiation of XCI[3,5,6,13,14], the induction of X-linked gene silencing is largely unaffected when the *Xist*-interacting RNA-binding proteins (RBPs) PTBP1, MATR3, TDP-43 or CELF1 are depleted (Extended Data Fig. 1a–c), raising the question of what role(s) these proteins have in XCI (Supplementary Note 1). Notably, in addition to their known functions in RNA processing[9–12], these RBPs can form higher-order assemblies—particularly when concentrated by RNAs containing multivalent protein-binding sites[15–17]. Because *Xist* contains several highly repetitive sequences[7], we explored interactions between *Xist*, PTBP1, MATR3, CELF1 and TDP-43 that might create a higher-order assembly within the Xi and thereby contribute to the formation of the Xi compartment.

We first examined whether the depletion of PTBP1, MATR3, CELF1 or TDP-43 affects *Xist* localization. Short interfering RNA (siRNA)-mediated knockdown of each factor during XCI initiation in female differentiating embryonic stem cells (ES cells) revealed considerable nuclear dispersal of *Xist* and defects in the *Xist*-dependent accumulation of the Xi mark H3K27me3 (histone 3 lysine 27 trimethylation)[18,19], with only small changes in *Xist* transcript or splicing levels (Extended Data Figs. 1d, e, 2a–f). PTBP1 knockdown in ES cells expressing *Xist* from an inducible cDNA transgene lacking introns resulted in similar dispersal of the *Xist* signal in RNA fluorescence in situ hybridization (FISH) experiments (Extended Data Fig. 2g). These findings demonstrate that these four RBPs mediate *Xist* localization on the forming Xi, independently of their RNA-processing activities.

To determine where on *Xist* these factors bind, we performed crosslinking immunoprecipitation followed by sequencing (CLIP–seq) analysis during XCI initiation. We identified a marked accumulation of

[1]Department of Biological Chemistry at the David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [2]Department of Biology and Center for Integrated Protein Science, LMU Munich, Munich, Germany. [3]École Normale Supérieure de Cachan, Université Paris-Saclay, Saclay, France. [4]Department of Microbiology, Immunology and Molecular Genetics, University of California Los Angeles, Los Angeles, CA, USA. [5]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [6]Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA, USA. [7]Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles, CA, USA. [8]Brain Research Institute, University of California Los Angeles, Los Angeles, CA, USA. [9]Graduate Program in the Biosciences, University of California Los Angeles, Los Angeles, CA, USA. [10]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, CA, USA. . [11]Present address: The Gurdon Institute and Department of Genetics, University of Cambridge, Cambridge, UK. [12]Present address: Department of Neurological Surgery, University of California San Francisco, San Francisco, CA, USA. [13]Present address: Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL, USA. [14]Present address: Department of Oral Biology, College of Dentistry, University of Florida, Gainesville, FL, USA. [15]Present address: Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA. [16]Present address: Division of Biomedical Science, University of California Riverside, Riverside, CA, USA. ✉e-mail: dougb@microbio.ucla.edu; kplath@mednet.ucla.edu
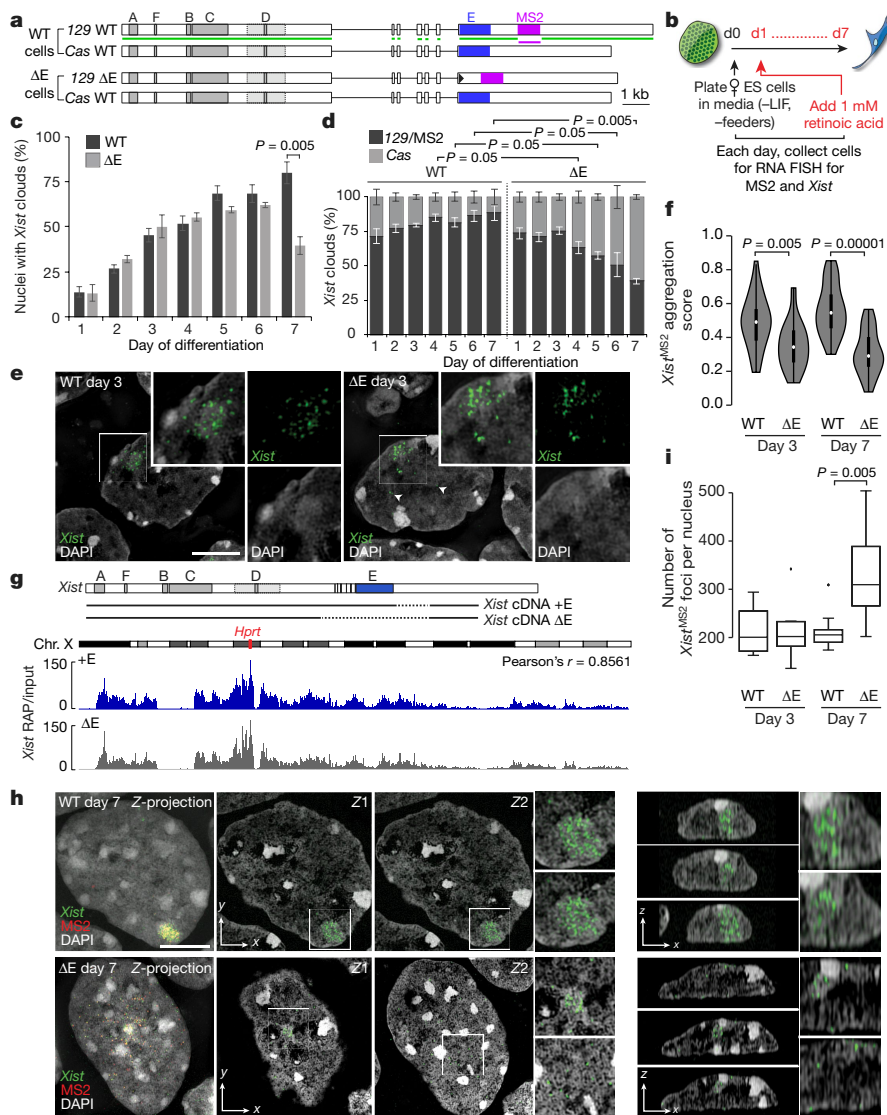
Fig. 1 | The E-repeat mediates *Xist* sequestration and controls the number of *Xist* foci. a, *Xist* alleles in female wild-type and ΔE ES cells. Green (*Xist*) and magenta (MS2) lines indicate FISH probes. WT, wild type. **b**, Schematic of the experimental procedure. **c**, The percentage of nuclei with an *Xist* cloud (*n* = 100) at the indicated days of wild-type and ΔE ES cell differentiation. Data are mean ± s.e.m across 3 replicates; two-tailed Student's *t*-test. **d**, The allelic origin of *Xist* clouds (*n* = 50). Data are mean ± s.e.m. across 3 replicates; two-tailed Student's *t*-test. **e**, Sections from 3D-SIM images showing *Xist* RNA FISH signals from the *Xist*^MS2(129) and *Xist*^ΔE,MS2(129) alleles at differentiation day 3 in wild-type and ΔE cells, respectively. Arrowheads indicate *Xist*^ΔE,MS2 foci located away from the *Xist* cloud. The insets show the enlargement of the marked regions. Right, the same images as in the inset but with the DAPI and *Xist* signals separated. Scale bar, 5 μm. **f**, Violin plots showing aggregation scores of *Xist*^MS2 clouds (*n* = 30) from one replicate in **d**. Violin plots depict median (white) and interquartile range (black), trimmed (grey) to represent data minimum and maximum values. Two-sample Kolmogorov–Smirnov test.

g, Top, tet-inducible *Xist* cDNA transgenes inserted into the *Hprt* locus in male ES cells used in RAP-seq experiments. Dashed lines indicate deleted regions. Bottom: RAP–seq profile of *Xist* containing the E-repeat (+E) and *Xist*^ΔE across the X chromosome after 6 h of doxycycline treatment in ES cells. Data are from one experiment. **h**, Left, 3D-SIM Z-projection of co-localizing *Xist* and MS2 RNA FISH signals from wild-type or ΔE cells at differentiation day 7, merged with DAPI. Scale bar, 5 μm. The next two panels show *Xist* and DAPI signals from two different *Z*-planes, and the smaller panels to the right show enlargements of the *Xist* signal from each *Z*-plane. Right, *Y*-plane sections through cells shown in the left panels, highlighting *Xist* localization relative to the nuclear lamina. Enlargements of the areas containing *Xist* are shown on the right. **i**, Box plot of the distribution of the number of *Xist* RNA foci from the wild-type *Xist* allele or the *Xist*^ΔE,MS2(129) allele (*n* = 10). Horizonal lines denote the median, whiskers indicate 1.5 times the interquartile range, dots represent outliers. Two-sample Kolmogorov–Smirnov test.

PTBP1, MATR3 and CELF1 reads over the E-repeat of *Xist*, which comprises more than fifty elements that are rich in C, U and G nucleotides, and that are predicted to serve as PTBP1-, MATR3- and CELF1-binding sites[20–22] (Extended Data Fig. 3a–c). We confirmed homomeric binding of recombinant (r)PTBP1 to the E-repeat RNA using an electrophoretic mobility shift assay (Extended Data Fig. 3d). Chromatin immunoprecipitation followed by sequencing (ChIP–seq) analysis of PTBP1 revealed a peak that is primarily located over the genomic E-repeat

region and appears upon induction of *Xist* expression (Extended Data Fig. 3a), indicating that PTBP1 engages *Xist* co-transcriptionally. The *Xist* CLIP–seq profiles of PTBP1 and PTBP2 (the neural homologue of PTBP1) in differentiated cells were markedly similar to that of PTBP1 during XCI initiation, and TDP-43 in the embryonic mouse brain displayed strongest binding at the 3′ end of the E-repeat, where multiple (GU)$_n$ tracts presumably serve as binding motifs[11] (Extended Data Fig. 3a, c). Together, these data show that the E-repeat serves as a multivalent
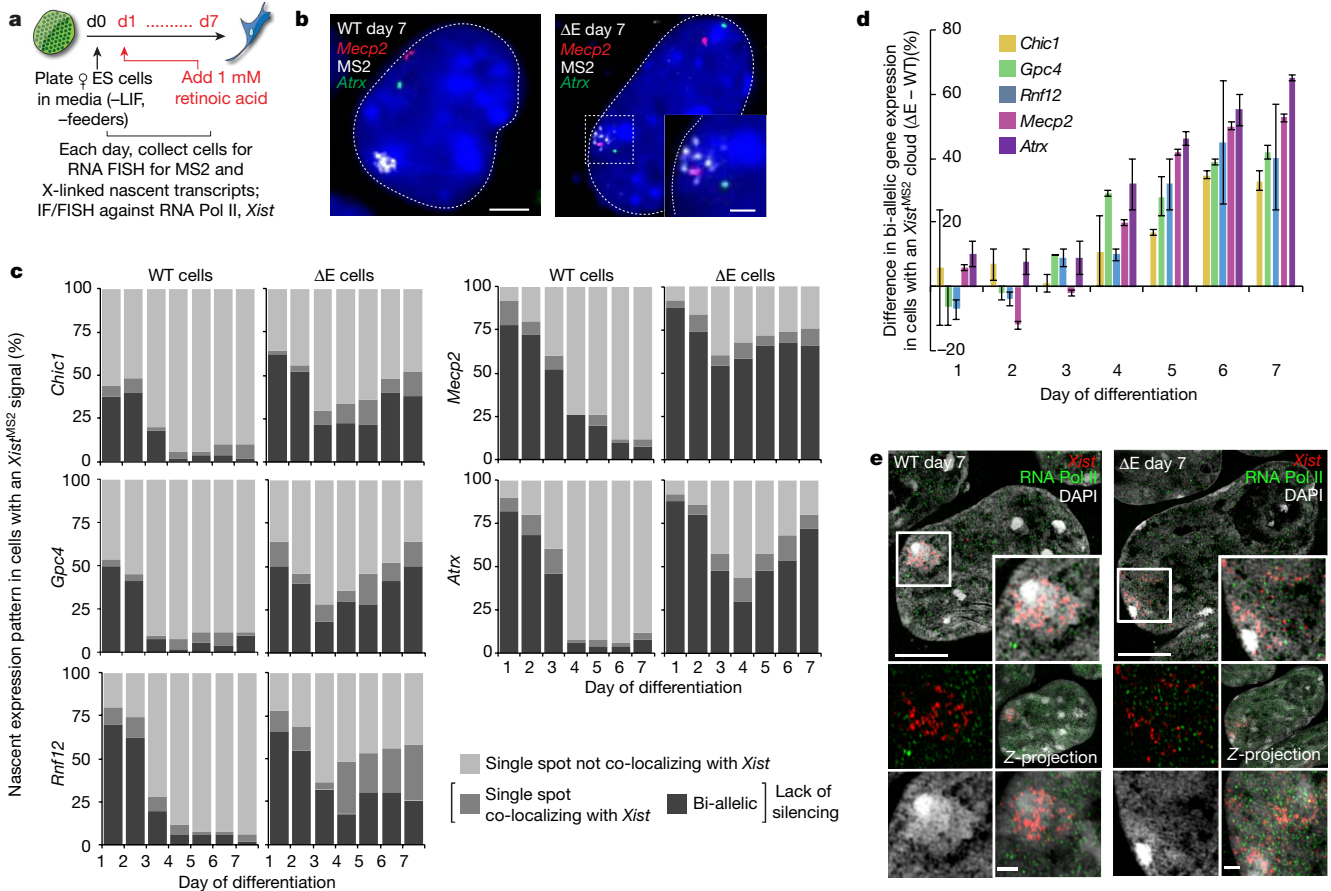
Fig. 2 | The E-repeat establishes heritable gene silencing. a, Schematic of the experimental procedure. **b**, Epifluorescence images showing the predominant nascent expression pattern of X-linked genes *Mecp2* (red) and *Atrx* (green) in wild-type (mono-allelic expression from the active X chromosome, Xa) and ΔE cells (bi-allelic expression) with a *Xist*-MS2 signal (white) at differentiation day 7. Inset, an enlargement of the boxed region, highlighting the fainter, dispersed *Xist*^ΔE^-MS2 signal. Scale bars, 5 μm (main), 1 μm (inset) **c**, Quantification of nascent expression patterns of the indicated X-linked genes in wild-type and ΔE cells displaying an *Xist*^MS2^-coated X chromosome (*n* = 50), across 7 days of differentiation. Results were replicated three times. **d**, The mean percentage difference in bi-allelic nascent gene expression (lack of silencing) between ΔE cells and wild-type cells with an *Xist*^MS2^ cloud, across 7 days of differentiation. Data are mean ± s.e.m. **e**, 3D-SIM sections through wild-type and ΔE cells expressing *Xist*^MS2^ at differentiation day 7, stained for RNA Pol II (green) and DAPI (grey) and probed for *Xist* (red). The inset shows a magnification of the indicated region. The four small images below each large image are as follows, clockwise from top left: same as inset but without DAPI; z-projection of the whole nucleus; z-projection of the inset; same as inset showing only DAPI. Scale bars, 5 μm (main); 1 μm (insets).

binding platform for PTBP1, MATR3, CELF1 and TDP-43; that binding of TDP-43 and PTBP1 to the E-repeat persists after XCI initiation is complete; and that members of the same protein family can replace PTBP1 on *Xist*.

Next, we asked whether recruitment of PTBP1, MATR3, CELF1 or TDP-43 by *Xist* could be detected by microscopy within the Xi during XCI initiation and after transition to the *Xist*-independent phase of XCI after day 3 of differentiation[8]. We observed an accumulation of CELF1 on the Xi, which increases in intensity from day 3 to day 7 of female ES cell differentiation, and noted a mesh-like pattern of PTBP1 localization within the Xi territory of some cells at day 7 of differentiation (Extended Data Fig. 4a–e). Although MATR3, TDP-43 and PTBP1 did not enrich in the Xi in most cells, they were not depleted (Extended Data Fig. 4e–h, Supplementary Note 2). We therefore conclude that PTBP1, MATR3 and TDP-43 are present—and CELF1 gradually concentrates—within the Xi-territory; findings that are consistent with the time-dependent formation of a spatially concentrated protein assembly.

If PTBP1, MATR3, CELF1 and TDP-43 control the accumulation of *Xist* on the Xi, loss of the E-repeat should disrupt XCI by reducing *Xist* enrichment within the X-chromosome territory. In support of this hypothesis, it has been shown that *Xist* exon 7—which contains the E-repeat—is required for persistent localization of *Xist* on the Xi in differentiating

ES cells[23]. We tested this possibility by deleting the E-repeat within *Xist* on the 129 allele—which also contains 11 copies of an MS2-RNA tag within *Xist*—in a polymorphic 129 × Cas female mouse ES cell line, yielding the *Xist*^ΔE,MS2(129)/WT(Cas)^ genotype (ΔE ES cells) (Fig. 1a, Extended Data Fig. 5). RNA FISH experiments revealed that the number of cells containing an *Xist*-coated X chromosome increased gradually until differentiation day 4 in both wild-type and ΔE cells (Fig. 1b, c). The proportion of ΔE cells with an *Xist* enrichment then declined compared to wild-type cells, with a significant reduction of approximately 50% by day 7 (Fig. 1c). This reduction was specific to the *Xist*^ΔE,MS2(129)^ allele (hereafter denoted *Xist*^ΔE^)—as revealed by RNA FISH experiments against the MS2 tag (Fig. 1d)—and there was no significant difference in the abundance or the half-life of the *Xist*^MS2^ allele in ΔE cells compared to the wild type (Extended Data Fig. 6a, b). RNA FISH experiments against an intronic *Xist* sequence labelled the nascent transcription site and not the *Xist* cloud (Extended Data Fig. 6c), indicating that the *Xist*^MS2^ RNA that coats the Xi in wild-type and ΔE cells is processed. The loss of the *Xist* accumulation in ΔE cells over the X territory is therefore not a consequence of decreased *Xist* abundance, splicing defects or reduced RNA stability.

A closer inspection of *Xist* localization at differentiation day 3 showed that *Xist*^ΔE^ enriched over the X chromosome, with aggregation
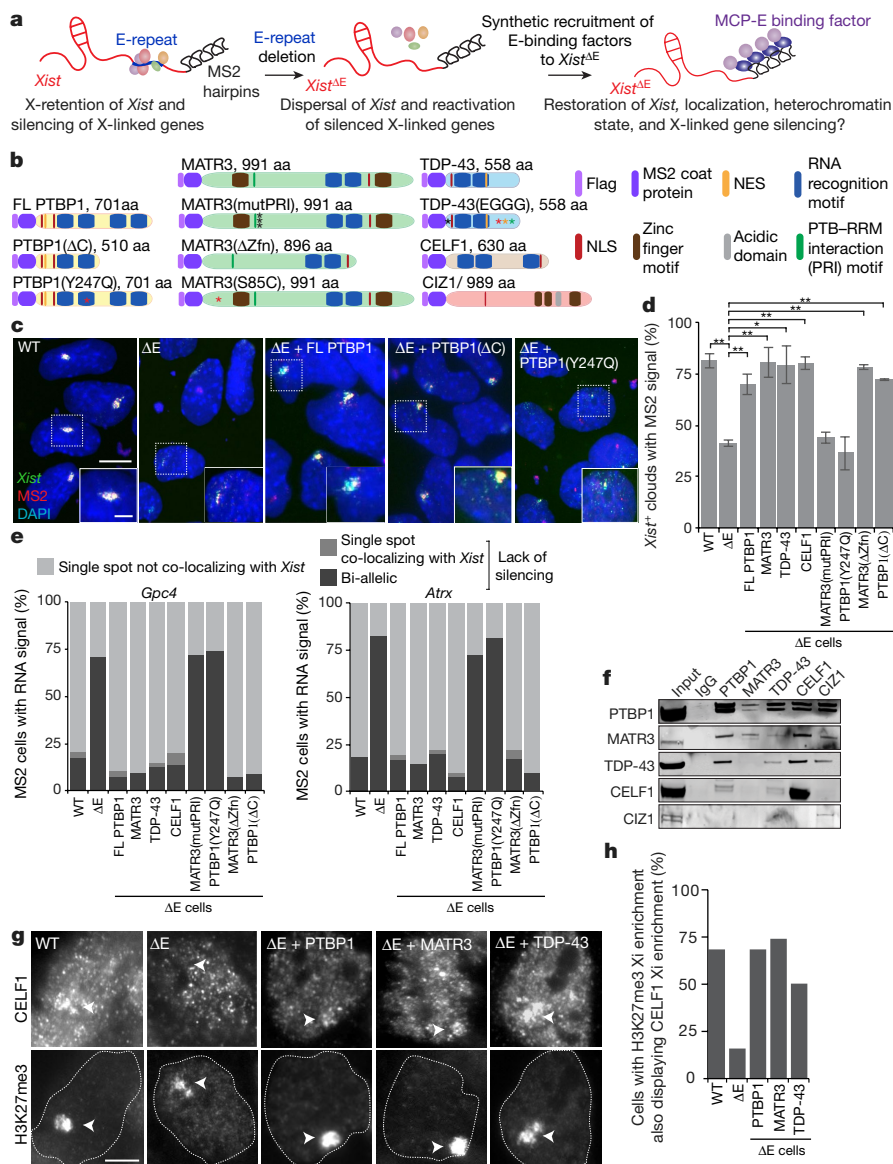
Fig. 3 | PTBP1, MATR3, TDP-43 and CELF1 confer gene silencing and *Xist* sequestration functions on the E-repeat. a, MCP-fusion protein rescue approach for *Xist*^ΔE,MS2. **b**, Illustration of Flag-tagged MCP-fusion proteins and mutants. Point mutations are indicated with asterisks. Length of fusions includes the Flag and MCP sequences. **c**, Representative epifluorescence images from RNA FISH experiments against *Xist*-MS2 in day 7 differentiated wild-type, ΔE, and ΔE lines expressing variants of MCP–PTBP1 fusion proteins. The inset shows the enlargement of the marked region. Scale bars, 10 μm. **d**, Histogram showing the proportion of nuclei with an *Xist* FISH signal (*n* = 80) that also displayed a co-localizing MS2 signal at differentiation day 7 in wild-type, ΔE, or ΔE lines expressing the indicated MCP fusion proteins. Data are mean ± s.e.m from two independent experiments; *P < 0.05, **P < 0.005, two-tailed Student's *t*-test. **e**, Quantification of nascent *Gpc4* or *Atrx* expression

pattern in cells expressing *Xist*-MS2 (*n* = 50) at differentiation day 7. The experiment was repeated twice with similar results. **f**, Immunoprecipitation of PTBP1, MATR3, CELF1, TDP-43 and CIZ1 from ES cell extracts (no RNase) and detection of co-precipitated proteins by immunoblotting, using the same antibodies. Similar results were obtained from three independent trials; for source data see Supplementary Fig. 1. **g**, Representative epifluorescence images showing the CELF1 pattern in wild-type, ΔE or ΔE-rescue cell lines expressing MCP–PTBP1, MCP–MATR3 or MCP–TDP-43 at day 7 of differentiation. Arrowheads indicate the Xi marked by H3K27me3-enrichment. Scale bars, 5 μm. **h**, Histogram showing the percentage of wild-type, ΔE or ΔE-rescue cell lines with H3K27me3 Xi enrichment that also display a co-localizing accumulation of CELF1 at differentiation day 7 (*n* = 50, from one experiment).

measurements (see Methods) revealing only a modest defect in the localization of *Xist*^ΔE compared to wild-type *Xist* (Fig. 1e, f). RNA antisense purification followed by sequencing (RAP–seq)[2] revealed highly correlated patterns of *Xist* association across the X chromosome for wild-type *Xist* and *Xist*^ΔE (Fig. 1g), indicating that the E-repeat is not involved in the initial transfer of *Xist* across the X chromosome. However, at day 7, we found significant dispersal of *Xist*^ΔE within the nucleus compared to the wild type, often localizing to the nuclear lamina (Fig. 1f, h, Extended Data Fig. 6d–j). Super-resolution three-dimensional

structured illumination microscopy (3D-SIM) imaging additionally revealed a significant increase in the number of individual *Xist*^MS2 foci in ΔE cells compared to wild-type cells at differentiation day 7; a difference that was not seen on day 3 (Fig. 1i). As this increase occurred without an increase in the number of *Xist*^ΔE transcripts compared to wild-type *Xist* transcripts (Extended Data Fig. 6a), these data support a model in which the E-repeat is required for the integration of multiple *Xist* transcripts into individual *Xist* foci and for stabilizing these foci within the X-chromosome compartment.
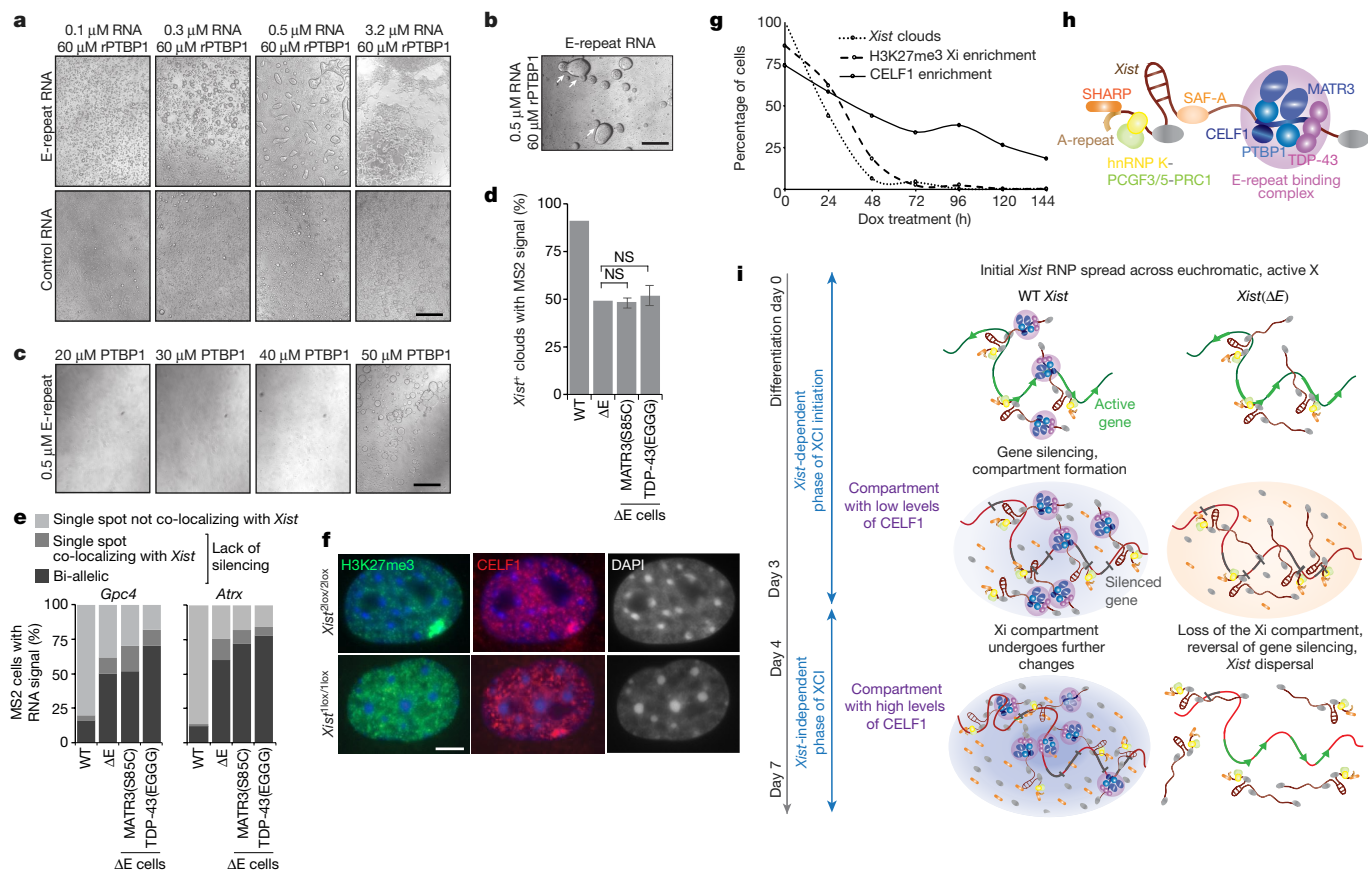
Fig. 4 | Self-association of E-repeat-binding RBPs is critical for formation of the Xi compartment. a, Bright-field images of droplets formed with rPTBP1 and different concentrations of E-repeat RNA or control RNA. Scale bar, 100 μm. **b**, Bright-field image of droplets undergoing fusion (arrows). Scale bar, 50 μm. **c**, Bright-field images of droplets formed with 0.5 μM E-repeat RNA and decreasing amounts of rPTBP1. Scale bar, 100 μm. For **a**–**c**, images were taken after 40 min. **d**, The percentage of nuclei with an Xist FISH signal (n = 100) at differentiation day 7 that also displayed a co-localizing MS2 signal in wild-type, ΔE or ΔE lines expressing the indicated MCP fusion proteins. Data are mean ± s.e.m. from two independent experiments; two-tailed Student's t-test. **e**, Histograms showing nascent Gpc4 or Atrx expression patterns in cells described in **d** expressing Xist-MS2 (n = 50, from one experiment). **f**, Representative images showing H3K27me3, CELF1 and DAPI staining in Xist^2lox/2lox, Rosa26^M2rtTA/tetO-Cre mouse embryonic fibroblasts (MEFs) before (2lox/2lox) or 96 h after (1lox/1lox) the addition of doxycycline to induce Xist excision. Scale bar, 5 μm. **g**, Graph showing the percentage of MEFs described in **f** that that show Xi-enrichment of H3K27me3 or CELF over a 144 h time course of doxycycline (dox) treatment

(n = 50, from one experiment). **h**, Illustration of the E-repeat-bound Xist ribonucleoprotein complex (RNP) with PTBP1, CELF1, MATR3 and TDP-43 binding to the E-repeat and undergoing additional protein–protein interactions. Other Xist-interactors are indicated. **i**, Model of Xi-compartment formation via protein condensation (Supplementary Note 6). Wild-type Xist: upon differentiation, the Xist RNP spreads across the X chromosome and induces the formation of a higher-order assembly by recruiting additional protein molecules into the Xist territory through extensive homo- and heterotypic protein–protein interactions (purple oval). We postulate that the condensate, in addition to E-repeat-interactors, integrates other proteins (grey) including SHARP[3] (orange). The assembly changes over time as indicated by low and high CELF1 levels and the increased purple coloring of the oval. Xist^ΔE: without the E-repeat, Xist localization and X-linked gene silencing initiate normally (middle), potentially through non-E-repeat-dependent protein condensation events (orange oval); however, they cannot be reinforced later, despite the Xist-independence of XCI at this point.

Consistent with defects in Xist localization, we observed lower H3K27me3 enrichment and reduced chromatin compaction over the Xi territory at differentiation day 7 in ΔE cells, despite normal establishment at day 3 (Extended Data Fig. 7a–e). The Xist^ΔE localization phenotype arises as the enrichment of the PRC2 complex on the Xi decreases[18,19] (Extended Data Fig. 7f), suggesting that it is associated with a reorganization of the X-chromosome compartment (Supplementary Note 3). Together, these results reveal a transition in the mechanisms that enrich Xist on the X chromosome during XCI initiation—switching from a largely E-repeat-independent phase to an E-repeat-dependent phase.

Because the control of Xist localization switches upon transition to the Xist-independent-phase of XCI initiation[8], we addressed whether X-linked gene silencing was affected by loss of the E-repeat. We examined nascent transcripts from five X-linked genes that are subject to XCI: Gpc4, Rnf12 (also known as Rlim), Mecp2, Chic1 and Atrx (Fig. 2a, b).

We observed little difference in the extent of gene silencing between wild-type and ΔE cells early during differentiation (Fig. 2c, d, Extended Data Fig. 8). However, at later stages of differentiation (days 4–7), cells expressing Xist^ΔE failed to maintain silencing of these five genes (Fig. 2c, d, Extended Data Fig. 8). Moreover, RNA Pol II—which was excluded from the Xist^ΔE-marked territory during early differentiation—intermingled with the Xist^ΔE foci at later times (Fig. 2e, Extended Data Fig. 7g–i). The E-repeat is therefore essential for sustaining Xist coating, silencing of X-linked genes and exclusion of RNA Pol II beyond the initial wave of transcriptional shutoff. Thus, the Xist-independent state of XCI initiation[8] is not established in the absence of the E-repeat, demonstrating that the E-repeat is involved in generation of the epigenetic memory for gene silencing by Xist.

To evaluate whether there is a causal relationship between the E-repeat-binding RBPs, Xist localization and gene silencing, we synthetically fused PTBP1, MATR3, CELF1 or TDP-43 to the MS2-coat protein

Article

(MCP) in order to recruit these proteins to $Xist^{\Delta E}$ via the 11×MS2-tag (Fig. 3a, Extended Data Fig. 9a–d). Continued expression of the MCP–PTBP1, MCP–MATR3, MCP–TDP-43 or MCP–CELF1 fusion proteins during differentiation in ΔE cells rescued $Xist$ localization, silencing of $Gpc4$ and $Atrx$, and H3K27me3 enrichment on the $Xist^{\Delta E,MS2(129)}$-associated X chromosome at differentiation day 7 (Fig. 3b–e, Extended Data Fig. 9e, f). These data demonstrate that the E-repeat controls $Xist$ localization, gene silencing and heterochromatin formation through interaction with the proteins PTBP1, MATR3, TDP-43 and CELF1.

Next, we addressed whether PTBP1, MATR3, TDP-43 and CELF1 act together to control these processes. Making use of a known direct interaction between PTBP1 and MATR3[20], we found that MCP–MATR3 harbouring a mutant PTBP1–RRM interaction (PRI; where RRM is RNA recognition motif) sequence (denoted MATR3(mutPRI))[20] partially rescued H3K27me3 enrichment, but was unable to rescue the $Xist$ localization and gene silencing defects observed upon loss of the E-repeat (Fig. 3b, d, e, Extended Data Fig. 9c–f). Similar results were observed with the converse mutation in PTBP1, a tyrosine-to-glutamine substitution at position 247 (Y247Q)[20] that prevents the interaction of PTBP1 with MATR3 (Fig. 3b–e, Extended Data Fig. 9c–e). These findings are supported by co-immunoprecipitation experiments demonstrating that PTBP1, MATR3, CELF1 and TDP-43 co-precipitate one another in the presence of RNA, whereas only PTBP1 and MATR3 robustly interact after RNase treatment (Fig. 3f, Extended Data Fig. 9g). These data indicate that a specific direct interaction between PTBP1 and MATR3 is critical for XCI, and consequently show that these proteins act non-redundantly in the XCI process. Furthermore, the finding that CELF1 enriches on the Xi in ΔE rescue cells that express MCP–PTBP1, MCP–MATR3 or MCP–TDP-43 (Fig. 3g, h) indicates that each of these RBPs can initiate the formation of a heteromeric protein assembly within the Xi.

The protein CIZ1 was previously suggested to anchor $Xist$ to chromatin through the E-repeat[24,25]. Although PTBP1 and MATR3 can interact with CIZ1, the expression of MCP–CIZ1 in ΔE cells did not rescue $Xist$ cloud formation or X-linked gene silencing (Fig. 3f, Extended Data Figs. 9g, 10a–e). Moreover, the Xi accumulation of CIZ1 that was observed in wild-type cells was not detected in ΔE cells expressing MCP–PTBP1, MCP–MATR3 or MCP–TDP-43 (Extended Data Fig. 10f, Supplementary Note 4). This indicates that the rescue of $Xist^{\Delta E}$ phenotypes by PTBP1, MATR3, TDP-43 and CELF1 is independent of CIZ1 and that distinct functional complexes assemble on the E-repeat. A bivalent MCP–GFP–MCP fusion protein was also unable to rescue the $Xist$ localization and silencing defects in ΔE cells (Extended Data Fig. 10g–k)—consistent with linkages formed by the four factors not simply tethering $Xist$ transcripts together.

We next aimed to define additional specific activities conferred by the recruited proteins that could facilitate the compartmentalization of $Xist$ and downstream events in XCI. By using a mutated version of MATR3 lacking both zinc finger domains (MATR3(ΔZfn)), we found that rescue of the $Xist^{\Delta E}$ phenotypes by MATR3 is independent of its zinc fingers (Fig. 3b, d, e, Extended Data Fig. 9c–f). We also noted that the expression of PTBP1 lacking RRMs 3 and 4 (denoted PTBP1(ΔC)) in ΔE cells rescued defects resulting from loss of the E-repeat; however, closer inspection of the $Xist^{\Delta E}$ clouds binding MCP–PTBP1(ΔC) revealed dispersed $Xist$ foci within the nucleus (Fig. 3b–e, Extended Data Fig. 9c–e). This finding implicates binding valency as a functional parameter of the PTBP1–$Xist$ assembly.

The formation of an $Xist$ territory containing PTBP1 was of interest given that PTBP1 can undergo liquid–liquid de-mixing in vitro when incubated at high concentrations with a binding RNA[15]. We therefore asked whether rPTBP1 forms liquid droplets upon interaction with the $Xist$ E-repeat. The addition of 3.2 μM E-repeat RNA to 60 μM rPTBP1 produced aggregate-like assemblies[15,16], whereas the addition of lower concentrations of RNA (0.1–0.5 μM) resulted in droplets that resembled phase-separated liquids and could fuse with each other (Fig. 4a, b, Extended Data Fig. 11a, b, Supplementary Note 5). By contrast, smaller droplets were produced by a control RNA (containing five short CU tracts that could bind PTBP1), with no observed aggregation at the highest RNA concentration (Fig. 4a). These findings indicate that the multivalent binding of PTBP1 to the E-repeat strongly promotes its condensation. A lack of droplet formation at near-physiological concentrations of rPTBP1 (20–40 μM) suggested that additional proteins promote the E-repeat-induced condensation of PTBP1 in vivo (Fig. 4c, Extended Data Fig. 11c), which is consistent with the interdependence of the functions of MATR3, PTBP1, CELF1 and TDP-43, as described above. We tested this idea by adding 20 μM rCELF1 to solutions containing 0.5 μM E-repeat RNA and varying concentrations of rPTBP1. Whereas aggregates formed at high rPTBP1 concentrations, lowering the concentration resulted in decreased aggregate sizes until, at 20 μM, they resolved into small spherical structures (Extended Data Fig. 11d, e). These observations are consistent with the formation of a higher-order protein condensate in the Xi that forms via the multivalent binding of several RBPs to the E-repeat, and suggest that the involvement of multiple RBPs lowers the concentration of each factor that is required for condensate induction.

The in vitro data suggested that the self-assembly of E-repeat-interacting proteins is required for their function in XCI. To explore this idea further, we assessed whether the self-assembly of TDP-43 affected XCI. TDP-43 forms higher-order complexes that undergo liquid–liquid phase separation, and this activity is reduced by several mutations: S48E, W334G, W385G and W412G[11]. Unlike wild-type TDP-43, the fusion protein MCP–TDP-43(EGGG)—containing these four mutations—did not rescue phenotypes associated with $Xist^{\Delta E}$ (Figs. 3b, 4d, e, Extended Data Fig. 11f–h); this suggests that self-association of TDP-43 permits the few available TDP-43 sites to support recruitment of multiple TDP-43 monomers. Similar results were obtained using a MATR3(S85C) mutant, which has previously been shown to impair both droplet formation and TDP-43 recruitment in comparison with wild-type MATR3[17] (Fig. 3b, 4d, e, Extended Data Fig. 11f–h). We therefore conclude that, through high-density binding to the E-repeat, $Xist$ concentrates PTBP1, MATR3, TDP-43 and CELF1, which use homo- and heterotypic interactions to establish a physical condensate that compartmentalizes $Xist$ and enforces X-linked gene silencing.

Our results suggested that the condensate established by the E-repeat is crucial for the $Xist$-independent phase of XCI after day 3 of differentiation[8], leading to the hypothesis that this condensate—containing PTBP1, MATR3, TDP-43 and CELF1—can be retained in the Xi in the absence of $Xist$. To test this idea, we confirmed that CELF1 is enriched in the Xi in primary female mouse embryonic fibroblasts carrying a conditional $Xist$ allele (Fig. 4f). Upon $Xist$ deletion, loss of the H3K27me3 Xi-accumulation closely followed the loss of $Xist$ over time (Fig. 4f, g, Extended Data Fig. 11i–k). Notably, CELF1 enrichment remained in 25–40% of cells even after the enrichment of $Xist$ or H3K27me3 in the Xi became undetectable (Fig. 4g, Extended Data Fig. 11i–k). CELF1 enrichment was dependent upon PTBP1, MATR3 and TDP-43, as their depletion in the absence of $Xist$ resulted in fewer cells with CELF1 Xi-accumulation (Extended Data Fig. 11l–q).

We conclude that the protein condensate seeded by the E-repeat is stable without $Xist$ and is critical for the enforcement of silencing during the $Xist$-independent phase of XCI in differentiating ES cells. Our findings uncover a mechanism for the persistence of a functional RNA-seeded nuclear compartment, and reveal an unanticipated mechanism for RBP-mediated gene regulation and epigenetic memory (Fig. 4h, i, Supplementary Note 6).

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/s41586-020-2703-0.

1. Strom, A. R. & Brangwynne, C. P. The liquid nucleome – phase transitions in the nucleus at a glance. *J. Cell Sci*. **132**, jcs235093 (2019).
2. Engreitz, J. M. et al. The *Xist* lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
3. McHugh, C. A. et al. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
4. Minajigi, A. et al. A comprehensive *Xist* interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276 (2015).
5. Chu, C. et al. Systematic discovery of *Xist* RNA binding proteins. *Cell* **161**, 404–416 (2015).
6. Galupa, R. & Heard, E. X-chromosome inactivation: a crossroads between chromosome architecture and gene regulation. *Annu. Rev. Genet*. **52**, 535–566 (2018).
7. Brockdorff, N. Local tandem repeat expansion in *Xist* RNA as a model for the functionalisation of ncRNA. *Noncoding RNA* **4**, 28 (2018).
8. Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* **5**, 695–705 (2000).
9. Keppetipola, N., Sharma, S., Li, Q. & Black, D. L. Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. *Crit. Rev. Biochem. Mol. Biol*. **47**, 360–378 (2012).
10. Coelho, M. B., Attig, J., Ule, J. & Smith, C. W. J. Matrin3: connecting gene expression with the nuclear matrix. *WIREs RNA* **7**, 303–315 (2016).
11. Prasad, A., Bharathi, V., Sivalingam, V., Girdhar, A. & Patel, B. K. Molecular mechanisms of TDP-43 misfolding and pathology in amyotrophic lateral sclerosis. *Front. Mol. Neurosci*. **12**, 25 (2019).
12. Beisang, D., Bohjanen, P. R. & Vlasova-St. Louis, I. A. in *Binding Protein* (ed. Abdelmohsen, K.) Ch. 8 (InTech, 2012).
13. Pintacuda, G. et al. hnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish Polycomb-mediated chromosomal silencing. *Mol. Cell* **68**, 955–969.e10 (2017).
14. Moindrot, B. et al. A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for *Xist* RNA-mediated silencing. *Cell Rep*. **12**, 562–572 (2015).
15. Li, P, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
16. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol*. **18**, 285–298 (2017).
17. Gallego-Iradi, M. C. et al. N-terminal sequences in Matrin 3 mediate phase separation into droplet-like structures that recruit TDP43 variants lacking RNA binding elements. *Lab. Invest*. **99**, 1030–1040 (2019).
18. Plath, K. et al. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135 (2003).
19. Silva, J. et al. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 Polycomb group complexes. *Dev. Cell* **4**, 481–495 (2003).
20. Coelho, M. B. et al. Nuclear matrix protein Matrin3 regulates alternative splicing and forms overlapping regulatory networks with PTB. *EMBO J*. **34**, 653–668 (2015).
21. Han, A. et al. De novo prediction of PTBP1 binding and splicing targets reveals unexpected features of its RNA recognition and function. *PLoS Comput. Biol*. **10**, e1003442 (2014).
22. Marquis, J. et al. CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding. *Biochem. J*. **400**, 291–301 (2006).
23. Yamada, N. et al. *Xist* exon 7 contributes to the stable localization of *Xist* RNA on the inactive X-chromosome. *PLoS Genet*. **11**, e1005430 (2015).
24. Ridings-Figueroa, R. et al. The nuclear matrix protein CIZ1 facilitates localization of *Xist* RNA to the inactive X-chromosome territory. *Genes Dev*. **31**, 876–888 (2017).
25. Sunwoo, H., Colognori, D., Froberg, J. E., Jeon, Y. & Lee, J. T. 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).

Article

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture

All mouse ES cell lines were cultured in knockout DMEM (Life Technologies) supplemented with 15% FBS (Omega), 2 mM L-glutamine (Life Technologies), 1× NEAA (Life Technologies), 0.1 mM β-mercaptoethanol (Sigma), 1× penicillin/streptomycin (Life Technologies), and 1,000 U ml⁻¹ murine LIF (homemade) on 0.3% gelatinized plates (porcine skin gelatin, Sigma) pre-plated with irradiated male DR4 feeders (homemade from day 14.5 embryos, with appropriate protocols in place ensuring the ethical treatment of animals, approved by the UCLA Institutional Animal Care and Use Committee, known as the Chancellor's Animal Research Committee (ARC), 2007-180-41). For 3D-SIM microscopy experiments, ES cells were maintained in 2i culture conditions without feeders, before differentiation[26]. No differences in results upon cell differentiation were observed between the ES cell propagation conditions. ES cells were maintained as small colonies and passaged with trypsin and single-cell dissociation at 80% confluency. Mycoplasma tests (Lonza) are routinely conducted on cells cultured in the laboratory. Additionally, DAPI staining of the cells used in the study did not indicate any mycoplasma contamination.

Female ES cell differentiation

Female wild-type F1 2-1 MS2[129] (and derivatives thereof)[27] were trypsinized to single cells and counted. Cells were seeded in 2 ml of MEF medium (DMEM (Invitrogen) supplemented with 10% FBS (Omega), 2 mM L-glutamine (Life Technologies), 1× NEAA (Life Technologies), 0.1 mM β-mercaptoethanol (Sigma) and 1× penicillin/streptomycin (Life Technologies)) at a density of 20,000–200,000 cells per 4 cm² (depending on the experiment) on tissue culture plates for western blotting or onto 18 mm sterile glass coverslips for immunofluorescence (IF)/FISH experiments, both of which were pre-coated with sterile 0.3% gelatin (porcine skin gelatin, Sigma) or Matrigel (Corning, diluted 1:100). At 24 h post-seeding, the culture medium was changed and supplemented with 1 µM all-*trans* retinoic acid (Sigma), which was changed daily thereafter until the cells were collected for analysis.

Female MEF culture

Female MEFs (*Xist*[2lox/2lox], *Rosa26*[M2rtTA/tetO-Cre-recombinase])[28] were maintained in MEF medium. To delete *Xist*, cells were treated with 2 µg ml⁻¹ doxycycline (Sigma) for up to 144 h to induce expression of Cre-recombinase.

Male ES cell culture

Male ES cells were maintained as described in the section 'Cell culture'. To express *Xist*, ES cells were trypsinized to single cells and counted. Cells were seeded in 2 ml of mouse embryonic cell media at a density of 20,000–200,000 cells per 4cm² (depending on the experiment) on tissue culture plates for western blotting and RNA collection or onto 18 mm sterile glass coverslips for IF/FISH experiments, both of which were pre-coated with sterile 0.3% gelatin (porcine skin gelatin, Sigma) or Matrigel (Corning, diluted 1:100 in cold DMEM media). For knockdown experiments, siRNAs were added upon plating (see section 'siRNA treatments'). For *Xist* expression, doxycycline (Sigma) was added to a final concentration of 2 µg ml⁻¹ for 6–24 h, depending on the experiment.

RNA FISH

FISH against *Xist* RNA was performed using both RNA and DNA probes. FISH against the MS2-insert, *Atrx*, *Gpc4*, *Mecp2*, *Rnf12* and *Chic1* was performed using DNA probes. In undifferentiated ES cells, the DNA probe against *Xist* additionally detects *Tsix*.

RNA probe preparation.

Strand-specific RNA probes were generated using a T3 in vitro transcription kit (Promega) in the presence of Chromatide AlexaFluor-UTP (ThermoFisher). Six transcription templates (about ~700 nt) were generated from *Xist* exon 1 (Primers UCLA 1416–1429, Supplementary Table 1), and used in transcription reactions containing 0.5 mM ATP, CTP, GTP, 0.1 mM UTP and 0.05 mM Chromatide Alexa Fluor 488-UTP (Life Technologies) along with 1× T3 transcription buffer supplemented with 10 mM DTT, 500U RNase inhibitor, 170U T3 RNA polymerase and 5 µg of pooled template DNA in a final volume of 500 µl at 37 °C overnight in the dark. The transcription reaction was treated with 15U RNase-Free DNase for 15 min at 37 °C before probe purification. To purify the probes, 1/3 of the transcription reaction was loaded on a pre-spun (700*g*, 5 min) Chromaspin-100 column (Clontech) and centrifuged (700*g*, 5 min). The eluates were combined and precipitated with 100% EtOH in the presence of 100 mg tRNA and 1/10 volume of sodium acetate (Sigma). We sometimes also purified RNA probes using a 2.5× volume of AMPure beads (Thermo Fisher 09-981-123, reconstituted according to ref. [29]), which were washed twice on a magnet with 80% ethanol before elution of the probes from the beads with 50 µl water, followed by ethanol precipitation. The RNA pellet was washed twice in 70% ethanol, resuspended in 400 µl of RNase-free water, to which 1 ml EtOH was added for storage at −20 °C. To make the final probe mix, 1/7 of the Probe/EtOH solution was added to 90 µl salmon sperm DNA (Sigma), 90 µl mouse Cot1 DNA (Life Technologies), 40 µl 3 M RNase-free sodium acetate (Sigma), 40 µl 10 mg ml⁻¹ tRNA (Life Technologies) and 1 ml EtOH. After vigorous shaking, the solution was centrifuged at maximum speed for 10 min. The pellet was washed once with 70% EtOH and then once with 100% EtOH, allowed to dry completely, and then resuspended in 200 µl deionized formamide (VWR) and 200 µl 2× hybridization buffer (20% dextran sulfate (Sigma), 4× SSC (Ambion), 0.1 M NaH₂PO₄). Probes were stored at −80 °C and denatured at 95 °C for 5 min before use.

DNA probe preparation.

For 3D-SIM and Airyscan experiments, FISH probes were labelled by nick translation as described previously[30] using p15 cDNA plasmid as template and home-labelled Atto488-, Cy3- or Texas Red-conjugated dUTPs[31]. For all other experiments, DNA probes were synthesized using the CGH Bioprime Array Kit (Thermo Fisher) according to the manufacturer's instructions. In brief, a 40 µl solution containing 100 ng of template DNA was denatured in the presence of 1× random primers at 95 °C for 5 min and snap-cooled on ice. Five microlitres of nucleotide mix, 5 µl of 488-, 555-, or 594- dUTP or dCTP chromatide fluorophore (Life Technologies) and 5U Klenow exo-enzyme were then added and incubated in the dark at 37 °C for 6 h, after which an additional 5U of Klenow exo-enzyme was added. The reaction was incubated at 37 °C overnight, quenched with 10 µl stop solution, and then purified over a Chromatide-100 column or AMPure beads as described in the section 'RNA probe preparation'. The eluate was precipitated in the presence of 100 mg yeast tRNA (Life Technologies) and sodium acetate (Sigma). The final DNA probe mix was then prepared as in the section 'RNA probe preparation' to yield 400 µl of probe solution in formamide/hybridization buffer.

The MS2 DNA template for DNA probe preparation was PCR-amplified from genomic DNA purified from wild-type F1 2-1 MS2[129] female ES cells (see Supplementary Table 1 for primers). For *Xist*, the DNA probe was synthesized using a full-length mouse *Xist* cDNA plasmid (p15A-31-17.9kb *Xist*, unpublished). The intron probe in Extended Data Fig. 6c was against intron 1, which is the longest intron within the gene. We were unable to get probes against other introns to work well in this assay, presumably owing to their short length and labile nature. Probes against X-linked genes were synthesized using BACs RP23-467J21 (*Gpc4*), RP23-265D6 (*Atrx*), WIBR1-2150D22 (*Chic1*), WIBR1-2704K12 (*Rnf12*) and W11-894A5 and W11-1189K18 (*Mecp2*) (all obtained from CHORI-BACPAC). Note that the use of two BACs for the *Mecp2* probe

sometimes resulted in nascent FISH signals from one X chromosome that appeared as doublets (see Fig. 2b, wild-type panel).

RNA FISH procedure for epifluorescence microscopy. Culture medium was changed 10 min before cell collection to remove dead cells and stimulate transcription. Upon collection, culture medium was aspirated, and coverslips were gently rinsed twice with cold 1× PBS. Coverslips were then transferred to a new culture dish containing 1× PBS, which was then aspirated, and the cells were fixed in 4% paraformaldehyde (PFA) (Electron Microscopy Sciences) in 1× PBS for 10 min at room temperature (RT) under standard laboratory safety practices. After fixation, the cells were permeabilized in 0.5% Triton X-100 (Acros) in 1× PBS with 2 mM vanadyl ribonucleoside complex (NEB) for 10–20 min on ice. Coverslips were then stored in 70% ethanol at −20 °C for 1 h or until samples from all time points had been collected. Before hybridization with the probe, the coverslips with cells were brought back to 4 °C and serially dehydrated by 5-min incubations in ice-cold 80%, 95% and 100% ethanol. Coverslips were removed from 100% ethanol and allowed to air dry before incubation with probe for 48 h at 37 °C in a sealed chamber humidified with 2× SSC/50% formamide. For RNA probes, coverslips were washed for 3 × 5 min in 50% formamide (Fisher)/2× SSC (Ambion) and 3 × 5 min in wash buffer II (10 mM Tris, 0.5 M NaCl, 0.1% Tween-20), before a 45 min incubation with 25 µg ml⁻¹ RNaseA (Thermo Fisher) in wash buffer II at 37 °C. After RNaseA treatment, coverslips were washed for 2 × 5 min in wash buffer II, 2 × 5 min in 50% formamide/2× SSC, 3 × 5 min in 2× SSC and 3 × 5 min in 1× SSC before briefly drying excess 1× SSC off and mounting with Vectashield mounting media lacking DAPI (Vector Labs). Coverslips were sealed with Biotium Covergrip coverslip sealant (Thermo Fisher). For DNA probes, coverslips were washed for 3 × 5 min in 50% formamide/2× SSC, 3 × 5 min in 2× SSC and 3 × 5 min in 1× SSC before mounting. A 1:10,000 dilution of DAPI (0.5 mg ml⁻¹) was included in all penultimate 1× SSC washes. All washes were conducted at 42 °C, cells were protected from light. All procedures were performed, and used reagents disposed of, according to standard laboratory safety procedures.

The *Xist* RNA FISH probe used in our study covers the ~17.9 kb exonic regions of *Xist*. The MS2 tag is ~1.1 kb long and the MS2 FISH probe was designed to cover the entirety of the tag (see Fig. 1a). Differences in the length of sequence targeted by these probes made the *Xist* probe signal much brighter than the MS2 probe signal when visualized microscopically. Consequently, in our RNA FISH experiments, we used both probes to differentiate between the *cas* (detected by the *Xist* probe only) and the *129* allele (detected by the *Xist* and MS2 probes) in both wild-type and ΔE cells. Using the *Xist* probe allowed for better detection of the extent of dispersal of the *Xist*ᐟᴱ transcripts, which was important for our aggregation score calculations (see Extended Data Fig. 1e and section '*Xist* aggregation analysis').

In the RNA FISH assay for the nascent transcripts of X-linked genes (Fig. 2b, c, d, Extended Data Fig. 8), the presence of two nuclear nascent transcript foci (or spots) is indicative of bi-allelic expression of the respective X-linked gene as is observed in undifferentiated ES cells that do not express *Xist* and have not yet initiated XCI (see Extended Data Fig. 8a, b). In cells expressing *Xist*, one focus on the X chromosome lacking *Xist* indicates silencing (see Fig. 2b, wild-type cell). Conversely, we interpreted a single focus co-localizing with the X chromosome expressing *Xist* (or MS2) as a lack of silencing. Cells expressing *Xist* with bi-allelic X-linked gene expression were also considered to have defective silencing (see Fig. 2b, ΔE cell).

RNA FISH procedure for 3D-SIM and improved-resolution microscopy. All coverslips were processed according to ref. [32].

Immunofluorescence staining
The cell culture medium was changed 10 min before collection. Upon collection, culture medium was aspirated, and coverslips were gently rinsed twice with cold 1× PBS. Coverslips were then transferred to a new culture dish containing cold 1× PBS. If cells were CSK treated (MS2-CP-GFP expressing wild-type F1 2-1 MS2¹²⁹ ES cells; Extended Data Fig. 9b), then coverslips were gently treated with 1 ml (added dropwise) ice-cold CSK buffer (100 mM NaCl, 300 mM sucrose, 3 mM MgCl₂, 10 mM PIPES pH 6.8) and incubated on ice for 30 s before aspiration. Coverslips were then similarly treated with 1 ml ice-cold CSK-Trt Buffer (CSK+0.5% Triton X-100) for 30 s, followed with a second ice-cold CSK treatment. Coverslips were then processed as described in ref. [33]. See Supplementary Table 2 for antibody information.

Immunofluorescence staining combined with RNA FISH
Where immunostaining and RNA FISH were combined, immunostaining preceded FISH.

Combined staining for epifluorescence microscopy. The immunostaining protocol was followed as outlined above, but coverslips were not mounted. Instead, after the last round of washes (omitting DAPI in the penultimate wash), coverslips were re-fixed in 4% PFA in 1× PBS for 10 min at RT and then dehydrated through a 70–85–95–100% ice-cold ethanol series before overnight incubation with probe as described above in the section 'RNA FISH procedure for epifluorescence microscopy'.

Combined staining for 3D-SIM and improved-resolution microscopy. All coverslips were processed according to refs. [32,33].

Plasmid construction and cell line generation
Xistᐟᴱ targeting construct. To create the targeting vector pCR2.1-Puro-*Xist*ΔE, 3 kb upstream and 1.2 kb downstream of the mouse *Xist* E-repeat were PCR-amplified from mouse genomic DNA using primers WRM163-166, modified for In-Fusion cloning (Clontech) using Kapa polymerase (Kapa biosystems) according to the manufacturer's instructions. The upstream homology arm was integrated at the EcoR1 site and the downstream homology arm at the BamH1 site, in a four-piece InFusion cloning reaction, into a vector containing a floxed puromycin resistance cassette (PCR2.1-*loxP*-pGK-Puro-pA-*loxP*). Positive recombinants were identified by restriction digest with HindIII and sequencing.

E-repeat deletion in wild-type F1 2-1 MS2¹²⁹ ES cells. The *Xist* E-repeat was deleted in female wild-type F1 2-1 MS2¹²⁹ ES cells derived from an F1 cross of mice from pure bred *129* and *castaneous* background, and then targeted to contain a 11× tandem repeat of the MS2 hairpin located 1.2 kb downstream of the E-repeat[27] via homologous recombination. The wild-type F1 2-1 MS2¹²⁹ female ES cells also harbour an M2-reverse tetracycline TransActivator (M2rtTA) cassette within the Rosa26 locus that confers neomcyin resistance on the cells. Cells obtained from half of a confluent T75 flask of wild-type F1 2-1 MS2¹²⁹ female ES cells were electroporated with 40 µg of PciI linearized PCR2.1-Puro *Xist*ΔE targeting plasmid (800 V, 0.2 ms, 4 mm cuvette, Biorad X-Cell electroporation module) and plated at varying dilutions on 10 cm plates of confluent irradiated DR4 feeders. Then, 36 h after plating, the cells were selected with 1 µg ml⁻¹ puromycin for 10 days. One hundred clones were picked, expanded and subjected to Southern blot analysis using a Sac1 digest and an external probe (amplified using primers WRM193/194 (Supplementary Table 1) as outlined in Extended Data Fig. 5. The positive clone number 35 was expanded in culture, then transfected with a Cre-recombinase plasmid using Lipofectamine 2000 according to the manufacturer's protocol (Thermo Fisher), to delete the floxed puromycin resistance cassette. Transfected cells were serially diluted, 100 clones were picked, expanded and replica-plated for growth in the presence or absence of puromycin. Sub-clone number 96 was sensitive to puromycin. PCR analysis of genomic DNA confirmed the deletion of the puromycin cassette with primers APJ439/440 (Supplementary Table 1). Subsequent Southern blot analysis and sequencing of wild-type *Xist*

and $Xist^{\Delta E}$ PCR amplicons from genomic DNA (intron 6 to exon 7 using APJ248/631 (Extended Data Fig. 5, Supplementary Table 1)) showed that the ΔE targeting construct integrated on the *129* allele of *Xist* upstream of the MS2 tag, preserving the 3' splice site of intron 6, to yield the heterozygous E-repeat deletion ES cell line $Xist^{\Delta E,MS2(129)/WT(Cas)}$ (ΔE ES cells) (Extended Data Fig. 5 and data not shown). Sequencing of the exon 6–exon 7 RT–PCR amplicon (obtained from cDNA of differentiated ES cells) derived from the 129^{MS2} *Xist* transcript revealed the use of a cryptic 3' splice site downstream of the *loxP* site (Extended Data Fig. 5). The use of the cryptic splice site extended the E-repeat deletion within the *Xist* transcript (as initially designed) by 42 nt and removed the *loxP* site and additional vector sequences present in the genomic DNA from mature $Xist^{\Delta E}$ transcripts, resulting in a scar-less ligation of the 3' terminus of exon 6 to nucleotide 1479 of exon 7 (Extended Data Fig. 5 and data not shown). We ensured that ΔE ES cells maintained two X chromosomes throughout the targeting process and differentiated equally to wild type as judged by changes in morphology, and loss of NANOG and *Tsix* expression upon induction of differentiation (Extended Data Figs. 5, 8).

Engineering of wild-type and ΔE ES cells with a FLP-FRT recombination platform for rescue experiments. Wild-type F1 2-1 MS2^{129} ES cells and ΔE ES cells as described in the section 'E-repeat deletion in wild-type F1 2-1 MS2^{129} ES cells' (half of a confluent T75 flask) were electroporated with 40 µg of Fsp1 linearized Flp-IN homing plasmid (that integrates a FRT landing site downstream of the Col1A locus and carried a puromycin resistance cassette for selection[34]) at 800 V, 0.2 ms, 4 mm cuvette using a Biorad X-Cell electroporation module before being serially diluted on 10 cm plates, pre-coated with irradiated DR4 feeders. At 36 h after plating, the cells were selected with 2 µg ml⁻¹ puromycin for 10 days after which 200 clones were picked and expanded. Genomic DNA was isolated and digested with EcoRI, before being subjected to Southern blot analysis with the Col1A Xba/Pst1 3' probe. Positive clones 1–61 (wild-type) and 137 (ΔE) were used for all subsequent experiments (Extended Data Fig. 9a).

Generation of Flp-In plasmids encoding Flag–MS2–CP fusion proteins. The MS2 coat protein (MCP) coding sequence was PCR-amplified with a forward primer encoding a 3×Flag tag downstream of a Kozak-ATG start signal from the pHMM vector (Addgene, 67717). The reverse primer contained an in-frame Nhe1 site (primers APJ526/570 (Supplementary Table 1)) such that any fragment ligated into the site would be expressed in frame with the MCP protein, separated by a three-amino acid (Gly-Leu-Gly) linker. The Flag–MCP–Nhe1 fragment was inserted into the EcoRI site of the pBS32 vector using Infusion cloning. This vector is similar to the pgkATGfrt vector described in ref. [34] except that the tet-inducible promoter was replaced with a CAGGS promoter, enabling constitutive expression of the fusion protein. The coding sequence for each protein (GFP, PTBP1, MATR3, TDP-43 and CELF1) was PCR-amplified from cDNA with infusion overhangs, or synthesized (Genewiz) and ligated into the Nhe1 site of the pBS32–Flag–MCP parent plasmid using InFusion cloning (Clontech). The PTBP1(Y247Q), MATR3(mutPRI) and MATR3(ΔZfn) mutants were generated using primer-directed mutagenesis. The wild-type PRI sequence (GILGPPP) was mutated to create the mutant PRI sequence (GAAAPPA)[20]. The coding sequences for the CELF1, MATR3(S85C), TDP43(EGGG) and MS2CP–GFP–MS2CP fusions were synthesized (Genewiz). All plasmids were verified by sequencing.

The ΔC-terminal PTBP1 fragment that is fused to Flag–MCP in our rescue system comprises the first 299 amino acids of PTBP1—which includes the first two RRMs as well as the MATR3 interaction site—followed by 68 amino acids that are out of frame, and do not encode a functional linker region. A premature stop codon terminates the protein at residue 367.

Generating wild-type and ΔE ES cells expressing Flag–MCP–fusion proteins via Flp-In recombination. A total of 33 µg of the pBS32 plasmid DNA encoding the Flag–MCP fusion proteins and 26 µg of plasmid encoding the flpase FlpO were electroporated into wild-type ES cells carrying the FRT homing site (clone 1–61) for the GFP fusion and ΔE ES cells with the FRT homing site (clone 137) for all other fusion constructs (1/2 of a confluent T75 flask of ES cells per electroporation) (Extended Data Fig. 9a). Cells were plated on confluent irradiated DR4 feeders in a 10 cm dish and, 36 h after plating, selected with 170 µg ml⁻¹ hygromycin for 14 days, after which all colonies were picked and expanded. The resulting clones were tested for protein expression by immunoblot of lysates (RIPA buffer in 1× SDS lysis buffer (Thermo Fisher)) using an anti-Flag antibody as well as antibodies against the respective fusion protein (Supplementary Table 2). Immunostaining confirmed nuclear localization of all fusion proteins that failed to rescue the phenotypes associated with loss of the E-repeat. All clones used maintained two X chromosomes, as determined by FISH against *Tsix* in undifferentiated cells. For all rescue experiments, at least two clones were analysed, which revealed that the data are robust. Owing to space limitations, often the results from only one rescue clone per protein or mutant are shown.

Generation of tet-inducible Xist$^{\Delta Tsix}$ V6.5 male ES cells. Tet-On *Xist* male V6.5 ES cells carrying a tet-inducible promoter in place of the endogenous *Xist* promoter and a M2rtTA trans-activator as well as puromycin resistance in the R26 locus[2] (1/2 of a confluent T75 flask) were electroporated with 40 µg of Not1 linearized paa2Δ1.7 plasmid DNA[35] (800 V, 0.2 ms, 4 mm cuvette using a Biorad X-Cell electroporation machine) and plated on confluent irradiated DR4 feeders, to stop *Tsix* expression. Then, 36 h after plating, the cells were selected with Neomycin/G418 for 10 days after which 100 clones were picked and subjected to Southern blot analysis as described in ref. [35] (data not shown). Positive clone 70 was used for the PTBP1 ChIP–seq experiments.

siRNA treatments

Silencer Select siRNAs (Thermo Fisher) against PTBP1 (s72337), MATR3 (s69629), CELF1 (s64632) and TDP-43 (s106686) were diluted to 20 nM in 1× siRNA buffer (60 mM KCl, 6 mM HEPES pH 7.5 0.2 mM MgCl₂), aliquoted and stored at −80 °C until further use. Under sterile conditions at RT, 2.5 µl of 20 nM siRNA were added to 80 µl of fresh Opti-MEM solution (Gibco). siRNA MAX transfection reagent (1.6 µl, Life Technologies) was added to 80 µl Opti-MEM solution and subsequently added to the siRNA/opti-MEM solution after 5 min of incubation. The resulting solution was mixed by pipetting and left to incubate at RT for 20 min. The solution was then added to 200,000 cells in 0.8 ml of culture medium and plated in 1 well of a 12 well plate on 18-mm gelatinized coverslips and left overnight at 37 °C. For female ES cells undergoing differentiation, cells were plated in MEF medium and after 24 h, the culture medium was changed (with the addition of 1 µM all-*trans* retinoic acid, Sigma) and a second round of siRNA treatment was performed. Knockdown efficiency was assessed by immunoblotting (Supplementary Table 2).

Immunoblotting

Cells were collected by trypsinization, pelleted (1,000*g*, 5 min), resuspended in 500 µl 1× PBS to wash, and re-pelleted. The washed cell pellet was lysed in 5 pellet volumes of RIPA buffer and 40U benzonase (Novogen) and incubated at 4 °C overnight. The lysate was centrifuged at maximum speed to pellet the remaining insoluble material and the supernatant was transferred to a new tube and mixed with 4× Novex sample buffer containing 5% 14.3 M β-mercaptoethanol (Sigma) to a final concentration of 1×. The samples were then denatured for 5 min at 95 °C and loaded onto a 4–12% Novex Bis-Tris acrylamide gel with 1× MES running buffer (Life Technologies) run at 120 V for 1.5–2 h. The gels were transferred to a protran BA-85 nitrocellulose membrane (Whatman) using a Novex XCell II transfer system for 2 h at 30 V, 4 °C (or overnight at 4 °C at 10 V) in transfer buffer (25 mM Tris-HCl, 192 mM glycine, 20% methanol). Membranes were probed with primary antibody (Supplementary Table 2) in 1× Odyssey blocking buffer (LI-COR)

overnight at 4 °C, washed for 3 × 5 min in PBS+0.2% Tween-20 (Thermo Fisher) and then incubated with appropriate secondary antibodies (1:10,000 dilution, Odyssey 700 and 800 nm antibodies) in the dark at room temperature for 30 min before being washed again and scanned on a LI-COR infrared imaging system.

Co-immunoprecipitation

For co-immunoprecipitation experiments, rabbit IgG and antibodies against PTBP1, MATR3, CELF1, CIZ1 and TDP-43 (Supplementary Table 2) were crosslinked to ProteinG-Dynabeads (Thermo Fisher) using the protocol provided by Abcam (http://www.abcam.com/protocols/cross-linking-antibodies-to-beads-protocol) with minor modifications. In brief, 20 µl of bead slurry was isolated on a magnet and washed for 3 × 5 min in 5 volumes of 1× PBS. Beads were then washed once in 5 volumes of binding buffer (100 µl, 1× PBS containing 1 mg ml⁻¹ of BSA (NEB)) for 10 min and incubated in 100 µl binding buffer supplemented with 5 µg of rabbit IgG or antibodies against PTBP1, MATR3, CELF1, CIZ1 or TDP-43. Samples were rotated for 1 h at 4 °C. Beads were then washed in binding buffer for 5 min, followed by an additional 5-min wash in 1× PBS. Next, the antibody was crosslinked by incubating in 100 µl of 1× PBS solution containing 0.2 M triethanolamine (Sigma) and 6.5 mg ml⁻¹ dimethyl pimelimidate (DMP) (Sigma) pH 8.5 for 30 min with rotation at room temperature. Beads were then washed in 250 µl 0.2 M triethanolamine in 1× PBS for 5 min. DMP incubation and wash steps were repeated twice more before samples were quenched in 100 µl of 50 mM ethanolamine in 1× PBS for 5 min. The quenching step was repeated, and excess non-crosslinked antibody removed with 2 × 10 min incubations in fresh 1 M glycine pH 3.0. Beads were washed in 1× PBS for 3 × 5 min before use in immunoprecipitations. Immunoprecipitations were performed under non-denaturing conditions according to the Abcam protocol (http://www.abcam.com/ps/pdf/protocols/immunoprecipitation%20protocol%20(ip).pdf). Plates (4 × 15cm) of confluent wild-type F1 2-1 MS2[129] female ES cells were lysed by pipetting in 3 ml of lysis buffer (10 M Tris-HCl pH8, 137 mM NaCl, 1% NP40, 2 mM EDTA) supplemented with 1× Complete EDTA-free Protease Inhibitors (Roche) and incubated for 1 h on ice with or without RNase (10 µg ml⁻¹ RNase A) (Thermo Fisher). Lysate was centrifuged at 4 °C at maximum speed in a tabletop microfuge for 15 min to pellet insoluble material. The supernatant was transferred to new tubes and precleared with 20 µl of washed ProteinG Dynabeads per 1 ml of lysate with rotation at 4 °C for 1 h. Then, 500 µl of lysate was added to each crosslinked antibody-proteinG Dynabead prep and rotated at 4 °C overnight. The next day, crosslinked antibody–proteinG Dynabeads were isolated on a magnet and washed for 4 × 5 min in ice-cold wash buffer (10 mM Tris-HCl pH 7.4, 1 mM EDTA, 1 mM EGTA, 150 mM NaCl, 1% TritonX-100) supplemented with 1× Complete EDTA-free Protease Inhibitors. The co-purified proteins were eluted by boiling in 1× NuPage Protein Loading buffer (Thermo Fisher) supplemented with 5% β-mercaptoethanol, at 95 °C for 5 min. Samples were assessed by immunoblotting. The input represents 4% of lysate added per immunoprecipitate. 1/4 of the eluate was loaded per lane.

In vitro RNA transcription

For several in vitro experiments (Droplet assays, electrophoretic mobility shift assays (EMSAs)), RNAs encoding the E-repeat and other sequences were obtained by in vitro transcription (IVT). Templates for IVT were amplified from DNA using KAPA polymerase according to the manufacturer's instructions (KAPA Biosystems), and then gel-purified and concentrated over AMPure beads (homemade)[29]. See Supplementary Table 1 for primer information. RNA was transcribed and UREA-PAGE purified as described in ref. [36]. For biotinylated RNAs, Biotin–UTP (Ambion) comprised 18% of the total UTP.

Droplet assays

rPTBP1 purification. Recombinant 6×-His tagged PTBP1 was expressed by IPTG induction from plasmid pQE-80L-PTBP4 (human PTBP1, isoform 4) (Douglas Black Lab) in BL21 bacterial cultures and purified using Ni-NTA agarose (Invitrogen) according to the manufacturer's instructions. The purified protein was dialysed and stored in buffer DG (20 mM HEPES-KOH pH 7.9, 80 mM K glutamate, 20% glycerol, 2.2 mM MgCl₂, 1 mM DTT and 0.1 mM PMSF) at a stock concentration of 36 mg ml⁻¹.

rCELF1 purification. Recombinant 6×-His-tagged CELF1 was expressed by IPTG induction from plasmid pET28a-CELF1 (human) in Rosetta bacterial cultures and purified over His-Trap and Superdex 200 gel filtration columns. Purified protein was concentrated and stored in a buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 10% glycerol at a stock concentration of 5 mg ml⁻¹. pET28a-CELF1 was constructed via In-fusion, using a fragment encoding the CELF1 coding region (see Supplementary Table 1 for primers) into the PET-28a plasmid. The CELF1 coding region was amplified from a fragment synthesized by Genewiz with primers modified for the pET-28a plasmid. The sequence of the plasmid was verified before use.

Droplet assays. Droplets (10 µl) were assembled in 1.5 ml Eppendorf tubes as described in ref. [37]. In brief, 5 µl of a 2× buffer containing 200 mM NaCl, 40 mM imidazole, 2 mM DTT and 20% glycerol was supplemented with the E-repeat or control IVT RNA (varying concentrations), rPTBP1 (to a maximum concentration of 60 µM) and/or rCELF1 (maximum concentration of 38 µM) and water to 10 µl (final volume). The solution was mixed by pipetting and transferred to one well of an 8-well glass chamber slide (Ibidi) that had been pre-coated with 3% BSA, washed 3 times with RNase-Free water and dried. Droplets were imaged at 10×–20× magnification.

Electrophoretic mobility shift assays

EMSAs were performed as described in ref. [38] except that 40,000 counts per million of 5′ end labelled RNA was used per condition.

Quantitative RT–PCR and actinomycin D treatment

In several experiments we determined the levels of *Xist* by RT–PCR. For experiments with actinomcyin D treatment, the drug was dissolved in DMSO at 1 mg ml⁻¹ and added to the culture medium to a final concentration of 1 µg ml⁻¹. For PCR with reverse transcription (RT–PCR), cells were collected in 1 ml TRIzol (Thermo Fisher), after culture medium removal and washing with PBS. RNA was purified over RNAeasy columns (Qiagen). Total RNA (1 µg) was used in a reverse-transcription (RT) reaction with SuperScript III and an appropriate strand-specific reverse primer, according to the manufacturer's instructions (Thermo Fisher). One-twentieth of the RT reaction was used in a quantitative PCR reaction, using either 480 SYBR Green LightCycler PCR mix (Roche), SsoAdvanced Universal SYBR mix (Bio-Rad) or SYBR Green Master Mix (Applied Biosystems) and appropriate primers (see Supplementary Table 1), in triplicate reactions. RT–qPCR experiments were normalized against *Gapdh, Snrnp27* or *Rrm2* transcripts.

Crosslinking and immunoprecipitation of RNA and high-throughput sequencing (iCLIP–seq) for MATR3 and PTBP1

PTBP1, PTBP2 and TDP-43 iCLIP in differentiated cells was obtained from published datasets[39,40]. PTBP1 and MATR3 iCLIP experiments in ES cells were performed as described in ref. [41]. For iCLIP–seq, all washes were conducted for 5 min per wash, at 4 °C with ice cold buffers. Three confluent 15-cm plates of male tetO-*Xist* V6.5 (pSM33) ES cells[2] were used per immunoprecipitation upon 6 h of induction of *Xist* expression with 2 µg ml⁻¹ doxycycline, and crosslinking was performed at 100 mJ cm⁻² at 4 °C in a Stratalinker 1800 (Stratagene). Crosslinked cells were collected by scraping in cold 1× PBS and pelleted at 700g for 2 min. Cell pellets were lysed in ice cold lysis buffer (20 mM HEPES-KOH pH 7.5 (Sigma), 150 mM NaCl (Sigma), 0.6% Triton X-100 (Sigma), 0.1% SDS (Sigma), 1 mM EDTA (Gibco), and 0.5 mM DTT (Sigma)) and sonicated

in a bioruptor (Diagenode) for 2 × 15 min (30 s on, 30 s off) on high setting at 4 °C. Sonicated lysates were cleared by centrifugation at 20,000g, 5 min, 4 °C, supernatants transferred to 15 ml Falcon tubes and diluted in 5 volumes of buffer containing 20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.5 mM DTT, 1.25× complete protease inhibitors EDTA-free (Roche), 50 µg ml^{-1} yeast tRNA (Life Technologies) and 400 U RNaseOUT (Life Technologies). Samples were briefly mixed and rotated overnight at 4 °C. To prepare beads for pulldown, a magnet was used to isolate beads from 200 µl of proteinG–Dynabead slurry, which were then washed 3 times in WB$_{150}$ (20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.1% Triton-X100) and incubated overnight at 4 °C with 50 µg anti-MATR3 (Abcam, ab151714) or anti-PTBP1 (Abcam, ab5642) in 700 µl WB$_{150}$. Beads were washed three times in WB$_{750}$ (20 mM HEPES-KOH pH 7.5, 750 mM NaCl, 0.1% Triton-X100) and once with WB$_{150}$ (20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.1% Triton-X100) before incubation with lysate. After overnight incubation in lysate, beads were collected at the bottom of the Falcon tube with a magnet and the supernatant was removed. Beads were then transferred to a 1.5 ml Eppendorf tube with 1 ml of WB$_{150}$, washed five times in WB$_{750}$ and twice in PNK buffer (20 mM HEPES-KOH pH 7.5, 10 mM MgCl$_2$, 0.2% Tween-20). The immunoprecipitated RNA was fragmented in 100 µl of 1× MNase buffer (NEB) containing 5.0 µg of yeast tRNA that was pre-warmed to 37 °C in a thermomixer (Eppendorf) set to shake for 15 s on/15 s off at 750 rpm (or minimum speed required to prevent settling of the beads). 1× MNase buffer (50 µl) containing 60 gel units per ml (6 Kunz units per ml) of micrococcal nuclease (NEB M0247S) were added and incubated for exactly for 5 min. The reaction was stopped by the addition of 500 µl of EGTA buffer (20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 20 mM EGTA, 0.1% TritonX-100). The beads were then washed four times in EGTA buffer and twice in cold PNK buffer. The fragmented RNA was dephosphorylated in 100 µl of 1x FastAP buffer (Fermentas) containing 0.15 U µl^{-1} of fast alkaline phosphatase (Thermo Scientific, EF0651) and 0.2 U µl^{-1} of RNaseOUT (Life Technologies, 10777-019), incubated in a thermomixer for 90 min at 37 °C, 15 s shaking/20 s rest. Beads were washed four times in WB$_{750}$ and twice in cold PNK buffer. The dephosphorylated RNA was then ligated to a 3′ biotinylated linker RNA in 40 µl of buffer containing 1 mM ATP, 25% PEG4000 (Sigma, 202398), 0.5 U µl^{-1} T4 RNA ligase1 (NEB M0204S), 0.5 U µl^{-1} RNaseOUT, and 6.0 µM L3 linker (Supplementary Table 1). The ligation reaction was incubated in a thermomixer overnight at 16 °C, 15 s on/4 min off at a speed that prevents beads from settling. The next day, beads were washed four times in WB$_{150}$ and twice with cold PNK buffer. The RNA was then 5′ end labelled in 24 µl PNK wash buffer with 16 µl of 1× PNK buffer (NEB) containing 150 µCi of γ[^{32}P]ATP, 10U PNK and 1U µl^{-1} of RNaseOUT. The reaction was incubated in a thermomixer for 20 min at 37 °C set to shake for 15sec on/20 s off. The beads were then washed three times with WB$_{150}$. The immunoprecipitated complexes were eluted off the Dynabeads in 50 µl of buffer (100 mM Tris-HCl pH 7.5, 0.6% SDS, 5 mM EDTA, 50 mM DTT and 50 ng µl^{-1} yeast tRNA) incubated for 10 min at 85 °C, shaking continuously at 900 rpm. The elute was transferred to a new tube and the beads were rinsed with 1,200 µl of buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1.25× complete protease inhibitors (Roche), 50 ng µl^{-1} yeast tRNA and 0.1% Triton X-100) which was added to the first eluate. The combined eluates were centrifuged for 5 min at 4 °C at maximum speed and the supernatant transferred to a new tube to prevent carry over of any remaining Dynabeads. To prevent contamination with IgG heavy chain, which co-migrates with many proteins of interest, the biotinylated RNA–protein complexes were bound to monomeric avidin beads. To do this, 10 µl packed monomeric avidin agarose beads (Thermo Fisher) were washed three times with WB$_{150}$. Beads were pelleted after each wash by spinning in a swing bucket rotor at 1,000g, 4 °C (use of the swing bucket rotor helps prevent loss of agarose beads). One packed bead volume was mixed with an equal volume of WB$_{150}$ and 15 µl of the bead slurry was added to each combined eluate and rotated at 4 °C for 4 h. The beads were then pelleted as above

and washed three times with WB$_{150}$. After the final wash, the remaining 5–20 µl of supernatant was carefully removed with a p10 pipette. The complexes were eluted from the avidin beads by incubation in 30 µl of buffer (10 mM Tris-HCl pH 7.5, 10% glycerol, 2.2% SDS, 5 mM EDTA) at 85 °C for 10 min in a thermomixer shaking at 900 rpm. After centrifugation to pellet the beads, the supernatant was transferred to a new tube and mixed with 5 µl of 1× LDS sample buffer (Life Technologies) with 300 mM DTT. Samples were incubated at 90 °C for 10 min and then loaded on a pre-run (75 v, 10 min) NuPAGE Bis-Tris Gel (Life Technologies NP0307) with 1× MOPS running buffer and run for 10–15 min at 75 V and then 120 V until each sample is satisfactorily separated. The gel was then incubated in transfer buffer (25 mM Bis-Tris, 25 mM bicine, 1 mM EDTA pH 7.2, 20% methanol) for 5 min and then transferred onto a protran BA-85 nitrocellulose membrane using a semi-dry transfer apparatus (Biorad 170-3940) for 75 min at 400 mA (not exceeding 15 V). After completion of the transfer, the membrane was briefly washed in milli-Q water, wrapped in plastic film and exposed on a phosophoimager screen for 1 h. The regions of interest were then excised from the membrane and transferred to an Eppendorf tube. The RNA was eluted from the membrane by incubation in 300 µl of buffer (100 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM EDTA and 2 µg µl^{-1} proteinase K) for 30 min at 55 °C in a thermomixer, shaking continuously. 300 µl of pre-warmed buffer (100 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM EDTA, 7 M urea and 2 µg µΛ^{-1} proteinase K) was then added to the tube and incubated for a further 30 min at 55 °C. The supernatant was then transferred to a new tube and extracted with an equal volume of phenol:chloroform (5:1, pH 4.5). The separated aqueous phase was precipitated with 0.5 µl of Glycblue (Life Technologies), 60 µl sodium acetate pH 5.4 and 600 µl isopropanol overnight at −20 °C. The next day, the RNA was pelleted by centrifugation at 4 °C for 30 min at maximum speed. The pellet was then washed with 1 ml 75% EtOH before air-drying for 2 min and dissolved in 5.70 µl RNase-free water and left on ice for 5–10 min before being reverse transcribed. To do this, 0.5 µl of 10 mM dNTPs and 0.5 µl of 2 µM RT primer (Supplementary Table 1) were added to the RNA, mixed by pipetting and denatured for 5 min at 70 °C before snap-cooling on ice. The RT primers contain an 11 nt unique molecule identifier (UMI) used in sequence analysis (see section 'CLIP–seq analysis'). The sample was then equilibrated at 25 °C in a PCR machine before 3.5 µl of RT mix were added (2 µl 5× first strand buffer, 0.5 µl 100 mM DTT, 0.5 µl 100 U µl^{-1} Superscript III (Life Technologies) and 0.5 µl 40 U µl^{-1} RNaseOUT (Life Technologies)) and incubated for 5 min at 25 °C and then for 20 min at 42 °C, and then 20 min at 48 °C. The reverse transcription reaction was then transferred to a new Eppendorf tube containing 100 µl TE, 11 µl 3 M sodium acetate and 2.5 volumes of 100% EtOH. The cDNA was precipitated overnight at −20 °C, pelleted and washed as described above, dissolved in 5 µl RNase-free water and then mixed with 7.5 µl of formamide containing 10 mM EDTA, bromophenol blue and xylene cyanol tracking dyes. For size determination, ladder was prepared as follows: 2 µl GeneScan 500LIZ size marker (Life Technologies 4322682), 3 µl H$_2$O, 15 µl formamide containing 10 mM EDTA with no tracking dyes. The samples and ladder were denatured for 5 min at 85 °C and then loaded on a pre-run 5.5% (19:1 bisacrylamide:acrylamide) urea-PAGE gel (1× TBE, 7.5 M urea) for 20 min at 21 V. The gel was then scanned and a gel slice in the range of 70–120 nt was excised, chopped into 1-mm cubes and the cDNA eluted in 700 µl of TE buffer rotating at RT overnight. The next day, the cDNA was precipitated overnight as described above. The washed pellet was then dissolved in 6.7 µl of RNase free water and left on ice for 5–10 min before being transferred to a PCR tube. Subsequently, the RNA was circularized by addition of 1.5 µl of: 0.8 µl of CicrLigase II buffer, 0.4 µl 50 mM MnCl$_2$ and 0.3 µl of 100 U µl^{-1} CicrLigase II ssDNA ligase (Epicentre CL9021K) and incubated in a PCR machine at 60 °C for 60 min. The circularized cDNA was then digested with BamHI by addition of 30 µl of: 4 µl 10× FastDigest Buffer, 0.9 µl of 10 µM cut_oligo and 25.1 µl RNase-free water. This mix was incubated

for 4 min at 95 °C after which the temperature was decreased by 1 °C each minute until 37 °C after which 2 µl of FastDigest BamH1 (Thermo Scientific FD0054) was added and incubated at 37 °C for a further 30 min. The sample was transferred to an Eppendorf tube and pelleted as described above. The pelleted DNA was dissolved in 12 µl RNase free water. 2 µl was used to prepare a 42 µl PCR mix, containing 1× PFU buffer, 0.2 mM dNTPs, 0.2 µM P3 and P5 solexa primers, and 0.5 U PFU polymerase. A negative control containing water instead of cDNA was also prepared. The 42 µl reaction was then split into 4 × 10 µl reactions and PCR-amplified for 20, 24, 28 and 32 cycles (94 °C/3′; 94 °C/30 s; 63.5 °C/15 s; 72 °C/30 s; with final extension at 72 °C for 7 min). The PCR amplicons were run on a 2% agarose gel in 0.5× TBE/EtBr and the number of cycles required to produce 50–200 ng of PCR product from the remaining 10 µl of cDNA template was calculated. The PCR reaction was repeated using 10 µl of the remaining ssDNA template, run on a 2% gel as before and the 150–210 bp size range was excised and purified using Zymoclean Gel DNA recovery kit (Zymo Research D4007). DNA concentration was determined by qubit using the dsDNA Broad Range assay and prepared for sequencing on an Illumina HiSeq2000 machine using a single end 100 bp protocol.

Enhanced crosslinking and immunoprecipitation of RNA and high-throughput sequencing (eCLIP–Seq) for CELF1

eCLIP experiments against CELF1 (anti-CELF1 (ab129115)) were performed as described in ref. [42] with a few modifications. As with iCLIP, male tet-inducible -*Xist* V6.5 ES cells (pSM33)[2] were induced with 2 µg ml⁻¹ doxycycline for 6 h before crosslinking at 100 mJ cm⁻² at 4 °C in a Stratalinker 1800 (Stratagene). Cells were then processed according to the eCLIP protocol for both input and immunoprecipitated samples until DNA was obtained. We then followed the iCLIP protocol from the gel-purification of the cDNA through to amplification and purification of the DNA library. eCLIP samples were sequenced on an Illumina HiSeq2000 machine using the single end 50 bp protocol.

CLIP–seq analysis

CLIP–seq results were mapped using TopHat and processed with the publicly available fastq-tools, fastx-toolkit, Samtools, Bedtools, Deep-Tools and UCSC scripts[43–45]. The first 11 bases of each sequenced read correspond to a UMI, composed of a library-specific barcode (3 nt) flanked by four degenerate nucleotides. The UMI permitted removal of PCR duplicates from the total sequenced reads with the fastq-uniq command-line tool. The Fastx-toolkit was then used to clip 3′ adaptor sequences. Sequences shorter than 20 nt were discarded. Reads were then de-multiplexed and mapped to the iGenome mm9 reference genome by TopHat with high stringency settings. Library-depth normalized counts were generated, and data were converted to bigWig format to visualize tracks in IGV. Peaks were called using CLIPper (see ref. [45] and https://github.com/YeoLab/clipper) using the –superlocal option. Scripts are available at https://www.github.com/ShanSabri/iCLIP.

ChIP–seq

Plates (3 × 15 cm) of male tetOXist-Δ*Tsix* V6.5 ES cells (around 100 million cells) were used to prepare chromatin for PTBP1 ChiP–seq. Cells were induced for 0 h or for 20 h with 2 µg ml⁻¹ doxycycline to induce *Xist* before collection by trypsinization. Cells were then pelleted by centrifugation at 700*g* for 5 min at RT and resuspended in a total volume of 10 ml PBS. The wash step was repeated twice before resuspending in 10 ml of 1× PBS and transferring to a 50 ml Falcon tube to which disuccinimidyl glutarate (DSG; Pierce) in DMSO was added for crosslinking to a final concentration of 2 mM and incubated for 10 min at room temperature with gentle mixing. Cells were then pelleted, and the supernatant was discarded. Cells were re-suspended in 10 ml ES cell medium and incubated for 10 min with 1% formaldehyde (16% methanol free, Pierce) at room temperature with gentle. The reaction

was quenched by addition of freshly made 0.125 M glycine (Sigma) for 5 min at room temperature. Cells were pelleted and supernatant was discarded. Cells were washed twice in 50 ml PBS with protease inhibitors (Complete EDTA free, Roche) before being pelleted and flash-frozen in liquid nitrogen and stored at −80 °C. The frozen pellets were processed for ChIP–seq as described in ref. [46].

ChIP–seq analysis

Reads were mapped using BowTie to the iGenome mm9 reference genome. Duplicate reads were removed, and length extended to 49 nt. Normalized reads count were generated across 50-nt bins. Tracks were visualized in IGV in bigWig format.

RNA affinity purification

RAP was performed as described in ref. [2]. For the RAP–seq experiment, we used male T20 ES cells[47] carrying a homing site in the *Hprt* locus on the single X chromosome as well as a tetracycline-inducible transactivator in the R26 locus. The *Hprt* homing site includes a bidirectional, tetracycline-inducible promoter for expression of a control gene (*EGFP*) and of the *Xist* cDNA transgene introduced later by site-specific recombination, as well as a *loxP* site neighbouring the tet-promoter and linked to a truncated neomycin-resistance gene that lacks a promoter and translation initiation codon[34]. We integrated two different *Xist* cDNA transgenes into the homing site by electroporation of the respective *Xist* cDNA encoding plasmid and a Cre expression plasmid. The *Xist* transgene plasmid contained a promoter-less *Xist* sequence followed by a polyadenylation signal and a PGK promoter and translation initiation codon linked to a *loxP* site. Site-specific recombination of the *loxP* sites in the *Xist* cDNA plasmid and the homing site linked the translation initiation codon and *Pgk1* promoter to the *neo* gene, which restored the antibiotic resistance marker. A single copy of the *Xist* cDNA transgene was thus integrated under the control of the inducible promoter. In this study, we used an approximately 14.5kb *Xist* cDNA with either a 4122 nucleotide deletion between BstEii sites within the *Xist* cDNA, deleting the E-repeat and surrounding sequences, or a 1237 nucleotide deletion ending at a similar region within *Xist* and not including the E-repeat, which was generated by deleting internal sequences within the cDNA by SnaBI digestion and re-ligation (Fig. 1g and data not shown). Cells were induced with 2 µg ml⁻¹ dox for 6 h, before fixation for RAP–seq. RAP libraries were sequenced on the Illumina platform. Adaptor trimming was performed using cutadapt version 1.15 in paired-end mode with the following parameters: "-a AGATCGGAAGAGC -A AGATCGGAAG AGC". Read mapping was performed using bwa version 0.7.17-r1188 and samtools version 1.4 with the following command: "bwa mem -t 30 -T 0 ${INDEX} ${R1} ${R2} | samtools view -b - | samtools sort -O sam -T tmp -n - | samtools fixmate -O sam - - | samtools sort -O bam -T tmp - > ${BAM}". Read filtering was performed using samtools version 1.4 with the following command: "samtools view -b -q 30 ${BAM} | samtools rmdup - - | samtools view -b -L ${BED} - | samtools sort -O bam -T ${tmp} - > ${FILT_BAM}". The ${BED} variable is a path to a .bed file containing all chromosomes except for the *Xist* locus (chrX:103460216-103483359). Library-depth-normalized tracks were generated using bedtools v2.27.1 with the following commands: "bamToBed -i ${FILT_BAM} | sort -u -k 1,1 -k 2,2 -k 3,3 > ${FILT_BED} ; genomeCoverageBed -split -bga -scale `echo 1000000000/$(wc -l ${FILT_BED} | awk '{print $1}') | bc` -i ${FILT_BED} -g ${CHROM_SIZES} > ${BG} ; bedGraphToBigWig ${BG} ${CHROM_SIZES} ${BW}". Sample library-depth-normalized tracks were divided by the library-depth-normalized input (SRR850637 from ref. [2]) in R v.3.5 using rtracklayer 1.42.2 to import the tracks as Rle and export the divided tracks as bigwig. Input-normalized tracks were smoothed using deeptools v.3.4.3 and bedtools v.2.27.1 with the following command: "multiBigwigSummary bins -b ${NORM_BW} -out res.npz -bs 1000 -outRawCounts ${NORM_BG}; bedGraphToBigWig ${NORM_BG} ${CHROM_SIZES} ${SMOOTHED_NORM_BW}".

Article

Microscopy

Epifluorescence imaging. Cells with immunofluorescence and RNA FISH stainings were imaged using a Zeiss AxioImager M1 microscope with a 63× objective and acquired with AxioVision software. Epifluorescence images shown are sections and were analysed, merged and quantified using ImageJ or Adobe Photoshop.

3D-structured illumination microscopy (3D-SIM). 3D-SIM super-resolution imaging was performed on a DeltaVision OMX V3 system (Applied Precision, GE Healthcare) equipped with a 100 Å/-1.40 NA Plan Apo oil immersion objective (Olympus), Cascade II:512 EMCCD cameras (Photometrics) and 405, 488 and 593 nm diode lasers. Image stacks were acquired with a z-distance of 125 nm and with 15 raw images per plane (five phases, three angles). The raw data were computationally reconstructed with the soft-WoRx 6.0 software package (Applied Precision) using a wiener filter set at 0.002 and channel-specifically measured optical transfer functions using immersion oil with different refractive indices as described in ref. [33]. Images from the different channels were registered using alignment parameters obtained from calibration measurements with 0.2-μm diameter TetraSpeck beads (Invitrogen) as described in ref. [48].

Improved confocal microscopy. Improved confocal laser scanning microscopy was performed on a LSM880 platform equipped with 100×/1.46 NA or 63×/1.4 NA plan Apochromat oil objectives and 405/488 diode and 594 helium–neon lasers using the Airyscan detector (Carl Zeiss Microscopy). An appropriate magnification was used in order to collect image stacks from a region that encompassed the nucleus of interest thereby optimizing imaging time and reducing photobleaching. The pixel size and z-optical sectioning were set to meet the Nyquist sampling criterion in each case. Airyscan raw data were linearly reconstructed using the ZEN 2.3 software.

Quantitative image analysis

All image analysis steps were performed using Fiji or ImageJ[49,50], or IMARIS (Oxford Instruments).

Xist aggregation analysis. To quantify the aggregation of *Xist* clouds, images were taken as Z-stacks and transformed in a maximum intensity projection image to detect the entire *Xist* FISH signal in one plane. The background was removed using a rolling ball radius of 50 pixels. *Xist* RNA cloud areas were measured by creating a binary mask over the *Xist* RNA FISH signal for each analysed *Xist* cloud. The edges of each *Xist* cloud signal were determined by selecting a central pixel and all associated pixels of same intensity value (±5 units). The ImageJ FracLac[51] plugin was then used to calculate the area of a circle encompassing each cloud signal. The ratio of the *Xist* cloud area over its bounding circle area approximates the compaction of the *Xist* RNA cloud. Significant differences between wild-type and ΔE ES cell or siRNA-treated samples were tested with the non-parametric two-sample Kolmogorov–Smirnov test.

Imaris measurements. Raw z-stack 3D-SIM images were converted to an Imaris-compatible format using the Imaris File Converter module. Before analysis, all images were adjusted to ensure identical intensity/brightness levels. Using the Imaris MeasurementPro module, 50 linear 3D distances between 100 randomly chosen *Xist* foci were measured per cell, across 5 cells per sample.

CELF1 intensity plot profiles. Airyscan image stacks were imported into ImageJ and converted to 16-bit composites. The 3D-stacks were reduced into 2D images and 2 μm intensity line plots were used to extract the intensity profiles over the Xi enriched signal in the CELF1 channel. The same line-plot was used in a random nucleoplasmic region to select for the average nuclear CELF-1 intensities. The ratio of the top 10% intensities of the signals were plotted after diving over the nucleoplasmic signal.

Xi DAPI intensities quantification. Wide-field image stacks were generated from 3D-SIM raw data of H3K27me3 and DAPI stained cells by average projection of five consecutive phase-shifted images from each plane of the first angle and subjected to an iterative 3D deconvolution using soft-WoRx 6.0 software. The reconstructed image stacks were imported to ImageJ and converted to 8-bit files. In order to measure the Xi underlying DAPI intensity, binary masks from the H3K27me3 channel were created to define the Xi territory of day 7 wild-type and ΔE ES cell nuclei. A threshold was carefully applied selecting the border of the H3K27me3-enriched region that demarcates the Xi territory. Subsequently, the grey values of the corresponding masked region in the DAPI channel were extracted and plotted.

Segmentation of Xist RNA foci from 3D-SIM datasets. The 32-bit reconstructed 3D-SIM image stacks were imported into ImageJ. Grey values were shifted to the positive range and converted to 16-bit composites after subtracting the mode grey value to remove background noise. Segmentation of *Xist* RNA foci was performed by using the TANGO plugin[52] on ImageJ according to the pipelines described in ref. [33]. In brief, nuclear masks were created by using the nucleus processing chain. *Xist* foci were segmented by first pre-filtering with a TopHat filter with a radius of 1 pixel in all three dimensions (xyz), followed by a Laplace of Gaussian filter with a radius of 1 pixel (x, y, z). Segmentation of foci was performed using the spot detector 3D with Otsu auto-thresholding. Segmented objects were post-filtered with a size and edge filter of 5 pixels per spot and a signal-to-noise ratio greater than 2.

Amira reconstructions. 3D reconstructions were performed using Amira 2.3 (Mercury Computer Systems). Image stacks were imported into Amira as separate channels. *Xist* FISH or antibody stainings were reconstructed as surface renderings, while DAPI was reconstructed as volume rendering using the Volren module that enables visualization of intensity in colour maps.

Reporting summary

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

Data availability

All genomic data for *Xist* interactions and chromatin association have been deposited in the Gene Expression Omnibus (GEO) database under accession number GSE137305. Reagents are available upon request.

26. Kunath, T. et al. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895–2902 (2007).
27. Jonkers, I. et al. *Xist* RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. *Mol. Cell. Biol.* **28**, 5583–5594 (2008).
28. Pasque, V. et al. X chromosome reactivation dynamics reveal stages of reprogramming to pluripotency. *Cell* **159**, 1681–1697 (2014).
29. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).
30. Cremer, M. et al. in *The Nucleus* Vol. 463 (ed. Hancock, R.) 205–239 (Humana Press, 2012).
31. Henegariu, O., Bray-Ward, P. & Ward, D. C. Custom fluorescent-nucleotide synthesis as an alternative method for nucleic acid labeling. *Nat. Biotechnol.* **18**, 345–348 (2000).
32. Markaki, Y., Smeets, D., Cremer, M. & Schermelleh, L. Fluorescence in situ hybridization applications for super-resolution 3D structured illumination microscopy. *Methods Mol. Biol.* **950**, 43–64 (2013).
33. Kraus, F. et al. Quantitative 3D structured illumination microscopy of nuclear structures. *Nat. Protoc.* **12**, 1011–1028 (2017).
34. Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44**, 23–28 (2006).
35. Sado, T., Wang, Z., Sasaki, H. & Li, E. Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development* **128**, 1275–1286 (2001).
36. Pandya-Jones, A. & Black, D. L. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15**, 1896–1908 (2009).
37. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–219 (2015).

38. Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **20**, 1250–1257 (2013).

39. Vuong, J. K. et al. PTBP1 and PTBP2 serve both specific and redundant functions in neuronal pre-mRNA splicing. *Cell Rep.* **17**, 2766–2775 (2016).

40. Rogelj, B. et al. Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Sci. Rep.* **2**, 603 (2012).

41. Damianov, A. et al. Rbfox proteins regulate splicing as part of a large multiprotein complex LASR. *Cell* **165**, 606–619 (2016).

42. Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).

43. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

44. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).

45. Lovci, M. T. et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* **20**, 1434–1442 (2013).

46. Chronis, C. et al. Cooperative binding of transcription factors orchestrates reprogramming. *Cell* **168**, 442–459.e20 (2017).

47. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat. Genet.* **30**, 167–174 (2002).

48. Demmerle, J. et al. Strategic and practical guidelines for successful structured illumination microscopy. *Nat. Protoc.* **12**, 988–1010 (2017).

49. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

50. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

51. Karperian, A. *FracLac for ImageJ* http://rsb.info.nih.gov/ij/plugins/fraclac/FLHelp/Introduction.htm (1999–2013).

52. Ollion, J., Cochennec, J., Loll, F., Escudé, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840–1841 (2013).

Author contributions K.P., A.P.-J., Y.M. and D.L.B. conceptualized the project and A.P.-J. performed the experiments unless stated otherwise. Y.M. and T.C. performed experiments for 3D-SIM imaging and acquired and analysed 3D-SIM data, overseen by H.L. Y.M. acquired high-resolution images and performed image analysis on immunostained cells. J.S. performed all aggregation measurements, helped with EMSAs and analysed RAP–seq data. R.M., W.M. and A.C. helped to create ES cell deletion lines. S.Z. performed the initial PTBP1/2 iCLIP–seq experiments, A.D. helped A.P.-J. with iCLIP–seq experiments, S.S. and J.S. analysed CLIP–seq data, B.P. and C.C. performed and analysed CHIP–seq experiments, X.-J.W. purified rPTBP1 and rCELF1, and C.-K.C. performed RAP–seq experiments. A.P.-J., J.S., Y.M., T.C. and K.P. analysed data, A.P.-J., Y.M., J.S., M.G. and K.P. interpreted the data and contributed towards methodology and model creation, K.P., D.L.B, M.G. and H.L. acquired funding to support the project, A.P.-J. and K.P. administered the project and A.P.-J. and K.P. wrote the manuscript, including edits from all authors.

Article
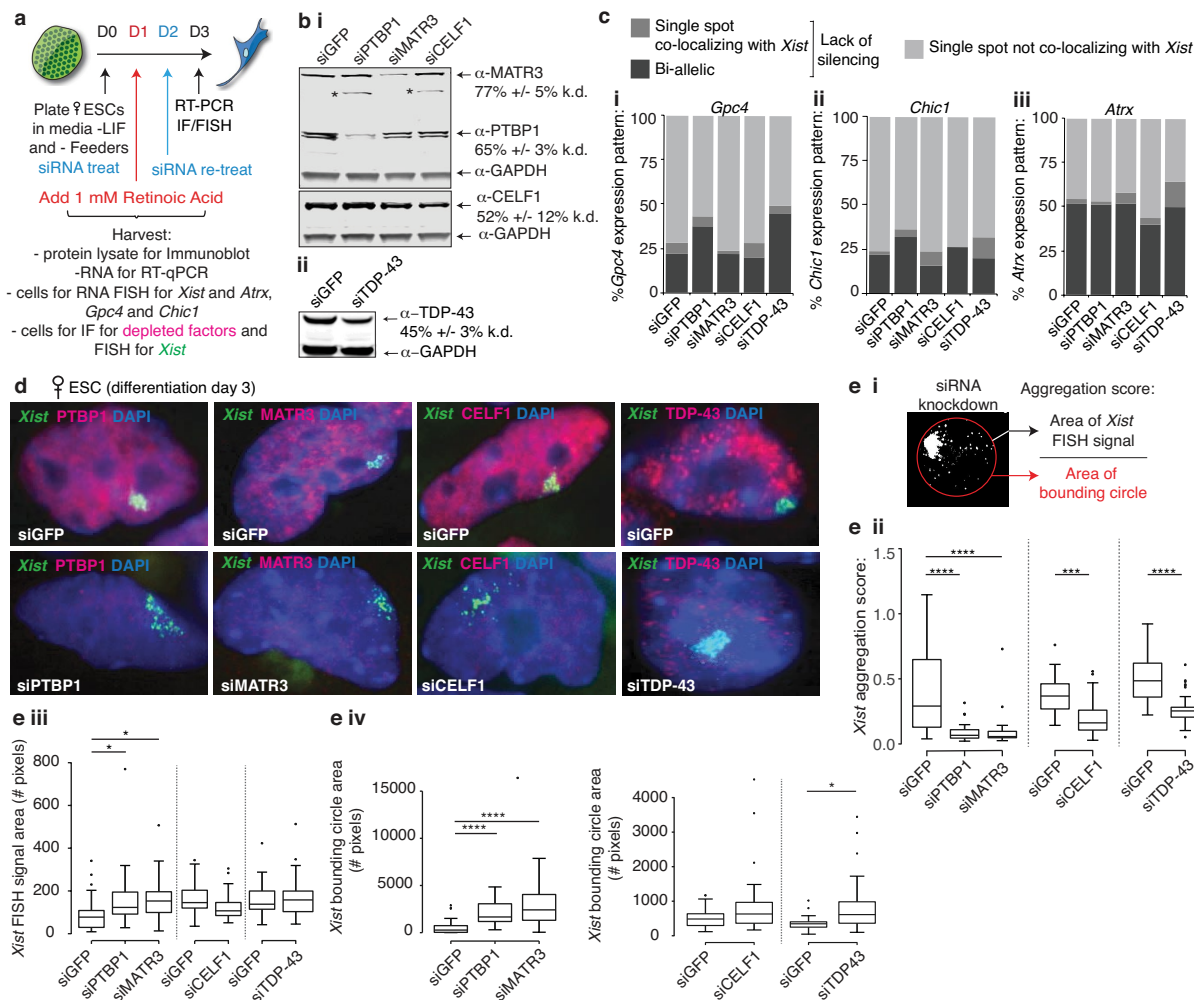
Extended Data Fig. 1 | Depletion of PTBP1, MATR3, CELF1 and TDP-43 does not strongly affect gene silencing during the *Xist*-dependent stage of XCI initiation. a, Experimental schematic. **b** (i), Immunoblot confirming the siRNA-mediated knockdown of PTBP1, MATR3, and CELF1, normalized to GAPDH. Asterisks indicate non-specific bands. (ii), As in (i), except for TDP-43. Error represents the s.e.m. from three independent experiments. For source data see Supplementary Fig. 1. **c** (i), Graph showing nascent transcription patterns of the X-linked gene *Gpc4* after 3 days of differentiation and knockdown of the indicated factor (spot refers to nascent transcription event on one chromosome) ($n = 50$ from 1 experiment). (ii), Same as (i) but for *Chic1*. (iii), Same as (i) but for *Atrx*. **d**, Representative images of siRNA-treated

differentiating cells immunostained for indicated proteins (red), probed for *Xist* (green) and DAPI stained (blue). **e** (i), Schematic for aggregation score calculation. (ii), Box plots showing *Xist* aggregation scores upon depletion of indicated proteins. Independent siGFP controls were used for CELF1 and TDP-43 experiments. (iii), Box plots showing the *Xist* mask values used to calculate the aggregation scores in (ii). (iv), Box plots showing the bounding circle area values encompassing the *Xist* mask used to calculate the *Xist* aggregation scores in (ii). For box plots in (ii)–(iv): ($n = 25$): *$P < 0.05$, ***$P < 0.0005$, ****$P < 0.00005$; two-tailed Kolmogorov–Smirnov test from one replicate in **b**. Horizontal lines denote the median, whiskers indicate 1.5× the interquartile range, dots represent outliers.
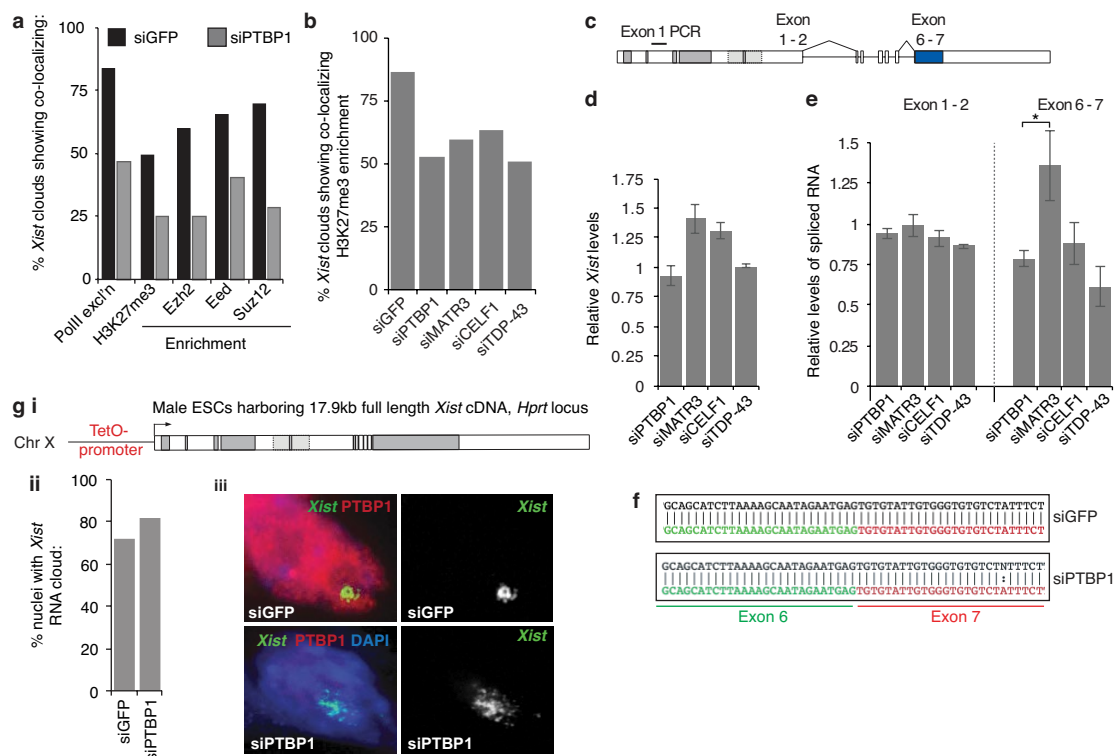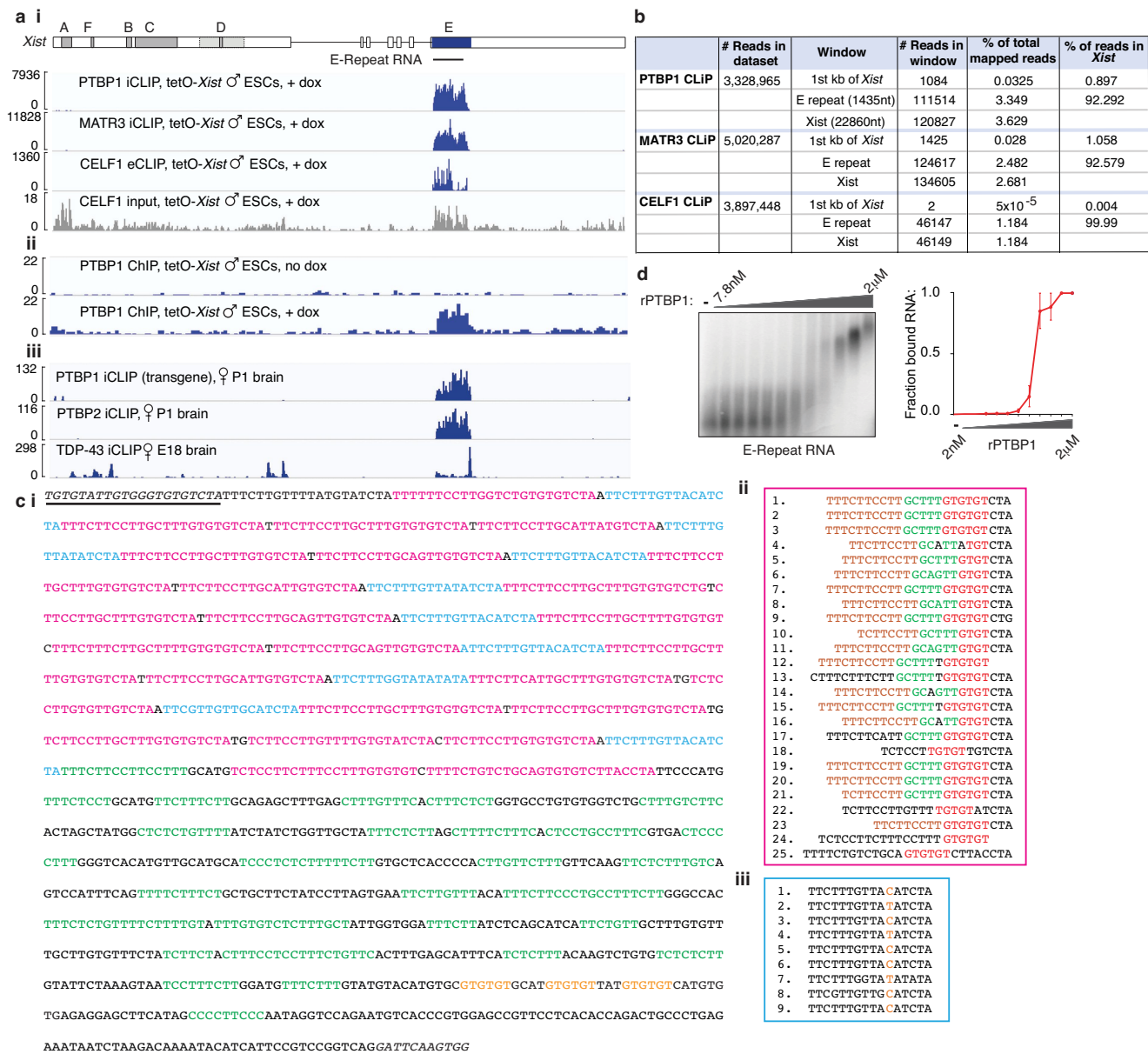
**Extended Data Fig. 2 | Depletion of PTBP1, MATR3, CELF1 and TDP-43 affects *Xist* localization during XCI initiation without strongly altering *Xist* processing. a**, Proportion of *Xist*-positive cells with co-localizing exclusion of RNA Pol II or enrichment of H3K27me3 or the PRC2 components EZH2, EED, SUZ12 on the Xi, in female ES cells differentiated for 3 days and treated with siGFP or siPTBP1 (*n* = 50 from one experiment). **b**, Percentage of *Xist*-positive cells with H3K27me3 Xi-enrichment in day 3 differentiated female ES cells treated with siGFP, siPTBP1, siMATR3, siTDP-43 or siCELF1 (*n* = 100, from one experiment). The siPTBP1 sample is independent from that in **a**. **c**, *Xist* splicing events assessed below. **d**, Histogram showing *Xist* abundance (exon 1 PCR above) upon siRNA-mediated knockdown of indicated RBPs in female ES cells at differentiation day 3. **e**, As in **d**, except for the abundance of spliced *Xist* exon 1–2 and exon 6–7 amplicons upon knockdown. For **d** and **e**, samples were

normalized against siGFP and *Snrnp27* RNA and assessed in triplicate from three independent experiments. Error bars represent s.e.m; *P < 0.05, two-tailed Student's t-test. **f**, Snapshot of expected spliced exon 6 (green) to exon 7 (red) sequence. Correct exon 6–7 ligation occurs after 72 h of siGFP or siPTBP1 treatment in differentiating female ES cells (black sequence) in two independent experiments. **g** (i), A tet-inducible full-length *Xist* cDNA transgene was inserted into the X-linked *Hprt* locus in male ES cells. (ii), Percentage of cells with an *Xist* cloud after 48 h of siPTBP1, and dox treatment starting at 24 h of siRNA treatment, in cells described in (i) (*n* = 80, from one experiment). (iii), Representative RNA FISH images of *Xist*, co-immunostained for PTBP1 and DAPI labelled, in cells described and treated as in (i), (ii). Note *Xist* dispersal upon PTBP1 knockdown, despite the absence of *Xist* splicing.

a

i

Xist — A F B C D E — E-Repeat RNA

PTBP1 iCLIP, tetO-*Xist* ♂ ESCs, + dox — 7936 / 0

MATR3 iCLIP, tetO-*Xist* ♂ ESCs, + dox — 11828 / 0

CELF1 eCLIP, tetO-*Xist* ♂ ESCs, + dox — 1360 / 0

CELF1 input, tetO-*Xist* ♂ ESCs, + dox — 18 / 0

ii

PTBP1 ChIP, tetO-*Xist* ♂ ESCs, no dox — 22 / 0

PTBP1 ChIP, tetO-*Xist* ♂ ESCs, + dox — 22 / 0

iii

PTBP1 iCLIP (transgene), ♀ P1 brain — 132 / 0

PTBP2 iCLIP, ♀ P1 brain — 116 / 0

TDP-43 iCLIP ♀ E18 brain — 298 / 0

b

	# Reads in dataset	Window	# Reads in window	% of total mapped reads	% of reads in *Xist*
PTBP1 CLiP	3,328,965	1st kb of *Xist*	1084	0.0325	0.897
		E repeat (1435nt)	111514	3.349	92.292
		Xist (22860nt)	120827	3.629	
MATR3 CLiP	5,020,287	1st kb of *Xist*	1425	0.028	1.058
		E repeat	124617	2.482	92.579
		Xist	134605	2.681	
CELF1 CLiP	3,897,448	1st kb of *Xist*	2	5×10^{-5}	0.004
		E repeat	46147	1.184	99.99
		Xist	46149	1.184	

d

rPTBP1: - 7.8 nM → 2 µM

E-Repeat RNA

Fraction bound RNA: 1.0 / 0.5 / 0.0 — 2nM → rPTBP1 → 2µM

c (i)

Extended Data Fig. 3 | PTBP1, MATR3, CELF1 and TDP-43 directly bind the *Xist* E-repeat, comprising a tandem array of 20–25nt C/U/G-rich elements. **a** (i), Top, diagram of the *Xist* genomic locus. The IVT E-repeat RNA used in **d** is indicated. Bottom, PTBP1, MATR3 and CELF1 i/eCLIP–seq profiles across the *Xist* locus in male tetO-*Xist* ES cells after 6 h of dox induction. CELF1 input profile is shown, read counts indicated on left. (ii), PTBP1 ChIP–seq profiles across the *Xist* locus before or after 20 h of dox treatment in male tetO-*Xist* ES cells. (iii), PTBP1, PTBP2 and TDP-43 iCLIP–seq profiles across the *Xist* locus in the female mouse brain. **b**, Table of mapping statistics for PTBP1, MATR3 and CELF1 i/eCLIP–seq data in **a**. Note that *Xist* is overexpressed in this experiment, which influences the number of reads mapping to the locus. **c** (i), The first 1,500 nt of exon 7 of *Xist* are shown, capturing the E-repeat. The sequence remaining after splicing of the *Xist*$^{\Delta E}$ transcript is underlined and italicized. The C/U/G

tandem repeats within the 5′ half of the E-repeat are indicated (pink-full and blue-truncated repeats) as are the CU-tracts (green) in the 3′ half. Potential TDP-43 sites are indicated in orange. (ii), Alignment of the 25 full C/U/G-tandem repeats (pink) from (i). Brown tracts encode putative PTBP1/MATR3 binding sites, red tracts putative CELF1/TDP-43 binding sites. (iii), Alignment of the nine truncated C/U/G-tandem repeats (blue) from (i). Orange coloured nucleotides are variable within each truncated repeat unit. **d**, Left: EMSA of IVT E-repeat RNA (see **a**) and either none, or increasing amounts of rPTBP1 (0, 1.95 nM, 3.9 nM, 7.8 nM, 15.6 nM, 31.3 nM, 62.5 nM, 125 nM, 250 nM, 500 nM, 1 µM and 2 µM). Right, quantification of the bound RNA fraction (dissociation constant, $K_d \approx 200$ nM, from two independent experiments, with s.e.m shown). For source data see Supplementary Fig. 1.
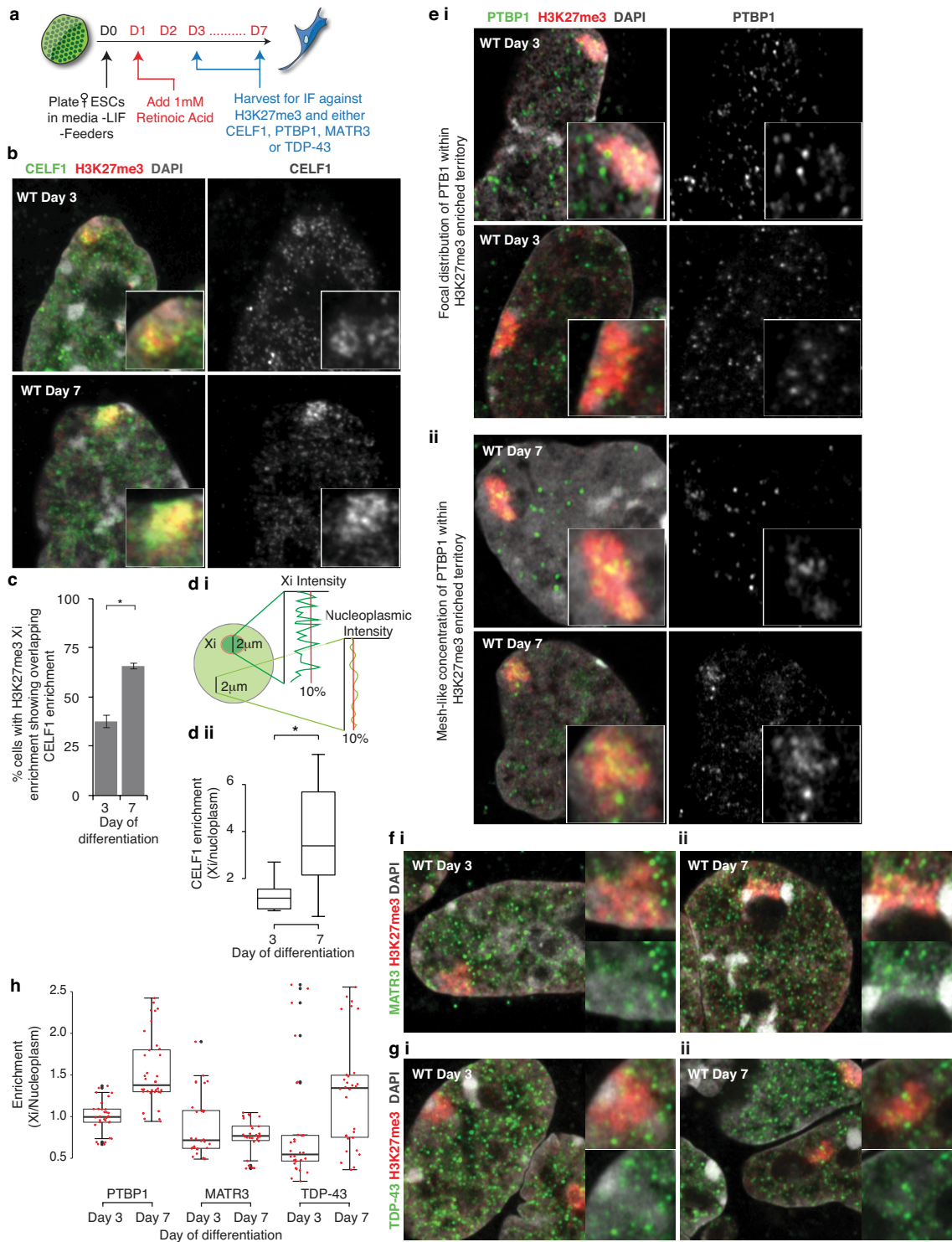
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | CELF1 and PTBP1 localize within the *Xist*-coated territory. a, Experimental schematic. **b**, Left, confocal-Airyscan sections of wild-type ES cells at differentiation day 3 and 7, immunostained for CELF1 and H3K27me3. Inset, enlargement of the Xi territory. Right, CELF1 staining in greyscale. **c**, Histogram showing the proportion of H3K27me3-marked Xi's with a co-localizing CELF1 enrichment. Error bars indicate s.e.m. ($n = 50$ from 3 coverslips across 2 independent differentiations); *$P < 0.05$, two-tailed Student's *t*-test. **d** (i), Intensity values for CELF1 fluorescence were recorded across a 2 μm line over the Xi (identified on the basis of the H3K27me3 Xi-staining) or within the nucleoplasm of the same nucleus in *z*-stack projections. (ii), Box plot showing the distribution of the ratio between the top 10% CELF1 Xi intensity values compared to the top 10% intensity values from the nucleoplasm ($n = 12$, from one experiment); *$P < 0.05$, two-sample Kolmogorov–Smirnov test. **e** (i), Left, As in **b**, but showing PTBP1 immunostaining at differentiation day 3. Right, PTBP1 staining in greyscale. (ii), As in (**i**), except at differentiation day 7. Note that these images highlight a mesh-like PTBP1 concentration within the Xi observed in a small fraction of cells, distinct from that observed in the nucleoplasm of these cells or from the pattern within the Xi at day 3. **f** (i), As in **e** (i), but showing MATR3 immunostaining and Xi-zoom ins. (ii), As in **e** (ii), except showing MATR3 immunostaining. **g**, As in **f**, except for TDP-43. **h**, As in **d** (ii), except showing data for PTBP1, MATR3 and TDP-43, ($n = 5$, from one experiment). Red dots, data points for the top 10% Xi/Nucleoplasmic intensity values from 5 cells. For box plots in **d** (ii) and **h**, horizontal lines denote the median, whiskers indicate 1.5× the interquartile range, dots represent outliers.
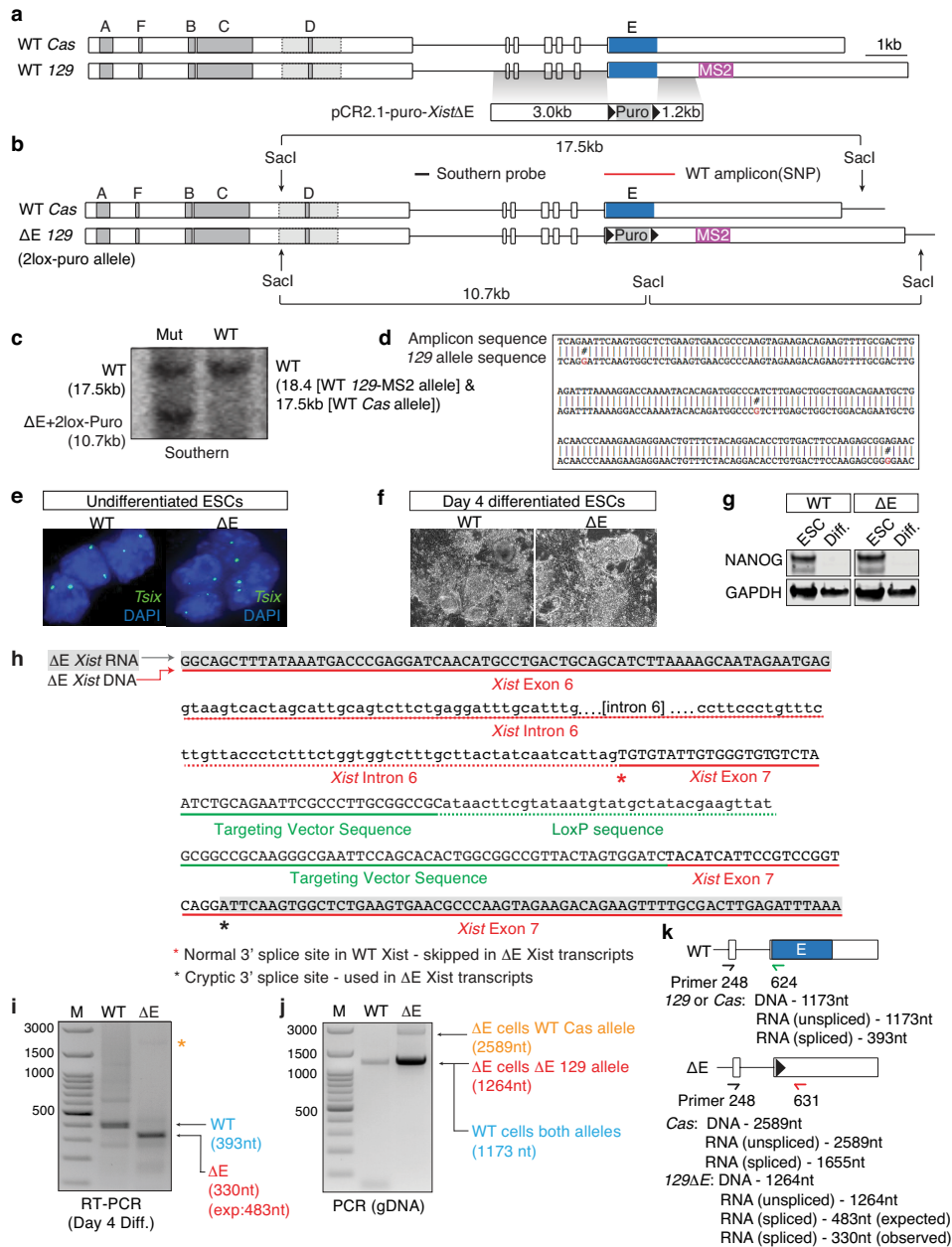
Extended Data Fig. 5 | ΔE ES cells undergo differentiation similar to wild-type ES cells and splicing of *Xist* intron 6 proceeds in the absence of the E-repeat. a, Homologous recombination strategy used to delete the *Xist* E-repeat in female ES cells. **b**, Southern blot strategy with a 5′ external probe for identification of deletion clones. **c**, Southern blot (described in **b**) on targeted ES cells with a *loxP*-flanked puromycin cassette in place of the E-repeat on one *Xist* allele. **d**, Sequencing analysis (black) of the wild-type *Xist*-PCR amplicon in ΔE cells (red line in **b**). *129*-allele SNPs are shown in red and do not match those in PCR amplicon, confirming E-repeat deletion on the *Xist*^MS2(129) allele. **e**, *Tsix* RNA FISH on undifferentiated wild-type and ΔE ES cells confirms the presence of two *Tsix* nascent transcription units, used as a proxy to confirm targeted cells maintain two X chromosomes. **f**, Bright-field images of wild-type and ΔE cells at day 4 of differentiation, showing that differentiating cells are morphologically similar. **g**, Immunoblot of differentiation day 2 wild-type and

ΔE cell lysate, showing equal loss of NANOG expression. **h**, Sequence of genomic and cDNA amplicons of the *Xist*^ΔE allele after puromycin cassette removal, confirming correct targeting and the use of a cryptic splice site in ΔE cells. **i**, Exon 6–7 RT–PCR amplicons generated from RNA isolated from day 4 differentiated wild-type (primers APJ248/624) or ΔE (primers APJ248/631) cells. The ΔE PCR amplicon was shorter than expected. Sequencing revealed a cryptic 3′ splice site downstream of the *loxP* site that extended the E-repeat deletion within the *Xist* transcript (but not the *Xist* genomic DNA) by 42 nt (see (**h**)). **j**, PCR amplicons from wild-type or ΔE genomic DNA using the same primers as in **i**. The intron 6-containing products can be amplified, indicating non-detection of intron 6-containing *Xist* transcripts is not due to amplification problems. **k**, Schematic outlining primers used to assess *Xist* DNA and RNA in **i** and **j**. For **c**, **g**, **i** and **j**, see Supplementary Fig. 1 for source data.
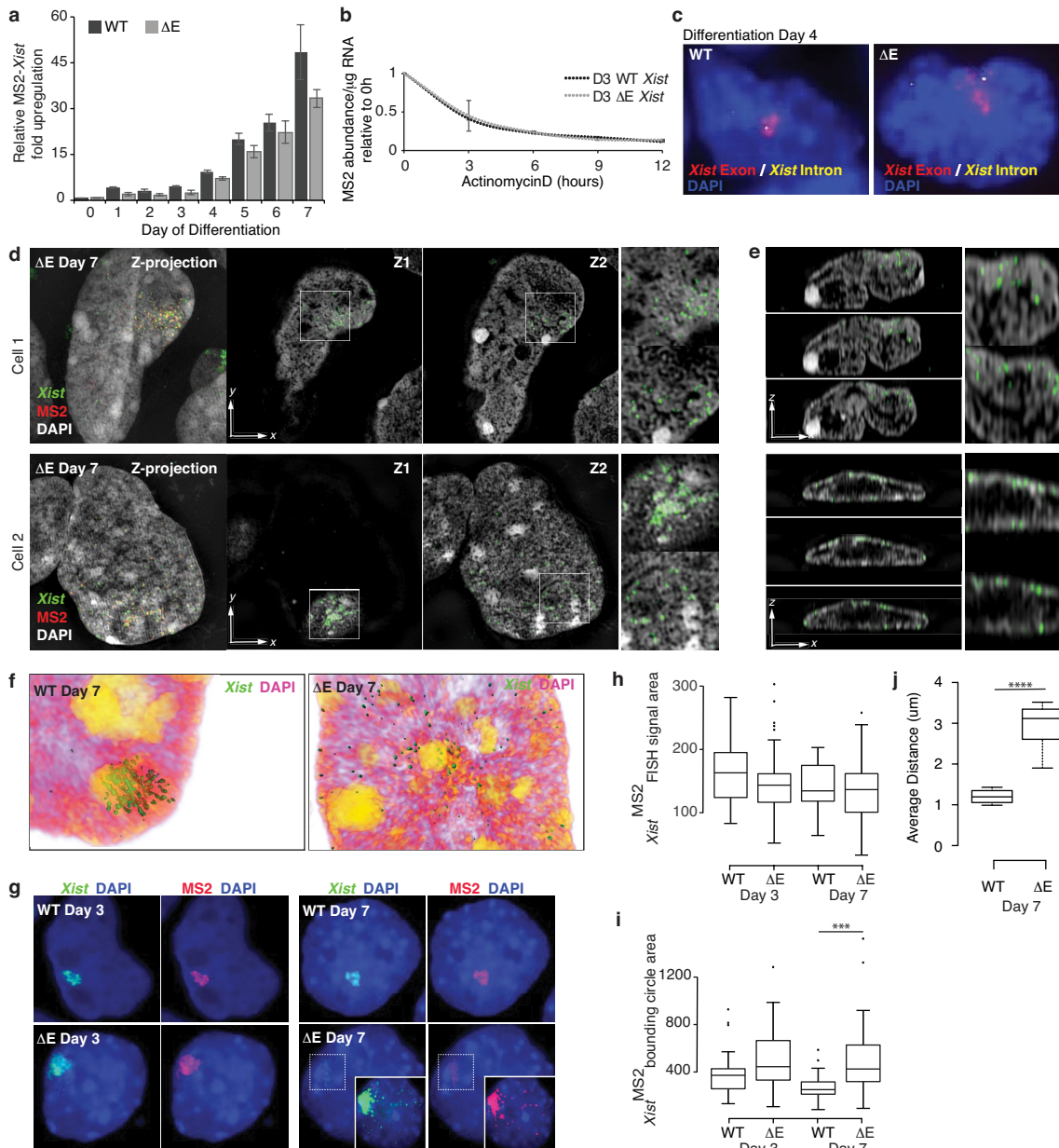
Extended Data Fig. 6 | Loss of the E-repeat does not affect *Xist* abundance, splicing or stability. a, RT–qPCR quantification of the fold upregulation of *Xist*[MS2] RNA during differentiation of wild-type or ΔE cells normalized against undifferentiated samples and an internal control (*Rrm2*). **b**, RT–qPCR measurements of *Xist*[MS2] RNA half-life (upon actinomycin D treatment) at day 3 of differentiation in wild-type or ΔE cells, calculated as MS2 transcript copy number per μg of total RNA. For **a** and **b**, error bars represent the s.e.m. (*n* = 3, measured in triplicate). Differences were not significant by two-tailed Student's *t*-test. **c**, Epifluorescence images of differentiation day 4 wild-type and ΔE cells probed for exonic regions of *Xist* (red) or *Xist* intron 1 (yellow), and DAPI stained, indicating that the *Xist*[ΔE] transcripts within the cloud are spliced. **d**, Same as Fig. 1h, except two additional *Xist*[ΔE]-expressing nuclei are shown. Scale bar, 5 μm. **e**, Same as Fig. 1h, except for the nuclei in **d**. Note aberrant localization of *Xist*[ΔE] at the nuclear lamina. **f**, 3D Amira reconstructions of the cells shown in Fig. 1h. **g**, Representative epifluorescence images of RNA FISH against *Xist* and MS2 with DAPI staining for comparison to super-resolution images in **d**, **e** and Fig. 1e, h. Inset, enhanced image of the marked area. **h**, Box plot showing the distribution of the area (in pixels) covered by the *Xist* RNA FISH signal, used to calculate the *Xist* aggregation score in Fig. 1f (*n* = 30, from one experiment). **i**, Same as **h** except showing distribution of the bounding circle area, (*n* = 30, from one experiment); ***P* < 0.00005, two-sample Kolmogorov–Smirnov test. **j**, Box plot of the average distance between *Xist* foci within *Xist*.[MS2] clouds in differentiation day 7 wild-type and ΔE ES cells, as measured by IMARIS. 50 measurements were made per cell, 5 cells per sample; *****P* < 0.000005, two-sample Kolmogorov–Smirnov test. For **h**–**j**, horizontal lines denote the median, whiskers indicate 1.5× the interquartile range, dots represent outliers.
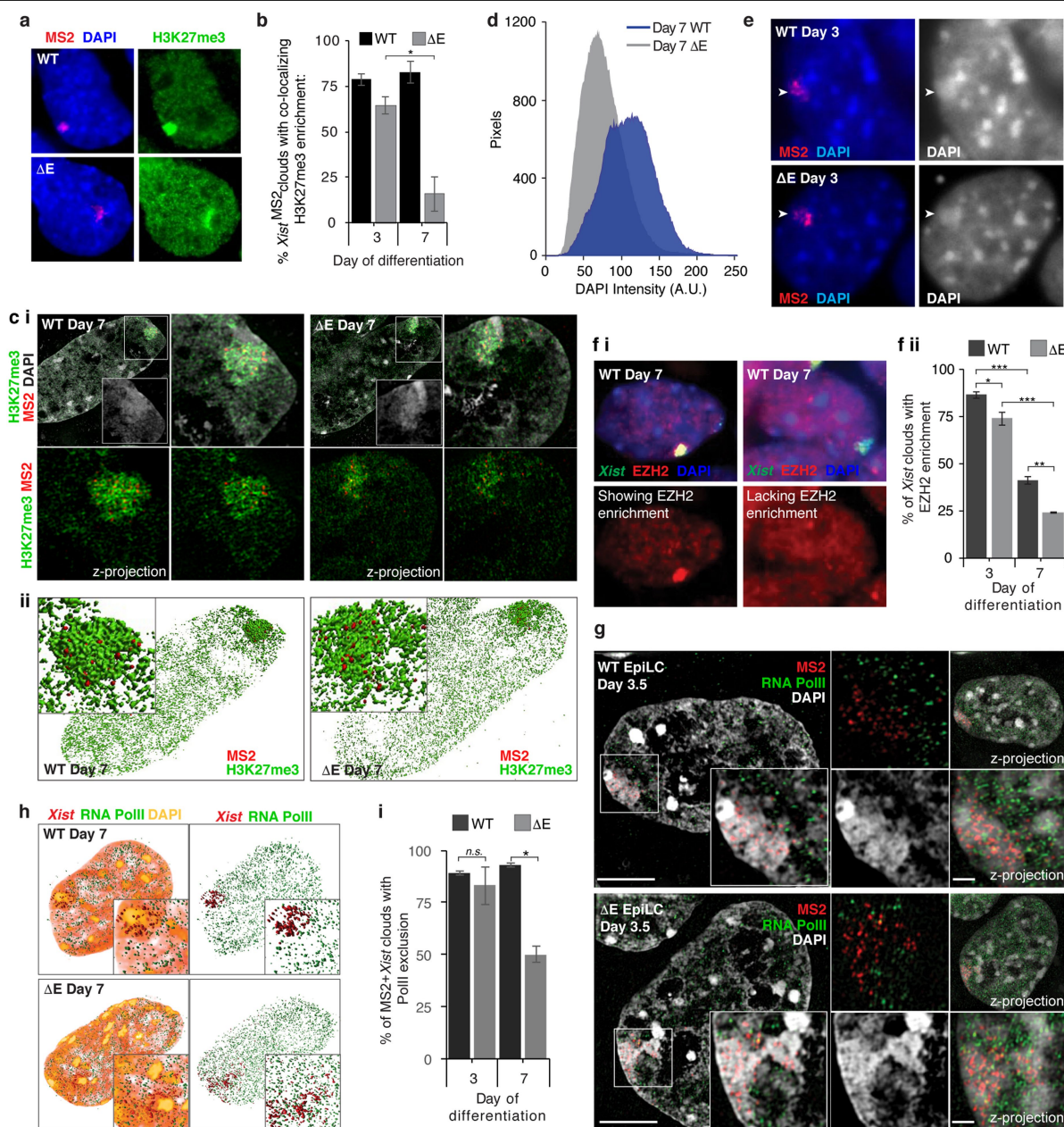
Extended Data Fig. 7 | The *Xist*^ΔE^-coated X chromosome displays decreased DAPI staining and less compact H3K27me3 accumulation at differentiation day 7. a, Epifluorescence images of cells immunostained for H3K27me3 and probed for MS2. **b**, Quantification of *Xist*^MS2^ RNA FISH clouds with a co-localizing accumulation of H3K27me3 at day 3 or 7 of differentiation in wild-type or ΔE cells (*n* = 60/coverslip, 3 coverslips over 2 experiments); *P = 0.05, two-tailed Student's *t*-test. **c** (i), Top left, 3D-SIM section of wild-type and ΔE cells at differentiation day 7 stained for H3K27me3 and DAPI and probed for MS2. Inset, DAPI staining of marked region. Right, magnification of inset area with (top) or without DAPI (bottom). Bottom left, *Z*-stack projection of inset without DAPI. (ii), 3D Amira reconstruction of images in (i). **d**, Graph showing the number of pixels with indicated DAPI fluorescence intensity from *Xist*^MS2^-expressing X chromosome in wild-type and ΔE cells, masked by H3K27me3 enrichment (*n* = 10, from one experiment). **e**, Epifluorescence images of wild-type and ΔE cells probed for MS2. Arrowheads point to the *Xist*

cloud and highlight the DAPI-bright staining for the X-territory. **f** (i), Epifluorescence images of wild-type cells stained for EZH2 and *Xist*, with (left) and without (right) EZH2 Xi-enrichment at differentiation day 7. (ii), Histogram of the percentage of *Xist* clouds with co-localized EZH2 enrichment (*n* = 60 per coverslip, 3 coverslips from 2 experiments), *P < 0.05, **P < 0.005, ***P < 0.0005, two-tailed Student's *t*-test. **g**, 3D-SIM sections through day 3.5 differentiated wild-type or ΔE ES cells (EpiLC differentiation), immunostained for RNA Pol II and probed for *Xist*, showing exclusion of RNA Pol II from the X-territory. Inset, signals derived from marked area. Small images: top left, same as inset without DAPI; bottom left, same as inset with only DAPI; top right, *Z*-stack projection of the cell; bottom right: *Z*-stack projection of the *Xist*-coated X chromosome. Scale bar, 5 μm; inset, 1 μm. **h**, 3D Amira reconstruction of cells in Fig. 2e. Inset, enlargement of the *Xist*^MS2^-expressing X. Right, same as left without DAPI. **i**, Quantification of RNA Pol II exclusion from *Xist*^MS2^-coated territory (*n* = 50 per coverslip, 2 coverslips from 1 experiment), *P = 0.05, two-tailed Student's *t*-test.

Article
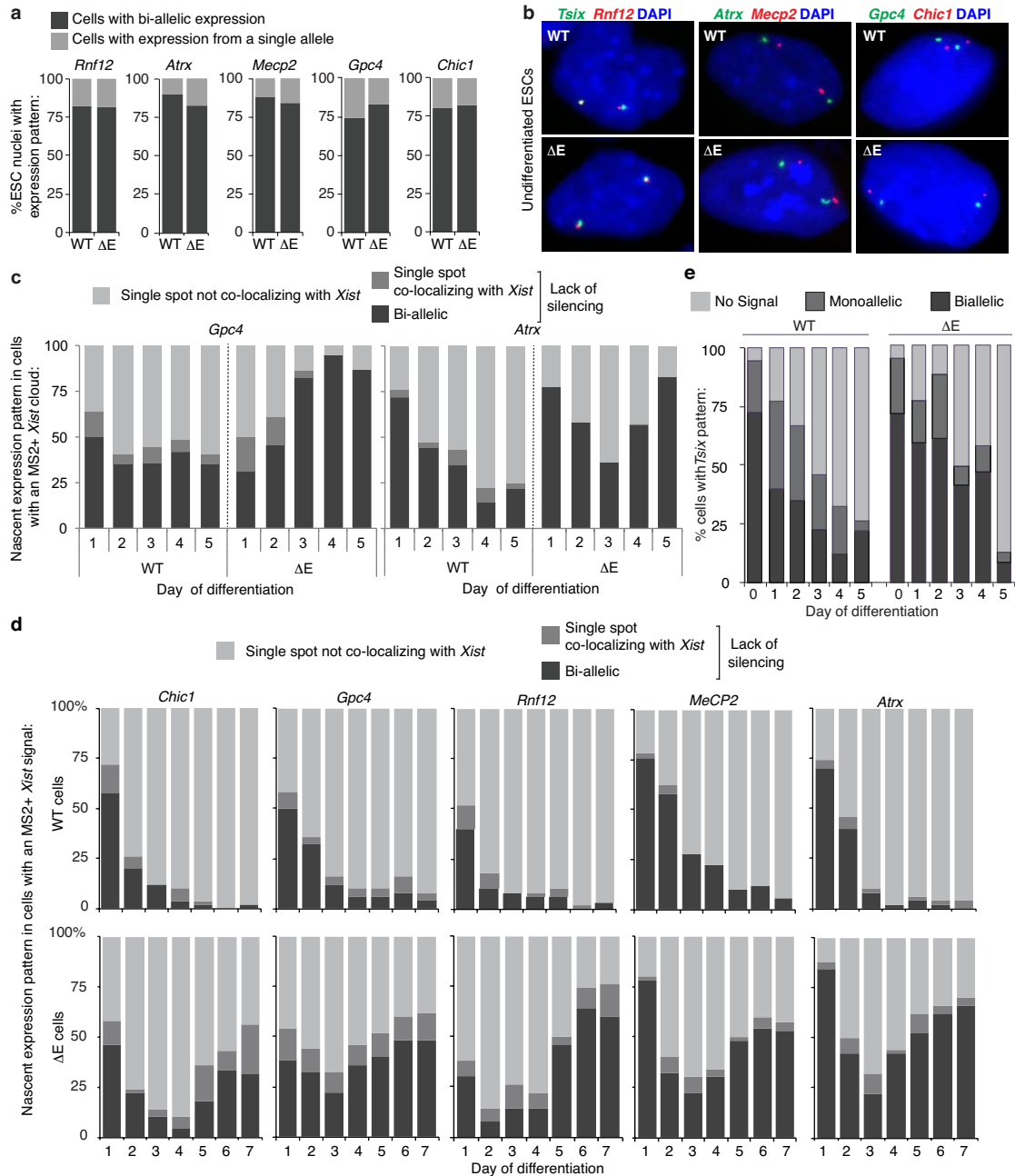
Extended Data Fig. 8 | Loss of the E-repeat prevents continued gene silencing in differentiating ES cells. a, Histograms of nascent transcription pattern of indicated X-linked genes (*Rnf12* (*Rlim*), *Atrx*, *Mecp2*, *Gpc4* and *Chic1*) in undifferentiated wild-type and ΔE ES cells, demonstrating that heterozygous deletion of the E-repeat does not interfere with X-linked gene expression in undifferentiated ES cells (*n* = 60, from one experiment). **b**, Representative epifluorescence images of cells counted in **a**. *Tsix*, the antisense transcript of *Xist*, was also detected here to identify both X chromosomes. Co-localized foci appear yellow. **c**, Histograms of nascent expression patterns of the X-linked genes *Gpc4* and *Atrx* in wild-type and ΔE cells displaying an *Xist*^MS2-coated X chromosome (*n* = 50), across 5 days of differentiation. These data were derived from an independent differentiation from that shown in Fig. 2c. **d**, Histograms of nascent expression patterns of indicated X-linked genes in wild-type and ΔE cells displaying an *Xist*^MS2-coated X chromosome (*n* = 50), across 7 days of differentiation derived from an independent differentiation from that shown in **c** and Fig. 2c. **e**, Histogram of nascent expression patterns of the X-linked gene *Tsix* in wild-type and ΔE cells across 5 days of differentiation. Note that these data were not scored relative to *Xist*^MS2 expression (that is, the monoallelic Tsix signal can be derived from either the *129* or *cas* allele (*n* = 70, except for the ΔE cells at day 5 with only 47 cells counted).
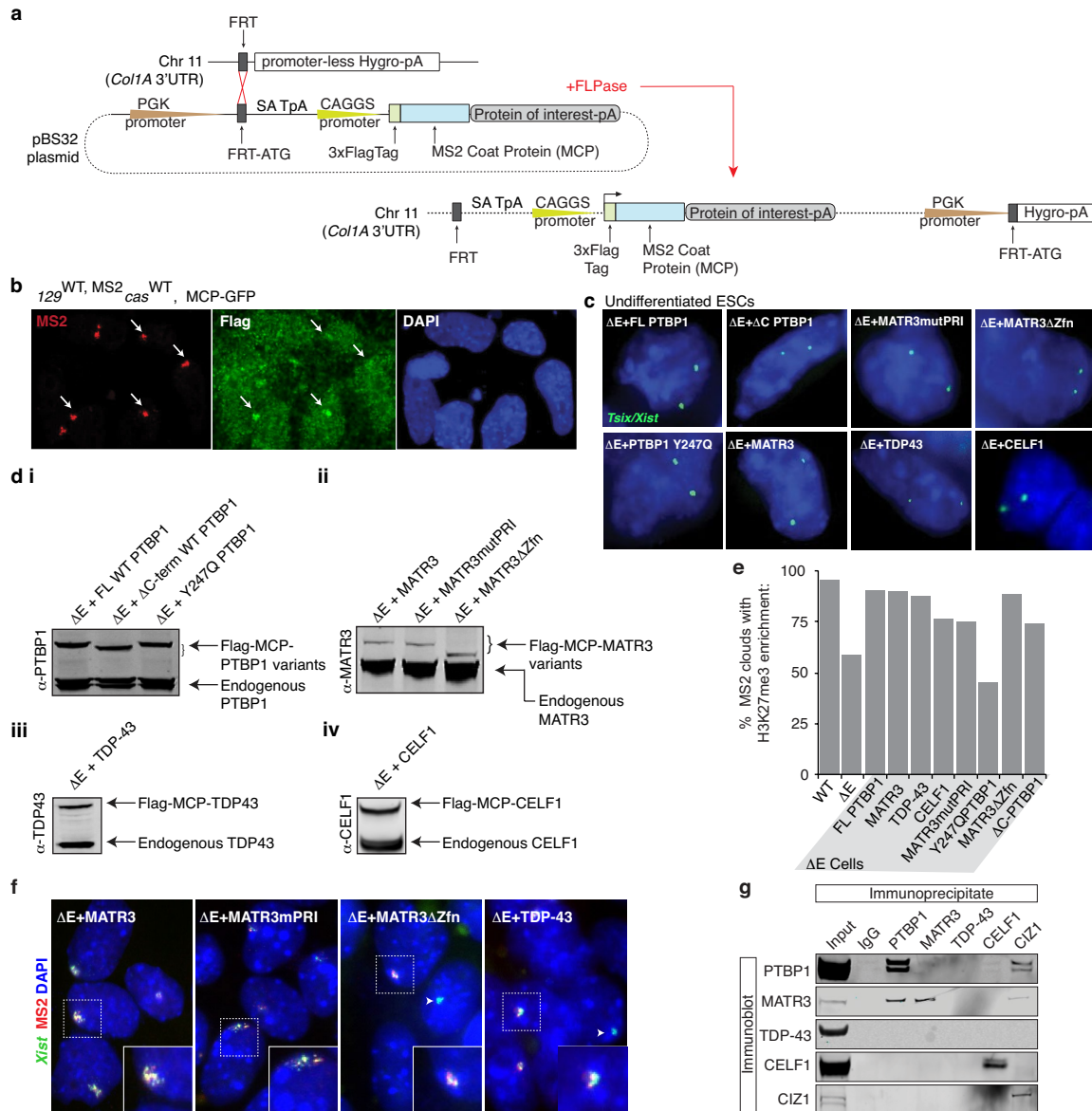
Extended Data Fig. 9 | A site-specific recombination-based approach to rescue phenotypes associated with loss of the E-repeat. a, Flp-In approach taken to constitutively express Flag-tagged MCP fusion proteins in ES cells (Methods). The Flag–MCP–GFP fusion protein was only expressed in wild-type ES cells. All other rescue constructs were expressed in ΔE ES cells. **b**, Flag–MCP–GFP fusion protein recruitment to *Xist*[MS2] in wild-type cells at differentiation day 7 shown with representative epifluorescence images. Arrows indicate MS2+*Xist*[129] clouds with co-localizing Flag–MCP–GFP enrichment. **c**, *Tsix* expression was used as a proxy to confirm presence of two X chromosomes in rescue ES cell lines. **d** (i), PTBP1-probed immunoblot on lysates from undifferentiated ΔE ES cells expressing full-length MCP–PTBP1 or MCP-PTBP1 mutants. (ii), As in **d** (i) except for MATR3 immunoblot for various MATR3 rescue lines. (iii), TDP-43-probed immunoblot on lysates from undifferentiated ΔE ES cells expressing MCP–TDP-43. (iv), CELF1-probed immunoblot on lysates from undifferentiated ΔE ES cells expressing MCP–CELF1. **e**, Histogram of the percentage of *Xist*[MS2] clouds that also show enrichment of H3K27me3 in wild-type or ΔE cells, or ΔE cells expressing the indicated MCP-fusion protein at differentiation day 7 ($n = 80$, from one experiment). **f**, Representative epifluorescence images of RNA FISH against *Xist* (green) and MS2 (red) in day 7 differentiated ΔE cell lines expressing the indicated variants of MCP fusion proteins. Inset, enlargement of the marked area. Arrowheads indicate wild-type *Xist* clouds in ΔE cells, derived from the *cas* allele. **g**, Immunoprecipitation of PTBP1, MATR3, CELF1, TDP-43 and CIZ1 from ES cell nuclear extracts (RNase treated) and detection of co-precipitated proteins with the same antibodies by immunoblotting (to accompany Fig. 3f). For images in **d** and **g**, see Supplementary Fig. 1 for source data.
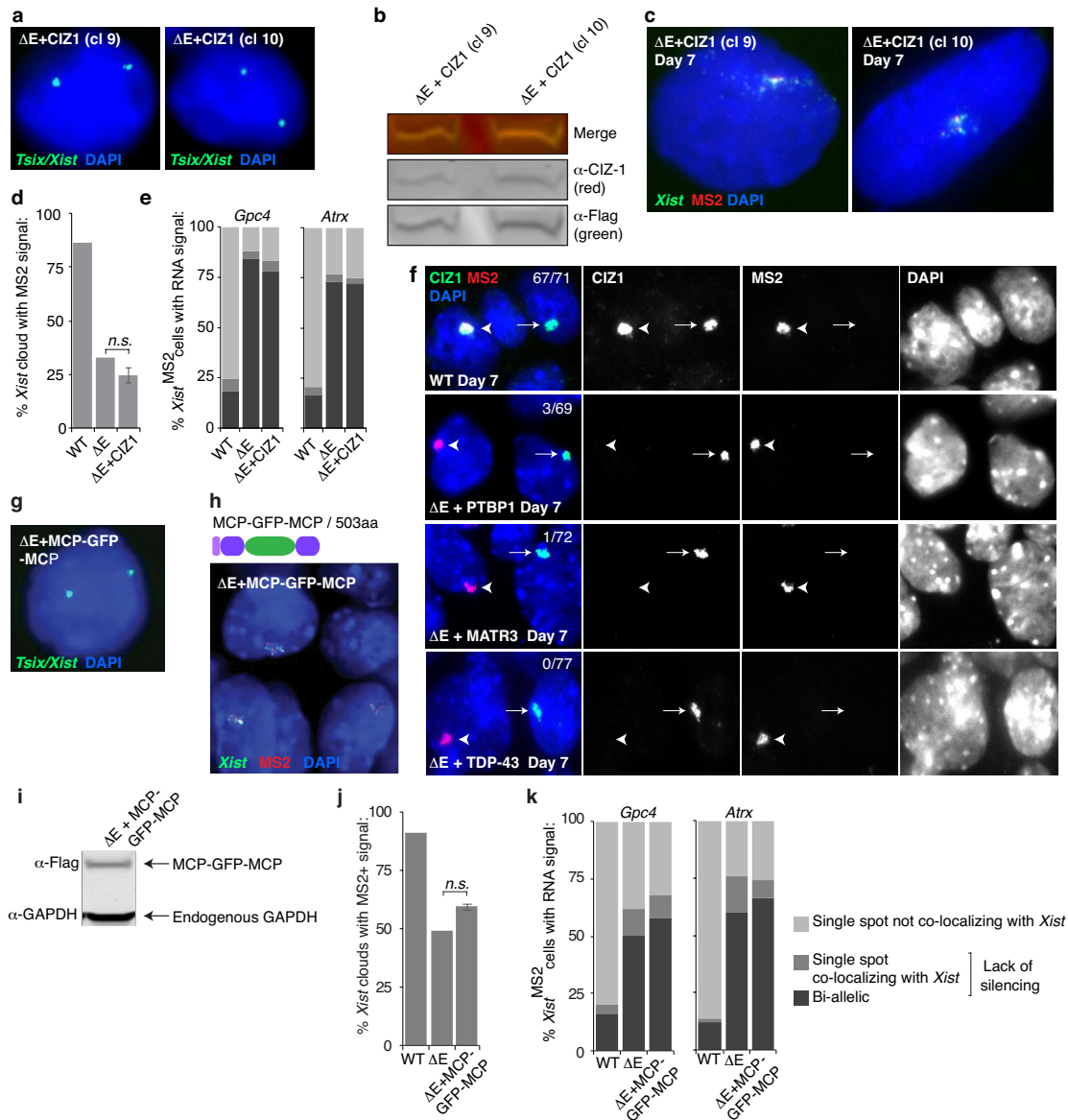
Extended Data Fig. 10 | Expression of MCP–CIZ1 or MCP–GFP–MCP does not rescue phenotypes due to loss of the E-repeat. a, RNA FISH images of *Tsix* transcripts for detection of two X chromosomes. Two ΔE MCP–CIZ1 ES cell clones (9 and 10) are shown. **b**, Immunoblot result for undifferentiated ΔE ES cell clones expressing MCP–CIZ1. **c**, Representative epifluorescence images of day 7 differentiated MCP–CIZ1-expressing ΔE clones, probed for *Xist* and MS2. **d**, Proportion of *Xist* clouds also displaying a co-localizing MS2 signal at differentiation day 7. The results for both CIZ1 rescue clones from one experiment were merged and the error bars represent s.e.m (*n* =120), *P*: not significant, two-tailed Student's *t*-test. **e**, Quantification of nascent *Gpc4* or *Atrx* expression patterns in wild-type, ΔE, or ΔE cells expressing MCP–CIZ1 (clone 9) displaying *Xist*^MS2 expression, at differentiation day 7 (*n* = 50, from one experiment). See **k** for legend. **f**, Representative epifluorescence images of in wild-type, ΔE or indicated ΔE rescue cell lines at differentiation day 7 immunostained for CIZ1 and probed for MS2. Arrowheads indicate rescued

cloud from the ΔE *Xist*^MS2 allele. Fraction of MS2+*Xist* clouds showing CIZ1 enrichment is given. **g**, RNA FISH images of *Tsix* transcripts in ΔE MCP–GFP–MCP ES cells to demonstrate the presence of two X chromosomes. **h**, Representative epifluorescence images of day 7 differentiated MCP–GFP–MCP-expressing ΔE ES cells probed for *Xist* and MS2, and illustration of Flag-tagged MCP–GFP–MCP fusion protein (see Fig. 3b for key). **i**, Immunoblot against the Flag-tag and GAPDH using lysates from undifferentiated MCP–GFP–MCP ΔE ES cells. **j**, Histogram showing the proportion of nuclei with *Xist* FISH signal that also displayed a co-localizing MS2 signal at differentiation day 7 for indicated cell lines (*n* =100); *P*: not significant, two-tailed Student's *t*-test, using 2 independent MCP–GFP–MCP expressing clones from one experiment. **k**, Quantification of nascent *Gpc4* or *Atrx* expression patterns in cells displaying *Xist*^MS2 expression at differentiation day 7 (*n* = 50, from one experiment). For images in **b** and **i**, see Supplementary Fig. 1 for source data.
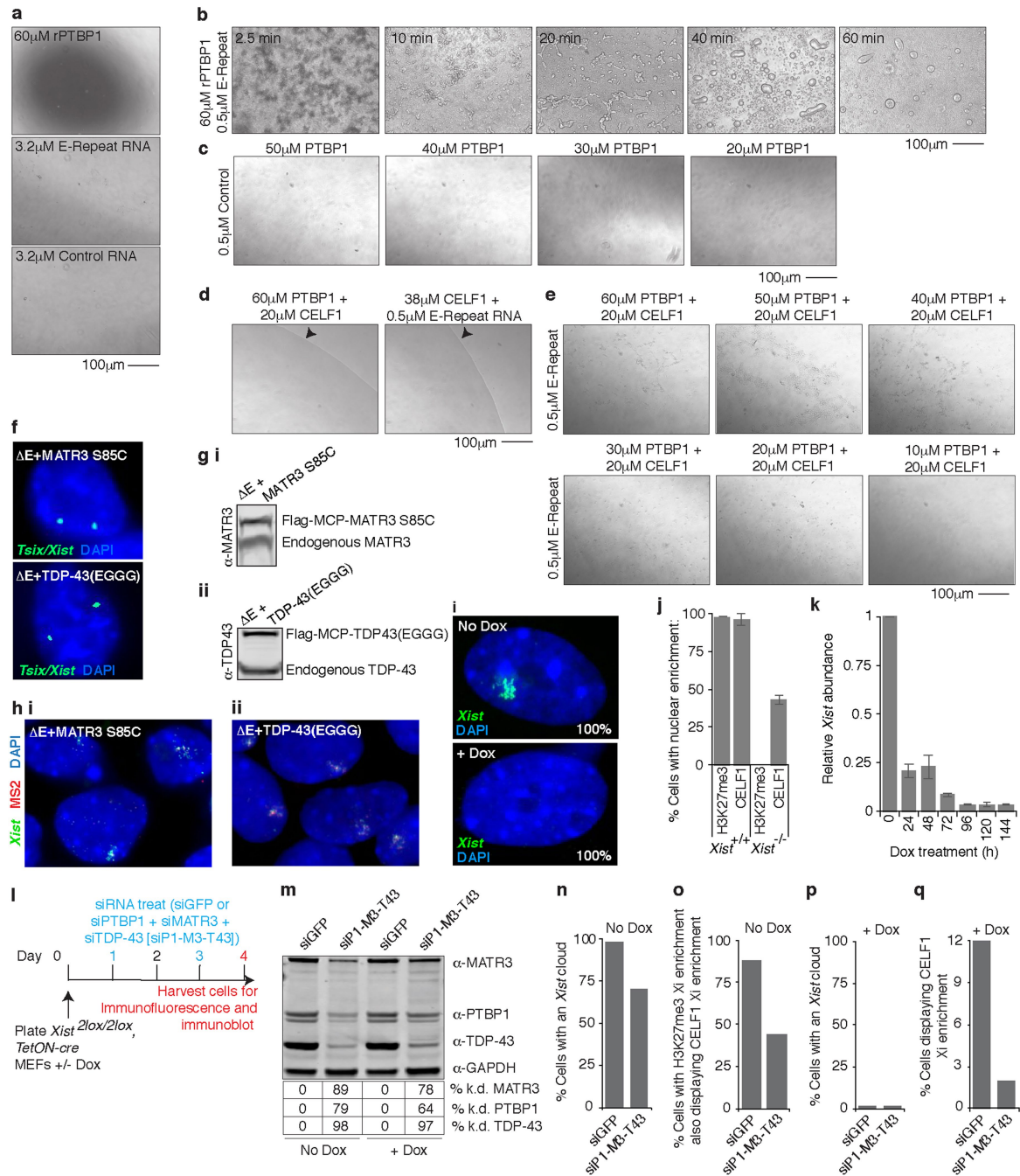
Extended Data Fig. 11 | CELF1 enhances droplet formation of PTBP1 with the E-repeat in vitro and mutations in MATR3 and TDP-43 that abrogate their self-association do not rescue ΔE phenotypes. a, Images showing lack of droplets with 60 µM rPTBP1, 3.2 µM E-repeat or control RNA at 40 min. **b**, Droplets formed from 60 µM rPTBP1 and 0.5 µM E-repeat RNA over time. **c**, Same as **a** except with 0.5 µM control RNA and different concentrations of rPTBP1 (40 min). **d**, Same as **a** except with 60 µM rPTBP1 and 20 µM rCELF1, or 38 µM rCELF1 with 0.5 µM E-Repeat RNA. Arrowheads indicate solution boundary with sample on left. **e**, Bright-field images showing aggregate-like formations of 20 µM rCELF1, 0.5 µM E-repeat RNA with varied concentrations of rPTBP1. **f**, RNA FISH images of *Tsix* transcripts in indicated ES cell lines to show presence of both X chromosomes. **g** (i), MATR3 immunoblot on extracts from ΔE ES cells expressing MCP−MATR3(S85C). (ii), TDP-43 immunoblot on ΔE ES cells expressing MCP−TDP-43(EGGG). **h**, Epifluorescence images of day 7 differentiated ΔE cells expressing MCP−MATR3(S85C) (i) or MCP−TDP-

43(EGGG) (ii) probed for *Xist* and MS2. **i**, MEFs (*Xist*[2lox/2lox], R26[M2rtTA/tetO-Cre]) probed for *Xist*, before or after dox treatment (96 h). Percentage of cells with displayed *Xist* pattern is given (*n* = 50, two biological replicates). **j**, Histogram showing percentage of MEFs with H3K27me3 or CELF1 Xi-enrichment under conditions described in **i**. Error bars represent s.e.m, (*n* = 50, from two biological replicates). **k**, Histogram showing relative *Xist* abundance over time of dox treatment for cells in **i** (see Fig. 4g.). **l**, Experimental schematic for knockdown experiment in **m**−**q**. **m**, Immunoblot showing knockdown of indicated factors in the experiment described in **l**. **n**, Percentage of MEFs (no dox) with an *Xist* cloud for indicated knockdowns (*n* = 50, from one experiment). **o**, Percentage of MEFs (no dox) with an Xi-enrichment of H3K27me3 that show a co-localizing accumulation of CELF1 (*n* = 50, from one experiment). **p**, Same as **n** except with dox treatment. **q**, Percentage of MEFs with CELF1 enrichment (*n* = 50, from one experiment). For images in **g** and **m**, see Supplementary Fig. 1 for source data.

nature research

Corresponding author(s):	Douglas Black and Kathrin Plath
Last updated by author(s):	May 20, 2020

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	No computer codes were used to collect the data used in this study. Codes used for analyses are listed in the methods and where necessary (if not publicly available) have been posted on GitHub.
Data analysis	Statistical tests (Kolmogrov-Smirnov tests, Pearsons r, Two-tailed students t-test) were performed using standard packages available in R (used r-studio for Mac). For CLIP, RAP and ChIP data, alignments were done using the publicly available BowTie package. Pre-alignment processing for CLIP data used publicly available packages: fastq-tools, fastx-toolkit, Samtools, Bedtools, DeepTools and UCSC scripts. Peak Calls for ChIP data was done using the publicly available package MACS2. Image analysis was performed on FIJI/ImageJ. This information has been included in the methods section of the paper.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Data generated during has been deposited in a public repository GEO. The account is currently private but the accession number is provided in manuscript and the

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	For image analyses, samples were analyzed in duplicate or triplicate (from independent experiments). Usually 50-100 units (Cells, Xist clouds) were counted per sample. This number was chosen based upon previous standards in the field and experimental measurements of the number of units needed to statistically test the observation under study.
Data exclusions	No data were excluded from this study.
Replication	1) For data assessed by microscope, experiments were repeated by the same (Amy Pandya-Jones), and different (Yolanda Markaki, Tsotne Chitiashvilli) researchers, with the same results. 2) For CLIP data, the experiments were independently repeated or matched previously published datasets from independent groups. 3) For all other experiments, reproducibility was established by having biological replicates or technical replicates (where biological replicates were not available).
Randomization	Not relevant to this study.
Blinding	For analysis of CLIP data, the data was blinded. Data was not blinded for all other experiments.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Antibodies

Antibodies used	The SI section of the submitted paper contains a table listing all this information
Validation	Most antibodies were validated using western blot. For those used in identification of the Inactive X-Chromosome, validation was done by Immunofluorescent analysis - with enrichment on the inactive X-Chromosome as a readout.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The parent F1 2-1 MS2 female ESC cell line was initially obtained from the lab of Rudoph Jaenisch where it was made, as was the v6.5 male parents ESC line. All genetic manipulations were performed by the authors as described in the methods section of the manuscript. These cell lines are not on the ICLAC cross-contamination list.
Authentication	Authentication of cell lines is not necessary as these cell lines are not on the ICLAC cross-contamination list. Genetic manipulations made for this study were confirmed by Southern blot and DNA sequencing as described in the methods section of the manuscript. In addition, cell lines in which synthetic or fusion proteins were expressed, authentication was

	performed by immunoblot.
Mycoplasma contamination	Parent cell lines were tested for Mycoplasma and declared negative. The derived cell lines were not tested for Mycoplasma.
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

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.*	Sequencing Data has been submitted to GEO under record # GSE137305. While it remains in private status the following token can be used for reviewer access: izozcqsinlgldst
Files in database submission	PTBP1_iclip_Xist.bw bigWig MATR3_iclip_Xist.bw bigWig CELF1_eclip_Xist.bw bigWig CELF1_input_Xist.bw bigWig PTBP1chip_NoDox_Xist.bw bigWig PTBP1chip_Dox_Xist.bw bigWig PTBP1_iclip_Xist.fastq.gz Fastq MATR3_iclip_Xist.fastq.gz Fastq CELF1_eclip_Xist.fastq.gz Fastq CELF1_input_Xist.fastq.gz Fastq PTBP1chip_NoDox_Xist.fastq.gz Fastq PTBP1chip_Dox_Xist.fastq.gz Fastq RAP +E: 162_v_SRR850637^smoothed.bw bigWig RAP -E: 5A1_v_SRR850637^smoothed.bw bigWig RAP +E: 162_R1.fastq.gz 162_R2.fastq.gz Fastq RAP -E: 5A1_R1.fastq.gz 5A1_R2.fastq.gz Fastq
Genome browser session (e.g. UCSC)	https://genome.ucsc.edu/s/PTBP1_MATR3_CELF1/PTBP1_MATR3_CELF1

Methodology

Replicates	ChIP data was reproduced twice with close agreement - both showing a peak across the E-repeat in Xist. PTBP1 and MATR3 CLIP experiments were performed once. PTBP1 CLIP results match published reports using the similar conditions and cell lines. CELF1 CLIP was performed twice, second replicate was in done under slightly different Xist induction conditions so not included in this study, but both replicates matched closely with strong a strong CELF peak across the 5' region of the Xist E repeat. RAP-seq experiments were performed once.
Sequencing depth	PTBP1 CLIP: 3.5M unique reads, 100bp single end MATRIN3 CLIP: 5.3M unique reads, 100bp single end CELF1 CLIP: 8.9M unique reads, 51bp single end CELF1 Input 5.7M unique reads, 51bp single end PTBP1 ChIP No Dox: 11.1M unique reads, 51bp single end PTBP1 ChIP Dox: 10.5M unique reads, 51bp single end RAP +E: ~20M reads, 41 and 33bp, paired end. RAP - E: ~20M reads, 41 and 33bp, paired end.
Antibodies	For PTBP1 CLIP and ChIP experiments, abcam ab5642 was used. For the MATRIN3 CLIP experiment Abcam ab151714, clone EPR10635(B) was used. For the CELF1 CLIP experiment Abcam ab129115, clone EPR8298(B) was used.
Peak calling parameters	ChIP peaks in Xist were called with MACS2 as described in the methods section of the manuscript. CLiP peaks were not called. RAP Peaks were not called.
Data quality	Data quality was controlled for using aligner specific parameters, including excluding reads shorter than 20 nt.
Software	Basecalls performed using CASAVA version 1.4, ChIP-seq, RAP-seq and CLIP-seq reads were aligned to the mm9 or mm10 genome assembly using Bowtie2 (Langmead et al. 2012) with only those reads that aligned to a unique position with no more than two sequence mismatches were retained for further analysis. CHIP-seq Peaks were called with MACS2. CHIP-, RAP- and CLIP-seq analysis is detailed in the methods.