10500–10517 *Nucleic Acids Research, 2020, Vol. 48, No. 18* *Published online 28 September 2020*
doi: 10.1093/nar/gkaa789

Elements at the 5′ end of *Xist* harbor SPEN-independent transcriptional antiterminator activity

Jackson B. Trotman [1,†], David M. Lee [1,2,†], Rachel E. Cherney [1,2], Susan O. Kim[1], Kaoru Inoue[1], Megan D. Schertzer [1,2], Steven R. Bischoff [3], Dale O. Cowley[3] and J. Mauro Calabrese [1,*]

[1]Department of Pharmacology and Lineberger Comprehensive Cancer Center, Chapel Hill, NC 27599, USA, [2]Curriculum in Genetics and Molecular Biology, Chapel Hill, NC 27599, USA and [3]Animal Models Core, University of North Carolina at Chapel Hill, 120 Mason Farm Road, Chapel Hill, NC 27599, USA

Received April 23, 2020; Revised August 20, 2020; Editorial Decision September 06, 2020; Accepted September 12, 2020

ABSTRACT

The *Xist* lncRNA requires Repeat A, a conserved RNA element located in its 5′ end, to induce gene silencing during X-chromosome inactivation. Intriguingly, Repeat A is also required for production of *Xist*. While silencing by Repeat A requires the protein SPEN, how Repeat A promotes *Xist* production remains unclear. We report that in mouse embryonic stem cells, expression of a transgene comprising the first two kilobases of *Xist* (*Xist*-2kb) causes transcriptional readthrough of downstream polyadenylation sequences. Readthrough required Repeat A and the ∼750 nucleotides downstream, did not require SPEN, and was attenuated by splicing. Despite associating with SPEN and chromatin, *Xist*-2kb did not robustly silence transcription, whereas a 5.5-kb *Xist* transgene robustly silenced transcription and read through its polyadenylation sequence. Longer, spliced *Xist* transgenes also induced robust silencing yet terminated efficiently. Thus, in contexts examined here, *Xist* requires sequence elements beyond its first two kilobases to robustly silence transcription, and the 5′ end of *Xist* harbors SPEN-independent transcriptional antiterminator activity that can repress proximal cleavage and polyadenylation. In endogenous contexts, this antiterminator activity may help produce full-length *Xist* RNA while rendering the *Xist* locus resistant to silencing by the same repressive complexes that the lncRNA recruits to other genes.

INTRODUCTION

The long noncoding RNA (lncRNA) *Xist* functions *in cis* to silence nearly all genes along the 165-megabase (Mb) X chromosome as part of X-chromosome inactivation (XCI), the dosage compensation process that occurs early during the development of eutherian mammals. At the level of sequence composition, *Xist* is notable for the presence of several internal domains of tandem repeats, which each help *Xist* achieve its repressive function by recruiting distinct subsets of RNA-binding proteins (1–3). One such repeat is found in the first thousand nucleotides of *Xist* and is called 'Repeat A'. Repeat A consists of eight to nine tandemly arrayed, 50-nucleotide-long repeating elements that each harbor a degenerate U-rich region followed by a distinctive GC-rich region that is highly conserved among eutherians (2–5).

Gene silencing induced by *Xist* has been shown to depend on Repeat A as well as the Repeat-A-binding protein SPEN. Recruitment of SPEN to the X chromosome via Repeat A likely induces gene silencing by locally activating and/or recruiting various corepressor and histone deacetylase complexes (6–9). Concordantly, deletion of Repeat A or SPEN each results in failure of XCI (5–14), and robust associations between SPEN and Repeat A have been detected *in vivo* (6,15,16). *In vitro*, SPEN binds to single-stranded regions of Repeat A that are located directly adjacent to its structured, GC-rich segments (12,17).

*To whom correspondence should be addressed. Tel: 919 843 3257; Email; jmcalabr@med.unc.edu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Present addresses:
Susan O. Kim, Biochemistry, Cell and Developmental Biology Graduate Program, Emory University, Atlanta, GA 30322, USA.
Kaoru Inoue, National Institute for Environmental Health Sciences, Research Triangle Park, NC 27709, USA.
Megan D. Schertzer, New York Genome Center, New York, NY 10013, USA.

Although Repeat A and SPEN are both necessary for XCI, *Xist* transgenes that contain Repeat A but lack essentially all other *Xist* sequence downstream are unable to induce chromosome-level silencing, implying either (i) that SPEN requires sequence elements in addition to Repeat A to associate with *Xist* or (ii) that *Xist* sequence elements in addition to a SPEN-bound Repeat A are needed to induce chromosome-level silencing (4,5,18).

Moreover, despite clear links between SPEN and Repeat A, questions remain regarding the mechanism through which Repeat A carries out its functions. Most notably, deletion of Repeat A from the endogenous *Xist* locus causes not only a failure of XCI but a dramatic reduction in the abundance of full-length, spliced *Xist* (13,14,19,20). Similar reductions in *Xist* RNA abundance have been observed upon Repeat A deletion in transgenic contexts (11,18,21–23), seemingly due to a defect in the production of nascent *Xist* and not due to a reduction in RNA stability (22). Current evidence is at odds as to whether the loss of SPEN reduces or has no impact on *Xist* RNA abundance (9,11). Similarly, a seminal study found that the repressive function of an *Xist* transgene lacking Repeat A can be restored by appending Repeat A to its 3′ end, suggesting that the function of Repeat A is independent of its position within *Xist* (5). In contrast, a recent report found loss of *Xist* abundance and *Xist*-induced silencing when Repeat A was moved from the 5′ end of *Xist* to positions further downstream, suggesting the opposite conclusion—that in addition to the sequence of Repeat A, its 5′-proximal position may be important for the production and function of *Xist* (24). The basis for this apparent difference is unclear.

We recently developed a transgenic assay that recapitulates certain aspects of Repeat-A-dependent gene silencing, which we called TETRIS (Transposable Element to Test RNA's effect on transcription *in cis*; (25)). In TETRIS, a doxycycline-inducible transgene that comprises the first two kilobases (kb) of *Xist* and contains Repeat A (*Xist*-2kb) is positioned in convergent orientation relative to a luciferase reporter gene in the context of a piggyBac cargo plasmid. After insertion of the plasmid into chromatin via the piggyBac transposase, induction of *Xist*-2kb expression results in an 80–90% reduction of luciferase activity relative to uninduced cells. We demonstrated that repression of luciferase activity by *Xist*-2kb in this assay requires the same sequence motifs within Repeat A that are required for transcriptional repression by full-length *Xist*—its GC-rich portions but not its U-rich spacer sequences—as well an adjacent region that harbors structured elements and intervening sequence (5,25,26).

Given the unresolved questions surrounding Repeat A, we sought to investigate the mechanism of repression induced by *Xist*-2kb in the TETRIS assay as well as in a transgenic, single-copy insertion assay that is analogous to assays previously employed to identify seminal aspects of *Xist* biology (4–12,18,21,26,27). Through a circuitous series of experiments, we unexpectedly found that in these transgenic contexts, Repeat A and nearby sequence harbor transcriptional antiterminator activity that does not require the silencing cofactor SPEN. Furthermore, despite associating with SPEN and with chromatin, RNA produced from the *Xist*-2kb transgene was unable to induce robust local or long-distance transcriptional silencing. Instead, long-distance silencing required synergy between elements located in the first two kilobases of *Xist* and regions downstream. Our findings underscore a surprising level of interdependency between domains within *Xist,* highlight a new potential function of Repeat A, and hint at the existence of a broader class of RNA elements that may function to promote transcriptional processivity in mammals.

MATERIALS AND METHODS

Embryonic stem cell culture

E14 mouse embryonic stem cells (ESCs; kind gift of D. Ciavatta) were cultured at 37°C in a humidified incubator at 5% CO_2. Medium was changed daily and consisted of DMEM high glucose plus sodium pyruvate, 15% ESC qualified fetal bovine serum, 0.1 mM non-essential amino acids, 100 U/ml penicillin–streptomycin, 2 mM L-glutamine, 0.1 mM 2-mercaptoethanol, and 1:500 LIF conditioned media produced from Lif-1Cα (COS) cells (kind gift of N. Hathaway). ESCs were split at an approximate ratio of 1:6 every 48 h. *Rosa26*-RMCE cells were grown on gamma-irradiated mouse embryonic fibroblast (MEF) feeder cells plated at ∼1.5 million cells per 10-cm plate. Prior to harvesting of RNA for sequencing, *Rosa26*-RMCE cells were passaged twice off of MEF feeder cells with a 40-min pre-plate each passage and grown in 70% MEF-conditioned medium supplemented as above.

TETRIS line generation

TETRIS lines were made as described in (25). Briefly, 500,000 E14 cells were seeded in a single well of a six-well plate and transfected 24 h later with 0.5 μg TETRIS cargo plasmid, 0.5 μg *rtTA*-cargo plasmid, and 1 μg of pUC19-piggyBAC transposase plasmid using Lipofectamine 3000 (Invitrogen) according to manufacturer instructions. Cells were selected for 7–9 days with puromycin (2 μg/ml) and G418 (200 μg/ml) beginning 24 h after transfection.

TETRIS luminescence assays

For each independent TETRIS cell line, six wells of a 24-well plate were seeded at 100,000 cells per well. Three of the six wells were induced with 1 μg/ml doxycycline (Sigma) beginning when the cells were plated; the remaining three wells served as 'no dox' controls. After 48 h, the cells were washed with PBS and lysed with 100 μl of passive lysis buffer (Promega) and luciferase activity was measured using Bright-Glo Luciferase Assay reagents (Promega) on a PHERAstar FS plate reader (BMG Labtech). Luciferase activity was normalized to total protein concentration in the lysates measured via Bradford assay (Bio-Rad). See Supplementary Table S1 for additional information regarding experimental replicates.

RNA isolation, subcellular fractionation, and RT-qPCR

RNA was isolated using Trizol according to manufacturer protocol (Invitrogen). Subcellular fractionation was performed essentially as in (28,29). For RT-qPCR assays, equal

amounts of RNA (0.5–1 μg) were reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) with a pool of random primers or a single, strand-specific primer (Supplementary Table S2). qPCR was performed using iTaq Universal SYBR Green (Bio-Rad) and custom primers (Supplementary Table S2) on a Bio-Rad CFX96 system with the following thermocycling parameters: initial denaturation at 95°C for 10 min; 40 cycles of 95°C for 15 s, 60°C for 30 s, and 72°C for 30 s followed by a plate read. Data were normalized (usually to the average of the 'no dox' control measurements) using Microsoft Excel and plotted using GraphPad Prism 8. See Supplementary Table S1 for information regarding experimental replicates. Complete details can be found in the Supplementary Methods.

Stellaris single-molecule sensitivity RNA FISH

Custom Stellaris FISH probes were designed against the first 2 kb of *Xist*, firefly luciferase (*luc2* in pGL4.10; Promega), and the hygromycin resistance gene (*HygroR*) using the Stellaris RNA FISH Probe Designer (Biosearch Technologies, Inc.) and labeled with Quasar 670 (*Xist*-2kb) or 570 (*Luc* and *HygroR*) dye. FISH was performed as described in (30). Cells were grown for 2 days on glass coverslips before fixing with 4% formaldehyde and permeabilizing overnight with 75% ethanol at 4°C. Coverslips were washed briefly and then incubated overnight at 37°C with the labeled probes. Coverslips were washed twice, with the second wash containing DAPI. After mounting and curing coverslips with Prolong Gold or Vectashield, multi-channel Z-stack images were obtained with an Olympus BX61 wide-field fluorescence microscope. The images shown are maximum intensity projections produced with ImageJ. Complete details can be found in the Supplementary Methods.

Northern blots

Northern blots were performed using a protocol adapted from (31). Thermo Fisher Millennium RNA Markers and samples containing 15 μg RNA were run on 1% agarose gels containing formaldehyde (Thermo Fisher NorthernMax Denaturing Gel Buffer) and ethidium bromide. After washing the gel twice each with water and then 10× SSC, RNA was transferred to a GE Healthcare Hybond-N membrane via capillary transfer overnight. The transferred RNA was crosslinked to the membrane with UV light and imaged (ethidium signal) with a Bio-Rad ChemiDoc MP Imaging System in Flamingo mode. The membrane was incubated at 42°C overnight in Thermo Fisher ULTRAhyb-Oligo containing a ^{32}P-radiolabeled single-stranded DNA oligo probe (see Supplementary Table S2 for sequences). The membrane was washed and then exposed to a storage phosphor screen overnight before acquiring images with a Amersham Typhoon 5 Biomolecular Imager. Complete details can be found in the Supplementary Methods.

RNA immunoprecipitation

RNA immunoprecipitation experiments were performed using a protocol from (32). Mouse ES cells were fixed with 0.3% formaldehyde for 30 min, quenched with 125 mM glycine, washed with PBS, and snap-frozen in liquid nitrogen as 10-million-cell aliquots to store at −80°C. Fixed cell pellets were thawed on ice, resuspended in RIPA buffer, and sonicated twice for 30 s using a Sonics Vibra-Cell VX130. A small portion of each sonicated and centrifuge-cleared lysate was stored at −20°C for processing as input. The remaining lysate was added to Santa-Cruz Protein A/G agarose beads pre-conjugated with appropriate antibody, and the mixtures were incubated end-over-end at 4°C overnight. Beads were washed once in fRIP Buffer, three times in ChIP Buffer, once in High-Salt Buffer, and once in LiCl Wash Buffer (all at 4°C; see Supplementary Methods for composition of buffers). After washing, beads and input samples were brought to the same volume in a solution containing proteinase K, and to remove protein-RNA crosslinks, samples were incubated 1 h at 42°C, 1 h at 55°C, and 30 min at 65°C. RNA was purified from the aqueous phase of Trizol extractions using the Zymo RNA Clean & Concentrator-5 kit with on-column DNase I digestion. Complete details can be found in the Supplementary Methods.

Recombinase-mediated cassette exchange (RMCE)

A male F1-hybrid mouse ESC line (derived from a cross between C57BL/6J (B6) and CAST/EiJ (Cast) mice; kind gift of T. Magnuson) was made competent for RMCE by insertion of a custom homing cassette into the *Rosa26* locus on chromosome 6 via homologous recombination. *Xist* transgenes were cloned via PCR or recombineering into a custom RMCE-cargo vector and then electroporated along with a plasmid expressing Cre-recombinase into RMCE-competent cells using a Neon Transfection System (Invitrogen). Individual colonies were picked and genotyped, then rendered doxycycline-sensitive using the *rtTA*-expression cassette from (33). Complete details can be found in the Supplementary Methods.

RNA sequencing and analysis

RNA-seq libraries were prepared using the RNA HyperPrep Kit with RiboErase (Kapa Biosciences) and sequenced on an Illumina NextSeq 500 machine using a 75-cycle high output NextSeq kit (Illumina). Sequencing reads were aligned and processed essentially as in (34,35). Differential expression analysis was performed with DESeq2 (36). Complete details can be found in the Supplementary Methods.

RESULTS

Xist-2kb represses expression of adjacent genes by reading through polyadenylation sequences

TETRIS is a transgenic assay that allows the sequence of a lncRNA to be manipulated in a plasmid and then tested for its ability to repress adjacent reporter genes in a chromosomal context. The assay employs the piggyBac transposase to insert a cassette containing a doxycycline-inducible lncRNA gene, a luciferase reporter gene (*Luc*), and a puromycin resistance gene (*PuroR*) into the genomes of transfected cells (Figure 1A, Supplementary Figure

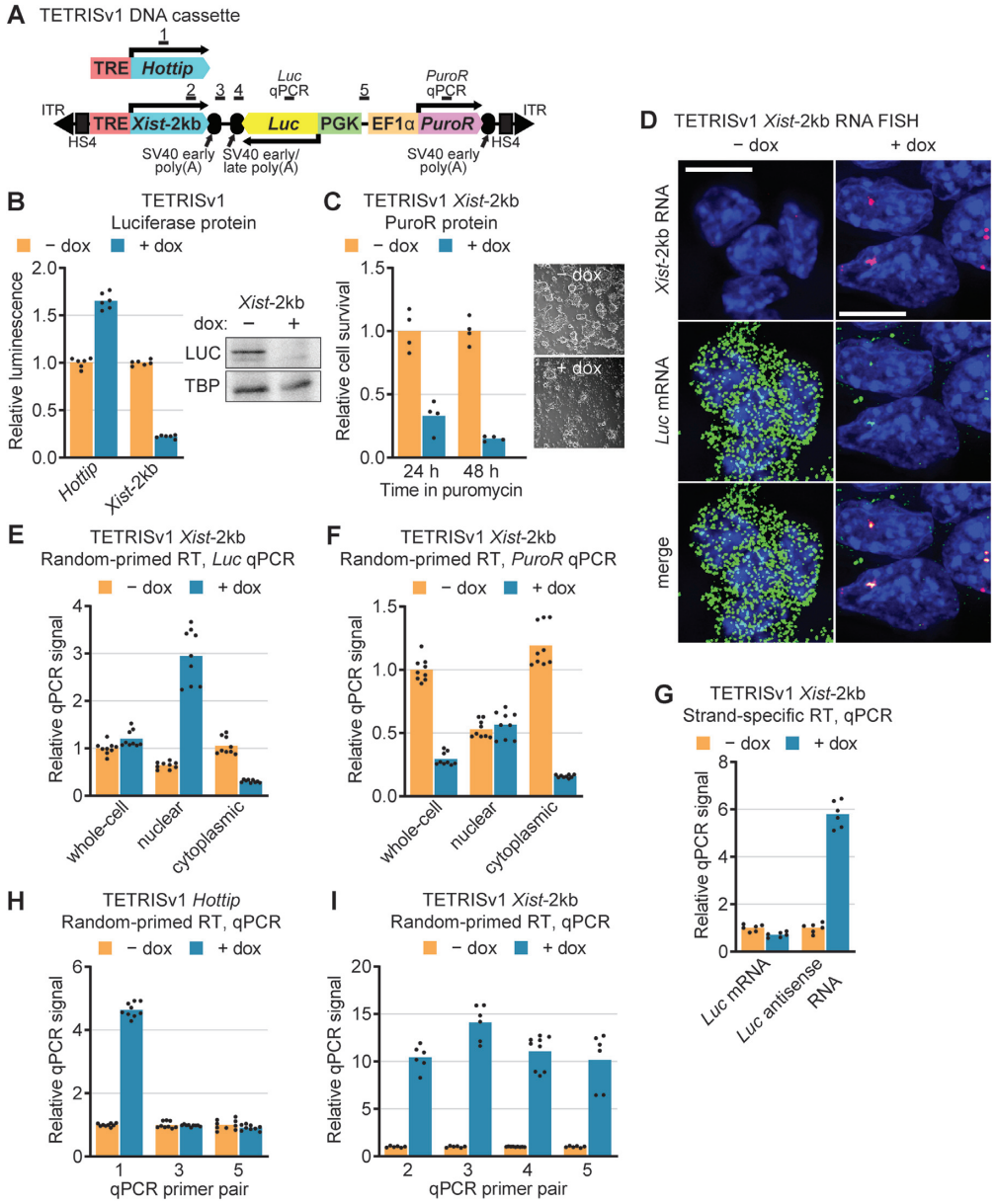

Figure 1. *Xist*-2kb represses expression of adjacent genes by reading through polyadenylation sequences. (**A**) Schematics of TETRISv1 DNA cassettes. ITR, inverted tandem repeat recognized by piggyBac transposase; HS4, chicken β-globin insulator sequence; TRE, tetracycline response element (doxycycline-inducible promoter); *Xist*-2kb, nucleotides 1–2016 of mouse *Xist*; SV40, simian virus 40 (polyadenylation sequences); *Luc*, firefly luciferase gene; PGK and EF1α, constitutive promoters; *PuroR*, puromycin resistance gene. Numbered bars indicate regions targeted by qPCR primer pairs. (**B**) Luminescence assay (left) and western blot (right) showing effects of *Hottip* and *Xist*-2kb induction on luciferase protein expression. (**C**) Relative survival of TETRISv1 *Xist*-2kb cells grown for 24 or 48 h in the presence of puromycin after 48 h of treatment with or without doxycycline. Representative images show differences in cell survival after 48 h in puromycin. (**D**) Stellaris single-molecule, strand-specific RNA FISH shows *Xist* RNA (red) and *Luc* mRNA (green) in TETRISv1 *Xist*-2kb cells. DAPI-stained nuclei are blue. Scale bar = 10 μm. (**E**, **F**) RNA was prepared from TETRISv1 *Xist*-2kb cells directly (whole-cell) or following subcellular fractionation. Reverse transcription was performed with a pool of random oligos, and quantities of RNA corresponding to the *Luc* (E) or *PuroR* (F) gene in each sample were measured via qPCR and normalized to the whole-cell – dox value. (**G**) Whole-cell RNA from TETRISv1 *Xist*-2kb cells was reverse-transcribed using strand-specific primers targeting either the *Luc* mRNA or RNA produced from transcription occurring in the opposite direction (*Luc* antisense RNA). Quantities of RNA in each sample were measured with the same qPCR primer pair (relative to the – dox average for each strand-specific RT). (**H**) Nuclear RNA from TETRISv1 *Hottip* cells was reverse transcribed with random primers, and doxycycline-dependent changes in RNA produced from the *Hottip* gene and from the *Hottip-Luc* and *Luc-PuroR* intergenic regions were measured via qPCR with the primer pairs shown in (A). (**I**) Analysis as in (H) but with nuclear RNA from TETRISv1 *Xist*-2kb cells. For all panels, cells were treated with or without 1 μg/ml doxycycline for 48 h prior to harvesting. Unless stated otherwise, numerical values are shown relative to the – dox value for each cell identity or qPCR primer pair, which is set to one. Dots represent individual technical replicate measurements from a minimum of biological duplicate experiments, and bars represent the average value. See Supplementary Figure S2 for RT-qPCR control data. See Supplementary Table S1 for information regarding experimental replicates, Supplementary Table S2 for oligo sequences, and Supplementary Table S3 for raw RT-qPCR data.

S1A). Expression of non-repressive lncRNAs in TETRIS, such as *Hottip* (Figure 1B), typically causes a ∼1.5-fold increase in luciferase activity, a mild enhancer effect that we attribute to the proximity of the doxycycline-inducible TRE promoter and the *PGK* promoter that drives *Luc* gene expression (25). In contrast, induced expression of the first 2 kb of *Xist* (*Xist*-2kb) in TETRIS causes an 80 to 90% reduction of luciferase activity, which we confirmed by observing a loss of luciferase protein by western blot (Figure 1B). We also found that *Xist*-2kb induction repressed *PuroR* expression, impairing cell survival in the presence of puromycin (Figure 1C). Previously, we found that silencing by *Xist*-2kb depends on Repeat A and an additional ∼750 nucleotides of sequence located just downstream, implying that the mechanisms of *Xist*-2kb-induced repression in TETRIS are related to those employed by *Xist* in endogenous contexts (5,25).

We next performed controls to verify whether expression of *Xist*-2kb caused a level of transcriptional silencing commensurate with the observed reduction in luciferase protein levels. Indeed, we found via single-molecule sensitivity RNA FISH that, upon *Xist*-2kb expression, the number of luciferase mRNA molecules was greatly reduced throughout the cell (Figure 1D). We next analyzed RNA in subcellular fractions via reverse transcription (RT) with random primers followed by quantitative PCR (qPCR). In response to *Xist*-2kb expression, cytoplasmic qPCR signal for the *Luc* gene was reduced, in agreement with strand-specific RNA FISH (Figure 1D and E). However, in response to *Xist*-2kb expression, nuclear qPCR signal for the *Luc* gene increased, as if *Luc* had not been transcriptionally silenced (Figure 1E). Likewise, in response to *Xist*-2kb expression, cytoplasmic qPCR signal from the *PuroR* gene decreased while nuclear *PuroR* signal remained relatively constant, as if that gene too had not been silenced (Figure 1F).

To reconcile these seemingly conflicting results, we examined the possibility that transcription from the *Xist*-2kb transgene failed to terminate at its specified polyadenylation sequence, causing readthrough into the downstream *Luc* and *PuroR* genes. Transcriptional readthrough would be indistinguishable from genic transcription as measured by the random-primed RT-qPCR assays of Figure 1E and F and could explain the relative increase in nuclear RT-qPCR signal over the *Luc* and *PuroR* genes induced by expression of *Xist*-2kb. To test this hypothesis, we performed strand-specific RT using primers targeting either the *Luc* mRNA or the putative readthrough transcription product. Subsequent qPCR confirmed that *Xist*-2kb expression led to readthrough transcription over the body of the *Luc* gene (Figure 1G).

To further examine transcriptional readthrough in the TETRIS assay, we used random-primed RT and a series of qPCR primer pairs targeting multiple sites within the TETRIS cassette, including primer pairs located downstream of *Xist*-2kb's SV40 early polyadenylation sequence (numbering per Figure 1A). Expression of the control lncRNA *Hottip* caused no change in qPCR signal downstream of the SV40 early polyadenylation sequence (Figure 1H). Importantly, this *Hottip* control lacked an endogenous polyadenylation sequence, underscoring that the SV40 early polyadenylation sequence in the TETRIS construct was indeed functional. On the other hand, expression of *Xist*-2kb caused a 10-to-15-fold increase in qPCR signal between the lncRNA gene and the *Luc* gene (primer pairs 3 and 4) that was commensurate to the level of *Xist*-2kb induction (primer pair 2), consistent with robust transcriptional readthrough of the SV40 early polyadenylation sequence (Figure 1I). Additional qPCR targeting the region between the *Luc* and *PuroR* genes (primer pair 5) suggests that readthrough transcription originating from *Xist*-2kb extends through *Luc* and into the downstream *PuroR* gene (Figure 1I). Remarkably, in order for this latter scenario to occur, transcription originating from the *Xist*-2kb transgene would have had to continue through a second SV40 early polyadenylation sequence, which is encoded by the reverse complement of the SV40 late polyadenylation sequence that terminates transcription of the *Luc* gene (see (37); Figure 1A). Sanger sequencing confirmed that the polyadenylation sequences in the TETRIS *Xist*-2kb and *Hottip* cargo plasmids were identical (Supplementary Figure S1B). Thus, in the context of TETRIS, transcription of *Xist*-2kb appears to extend well beyond its expected polyadenylation site.

Xist-2kb readthrough transcription occurs beyond multiple different polyadenylation sequences and is inhibited by efficient splicing

We next examined whether transcriptional readthrough beyond *Xist*-2kb would occur in a second version of TETRIS (TETRISv2) in which two consecutive polyadenylation sequences (SV40 early and bovine growth hormone [BGH]) were placed downstream of *Xist*-2kb, followed by a 1.1-kb spacer that separates these polyadenylation sequences from the SV40 early/late polyadenylation sequence of the convergently oriented *Luc* gene (Figure 2A, Supplementary Figure S1A). As with the original version of TETRIS, we observed that in TETRISv2, induction of *Xist*-2kb—but not an empty vector control—reduced luciferase activity by ∼90% (Figure 2B). Even in this second vector, upon *Xist*-2kb induction, we observed robust RT-qPCR signal in the region between *Xist*-2kb and *Luc* (primer pairs 6 and 7) and in the region between *Luc* and *PuroR* (primer pair 5), which is located approximately 6.3 kb downstream of the transcription start site of the *Xist*-2kb gene (Figure 2C). Therefore, for *Xist*-2kb, transcription appeared to read through three near-consecutive polyadenylation sequences: SV40 early, BGH, and a second SV40 early sequence at the 3′ end of the *Luc* gene.

Next, we investigated whether *Xist*-2kb would promote readthrough of an endogenous polyadenylation sequence. To do so, we replaced the SV40 early and BGH polyadenylation sequences in TETRISv2 with the natural polyadenylation sequence found at the annotated 3′ end of the mouse *Xist* gene (TETRISv2.1, Figure 2D). This *Xist* polyadenylation sequence was effective in terminating transcription of a highly expressed, intronless *GFP* transgene (Figure 2E). However, with *Xist*-2kb, we still observed robust transcriptional readthrough beyond the *Xist* and early/late SV40 polyadenylation sequences (Figure 2F). Thus, *Xist*-

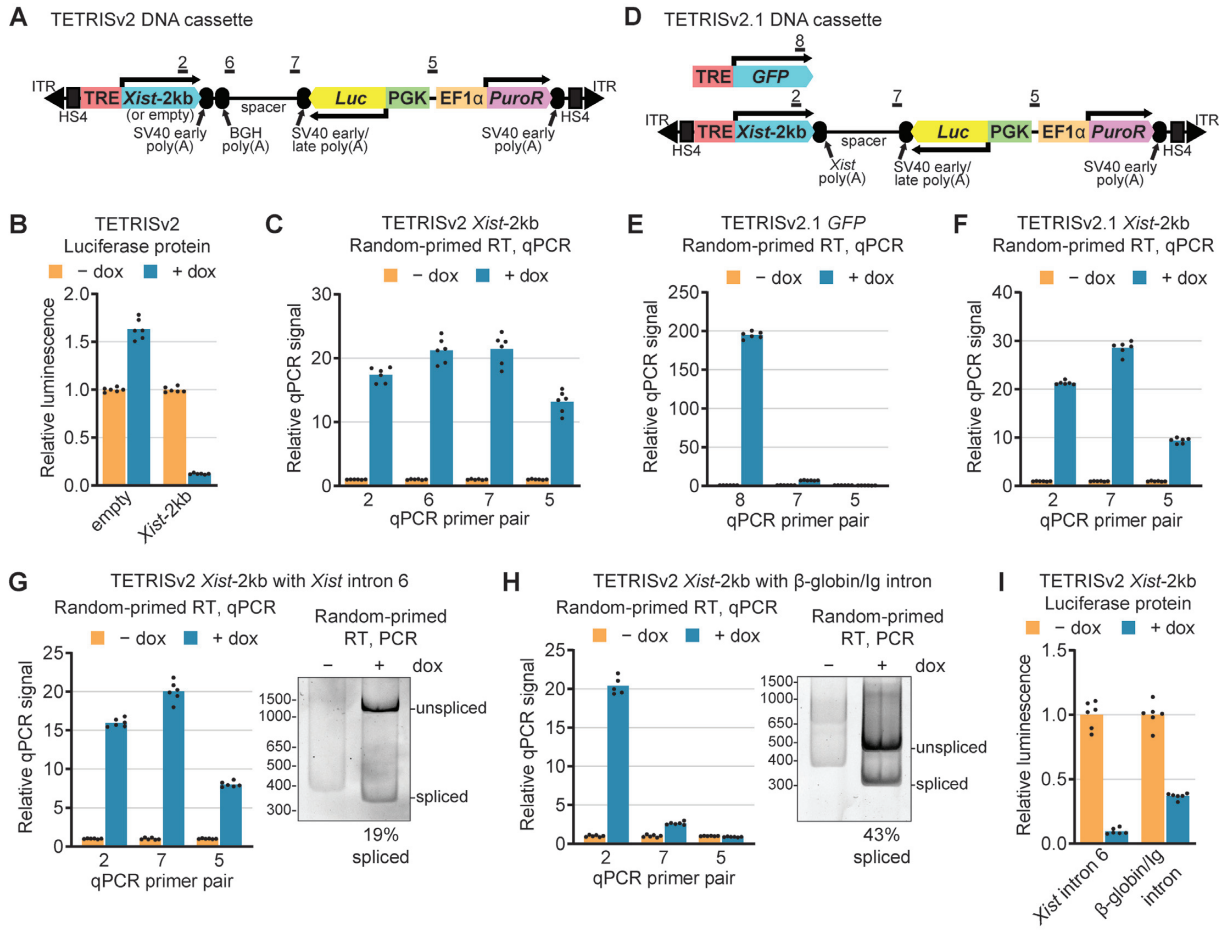

Figure 2. *Xist*-2kb readthrough transcription occurs beyond multiple different polyadenylation sequences and is inhibited by efficient splicing. (**A**) Diagram of the TETRISv2 DNA cassette and regions targeted by qPCR primer pairs. (**B**) Luminescence assay with TETRISv2 empty (no lncRNA cargo) or *Xist*-2kb cells. (**C**) Random-primed RT-qPCR with nuclear RNA from TETRISv2 *Xist*-2kb cells showing induction of readthrough transcripts beyond the polyadenylation sites. (**D**) Diagram of the TETRISv2.1 DNA cassette and regions targeted by qPCR primer pairs. (**E**, **F**) Random-primed RT-qPCR with whole-cell RNA from TETRISv2.1 GFP (E) and *Xist*-2kb (F) cells showing induced levels of readthrough transcript RNA beyond the polyadenylation sites. (**G**) Random-primed RT-qPCR measuring induced transcriptional readthrough (left) and intron-flanking RT-PCR to estimate splicing efficiency (right) of *Xist* intron 6 inserted in *Xist*-2kb of TETRISv2 (between nucleotides G1919 and G1920 of *Xist*-2kb, leaving a 367-nucleotide final exon before the expected cleavage site within the SV40 early polyadenylation sequence). Both assays were performed using whole-cell RNA. (**H**) Same as (G) but for TETRISv2 *Xist*-2kb cells with a chimeric β-globin/immunoglobulin intron inserted in the same site of *Xist*-2kb (40). (**I**) Luminescence assay with TETRISv2 *Xist*-2kb-intron cells from (G) and (H). For all panels, cells were treated with or without 1 μg/ml doxycycline for 48 h prior to harvesting. Numerical values are shown relative to the – dox average for each cell identity or qPCR primer pair, which is set to one. Dots represent individual technical replicate measurements from a minimum of biological duplicate experiments, and bars represent the average value. See Supplementary Figure S2 for RT-qPCR control data. See Supplementary Table S1 for information regarding experimental replicates, Supplementary Table S2 for oligo sequences, and Supplementary Table S3 for raw RT-qPCR data.

2kb promotes transcriptional readthrough of many different polyadenylation sequences, including one found at the 3′ end of the endogenous *Xist* gene.

Long-read sequencing has recently revealed that transcriptional readthrough of endogenous polyadenylation sites occurs frequently for a subset of transcripts that fail to undergo splicing (38,39). Because our *Xist*-2kb transgene lacks any introns, we sought to determine whether transcriptional readthrough would occur upon insertion of an intron in *Xist*-2kb. We created two separate TETRISv2 constructs where the same site in *Xist*-2kb was modified to include either the final 781-bp intron of mouse *Xist* or a 133-bp chimeric β-globin/immunoglobulin intron shown previously to be spliced out efficiently in transgenic contexts (40). RT-PCR with primers flanking the inserted introns showed that the chimeric β-globin/immunoglobulin intron was spliced out more efficiently than the *Xist* intron (∼43% versus ∼19%; Figure 2G and H). Consequently, while insertion of the *Xist* intron had little impact on *Xist*-2kb transcriptional readthrough, we observed that transcriptional readthrough was strikingly impaired by insertion of the chimeric β-globin/immunoglobulin intron (Figure 2G and H). As a control, we also observed that insertion of the *Xist* intron had no impact on the ∼90% repression of luciferase activity seen with intronless *Xist*-2kb (Figure 2B), whereas inserting the chimeric β-globin/immunoglobulin intron caused weaker levels of luciferase repression (∼60%), a likely consequence of reduced transcriptional interference over the *Luc* gene (Figure 2I). Together, these data suggest that efficient splicing may promote proper 3′-end processing

of a transgene that otherwise harbors strong antiterminator properties.

Transcriptional readthrough induced by *Xist*-2kb requires Repeat A and downstream sequence

Given that *Xist*-2kb but not control transgenes caused transcriptional readthrough in the context of the TETRIS assay, we hypothesized that *Xist*-2kb contains sequences that promote transcriptional readthrough of downstream polyadenylation sequences. Our prior data showed that in the original TETRIS assay (TETRISv1), repression of luciferase activity by *Xist*-2kb required the same sequence elements within Repeat A that are required by full-length *Xist* to induce chromosome-level transcriptional silencing: its GC-rich repeats but not its U-rich linkers (5,25). We therefore investigated whether transcriptional readthrough induced by *Xist*-2kb required these GC-rich repeats (Figure 3A). Expression of a mutant *Xist*-2kb lacking the GC-rich repeats of Repeat A displayed reduced levels of transcriptional readthrough relative to wild-type *Xist*-2kb (Figure 3B), as well as reduced steady-state levels upon induction by doxycycline (∼2-fold induction in Figure 2B vs. ∼10-fold induction in Figure 1I). Our previous work also showed that in TETRISv1, repression of luciferase activity by *Xist*-2kb was attenuated by deletion of a ∼750-nt region immediately downstream of Repeat A that contains elements predicted to form stable structures ('ss234') (Figure 3A; (25,26)). Accordingly, we found that deleting this region almost completely prevented transcriptional readthrough despite a ∼10-fold induction of the mutant *Xist*-2kb (Figure 3C). Lastly, deletion of Repeat A and the ss234 element together ('ΔrA234', Figure 3A) reduced the levels of *Xist*-2kb induction and readthrough transcripts (Figure 3D). We conclude that transcriptional readthrough induced by *Xist*-2kb requires Repeat A as well as a ∼750-nt downstream region.

To validate these results and observe the lengths of *Xist*-2kb readthrough transcripts, we analyzed RNA from untreated and doxycycline-treated TETRIS cells by northern blot, using single-stranded DNA probes targeting either *Xist*-2kb or RNA antisense to the luciferase mRNA product (Figure 3E and F). In both TETRISv1 and TETRISv2, 2349 bp separates *Xist*-2kb's transcriptional start site from its expected cleavage site in the downstream SV40 early polyadenylation sequence (41). Accordingly, the *Xist*-2kb probe detected a prominent band around 2.5 kb in doxycycline-induced samples, consistent with cleavage at this site and the addition of a poly(A) tail (Figure 3E, Supplementary Figure S3B). However, distinct bands between the 6-kb and 9-kb size markers were also detected. These bands suggest a portion of readthrough transcripts terminate at the SV40 early polyadenylation sequence at the end of the *PuroR* gene, 7166 bp away from *Xist*-2kb's transcriptional start site in TETRISv1 and 7735 bp away in TETRISv2. The higher-molecular-weight products were also detected with the *Luc* antisense-targeting probe, confirming that these RNAs contain both *Xist*-2kb sequence and sequence antisense to *Luc* mRNA (Figure 3F). As controls, *Hottip* and *Xist*-2kb lacking the ss234 region were detected with sequence-specific probes only at their expected sizes (Figure 3E and Supplementary Figure S3A). Moreover, the *Luc* antisense-targeting probe did not detect readthrough transcription products with *Hottip* or *Xist*-2kb Δss234 (Figure 3F), confirming efficient cleavage and polyadenylation of these lncRNAs. Thus, sequences within *Xist*-2kb harbor antiterminator activity that can suppress downstream 3′-end processing at strong and commonly used polyadenylation sequences, resulting in robust transcriptional readthrough into downstream sequences.

Xist-2kb induces mild levels of local transcriptional silencing in a SPEN-independent manner

Our experiments unexpectedly revealed transcriptional readthrough as a confounding factor in our ability to determine the extent to which *Xist*-2kb directly induced transcriptional silencing of adjacent reporter genes in the context of TETRIS. Upon *Xist*-2kb induction, our original strand-specific RNA FISH in Figure 1D showed a loss in *Luc* mRNA signal throughout the cell but also suggested that some portion of *Luc* transcripts remain at their site of transcription in the nucleus. These data raised the possibility that upon induction of *Xist*-2kb, transcriptional silencing of the *Luc* gene is incomplete and that the export of *Luc* mRNA to the cytoplasm may be impaired. To test this hypothesis, we performed strand-specific RT followed by qPCR on RNA extracted from cytoplasmic and nuclear fractions, in order to detect the effect that *Xist*-2kb expression has on RNA produced from both strands over the *Luc* gene (Supplementary Figure S4A). Upon expression of *Xist*-2kb, the levels of *Luc* mRNA decreased in the cytoplasm but remained unchanged in the nucleus (Supplementary Figure S4B), while *Xist*-2kb readthrough transcripts (which are antisense to the *Luc* gene) were exclusively nuclear, as expected (Supplementary Figure S4B). These data indicate that in the context of TETRIS, *Xist*-2kb expression does not completely silence *Luc* gene transcription, and also suggest that *Xist*-2kb readthrough transcription may impair *Luc* mRNA release from chromatin or its nuclear export.

We next sought to address the extent to which *Xist*-2kb functions as a transcriptional repressor in a setting in which transcriptional readthrough would not be a confounding factor. To do this, we excised the doxycycline-inducible lncRNA expression cassette from TETRISv2, re-circularized the vector, and re-cloned the lncRNA expression cassette downstream of and in the same orientation as the *PuroR* gene, such that induction of *Xist*-2kb would not cause any readthrough transcripts to traverse either the *Luc* or *PuroR* genes (Figure 4A). In this new version of TETRIS (TETRISv3), expression of *Xist*-2kb caused only a ∼2-fold reduction of luciferase protein activity relative to the empty vector control (Figure 4C and D). RT-qPCR indicated that the ∼2-fold repression of luciferase protein activity in TETRISv3 was coincident with a ∼2-fold reduction of nuclear *Luc* mRNA levels, indicating that *Xist*-2kb induced a mild amount of transcriptional silencing in this setting (Figure 4E). Similarly weak levels of transcriptional silencing were observed with *PuroR* mRNA (Figure 4F), and the nucleocytoplasmic ratios of both *Luc* and *PuroR* mRNAs were unchanged by *Xist*-2kb expression (Figure 4G and H). These results are consistent with the notion that

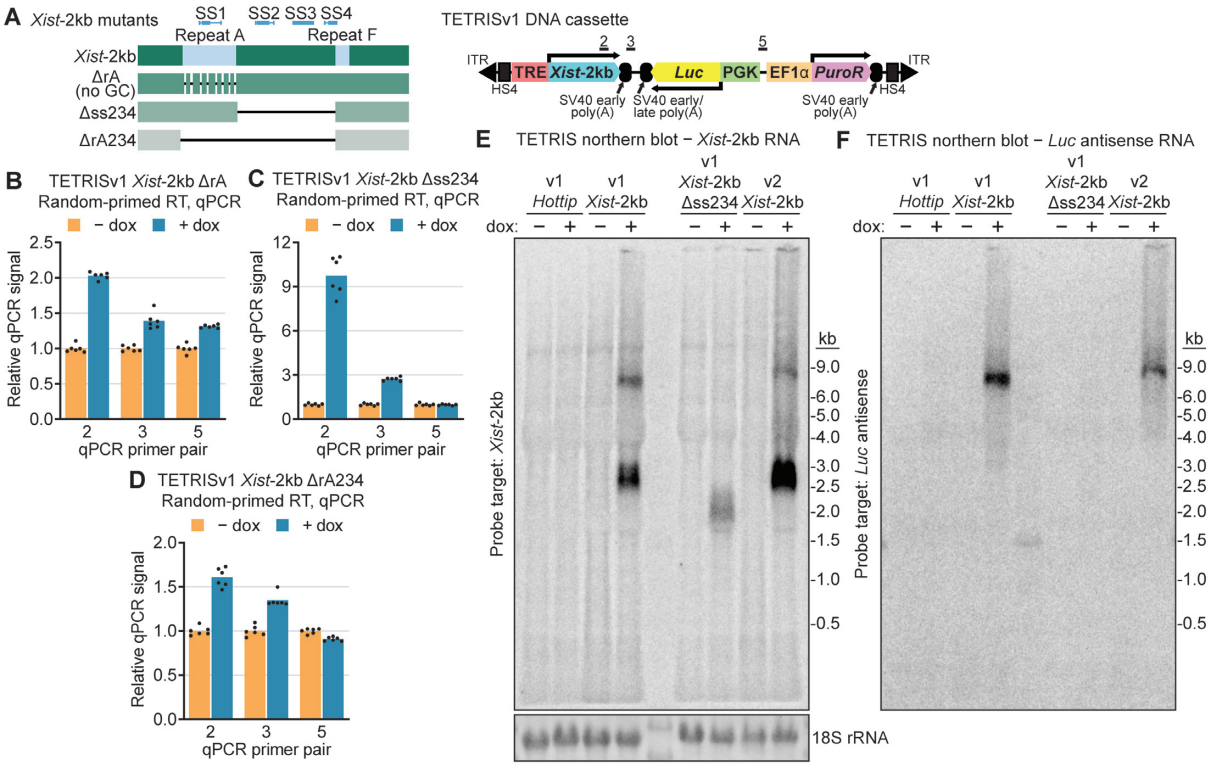

Figure 3. Transcriptional readthrough induced by *Xist*-2kb requires Repeat A and downstream sequence. (**A**) Left: diagram of *Xist*-2kb showing the location of Repeat A, Repeat F, stably structured elements (SS; (26)), and mutants that reduce repressive activity in TETRIS (25). Sequences of each mutant can be found in (25). The deleted ss234 region corresponds to nt 733–1474 of *Xist*, chromosomal coordinates chrX:103481759–103482500 in mm10. Right: diagram of TETRISv1 DNA cassette with locations of primer pairs used in (B-D). (**B–D**) Nuclear RNA from cells harboring TETRISv1 *Xist*-2kb deletion mutants was reverse-transcribed with random primers and analyzed by qPCR with primer pairs targeting regions shown in (A). Numerical values are shown relative to the – dox average for each qPCR primer pair, which is set to one. Dots represent individual technical replicate measurements, and bars represent the average value. (**E**, **F**) Strand-specific northern blot analysis of total RNA prepared from cells harboring the indicated TETRIS expression cassettes. Following membrane transfer, ethidium-stained RNA was imaged to visualize evenness of loading (18S rRNA in (E)) and migration of RNA size markers. The membrane was probed first with a single-stranded DNA oligo targeting *Luc* antisense RNA (F), stripped and imaged to ensure removal of signal (not shown), and re-probed with a single-stranded DNA oligo targeting *Xist*-2kb (E). See Supplementary Figure S3 for additional images. For all panels, cells were treated with or without 1 μg/ml doxycycline for 48 h prior to harvesting. See Supplementary Figure S2 for RT-qPCR control data. See Supplementary Table S1 for information regarding experimental replicates, Supplementary Table S2 for oligo sequences, and Supplementary Table S3 for raw RT-qPCR data.

readthrough transcription is responsible for the bulk of silencing induced by *Xist*-2kb in TETRISv1, and highlight a surprisingly weak ability of *Xist*-2kb to silence genes at the transcriptional level in this transgenic context.

Because the protein SPEN is required for transcriptional silencing induced by full-length *Xist* (6–12), we sought to determine whether SPEN was required for the mild levels of transcriptional silencing induced by the *Xist*-2kb transgene in TETRISv3. We used CRISPR to delete ∼40 kb of the *Spen* gene in ESCs, including its major RNA-binding domains (as done in (12); Supplementary Figure S5A). This deletion is known to cause complete failure of XCI and is expected to comprise a null mutant (12). The deletion was confirmed in select clones by PCR of genomic DNA (Supplementary Figure S5B) and RT-qPCR, which showed the expected loss of *Spen* mRNA expression in the deleted region (Figure 4B, Supplementary Figure S5C). Using these *Spen* knockout ESCs, we then generated TETRISv3 cells inducibly expressing *Xist*-2kb or an empty expression cassette. The levels of silencing induced by *Xist*-2kb were ∼2-fold stronger in SPEN knockout clones compared to wild-type cells, unexpectedly suggesting that SPEN knockout may moderately increase the silencing ability of *Xist*-2kb in TETRIS (Figure 4C, E and F). However, SPEN knockout also eliminated the reproducible, 1.5- to 2-fold boost in luciferase activity typically seen upon addition of doxycycline to TETRIS-empty-vector control cells (see (25), Figures 2B and 4C). Thus, when normalizing relative to empty-vector control, the levels of *Luc* silencing induced by *Xist*-2kb in SPEN knockout clones were commensurate to those induced in wild-type ESCs (Figure 4C and D). From these data, we conclude that the ∼2-fold transcriptional silencing induced by the *Xist*-2kb transgene in the context of TETRIS is SPEN-independent, and may even be enhanced by knockout of SPEN.

Xist-2kb binds SPEN and RBM15, but these proteins are not required for transcriptional readthrough

We next sought to determine whether *Xist*-2kb expressed in TETRIS interacts with proteins known to bind this region in the context of full-length *Xist*. To answer this ques-

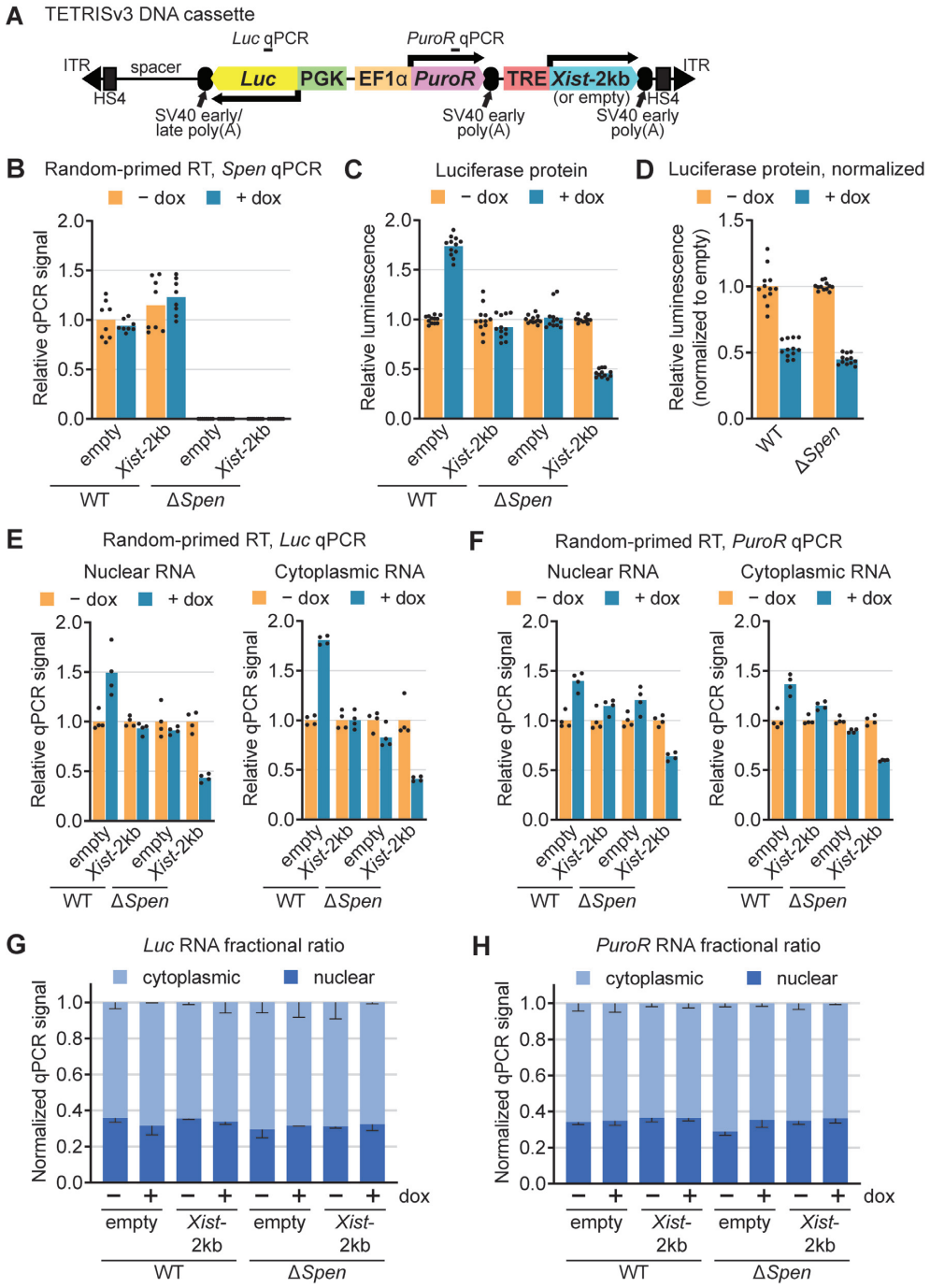

Figure 4. *Xist*-2kb induces mild levels of local transcriptional silencing in a SPEN-independent manner. (**A**) Diagram of TETRISv3 DNA cassette. (**B**) RT-qPCR showing levels of *Spen* mRNA in wild-type and Δ*Spen* cells harboring the TETRISv3 empty and *Xist*-2kb cassettes. See Supplementary Figure S5 for additional characterization of Δ*Spen* cells. (**C**) Luminescence assay showing relative luciferase protein levels in TETRISv3 empty and *Xist*-2kb cells treated with or without doxycycline for two days. (**D**) The *Xist*-2kb values in (C) are shown normalized to the corresponding empty values. (**E**, **F**) Random-primed reverse transcription of nuclear and cytoplasmic RNA from each cell line was followed by qPCR targeting either luciferase (E) or *PuroR* (F). For panels B, E, and F, values are normalized to the – dox average for WT empty cells. (**G**, **H**) The sum of nuclear and cytoplasmic qPCR signal for each cell line in (E-F) was normalized to one to depict nucleocytoplasmic ratios. Dots represent individual technical replicate measurements, and bars represent the average value. For all panels, cells were treated with or without 1 μg/ml doxycycline for 48 h prior to harvesting. See Supplementary Figure S2 for RT-qPCR control data. See Supplementary Table S1 for information regarding experimental replicates, Supplementary Table S2 for oligo sequences, and Supplementary Table S3 for raw RT-qPCR data.

tion, we first used an antibody-independent method called capture hybridization analysis of RNA targets coupled with mass spectrometry (CHART-MS), in which biotinylated DNA oligonucleotides are used to capture an RNA of interest and its bound proteins from nuclear extracts prepared from formaldehyde-crosslinked cells (42,43). Specifically, we performed CHART-MS to compare the proteins associated with wild-type *Xist*-2kb to those associated with a nonfunctional mutant version of *Xist*-2kb lacking Repeat A and the ∼750-nt ss234 region (ΔrA234, Figure 3A). We identified 22 proteins that associated with wild-type *Xist*-2kb in two independent replicate experiments but not with the ΔrA234 mutant *Xist* in the single replicate performed (Supplementary Table S4). The full list of proteins retrieved in each of the three experiments is shown in Supplementary Table S5, and Scaffold output data are available as Supplementary File 1. Two of the proteins enriched specifically by *Xist*-2kb were SPEN and RBM15, both of which associate with Repeat A in the context of full-length *Xist* (6–12,44). Also identified were proteins known to play roles in splicing and RNA export, including U2AF2, DHX15, SRSF6, NXF1 and SNRNP200, consistent with previous studies linking Repeat A to RNA processing (Supplementary Table S4; (14,16,18,19,45)). Using a formaldehyde-based RNA immunoprecipitation protocol (32), we confirmed that SPEN and RBM15 associated with RNA produced from the *Xist*-2kb transgene in a Repeat-A-dependent manner (Figure 5A and B).

Because transcriptional readthrough induced by *Xist*-2kb required Repeat A (Figure 3B and D), we hypothesized that transcriptional readthrough might also require the Repeat-A-binding proteins SPEN or RBM15. TETRISv1 assays performed in two independent SPEN deletion lines (Supplementary Figure S5) showed that SPEN deletion had no effect on luciferase repression induced by *Xist*-2kb (Figure 5C). Given that luciferase repression induced by *Xist*-2kb in TETRISv1 is largely the result of transcriptional readthrough, this result implies that SPEN is not required for transcriptional readthrough. Accordingly, subsequent RT-qPCR assays provided direct evidence for this notion (Figure 5D), although we did observe that in the absence of SPEN, lower levels of *Xist*-2kb were induced than in SPEN-expressing cells (∼4-fold in Figure 4D versus ∼10-fold in Figure 1I). To examine the requirement of RBM15 for transcriptional readthrough induced by *Xist*-2kb, polyclonal cell populations were generated that carried sgRNAs targeting a doxycycline-inducible Cas9 to numerous locations in exons of the *Rbm15* gene, leading to depletion of the protein product (Figure 5E, Supplementary Figure S5E). RBM15B, a paralog of RBM15, was also depleted individually and in combination with RBM15 (Figure 5E, Supplementary Figure S5F). EZH2, which has also been shown to associate with Repeat A (46–48), was also depleted (Figure 5E, Supplementary Figure S5D). Relative to control cells that expressed a non-targeting sgRNA, TETRIS luminescence assays in RBM15, RBM15B, RBM15+RBM15B, and EZH2 knockdown cell lines showed no hindrance of luciferase protein repression by *Xist*-2kb (Figure 5F). Therefore, SPEN, RBM15, RBM15B, and EZH2 do not appear to be required for transcriptional readthrough induced by *Xist*-2kb.

As single-copy transgenes, 5′ fragments of *Xist* cause transcriptional readthrough beyond their polyadenylation sites

Under the conditions we used to make standard TETRIS cell lines, approximately five copies of the TETRIS DNA cassette are randomly inserted into the genome of each cell that survives the selection process (25). Thus, while TETRIS is suitable to determine the local extent of silencing induced by lncRNAs, it cannot report on long-distance silencing. We therefore sought to determine the extent to which *Xist*-2kb and other *Xist* transgenes could promote transcriptional readthrough and long-distance transcriptional silencing by inserting the transgenes as a single copy into a defined chromosomal locus.

To this end, we established a recombinase-mediated cassette exchange (RMCE) system in the *Rosa26* locus of chromosome 6 (49). We created a *Rosa26*-targeting vector that contained a lox66 site, a *PuroR-ΔTK* fusion gene (50), and a lox2272 site followed by a BGH polyadenylation sequence (Supplementary Figure S6A). This targeting construct was electroporated into F1-hybrid, male ESCs that were derived from a cross between C57BL/6J (B6) and CAST/EiJ (Cast) mice that enable allelic analysis of gene expression. Southern blot was used to confirm insertion of the construct into the correct locus on the B6 allele of selected clones (Supplementary Figure S6B). In parallel, we created a cargo vector that contained a lox71 site, a lncRNA-expression cassette driven by a doxycycline-inducible promoter, a tandemly oriented and constitutively expressed hygromycin B resistance gene (*HygroR*) lacking a polyadenylation sequence, and a lox2272 site (Supplementary Figure S6C). Electroporation of the cargo vector along with Cre recombinase into our F1-hybrid RMCE cells, followed by positive selection with hygromycin B and negative selection with ganciclovir, generated a small number of surviving clones that harbored cargo vectors inserted in the desired orientation in *Rosa26* (Supplementary Figure S6D; not shown).

We employed this RMCE system to create four separate ESC lines that expressed different versions of inducible *Xist* transgenes from *Rosa26* (Figure 6A): one line expressed the *Xist*-2kb transgene, another expressed full-length *Xist* from its endogenous DNA sequence (including its introns and natural polyadenylation sequence), another expressed the first 5.5 kb of *Xist* (‘*Xist*-5.5kb’), which includes the Repeat B and Repeat C domains of *Xist* known to recruit Polycomb repressive complex 1 (PRC1 (18,19,27,51)), and a final line expressed *Xist*-2kb fused to the final two exons of *Xist* (‘*Xist*-2kb+6,7’, which includes the intron between exons 6 and 7). The final exon of *Xist* includes Repeat E and is essential for proper *Xist* localization (52–54). As a control, we created an ESC line that underwent recombination but lacked any *Xist* insertion (‘empty’). We then used piggyBac-mediated transgenesis to insert the *rtTA* gene into select clones of each genotype in order to allow doxycycline-inducible expression of each cargo RNA. RNA FISH confirmed that RNA produced from each *Xist* transgene was doxycycline-inducible and remained localized in the nucleus, presumably surrounding its site of transcription (Figure 6B). We note that despite several attempts, we were unsuccessful in cloning an *Xist*-5.5kb construct that contained all ∼36 repeats in Repeat B; the construct used for this study

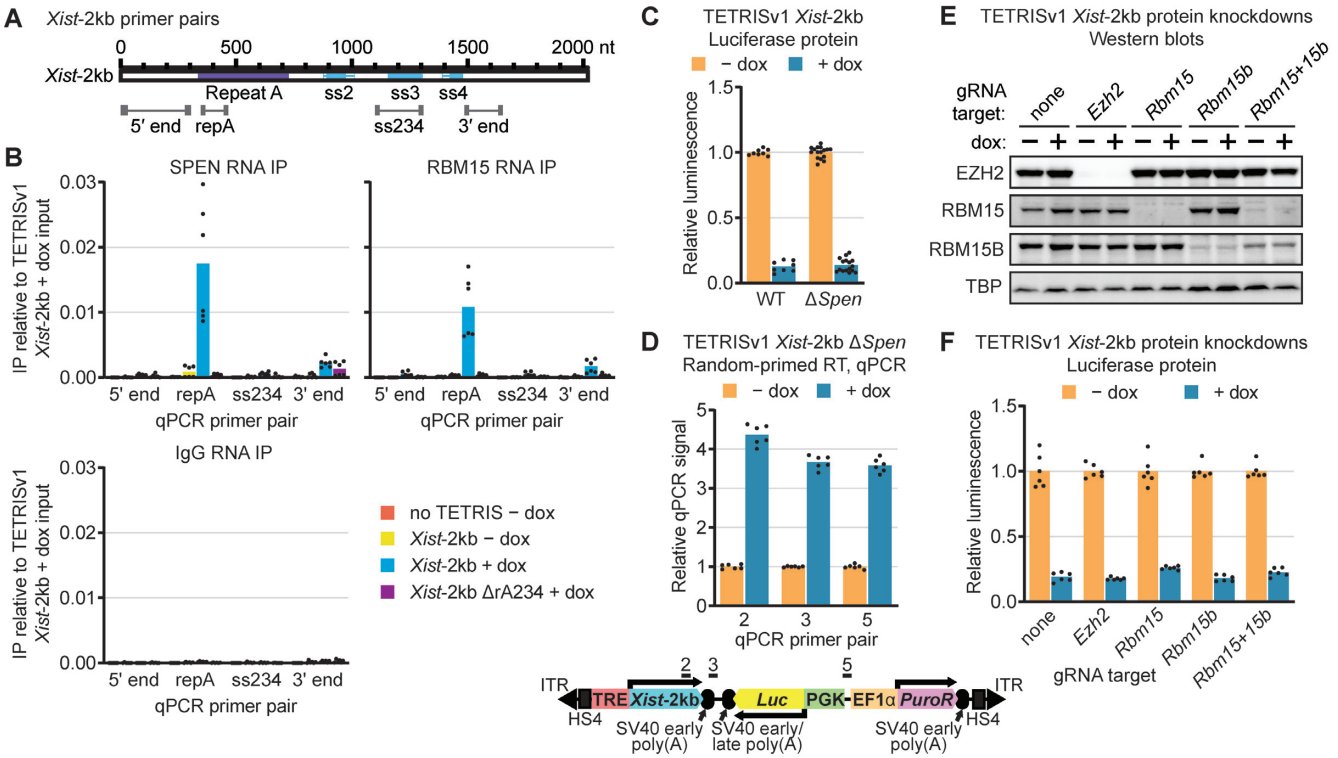

Figure 5. *Xist*-2kb binds SPEN and RBM15, but these proteins are not required for transcriptional readthrough. (**A**) Diagram of *Xist*-2kb showing locations of qPCR primer pairs. (**B**) Four different cell identities (cells without TETRIS cargo, TETRISv1 *Xist*-2kb cells treated with or without doxycycline, and TETRISv1 *Xist*-2kb ΔrA234 cells treated with doxycycline) were used for RNA immunoprecipitation with SPEN, RBM15, or IgG control antibodies, followed by random-primed RT and qPCR with primer pairs shown in (A). Values are relative to *Xist*-2kb + dox input. (**C**) Luminescence assay in wild-type and Δ*Spen* cells harboring TETRISv1 *Xist*-2kb cargo. (**D**) Nuclear RNA from Δ*Spen* cells harboring TETRISv1 *Xist*-2kb cargo was reverse-transcribed with random primers and analyzed by qPCR with primer pairs targeting regions shown in the diagram at the bottom. (**E**) Western blots demonstrating successful CRISPR-Cas9 targeting of *Ezh2*, *Rbm15*, *Rbm15b*, and both *Rbm15* and *Rbm15b* in cells harboring TETRISv1 *Xist*-2kb cargo. (**F**) Luminescence assays with CRISPR knockdown TETRISv1 *Xist*-2kb cells shown in (E). See Supplementary Figure S5 for characterization of Δ*Spen* cells and for locations of CRISPR gRNAs. Dots represent individual technical replicate measurements, and bars represent the average value, relative to the – dox value for each cell line. See Supplementary Figure S2 for RT-qPCR control data. See Supplementary Table S1 for information regarding experimental replicates, Supplementary Table S2 for oligo sequences, and Supplementary Table S3 for raw RT-qPCR data.

contained ∼14 repeats (Figure 6A and Supplementary Figure S7A).

To determine whether the *Xist* transgenes caused transcriptional readthrough of their downstream SV40 early polyadenylation sequence as well as to determine their effects on the expression of genes across chromosome 6, we induced the expression of each *Xist* transgene for three days and performed RNA-seq using RNA purified from cytoplasmic and nuclear fractions. In parallel, we treated empty-cargo ESCs with doxycycline for three days and sequenced RNA purified from cytoplasmic and nuclear fractions. Expression of the *Xist* transgenes was verified by examining reads mapping to the endogenous *Xist* locus (which is not expressed in these cells; Supplementary Figure S7B), and total counts were used to calculate RPKM values for each *Xist* transgene. Consistent with RNA FISH data in Figure 6B, RNA produced from each *Xist* transgene localized in the nucleus (Supplementary Figure S7C), and expression levels were only slightly lower for the *Xist* hypomorphs than for full-length *Xist* (Figure 6C).

We examined our RNA-seq data for evidence of *Xist*-dependent transcriptional readthrough in the RMCE setting. We aligned our RNA-seq data to an *in-silico* genome build that contained the transgenic vector features as they were inserted into *Rosa26* and then visualized the aligned reads in the Integrative Genomics Viewer (IGV; (55)). Consistent with our results above, when *Xist*-2kb, but not empty-cargo control, was induced, we observed robust transcriptional readthrough of the SV40 early polyadenylation sequence; this readthrough extended through the *PGK* promoter that drives expression of *HygroR*, through the BGH polyadenylation sequence at the 3′ end of *HygroR*, and through the bidirectionally functional triple-SV40-polyadenylation sequences designed to terminate expression of *Rosa26* (Figure 6D, regions shaded in grey). Small amounts of readthrough transcription appeared to extend through the *Rosa26* promoter and into the *Setd5* gene located just downstream. Readthrough transcription appeared more robust in cells expressing *Xist*-5.5kb but was barely apparent in cells expressing *Xist*-2kb+6,7 and was not apparent in cells expressing full-length *Xist*. In agreement with these data and the nuclear localization of the *Xist* transgenes (Supplementary Figure S7C), induction of *Xist*-2kb and *Xist*-5.5kb caused a dramatic shift of *HygroR* reads to the nucleus, suggesting the production of chimeric, chromatin-bound readthrough RNAs contain-

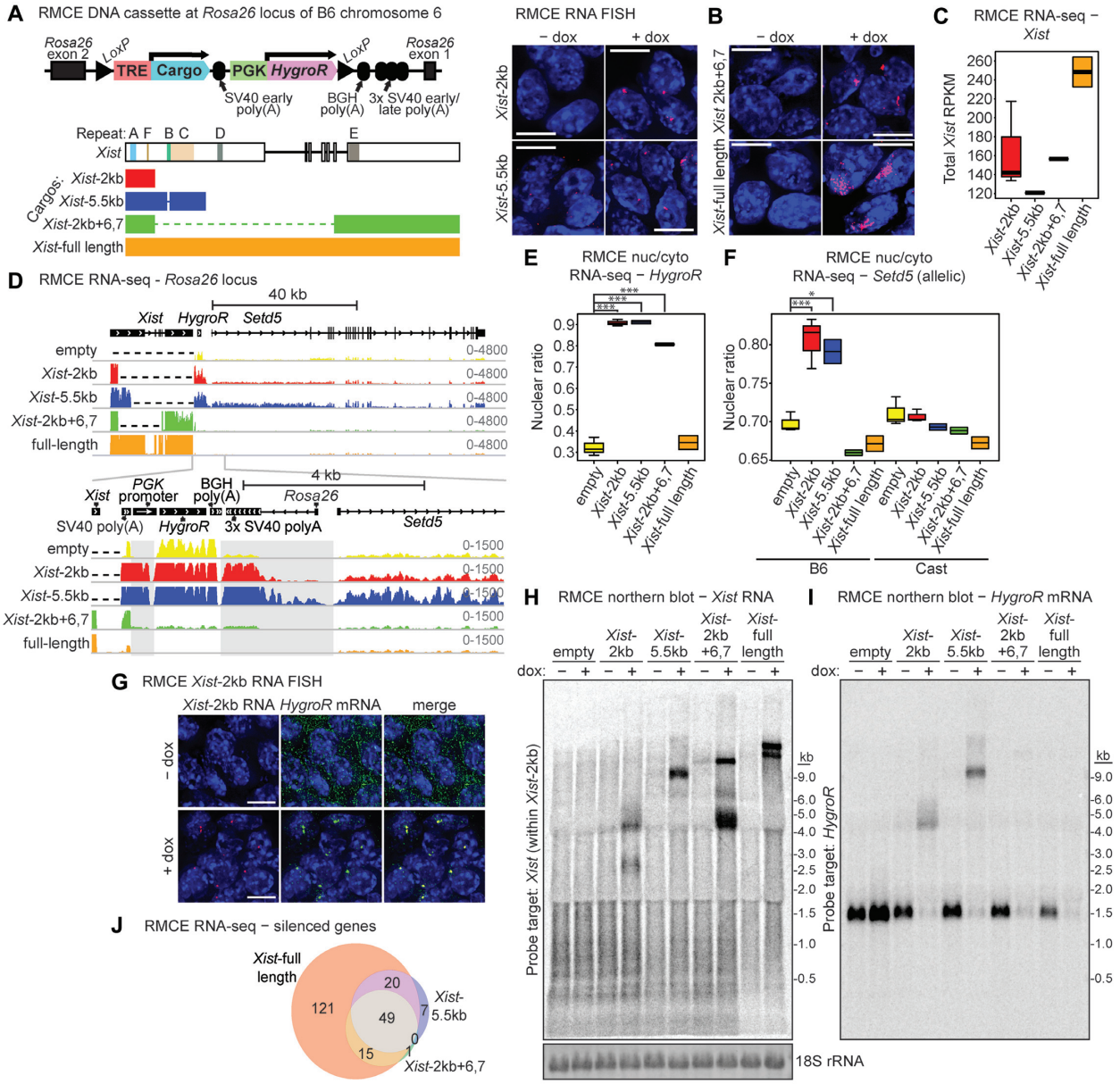

Figure 6. As single-copy transgenes, 5′ fragments of *Xist* cause transcriptional readthrough beyond their polyadenylation sites. (**A**) Diagram of recombination-mediated cassette exchange (RMCE) expression cassette inserted into the *Rosa26* locus on the B6 allele of chromosome 6 in B6/Cast F1 hybrid ES cells. Full-length *Xist* and three mutant transgene cargos were inserted into this expression cassette under a doxycycline-inducible promoter. *Xist*-5.5kb is missing a portion of Repeat B (Supplementary Figure S7A). *HygroR*, constitutively expressed hygromycin resistance gene. (**B**) Stellaris single-molecule, strand-specific RNA FISH for *Xist* RNA (red) in cells expressing *Xist*-2kb, *Xist*-5.5kb, *Xist*-2kb+6,7 or full-length *Xist* inserted at *Rosa26*, treated with or without doxycycline. DAPI-stained nuclei are blue. Scale bar = 10 μm. (**C**) *Xist* expression in each cell line following 3 d doxycycline treatment, determined via RNA-seq. Reads per kilobase per million (RPKM) values were calculated using reads aligned to the endogenous *Xist* locus divided by length of inserted transcript in kb divided by total aligned reads in millions for each dataset. (**D**) RNA-seq reads were aligned to a custom-made IGV build containing the single-insertion *Xist*/*HygroR* cassette within the *Rosa26* locus on chromosome 6. Top panel depicts this cassette upstream of the endogenous *Setd5* gene. Dashed lines within the *Xist* locus represent sequences not present in each version of the RMCE cassette. The bottom panel shows a closer view of RNA-seq reads extending beyond the SV40 polyadenylation sequence at the 3′ end of *Xist*, with regions of clear transcriptional readthrough highlighted in grey. Note that the y-axis is different between the two panels. (**E**) Nuclear ratio of reads mapping to *HygroR* following doxycycline treatment. (**F**) Nuclear ratio of allele-specific reads mapping to *Setd5* following doxycycline treatment. Note the difference in y-axis compared to (E). ***$P < 0.0001$; *$P < 0.01$, Tukey HSD post-hoc analysis of significant differences by ANOVA. (**G**) Stellaris single-molecule RNA FISH showing *Xist*-2kb RNA (red) and *HygroR* mRNA (green) in cells with *Xist*-2kb inserted at *Rosa26*. DAPI-stained nuclei are blue. Scale bar = 10 μm. (**H-I**) Northern blot analysis of RNA prepared from cells harboring the indicated RMCE expression cassettes. Following membrane transfer, ethidium-stained RNA was imaged to visualize evenness of loading (18S rRNA in (H)) and migration of RNA size markers (Supplementary Figure S8A). Membrane was probed first with a single-stranded DNA oligo targeting *HygroR* (I), stripped and imaged to ensure removal of signal (not shown), then re-probed with a single-stranded DNA oligo targeting *Xist* within *Xist*-2kb (H). (**J**) Venn diagram showing overlaps of significantly repressed genes in *Xist*-5.5kb, *Xist*-2kb+6,7 and full-length *Xist* cells. See Supplementary Table S6 for RNA-seq gene expression data and Supplementary Figures S6-S10 for additional information related to Figure 6. See Supplementary Table S1 for information regarding experimental replicates and Supplementary Table S2 for oligo sequences.

ing both *Xist* transgene and *HygroR* sequence (Figure 6E). Expression of *Xist*-2kb and *Xist*-5.5kb also caused mild, yet significant shifts of syntenic *Setd5* reads to the nucleus (Figure 6F), consistent with low-level readthrough extending into this gene. Inducing *Xist*-2kb+6,7 caused a less-pronounced nuclear shift in *HygroR* reads and no significant shift in *Setd5* reads, consistent with *Xist*-2kb+6,7 causing only trace amounts of transcriptional readthrough. Induction of full-length *Xist* had no effect on the nucleocytoplasmic distribution of *HygroR* or *Setd5* reads, reflecting efficient transcriptional termination of this transgene. Remarkably, strand-specific RNA FISH using probes designed to detect RNA produced from the *HygroR* gene gave an *Xist*-like staining pattern after induction of *Xist*-2kb, further supporting the occurrence of readthrough transcription and indicating that chimeric *Xist-HygroR* transcripts associate with chromatin in *in cis*, in an *Xist*-like manner (Figure 6G).

To corroborate these results, we used strand-specific northern blot analysis (Figure 6H and I, Supplementary Figure S8A). Cleavage occurring at the SV40 early polyadenylation site downstream of *Xist*-2kb would be expected to produce a 2261-nt RNA (excluding its poly(A) tail). About half of *Xist*-2kb signal migrated around 2.5 kb, suggesting that some level of cleavage occurs at this site (Figure 6H). However, the other half of *Xist*-2kb signal migrated around 4–5 kb. This longer RNA was also detected with a probe targeting *HygroR* (Figure 6I), suggesting that, in the RMCE setting, about half of *Xist*-2kb transcripts do not terminate at the expected polyadenylation site but instead continue downstream through the *HygroR* gene. Termination of the *Xist*-2kb-*HygroR* chimeric transcript at the BGH polyadenylation site downstream of *HygroR* or at the three SV40 late polyadenylation sites further downstream would produce RNAs 4155, 4501, 4755, and 5009 nt in length (disregarding poly(A) tail length), in agreement with the molecular weight of the larger product observed via northern blot.

For *Xist*-5.5kb, cleavage at its SV40 early polyadenylation site would be expected 5598 nt downstream of its transcription start site. However, no product of this size was detected (Figure 6H). Instead, nearly all signal migrated at an approximate size of 9 kb. This signal was also detected with the *HygroR*-specific probe, indicating that nearly all *Xist*-5.5kb transcripts continued downstream into the *HygroR* gene (Figure 6I). The ∼9-kb size estimated for this RNA slightly exceeds the lengths expected for cleavage at the BGH and SV40 late poly(A) sites (7492, 7838, 8092 and 8346 nt). For both *Xist*-2kb and *Xist*-5.5kb, smears of signal at even higher molecular weights are consistent with RNA-seq data suggesting readthrough into *Setd5* (Figure 6D). Additionally, upon expression of both *Xist*-2kb and *Xist*-5.5kb, the 1.5-kb *HygroR* mRNA signal was lost, suggesting that transcriptional readthrough from *Xist*-2kb and *Xist*-5.5kb interfered with transcription initiation at the *HygroR* promoter. These data indicate that *Xist*-2kb and *Xist*-5.5kb can robustly suppress downstream cleavage and polyadenylation to generate abundant transcriptional readthrough products that contain *HygroR* sequence. Furthermore, our RNA-seq and RNA FISH data suggest that, rather than being exported to the cytoplasm, chimeric *Xist-HygroR* RNAs are retained at their sites of transcription in an *Xist*-like manner (Figure 6E and G).

In contrast to *Xist*-2kb and *Xist*-5.5kb, induction of *Xist*-2kb+6,7 produced an RNA band of length greater than 9 kb (Figure 6H), in agreement with its expected size of 10,079 nt (excluding poly(A) tail length). A smaller band between 4 and 5 kb was also detected, likely due to alternative splicing within the final exon that removes ∼5–6 kb (Supplementary Figure S8B). Inducing full-length *Xist* (expected size of 18,001 nt from CMV TSS to annotated endogenous *Xist* 3′ end) resulted in a doublet band at high molecular weight (Figure 6H), also reflecting the production of two *Xist* isoforms due to alternative splicing (Supplementary Figure S8B). Induction of *Xist*-2kb+6,7 and full-length *Xist* both caused strong reductions in 1.5-kb *HygroR* mRNA signal (Figure 6I). Only a very faint signal migrating around the size of *Xist*-2kb+6,7 was detected with the *HygroR* probe, consistent with trace amounts of transcriptional readthrough suggested by RNA-seq (Figure 6D and E). No *HygroR* signal was detected at higher molecular weights upon expression of full-length *Xist*, consistent with efficient transcriptional termination at *Xist*'s natural polyadenylation sequence. Together, these results confirm a lack of robust transcriptional readthrough beyond *Xist*-2kb+6,7 and full-length *Xist* and suggest that the doxycycline-dependent loss in *HygroR* mRNA signal is due to repressive activity of the *Xist*-2kb+6,7 and full-length *Xist* RNAs.

Relative to longer *Xist* transgenes, *Xist*-2kb is severely deficient in its ability to induce long-distance transcriptional silencing

Next, we examined the extent to which the different *Xist* transgenes silenced gene expression on their syntenic chromosome (Supplementary Figure S9). In cells expressing full-length *Xist*, 2404 genes were differentially expressed relative to empty-cargo ESCs. On the B6 allele of chromosome 6—the chromosome that harbors the *Xist* transgene at the *Rosa26* locus—214 genes were differentially expressed (Supplementary Figure S9A). Of these, 205 genes shifted in the downward direction and 9 shifted in the upward direction, consistent with repression by full-length *Xist* (Supplementary Figure S9A). On the Cast allele of chromosome 6, only 34 genes changed, with similar numbers going up and down (21 and 13, respectively; Supplementary Figure S9B). Throughout the genome, near-equal numbers of up- and down-regulated genes were detected (1012 and 1163, respectively). Thus, expression of full-length *Xist* caused chromosome-level repression of genes *in cis* and genome-wide changes in gene expression, the latter presumably being due to secondary effects.

Expression of *Xist*-5.5kb and *Xist*-2kb+6,7 also caused gene silencing, but at a reduced level relative to full-length *Xist*. *Xist*-5.5kb expression led to differential expression of 98 genes on the B6 allele of chromosome 6, 76 of which were shifted down (Supplementary Figure S9C). *Xist*-2kb+6,7 expression led to differential expression of 88 genes on the B6 allele of chromosome 6, 65 of which were shifted down (Supplementary Figure S9E). The majority of genes repressed by these two transgenes were also repressed by full-

length *Xist* (69 of 76 for *Xist*-5.5kb and 64 of 65 for *Xist*-2kb+6,7; Figure 6J). Loss of nuclear as well as cytoplasmic RNA-seq signal for these genes indicated that, like full-length *Xist*, the hypomorphic *Xist* transgenes induced transcriptional silencing (Supplementary Figure S10). In contrast, when comparing cytoplasmic RNA levels between *Xist*-2kb and empty-cargo ESCs, we only detected two differentially expressed genes genome-wide (excluding *Xist*), both on the B6 allele of chromosome 6. These genes were *Creld1*, located 406 kb downstream of the *Rosa*26 locus, and *Trh*, located ∼20 Mb upstream (Supplementary Figure S9G). These data indicate that, relative to the longer *Xist* transgenes, *Xist*-2kb is severely deficient in its ability to induce long-distance transcriptional silencing. Deficient silencing by *Xist*-2kb in the RMCE setting is consistent with our observations using TETRIS (Figure 4 and Supplementary Figure S4) and with observations made previously (4,5,18).

DISCUSSION

Transcriptional silencing during XCI is thought to initiate when Repeat A recruits SPEN; SPEN recruitment, in turn, is thought to recruit and/or locally activate histone deacetylases and corepressor proteins to silence transcription over the X chromosome (6–12). Our data demonstrate that in mouse ESCs, recapitulating a SPEN-Repeat A interaction on chromatin is insufficient to induce robust local or long-distance transcriptional silencing by *Xist*, highlighting a gap in our understanding of the mechanism through which Repeat A functions to silence gene expression. We found that even though *Xist*-2kb associates with SPEN and localizes to chromatin in a manner that is indistinguishable from silencing-competent versions of *Xist*, it fails to induce robust transcriptional silencing, even of nearby genes. However, fusion of *Xist*-2kb to the next 3.5 kb of *Xist* (*Xist*-5.5kb), which in our construct included Repeat C and a portion of the essential PRC1-recruitment domain Repeat B (Figure 6A, Supplementary Figure S7A; (18,19,27,51)), or the fusion of *Xist*-2kb to the final two exons of *Xist* (*Xist*-2kb+6,7), which lacks a PRC1 recruitment domain but contains Repeat E and additional downstream sequence elements (26,52,53,56), both conferred near-equal transcriptional silencing capability in an isogenic context (Figure 6J, Supplementary Figures S9 and S10). Thus, in mouse ESCs, Repeat A is necessary for long-distance transcriptional silencing induced by *Xist*, but it is not sufficient. Silencing requires synergy between Repeat A and additional downstream sequence elements within *Xist*. It will be important to define the molecular mechanisms that underlie this synergy in future work.

During the course of our study, we also found, quite unexpectedly, that sequences within the 5′ end of *Xist* harbor transcriptional antiterminator activity that can read through of strings of strong polyadenylation sequences that include the SV40 early, bovine growth hormone, and natural *Xist* polyadenylation sequences. These observations were made using two reductionist systems similar in nature to those previously employed to identify seminal aspects of *Xist* biology (4–12,18,21,26,27). In the context of the TETRIS assay, we found that the antiterminator activity of *Xist*-2kb required the GC-rich portions of Repeat A and a 742-nt region located just downstream that contains elements predicted to form stable structures (17,26,57). In the context of our RMCE system, *Xist*-2kb again displayed transcriptional antiterminator activity, while *Xist*-5.5kb displayed even stronger antiterminator activity, suggesting sequences downstream of *Xist*'s first two kilobases may enhance suppression of cleavage and polyadenylation. Longer, spliced *Xist* transgenes, namely *Xist*-2kb+6,7 and full-length *Xist*, terminated at the SV40 early and/or natural *Xist* polyadenylation sites despite the fact that the *Xist*-2kb transgene alone was capable of reading through both of these sites in TETRIS. This suggests that the antiterminator activity harbored by the 5′ end of *Xist* loses its potency at longer range and/or the act of splicing dampens it. In support of the latter, we found that inserting an efficiently spliced intron into *Xist*-2kb strongly inhibits transcriptional readthrough in TETRIS. This result is consistent with recent long-read sequencing observations that suggest a direct relationship between efficient splicing and 3′-end processing of endogenous transcripts (38,39).

To date, we have yet to identify the protein mediators of *Xist*'s antiterminator activity. While we found that SPEN and RBM15 associated with *Xist*-2kb, neither was required for antitermination, nor were EZH2 or RBM15B. Mechanistically, we hypothesize that *Xist*'s 5′ antiterminator activity may be related to a process called telescripting, in which the U1 snRNP, independent of its role in splicing, prevents premature termination of pre-mRNAs at cryptic cleavage and polyadenylation sites, particularly in long genes (58–60). While the 5′ end of *Xist* does not harbor obvious U1 snRNA binding sites (61), we and others have found that Repeat A associates with protein components of the U1 and other snRNPs (Supplementary Table S4; (14,16,18)). Moreover, several proteins that were identified as *Xist* cofactors in recent genetic and proteomic screens—namely NXF1, SRRT and SCAF4—have known roles in suppressing premature cleavage and polyadenylation (6,10,16,62–64). Future work will determine whether these proteins or others mediate *Xist*'s antiterminator activity.

Moving forward, it will also be important to determine the endogenous function of the antiterminator activity we discovered in the 5′ region of *Xist*. We note that PolyA_SVM predicts that the *Xist* gene harbors polyadenylation sequences throughout its length, particularly in its unusually long first exon (Supplementary Figure S11 (65,66)). Thus, one possible function of the antiterminator activity is to prevent premature polyadenylation and promote production of full-length *Xist*. It is now appreciated that premature 3′-end processing and premature transcriptional termination of mammalian genes is pervasive ((67–76); also reviewed in (77)). Moreover, sequence-specific, RNA-mediated transcriptional antitermination is well-documented in bacteria (78), so it is conceivable that similar antiterminator elements exist in mammalian RNAs beyond *Xist*. Antiterminator activities such as the one we describe could counteract premature termination to upregulate transcription in various developmental and physiological contexts, particularly for longer genes. Indeed and intriguingly, a transcribed CpG island is known to be required for efficient elongation of the *Xist*-analogous lncRNA *Airn*,

and in a previous study, we found that this CpG island sequence repressed luciferase protein production when expressed as a portion of a longer transcript in TETRIS, implying that it too may harbor antiterminator activity (25,79). The sole CpG island in the *Xist* gene is also found within the antiterminator-essential ss234 region, suggesting a possible connection between transcription of CpG island DNA and antiterminator activity.

Beyond suppression of premature polyadenylation, the antiterminator activity we observed in this study may be just one consequence of a broader ability of Repeat A and other 5′ elements to promote transcription of *Xist* from its endogenous locus. In addition to known antiterminator proteins, Repeat A associates with many SR proteins, including SRSF1, 2, 3, 4, 5, 6, 7, 9, 10 and 11 (Supplementary Table S4 and (14,16,18)). In separate contexts, SR proteins have been shown to promote transcriptional elongation as well as form nuclear condensates that colocalize with the hyperphosphorylated, elongating form of RNA polymerase II (80–82). It is conceivable that proteins recruited by the 5′ region of *Xist* promote a local environment that is conducive to robust transcription, rendering the *Xist* locus resistant to silencing by the same repressive complexes it recruits to other genes. Indeed, despite the evidence that high levels of SPEN bind the endogenous *Xist* locus early during XCI, the *Xist* lncRNA does not silence its own gene (9). Moreover, prior work has shown that when nascent pre-mRNAs associate with high levels of PRC2, they attenuate their own transcription (83). While the *Xist* lncRNA presumably associates with higher levels of PRC1 and PRC2 than most other nuclear RNAs, it does not silence itself (6,18,19,27,46–48,51,52,54,84). Thus, we speculate that a need to derive antiterminator activity could have been a driving force in the evolution of the 5′ region of *Xist*, enabling it to recruit proteins whose functions are to promote robust expression of the *Xist* lncRNA in an otherwise highly repressive local environment. Indeed, it is known that active transcription prevents *Xist* from depositing Polycomb over its target genes on the X (8,27). Antiterminator elements that directly protect the *Xist* locus from SPEN-dependent silencing could indirectly protect the locus from Polycomb-dependent silencing, simply through the act of promoting *Xist* transcription.

Lastly, our study underscores the likelihood that Repeat A has functions beyond initiating gene silencing. While much work has centered on Repeat A's silencing activity, relatively little attention has been paid to the fact that, in both endogenous and transgenic contexts, Repeat A deletion causes significant reductions in the abundance of *Xist* (11,13,14,18–24). Even in the context of TETRIS, deletion of Repeat A caused a dramatic reduction in the steady-state levels of RNA produced from the *Xist*-2kb transgene (Figure 3). Considering the likelihood that the abundance of *Xist* is directly related to its potency (85), it remains difficult to isolate the changes that deletion of Repeat A imparts on *Xist*-induced gene silencing while accounting for the changes that the deletion imparts on *Xist* abundance. Furthermore, *in vitro*, SPEN has no obvious preference for binding wild-type Repeat A over non-functional versions of Repeat A whose internal secondary structures have been minimally disrupted by point mutation or a two-nucleotide deletion (12). These data support the notion that SPEN binding, and in turn, gene silencing, is not the only function of Repeat A.

Extending this idea, a recent study demonstrated that Repeat A cooperates with Repeat B to spread PRC1 and PRC2 over the inactive X, specifically during the early stages of XCI (86). In that work, the authors made use of a novel deletion in the 5′ region of *Xist* that surprisingly had no effect on *Xist* abundance. It is unclear why this particular deletion did not affect *Xist* abundance, whereas other near-identical deletions made in *Xist*'s 5′ region have affected its abundance; also unclear is the mechanism through which Repeat A cooperates with Repeat B to mediate recruitment of the PRCs (86). We again highlight that in our assays, equivalent levels of silencing are observed when a version of Repeat B/C or the final two exons of *Xist* are appended to the 5′ region of *Xist*. Whether the PRCs are required for silencing by both of these *Xist* transgenes remains to be determined, but at a minimum, our data suggest that during XCI, Repeat A cooperates with multiple regions in *Xist*, not only Repeat B.

Thus, much is left to learn about the roles and mechanisms of Repeat A during XCI. We have shown that in transgenic contexts, a 5′ portion of *Xist* that contains Repeat A essentially lacks gene silencing activity yet is a potent transcriptional antiterminator. These observations remain highly surprising and warrant further investigation.

DATA AVAILABILITY

All genomic data, including raw sequencing files and processed data files, are available at GEO accession number GSE120197. Genomic data are also available in wiggle tracks at UCSC genome browser: http://genome.ucsc.edu/s/davidlee/rosa26_xist_fractions_lee_2019.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank D. Ciavatta, N. Hathaway, M. Magnuson and T. Magnuson for reagents, Y. Xu for contributions to Figure 1B, the UNC Microscopy Services Laboratory, and the Mass Spectrometry Facility at the University of Massachusetts Medical School.

Author contributions: J.B.T., D.M.L. and J.M.C. conceived the study; D.O.C., S.R.B. and J.M.C. designed the RMCE strategy; J.B.T., D.M.L., R.E.C., S.O.K., K.I., M.D.S., S.R.B. and J.M.C. designed and performed experiments, and J.B.T., D.M.L. and J.M.C. wrote the paper.

FUNDING

National Institutes of Health (NIH) [GM121806, T32 GM007092 to D.M.L. and R.E.C.]; March of Dimes Foundation [Basil O'Connor Award #5100683]; National Cancer Institute (NCI) [T32 CA217824 to J.B.T., P30 CA016086 to Microscopy Services Laboratory]; Lineberger Comprehensive Cancer Center and UNC Department of Pharmacology (to J.M.C.). Funding for open access charge: National Institutes of Health.

Conflict of interest statement. D.O.C. is employed by, has equity ownership in, and serves on the board of directors of TransViragen, the company contracted by UNC-Chapel Hill to manage its Animal Models Core Facility. The authors declare no other competing interests.

REFERENCES

1. Brockdorff,N. (2018) Local tandem repeat expansion in Xist RNA as a model for the functionalisation of ncRNA. *Noncoding RNA*, **4**, 28.
2. Brockdorff,N., Ashworth,A., Kay,G.F., McCabe,V.M., Norris,D.P., Cooper,P.J., Swift,S. and Rastan,S. (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, **71**, 515–526.
3. Brown,C.J., Hendrich,B.D., Rupert,J.L., Lafreniere,R.G., Xing,Y., Lawrence,J. and Willard,H.F. (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, **71**, 527–542.
4. Minks,J., Baldry,S.E., Yang,C., Cotton,A.M. and Brown,C.J. (2013) XIST-induced silencing of flanking genes is achieved by additive action of repeat a monomers in human somatic cells. *Epigenetics Chromatin*, **6**, 23.
5. Wutz,A., Rasmussen,T.P. and Jaenisch,R. (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.*, **30**, 167–174.
6. Chu,C., Zhang,Q.C., da Rocha,S.T., Flynn,R.A., Bharadwaj,M., Calabrese,J.M., Magnuson,T., Heard,E. and Chang,H.Y. (2015) Systematic discovery of Xist RNA binding proteins. *Cell*, **161**, 404–416.
7. McHugh,C.A., Chen,C.K., Chow,A., Surka,C.F., Tran,C., McDonel,P., Pandya-Jones,A., Blanco,M., Burghard,C., Moradian,A. *et al.* (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*, **521**, 232–236.
8. Zylicz,J.J., Bousard,A., Zumer,K., Dossin,F., Mohammad,E., da Rocha,S.T., Schwalb,B., Syx,L., Dingli,F., Loew,D. *et al.* (2019) The implication of early chromatin changes in X chromosome inactivation. *Cell*, **176**, 182–197.
9. Dossin,F., Pinheiro,I., Zylicz,J.J., Roensch,J., Collombet,S., Le Saux,A., Chelmicki,T., Attia,M., Kapoor,V., Zhan,Y. *et al.* (2020) SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature*, **578**, 455–460.
10. Moindrot,B., Cerase,A., Coker,H., Masui,O., Grijzenhout,A., Pintacuda,G., Schermelleh,L., Nesterova,T.B. and Brockdorff,N. (2015) A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. *Cell Rep.*, **12**, 562–572.
11. Nesterova,T.B., Wei,G., Coker,H., Pintacuda,G., Bowness,J.S., Zhang,T., Almeida,M., Bloechl,B., Moindrot,B., Carter,E.J. *et al.* (2019) Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat. Commun.*, **10**, 3129.
12. Monfort,A., Di Minin,G., Postlmayr,A., Freimann,R., Arieti,F., Thore,S. and Wutz,A. (2015) Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.*, **12**, 554–561.
13. Hoki,Y., Kimura,N., Kanbayashi,M., Amakawa,Y., Ohhata,T., Sasaki,H. and Sado,T. (2009) A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development*, **136**, 139–146.
14. Royce-Tolland,M.E., Andersen,A.A., Koyfman,H.R., Talbot,D.J., Wutz,A., Tonks,I.D., Kay,G.F. and Panning,B. (2010) The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation. *Nat. Struct. Mol. Biol.*, **17**, 948–954.
15. Cirillo,D., Blanco,M., Armaos,A., Buness,A., Avner,P., Guttman,M., Cerase,A. and Tartaglia,G.G. (2016) Quantitative predictions of protein interactions with long noncoding RNAs. *Nat. Methods*, **14**, 5–6.
16. Graindorge,A., Pinheiro,I., Nawrocka,A., Mallory,A.C., Tsvetkov,P., Gil,N., Carolis,C., Buchholz,F., Ulitsky,I., Heard,E. *et al.* (2019) In-cell identification and measurement of RNA-protein interactions. *Nat. Commun.*, **10**, 5317.
17. Lu,Z.P., Zhang,Q.C., Lee,B., Flynn,R.A., Smith,M.A., Robinson,J.T., Davidovich,C., Gooding,A.R., Goodrich,K.J., Mattick,J.S. *et al.* (2016) RNA duplex map in living cells reveals Higher-Order transcriptome structure. *Cell*, **165**, 1267–1279.
18. Pintacuda,G., Wei,G., Roustan,C., Kirmizitas,B.A., Solcan,N., Cerase,A., Castello,A., Mohammed,S., Moindrot,B., Nesterova,T.B. *et al.* (2017) hnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to establish Polycomb-mediated chromosomal silencing. *Mol. Cell*, **68**, 955–969.
19. Colognori,D., Sunwoo,H., Kriz,A.J., Wang,C.Y. and Lee,J.T. (2019) Xist deletional analysis reveals an interdependency between Xist RNA and polycomb complexes for spreading along the Inactive X. *Mol. Cell*, **74**, 101–117.
20. Lee,H.J., Gopalappa,R., Sunwoo,H., Choi,S.W., Ramakrishna,S., Lee,J.T., Kim,H.H. and Nam,J.W. (2019) En bloc and segmental deletions of human XIST reveal X chromosome inactivation-involving RNA elements. *Nucleic Acids Res.*, **47**, 3875–3887.
21. Chow,J.C., Hall,L.L., Baldry,S.E., Thorogood,N.P., Lawrence,J.B. and Brown,C.J. (2007) Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 10104–10109.
22. Wang,Y., Zhong,Y., Zhou,Y., Tanaseichuk,O., Li,Z. and Zhao,J.C. (2019) Identification of a Xist silencing domain by Tiling CRISPR. *Sci. Rep.*, **9**, 2408.
23. Ha,N., Lai,L.T., Chelliah,R., Zhen,Y., Yi Vanessa,S.P., Lai,S.K., Li,H.Y., Ludwig,A., Sandin,S., Chen,L. *et al.* (2018) Live-Cell imaging and functional dissection of Xist RNA reveal mechanisms of X chromosome inactivation and reactivation. *iScience*, **8**, 1–14.
24. Lu,Z., Guo,J.K., Wei,Y., Dou,D.R., Zarnegar,B., Ma,Q., Li,R., Zhao,Y., Liu,F., Choudhry,H. *et al.* (2019) Structural modularity of the XIST ribonucleoprotein complex. bioRxiv doi: https://doi.org/10.1101/837229, 24 November 2019, preprint: not peer reviewed.
25. Kirk,J.M., Kim,S.O., Inoue,K., Smola,M.J., Lee,D.M., Schertzer,M.D., Wooten,J.S., Baker,A.R., Sprague,D., Collins,D.W. *et al.* (2018) Functional classification of long non-coding RNAs by k-mer content. *Nat. Genet.*, **50**, 1474–1482.
26. Smola,M.J., Christy,T.W., Inoue,K., Nicholson,C.O., Friedersdorf,M., Keene,J.D., Lee,D.M., Calabrese,J.M. and Weeks,K.M. (2016) SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 10322–10327.
27. Bousard,A., Raposo,A.C., Zylicz,J.J., Picard,C., Pires,V.B., Qi,Y., Gil,C., Syx,L., Chang,H.Y., Heard,E. *et al.* (2019) The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO Rep.*, **20**, e48019.
28. Bhatt,D.M., Pandya-Jones,A., Tong,A.J., Barozzi,I., Lissner,M.M., Natoli,G., Black,D.L. and Smale,S.T. (2012) Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell*, **150**, 279–290.
29. Wysocka,J., Reilly,P.T. and Herr,W. (2001) Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells. *Mol. Cell. Biol.*, **21**, 3820–3829.
30. Dunagin,M., Cabili,M.N., Rinn,J. and Raj,A. (2015) Visualization of lncRNA by single-molecule fluorescence in situ hybridization. *Methods Mol. Biol.*, **1262**, 3–19.
31. Tatomer,D.C., Liang,D. and Wilusz,J.E. (2017) Inducible expression of eukaryotic circular RNAs from plasmids. *Methods Mol. Biol.*, **1648**, 143–154.
32. Raab,J.R., Smith,K.N., Spear,C.C., Manner,C.J., Calabrese,J.M. and Magnuson,T. (2019) SWI/SNF remains localized to chromatin in the presence of SCHLAP1. *Nat. Genet.*, **51**, 26–29.
33. Gossen,M., Freundlieb,S., Bender,G., Muller,G., Hillen,W. and Bujard,H. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science*, **268**, 1766–1769.
34. Calabrese,J.M., Starmer,J., Schertzer,M.D., Yee,D. and Magnuson,T. (2015) A survey of imprinted gene expression in mouse trophoblast stem cells. *G3 (Bethesda)*, **5**, 751–759.
35. Calabrese,J.M., Sun,W., Song,L., Mugford,J.W., Williams,L., Yee,D., Starmer,J., Mieczkowski,P., Crawford,G.E. and Magnuson,T. (2012) Site-specific silencing of regulatory elements as a mechanism of X inactivation. *Cell*, **151**, 951–963.

36. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
37. Fitzgerald,M. and Shenk,T. (1981) The sequence 5′-AAUAAA-3′ forms parts of the recognition site for polyadenylation of late SV40 mRNAs. *Cell*, **24**, 251–260.
38. Herzel,L., Straube,K. and Neugebauer,K.M. (2018) Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res.*, **28**, 1008–1019.
39. Reimer,K.A., Mimoso,C., Adelman,K. and Neugebauer,K.M. (2020) Rapid and efficient co-transcriptional splicing enhances mammalian gene expression. bioRxiv doi: https://doi.org/10.1101/2020.02.11.944595, 12 February 2020, preprint: not peer reviewed.
40. Younis,I., Berg,M., Kaida,D., Dittmar,K., Wang,C. and Dreyfuss,G. (2010) Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. *Mol. Cell. Biol.*, **30**, 1718–1728.
41. Hart,R.P., McDevitt,M.A., Ali,H. and Nevins,J.R. (1985) Definition of essential sequences and functional equivalence of elements downstream of the adenovirus E2A and the early simian virus 40 polyadenylation sites. *Mol. Cell. Biol.*, **5**, 2975–2983.
42. Simon,M.D., Wang,C.I., Kharchenko,P.V., West,J.A., Chapman,B.A., Alekseyenko,A.A., Borowsky,M.L., Kuroda,M.I. and Kingston,R.E. (2011) The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 20497–20502.
43. Davis,C.P. and West,J.A. (2015) Purification of specific chromatin regions using oligonucleotides: capture hybridization analysis of RNA targets (CHART). *Methods Mol. Biol.*, **1262**, 167–182.
44. Patil,D.P., Chen,C.K., Pickering,B.F., Chow,A., Jackson,C., Guttman,M. and Jaffrey,S.R. (2016) m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature*, **537**, 369–373.
45. Sprague,D., Waters,S.A., Kirk,J.M., Wang,J.R., Samollow,P.B., Waters,P.D. and Calabrese,J.M. (2019) Nonlinear sequence similarity between the Xist and Rsx long noncoding RNAs suggests shared functions of tandem repeat domains. *RNA*, **25**, 1004–1019.
46. Cifuentes-Rojas,C., Hernandez,A.J., Sarma,K. and Lee,J.T. (2014) Regulatory interactions between RNA and polycomb repressive complex 2. *Mol. Cell*, **55**, 171–185.
47. Davidovich,C., Wang,X., Cifuentes-Rojas,C., Goodrich,K.J., Gooding,A.R., Lee,J.T. and Cech,T.R. (2015) Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol. Cell*, **57**, 552–558.
48. Zhao,J., Sun,B.K., Erwin,J.A., Song,J.J. and Lee,J.T. (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, **322**, 750–756.
49. Chen,S.X., Osipovich,A.B., Ustione,A., Potter,L.A., Hipkens,S., Gangula,R., Yuan,W.P., Piston,D.W. and Magnuson,M.A. (2011) Quantification of factors influencing fluorescent protein expression using RMCE to generate an allelic series in the ROSA26 locus in mice. *Dis. Model. Mech.*, **4**, 537–547.
50. Chen,Y.T. and Bradley,A. (2000) A new positive/negative selectable marker, puDeltatk, for use in embryonic stem cells. *Genesis*, **28**, 31–35.
51. Almeida,M., Pintacuda,G., Masui,O., Koseki,Y., Gdula,M., Cerase,A., Brown,D., Mould,A., Innocent,C., Nakayama,M. *et al.* (2017) PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science*, **356**, 1081–1084.
52. Ridings-Figueroa,R., Stewart,E.R., Nesterova,T.B., Coker,H., Pintacuda,G., Godwin,J., Wilson,R., Haslam,A., Lilley,F., Ruigrok,R. *et al.* (2017) The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.*, **31**, 876–888.
53. Yamada,N., Hasegawa,Y., Yue,M., Hamada,T., Nakagawa,S. and Ogawa,Y. (2015) Xist exon 7 contributes to the stable localization of Xist RNA on the inactive X-Chromosome. *PLos Genet.*, **11**, e1005430.
54. Sunwoo,H., Colognori,D., Froberg,J.E., Jeon,Y. and Lee,J.T. (2017) Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 10654–10659.
55. Robinson,J.T., Thorvaldsdottir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.
56. Sunwoo,H., Wu,J.Y. and Lee,J.T. (2015) The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E4216–E4225.
57. Liu,F., Somarowthu,S. and Pyle,A.M. (2017) Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat. Chem. Biol.*, **13**, 282–289.
58. Berg,M.G., Singh,L.N., Younis,I., Liu,Q., Pinto,A.M., Kaida,D., Zhang,Z., Cho,S., Sherrill-Mix,S., Wan,L. *et al.* (2012) U1 snRNP determines mRNA length and regulates isoform expression. *Cell*, **150**, 53–64.
59. Kaida,D., Berg,M.G., Younis,I., Kasim,M., Singh,L.N., Wan,L. and Dreyfuss,G. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, **468**, 664–668.
60. Oh,J.M., Di,C., Venters,C.C., Guo,J., Arai,C., So,B.R., Pinto,A.M., Zhang,Z., Wan,L., Younis,I. *et al.* (2017) U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat. Struct. Mol. Biol.*, **24**, 993–999.
61. Yin,Y., Lu,J.Y., Zhang,X., Shao,W., Xu,Y., Li,P., Hong,Y., Cui,L., Shan,G., Tian,B. *et al.* (2020) U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature*, **580**, 147–150.
62. Gregersen,L.H., Mitter,R., Ugalde,A.P., Nojima,T., Proudfoot,N.J., Agami,R., Stewart,A. and Svejstrup,J.Q. (2019) SCAF4 and SCAF8, mRNA Anti-Terminator proteins. *Cell*, **177**, 1797–1813.
63. Kainov,Y.A. and Makeyev,E.V. (2020) A transcriptome-wide antitermination mechanism sustaining identity of embryonic stem cells. *Nat. Commun.*, **11**, 361.
64. Chen,S., Wang,R., Zheng,D., Zhang,H., Chang,X., Wang,K., Li,W., Fan,J., Tian,B. and Cheng,H. (2019) The mRNA export receptor NXF1 coordinates transcriptional dynamics, alternative polyadenylation, and mRNA export. *Mol. Cell*, **74**, 118–131.
65. Tian,B. and Manley,J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.*, **18**, 18–30.
66. Cheng,Y., Miura,R.M. and Tian,B. (2006) Prediction of mRNA polyadenylation sites by support vector machine. *Bioinformatics*, **22**, 2320–2325.
67. Chiu,A.C., Suzuki,H.I., Wu,X., Mahat,D.B., Kriz,A.J. and Sharp,P.A. (2018) Transcriptional pause sites delineate stable nucleosome-associated premature polyadenylation suppressed by U1 snRNP. *Mol. Cell*, **69**, 648–663.
68. Dubbury,S.J., Boutz,P.L. and Sharp,P.A. (2018) CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature*, **564**, 141–145.
69. Elrod,N.D., Henriques,T., Huang,K.L., Tatomer,D.C., Wilusz,J.E., Wagner,E.J. and Adelman,K. (2019) The integrator complex attenuates Promoter-Proximal transcription at Protein-Coding genes. *Mol. Cell*, **76**, 738–752.
70. Erickson,B., Sheridan,R.M., Cortazar,M. and Bentley,D.L. (2018) Dynamic turnover of paused Pol II complexes at human promoters. *Genes Dev.*, **32**, 1215–1225.
71. Kamieniarz-Gdula,K., Gdula,M.R., Panser,K., Nojima,T., Monks,J., Wisniewski,J.R., Riepsaame,J., Brockdorff,N., Pauli,A. and Proudfoot,N.J. (2019) Selective roles of vertebrate PCF11 in premature and full-length transcript termination. *Mol. Cell*, **74**, 158–172.
72. Krebs,A.R., Imanci,D., Hoerner,L., Gaidatzis,D., Burger,L. and Schubeler,D. (2017) Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell*, **67**, 411–422.
73. Lee,S.H., Singh,I., Tisdale,S., Abdel-Wahab,O., Leslie,C.S. and Mayr,C. (2018) Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature*, **561**, 127–131.
74. Ni,T.K. and Kuperwasser,C. (2016) Premature polyadenylation of MAGI3 produces a dominantly-acting oncogene in human breast cancer. *Elife*, **5**, e14730.
75. Singh,I., Lee,S.H., Sperling,A.S., Samur,M.K., Tai,Y.T., Fulciniti,M., Munshi,N.C., Mayr,C. and Leslie,C.S. (2018) Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat. Commun.*, **9**, 1716.
76. Steurer,B., Janssens,R.C., Geverts,B., Geijer,M.E., Wienholz,F., Theil,A.F., Chang,J., Dealy,S., Pothof,J., van Cappellen,W.A. *et al.*

(2018) Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E4368–E4376.
77. Kamieniarz-Gdula,K. and Proudfoot,N.J. (2019) Transcriptional control by premature termination: a forgotten mechanism. *Trends Genet.*, **35**, 553–564.
78. Goodson,J.R. and Winkler,W.C. (2018) Processive antitermination. *Microbiol Spectr*, **6**.
79. Koerner,M.V., Pauler,F.M., Hudson,Q.J., Santoro,F., Sawicka,A., Guenzl,P.M., Stricker,S.H., Schichl,Y.M., Latos,P.A., Klement,R.M. *et al.* (2012) A downstream CpG island controls transcript initiation and elongation and the methylation state of the imprinted Airn macro ncRNA promoter. *PLoS Genet.*, **8**, e1002540.
80. Guo,Y.E., Manteiga,J.C., Henninger,J.E., Sabari,B.R., Dall'Agnese,A., Hannett,N.M., Spille,J.H., Afeyan,L.K., Zamudio,A.V., Shrinivas,K. *et al.* (2019) Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*, **572**, 543–548.
81. Ji,X., Zhou,Y., Pandit,S., Huang,J., Li,H., Lin,C.Y., Xiao,R., Burge,C.B. and Fu,X.D. (2013) SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell*, **153**, 855–868.
82. Lin,S., Coutinho-Mansfield,G., Wang,D., Pandit,S. and Fu,X.D. (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.*, **15**, 819–826.
83. Wei,C., Xiao,R., Chen,L., Cui,H., Zhou,Y., Xue,Y., Hu,J., Zhou,B., Tsutsui,T., Qiu,J. *et al.* (2016) RBFox2 binds nascent RNA to globally regulate polycomb complex 2 targeting in mammalian genomes. *Mol. Cell*, **62**, 875–889.
84. Hendrickson,D.G., Kelley,D.R., Tenen,D., Bernstein,B. and Rinn,J.L. (2016) Widespread RNA binding by chromatin-associated proteins. *Genome Biol.*, **17**, 28.
85. Schertzer,M.D., Braceros,K.C.A., Starmer,J., Cherney,R.E., Lee,D.M., Salazar,G., Justice,M., Bischoff,S.R., Cowley,D.O., Ariel,P. *et al.* (2019) lncRNA-Induced spread of polycomb controlled by genome architecture, RNA abundance, and CpG island DNA. *Mol. Cell*, **75**, 523–537.
86. Colognori,D., Sunwoo,H., Wang,D., Wang,C.Y. and Lee,J.T. (2020) Xist repeats A and B account for two distinct phases of X inactivation establishment. *Dev. Cell*, **54**, 21–32.