

## GENETICS

# Zscan4 binds nucleosomal microsatellite DNA and protects mouse two-cell embryos from DNA damage

Rajini Srinivasan<sup>1\*</sup>, Nataliya Nady<sup>1\*</sup>, Neha Arora<sup>1\*</sup>, Laura J. Hsieh<sup>2</sup>, Tomek Swigut<sup>1</sup>, Geeta J. Narlikar<sup>2</sup>, Mark Wossidlo<sup>1,3†</sup>, Joanna Wysocka<sup>1,4,5,6†</sup>

Zinc finger protein Zscan4 is selectively expressed in mouse two-cell (2C) embryos undergoing zygotic genome activation (ZGA) and in a rare subpopulation of embryonic stem cells with 2C-like features. Here, we show that Zscan4 specifically recognizes a subset of (CA)<sub>n</sub> microsatellites, repeat sequences prone to genomic instability. Zscan4-associated microsatellite regions are characterized by low nuclease sensitivity and high histone occupancy. In vitro, Zscan4 binds nucleosomes and protects them from disassembly upon torsional strain. Furthermore, Zscan4 depletion leads to elevated DNA damage in 2C mouse embryos in a transcription-dependent manner. Together, our results identify Zscan4 as a DNA sequence-dependent microsatellite binding factor and suggest a developmentally regulated mechanism, which protects fragile genomic regions from DNA damage at a time of embryogenesis associated with high transcriptional burden and genomic stress.

## INTRODUCTION

The development of multicellular animals is initially controlled by maternally deposited transcripts and switches to the zygotic genome during the first major developmental milestone called maternal-to-zygotic transition (1). During this transition, maternal mRNAs are degraded, and zygotic transcription begins. Zygotic genome activation (ZGA) is associated with massive induction of transcription and widespread changes in chromatin architecture (2), placing enormous stress on the embryo and requiring mechanisms that ensure genome integrity at this complex developmental time point (3, 4). In mouse embryos, ZGA occurs at the two-cell (2C) stage. First, a minor wave of zygotic transcription is initiated from long terminal repeats (LTRs) of a specific class of mouse retrotransposons, MERVL (murine endogenous retrovirus with leucine tRNA primer), leading to expression of a set of 2C-specific chimeric transcripts (5). Subsequently, during the G<sub>2</sub> phase of the 2C stage, the major wave of ZGA occurs, resulting in transcriptional up-regulation of thousands of genes (6).

One of the 2C-specific transcripts activated during the minor ZGA is *Zscan4* (Zinc finger and SCAN domain-containing 4), encoded by a highly homologous six-gene cluster (*Zscan4a-f*) arising from recent duplications in mice (7). In mouse embryonic stem cell (mESC) cultures, Zscan4 is expressed in 1 to 5% of cells and is one of the earliest markers of a unique subpopulation transitioning through a 2C-like state (8). These spontaneously arising 2C-like cells recapitulate a subset of molecular and developmental features of the totipotent 2C mouse embryo, including the activity of

MERVL elements and expression of chimeric 2C transcripts arising from their LTRs, thus offering a model system to study molecular functions of Zscan4 and other proteins encoded by the 2C-specific transcripts. Although only a small subset of mESCs are in the 2C-like state at any given time, all cells fluctuate in and out of this state (9). Moreover, during these fluctuations, more than 80% of 2C-like cells arise from Zscan4-expressing cells, and it was therefore postulated that *Zscan4* transcription is the first molecular marker in the transition from mESCs toward 2C-like cells (8).

Curiously, despite the remarkable developmental specificity of Zscan4 expression during the 2C stage, this protein is not essential for mouse preimplantation development or for transition into the 2C-like state in mESCs (7, 10). Instead, perturbation of Zscan4 function during development or in mESCs leads to implantation failure, telomere maintenance defects, and chromosomal rearrangements (7, 9). During cellular reprogramming, the 2C-like transcriptional network is transiently induced, and Zscan4 was implicated in promoting developmental competence of resulting iPSCs (induced pluripotent stem cells) (11). Furthermore, Zscan4 was recently linked to DNA hypomethylation of 2C-like cells through promoting degradation of maintenance DNA methyltransferase complex, which, in turn, facilitates telomere elongation (10). Overall, although Zscan4 has been postulated to act as a transcription factor, its genomic occupancy patterns or mode of action on chromatin remains enigmatic. Here, we used a combination of experiments in 2C-like cells, 2C mouse embryos, and in vitro reconstitutions to probe molecular functions of Zscan4. Unlike a typical transcription factor, Zscan4 does not associate with cis-regulatory elements such as enhancers, promoters, or other nuclease hypersensitive regions. Instead, it occupies a subset of (CA)<sub>n</sub> microsatellite repeats in their nucleosomal form. In vitro, Zscan4's zinc finger domain (ZnF) directly binds nucleosomes and protects them from disassembly in the presence of curaxin, a DNA-intercalating drug, previously suggested to induce torsional strain and promote Z-DNA formation at prone regions such as (CA)<sub>n</sub> microsatellites (12). In the mouse 2C embryos undergoing ZGA, knockdown of Zscan4 induces DNA damage response in a transcription-dependent manner, altogether suggesting that Zscan4 functions in protecting

<sup>1</sup>Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>2</sup>Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA. <sup>3</sup>Department of Cell and Developmental Biology, Medical University of Vienna, 1090 Vienna, Austria. <sup>4</sup>Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>5</sup>Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>6</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA.

\*These authors contributed equally to this work.

†Corresponding author. Email: wysocka@stanford.edu (J.W.); mark.wossidlo@meduniwien.ac.at (M.W.)

fragile microsatellite regions from torsional strain induced by massive up-regulation of transcription.

## RESULTS

### Zscan4 occupies CA-repeat sequences in mouse 2C-like cells

To investigate the function of Zscan4 in 2C-like cells, we generated clonal mESC reporter lines with *Zscan4* promoter (9) driving expression of either green fluorescent protein (GFP) alone (*Zprom::GFP*) or an N-terminally tagged GFP-Zscan4 fusion protein (*Zprom::GFP-Zscan4*). Expression of the transgene was confirmed by immunoblotting and transcriptome analysis (fig. S1, A to D). Consistent with the previous studies, we found by fluorescence-activated cell sorting (FACS) that GFP<sup>+</sup> cells comprised 1 to 5% of the population (fig. S1B) and were enriched for *Zscan4* and other 2C-specific transcripts (fig. S1, C and D). To profile direct targets of Zscan4 in 2C-like cells genome-wide, we conducted chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing [ChIP sequencing (ChIP-seq)] using a GFP antibody in a sorted population of 2C-like cells (GFP<sup>+</sup> population) from the *Zprom::GFP-Zscan4* transgenic line and the control *Zprom::GFP* line (Fig. 1A). We also profiled genomic occupancy of endogenous Zscan4 using a Zscan4 antibody in the *Zprom::GFP* line (GFP<sup>+</sup> population and control GFP<sup>-</sup> population; Fig. 1A).

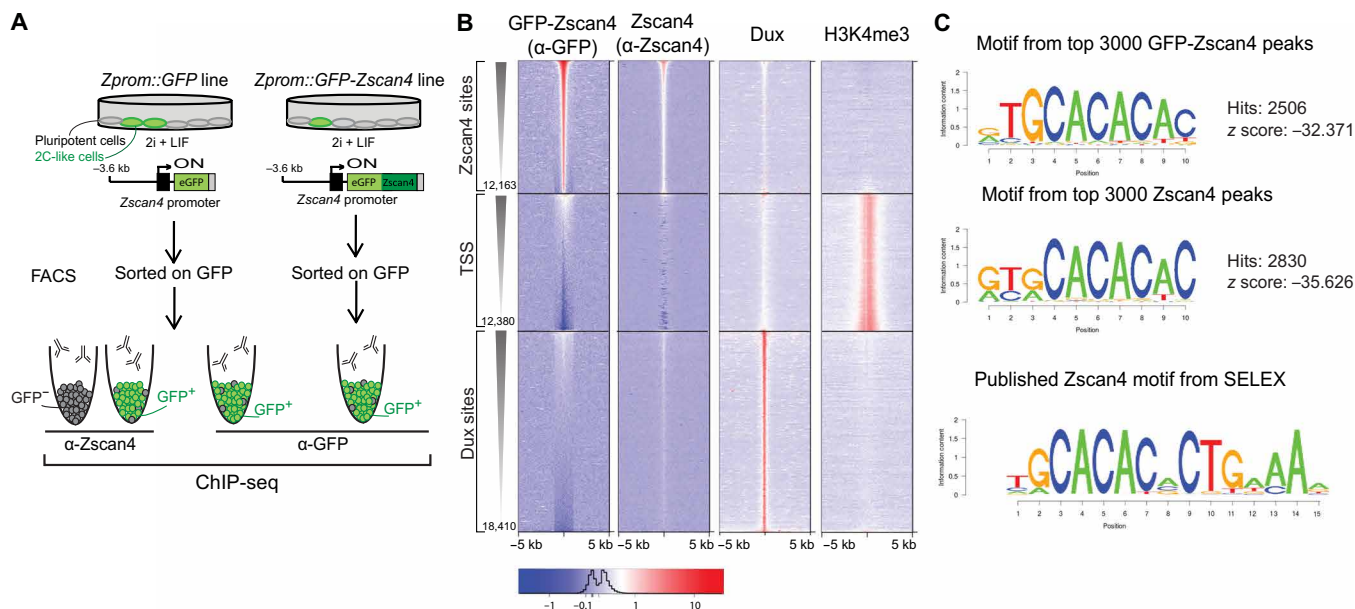
Heat map representation of each ChIP-seq experiment sorted by signal strength of GFP-Zscan4 showed concordance between genomic sites occupied by the endogenous and the transgenic Zscan4 (although GFP-Zscan4 ChIP enrichments were generally stronger than those of endogenous Zscan4; Fig. 1B). Next, to determine the genomic features of Zscan4 sites and its involvement in transcrip-

tional regulation of 2C-like cells, we compared the binding pattern of Zscan4 with (i) Dux, a major transcriptional driver of the 2C-like state, which binds and activates MERV1 elements and other distal cis-regulatory sites in this cellular state (13), and (ii) previously reported active transcriptional start sites (TSSs) in 2C embryos that are marked by H3K4me3 (14) (Fig. 1B). We found that Zscan4 binding was depleted at Dux binding sites and TSSs, indicating that Zscan4 and Dux bind to discrete distal elements that differ from canonical cis-regulatory elements in the 2C-like state and may play distinct roles during development.

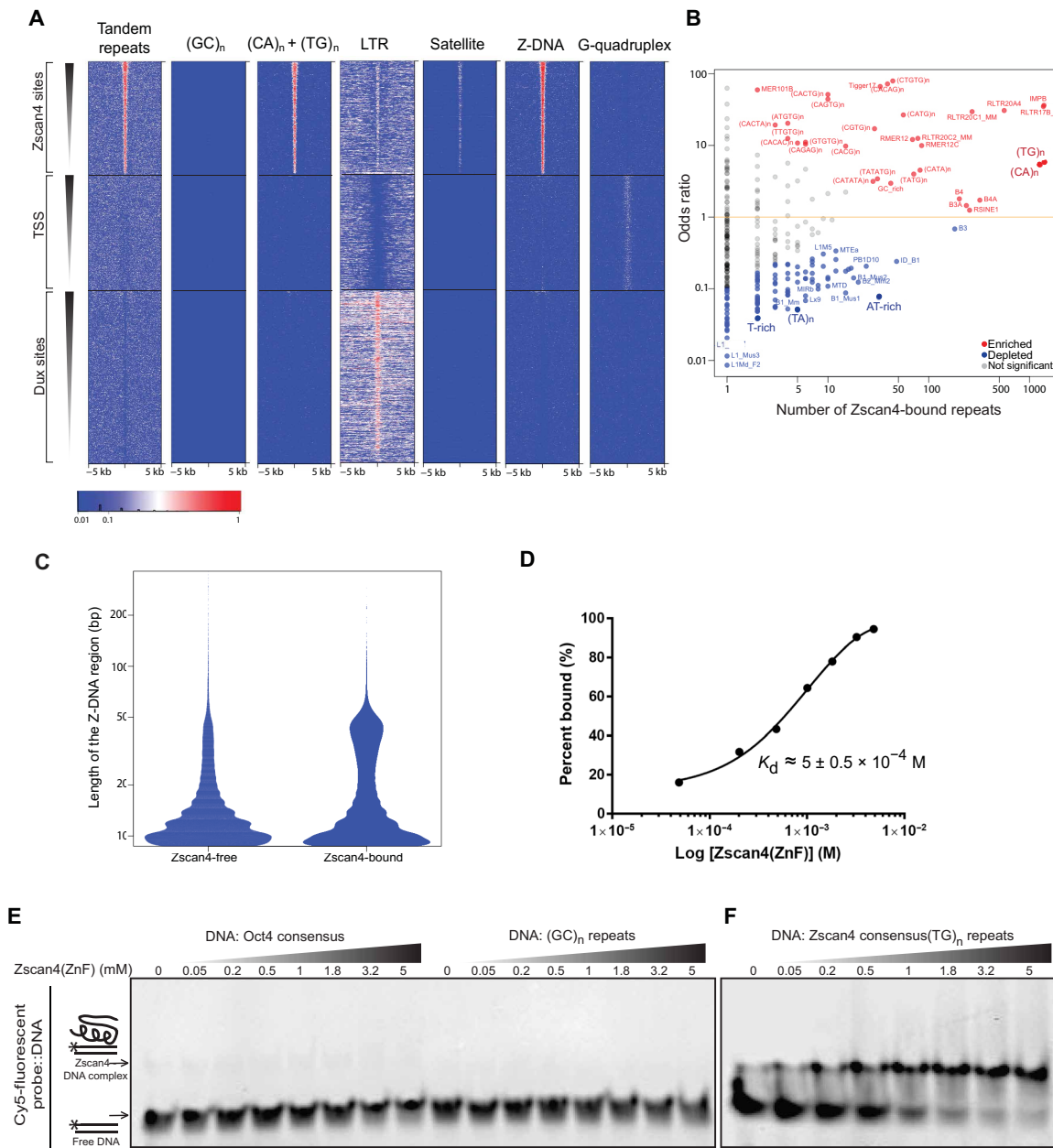
Examination of the underlying sequence from the top 3000 peaks in both Zscan4 ChIP-seq datasets identified a motif that encompassed CA dinucleotide repeats [which included an inverted dinucleotide (TG); Fig. 1C]. This sequence was similar to the previously identified Zscan4 motif from the in vitro SELEX (systematic evolution of ligands by exponential enrichment) experiments (15) (Fig. 1C).

### Zscan4 binds to microsatellite DNA in a sequence-specific manner

Given the presence of CA repeats within the Zscan4 motif and the fact that Zscan4 binding overlapped with neither active regions nor Dux sites, we next systematically investigated the enrichment of different repetitive element classes at Zscan4 targets. We found strong occupancy of Zscan4 at tandem repeats, specifically at short (TG)<sub>n</sub>/ (CA)<sub>n</sub> repeat sequences, but not at (GC)<sub>n</sub> repeats (Fig. 2, A and B). Under negative superhelical stress, these purine:pyrimidine dinucleotide repeats have the propensity to adopt a transient alternative DNA structure called Z-DNA, a rigid, left-handed helix conformation that cannot wrap around nucleosomes (16, 17). We found strong association between Zscan4 targets and computationally



**Fig. 1. Analysis of Zscan4 genomic occupancy in 2C-like cells.** (A) Schematic of the workflow describing the two mESC reporter lines and FACS strategy for ChIP-seq experiments. Reporter lines contain a transgene with a 3.6-kb region upstream from the *Zscan4* open reading frame driving either GFP (*Zprom::GFP*, left) or a GFP-Zscan4 transgene (*Zprom::GFP-Zscan4*; right). LIF, leukemia inhibitory factor. (B) Heat map representation of GFP-Zscan4, endogenous Zscan4, Dux (13), and H3K4me3 (14) ChIP-seq signal enrichments over indicated number of Zscan4 sites, active TSSs, and Dux sites ( $\pm 5$  kb from the center). Heat maps are sorted by the strength of GFP-Zscan4 ChIP-seq signals. The relative signal intensity is indicated in a color scale. Dux and H3K4me3 datasets are from (13, 14). (C) The top sequence motif recovered from the top 3000 peaks in Zscan4 ChIP-seq is highly similar to the GFP-Zscan4 motif and partially overlaps with the previously published SELEX motif (15). Logos for the consensus motifs were generated using SeqPos.

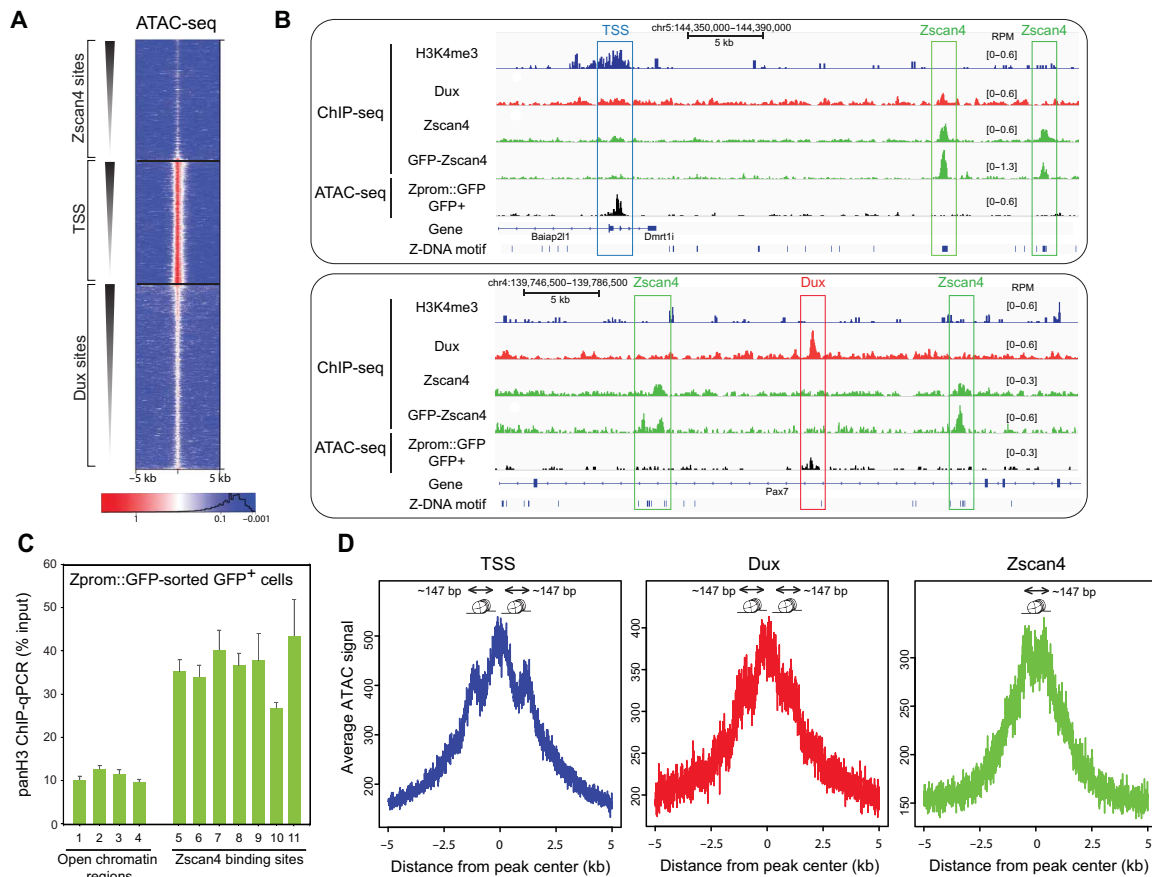


**Fig. 2. Zscan4 binds to DNA in a sequence-specific manner.** (A) Heat maps display density of repetitive elements and sequence features over Zscan4, active TSSs, and Dux sites. (B) Degree of enrichment/depletion of Zscan4 occupancy at indicated sequence repeats, with the strongest enrichment observed at (TG)<sub>n</sub> and (CA)<sub>n</sub> repeats and their variations. (C) Violin plot of the size distribution of the predicted Z-DNA prone regions, for Zscan-bound (right) and Zscan-free (left) instances. (D) Titration curve quantifying the binding affinity of purified Zscan4(ZnF) to its consensus TG repeat sequence in vitro.  $K_d$  is the mean value of three independent experiments. (E and F) EMSAs showing interaction of increasing amounts of Zscan4(ZnF) as indicated with a Cy5-labeled oligo containing an Oct4 consensus sequence, a (GC)<sub>n</sub> sequence (E), or a (TG)<sub>n</sub> repeat (F).

predicted Z-DNA prone regions but no detectable overlap with A-DNA or G-quadruplex DNA (Fig. 2, A and B). In addition, Zscan4-bound Z-DNA prone regions were typically longer [20 to 70 base pairs (bp)] than those devoid of Zscan4 (Fig. 2C). Similarly, Zscan4 binding was not detected at DNA transposons, Long INterspersed elements (LINEs), or Short INterspersed elements (SINEs; fig. S2A), and was specifically depleted from repetitive regions that include AT-rich and T-rich elements (Fig. 2B). While simple repeat sequences demonstrated the strongest enrichment, we also observed binding

at certain subclass of LTRs and satellite repeats (Fig. 2, A and B, and fig. S2A). On closer examination, these elements contained stretches of (CA)<sub>n</sub> or (TG)<sub>n</sub> tracts (fig. S2, A and B), underlining the specificity for (TG)<sub>n</sub>/(CA)<sub>n</sub> repeats.

To test direct binding of Zscan4 to the identified repeat sequence, we performed electrophoretic mobility shift assays (EMSAs) with short oligos (under conditions that favor the B-DNA form) and recombinant Zscan4 ZnF. We confirmed concentration-dependent association with (TG)<sub>n</sub>/(CA)<sub>n</sub> repeats [ $K_d$  (dissociation constant) ~ 0.5 mM;



**Fig. 3. Zscan4 associates with nucleosome-rich regions in 2C-like cells.** (A) Heat map of ATAC-seq signal from 2C-like cells FACS-sorted from the *Zprom::GFP* line. Signals were centered and sorted as in Fig. 1B. (B) Representative browser tracks illustrating ChIP-seq profiles from H3K4me3 (blue), Dux (red), endogenous and transgenic Zscan4 (green), and ATAC-seq (black). Z-DNA motif enrichment is shown at the bottom. Z-DNA motif predictions were downloaded from the non-B DB database (47). Colored rectangles highlight examples of Zscan4 binding sites (green), a Dux binding site (red), and an active TSS (blue). (C) ChIP-qPCR analysis in 2C-like cells (GFP<sup>+</sup>) FACS-sorted from the *Zprom::GFP* line measuring H3 occupancy, at a representative panel of Zscan4 binding sites and open chromatin regions, as determined by ATAC-seq. Error bars denote SD from three replicates. Primer sequences are provided in table S1. (D) Average ATAC-seq signal from reads > 147 bp, indicating nucleosome positioning at TSSs, Dux, and Zscan4 sites. Signal enrichment at the center of TSS and Dux sites indicates open chromatin with positional nucleosomes on either side, while a dip in signal at the center of Zscan4 binding site suggests nucleosomal protection.

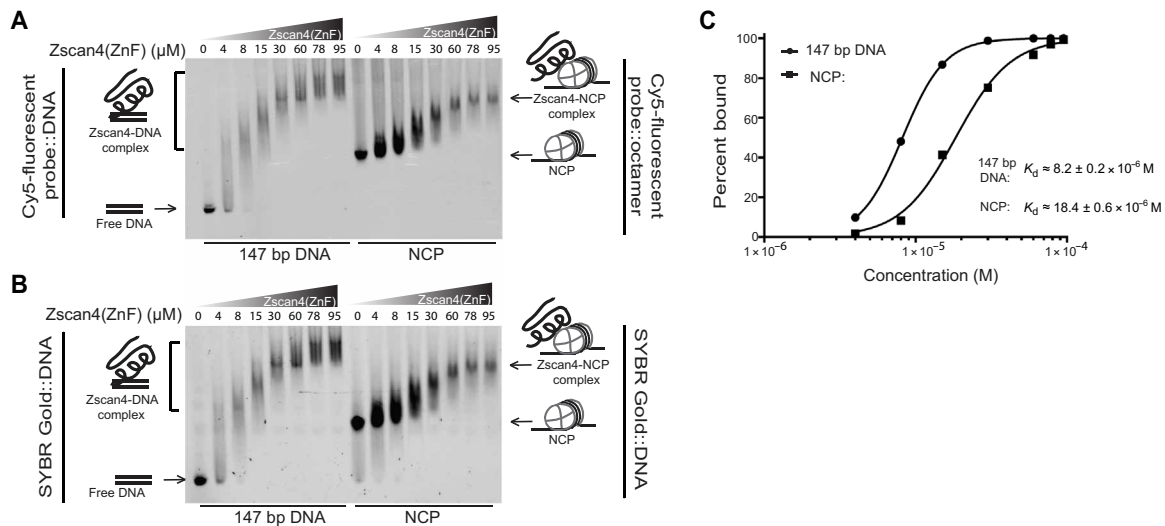
Fig. 2, D and F]. In contrast, we detected no binding at another dinucleotide sequence, (GC)<sub>n</sub>, or at the Oct4 consensus sequence (Fig. 2E), supporting sequence-specific recognition of DNA by Zscan4. Overall, our results reveal a unique binding profile for Zscan4 in 2C-like cells with an affinity for (TG)<sub>n</sub>/(CA)<sub>n</sub> microsatellite repeat sequences.

### Zscan4 occupies nucleosome-bound regions in 2C-like cells

Most transcription factor occupancy sites in mammalian genomes are associated with nuclease/transposase hypersensitive regions, indicative of nucleosomal depletion (18). In contrast, repetitive DNA is often transcriptionally silent and found within regions where the chromatin is more compacted. To examine chromatin accessibility at Zscan4 binding sites, we performed ATAC-seq [Assay for Transposase Accessible Chromatin using sequencing; (19)] in 2C-like cells with endogenous Zscan4 expression (GFP<sup>+</sup> population isolated from the *Zprom::GFP* mESCs; fig. S3A). As expected, TSSs in this population were highly accessible and, to a lesser degree, distal sites occupied by Dux were also associated with open chromatin (Fig. 3A and fig. S3B). In contrast, Zscan4 sites had very low ATAC-seq sig-

nal enrichments (Fig. 3A and fig. S3B) and typically did not overlap with clear ATAC-seq peaks, Dux occupancy, and H3K4me3 mark (representative examples shown in Fig. 3B), suggestive of nucleosome occupancy at these sites. Instead, Zscan4 peaks overlapped with wide stretches of putative Z-DNA-forming regions (Fig. 3B, bottom track), consistent with Zscan4 binding at Z-DNA prone (CA)<sub>n</sub> repeats (Fig. 2, A and B). Low transposase hypersensitivity over Zscan4 sites was corroborated by the average ATAC-seq signal profiles at top 1000 sites bound by Zscan4, TSS, or Dux in ChIP-seq (fig. S3B). To exclude the possibility that these low signals may be due to the Tn5 transposase sequence bias at highly repetitive (TG)<sub>n</sub>/(CA)<sub>n</sub> sites, we performed pan-H3 ChIP-quantitative polymerase chain reaction (qPCR) at select Zscan4 target sites in 2C-like cells (Fig. 3C). Consistent with the ATAC-seq data, Zscan4 binding sites have relatively higher histone H3 content, as compared to open chromatin regions.

To profile nucleosome positioning at Zscan4 binding sites, we analyzed ATAC-seq data using only reads consistent with (or longer than) the approximate length of DNA protected by a nucleosome, 147 nucleotides (nt). Both TSSs and Dux sites had overall similar



**Fig. 4. Zscan4 binds to nucleosomes in vitro.** (A and B) EMSAs showing the interaction of increasing amounts of Zscan4(ZnF), as indicated, with 147 bp of Cy5-labeled DNA sequence containing (TG)<sub>n</sub> consensus sequence (left) and with Cy5-labeled NCP, assembled with the same 147 bp of DNA (right). Detection with Cy5 fluorophore is shown in (A) and detection with SYBR Gold for visualization of nucleic acids is shown in (B). (C) Titration curve quantifying the binding affinity of purified Zscan4(ZnF) under identical conditions to the naked 147-bp DNA and NCP in vitro.  $K_d$  is the mean value of three independent experiments.

profiles, with relative depletion at the center and enrichment of +1 and -1 positional nucleosomes on either side (Fig. 3D). However, Zscan4 sites had a distinct profile, showing protection of ~147 nt at the center, suggestive of occupancy by a nucleosome (Fig. 3D). Although (TG)<sub>n</sub>/(CA)<sub>n</sub> microsatellite sequences bound by Zscan4 are susceptible to Z-DNA formation, nucleosomal occupancy at these sites suggests that in a substantial proportion of cells within the 2C-like population, they adopt the B-DNA conformation, as Z-DNA is rigid and disfavors octamer wrapping (17).

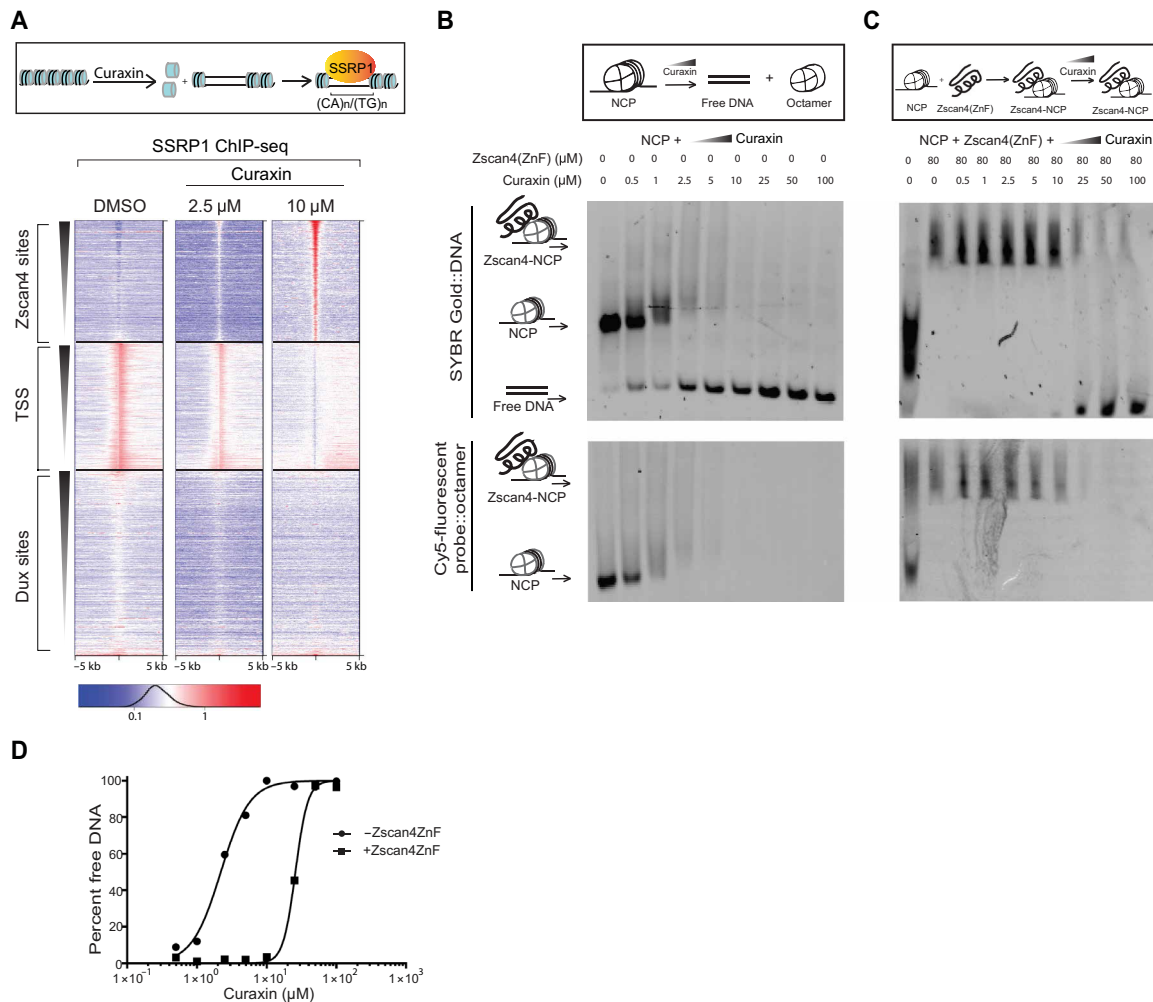
### Zscan4 directly binds nucleosomes

To demonstrate direct binding of Zscan4 to nucleosomes in vitro, we assembled recombinant nucleosome core particles (NCPs) that contained 147 bp corresponding to one of the strongest native Zscan4 binding sequences identified in our ChIP-seq experiments and an octamer with a Cy5-labeled histone H3. Zscan4(ZnF) binding to NCPs was detected by monitoring emergence of a size-shifted NCP band as visualized by the labeled octamer (Fig. 4A, right) or SYBR Gold-stained DNA (Fig. 4B, right). We compared binding to the corresponding 147-bp sequence of Cy5-labeled DNA alone under conditions identical to those used to detect NCP binding and visualized the interaction by Cy5 (Fig. 4A, left) or SYBR Gold-stained DNA (Fig. 4B, left). We detected a saturable interaction of Zscan4(ZnF) with both naked DNA and NCP, with dissociation constants of ~8 and ~18 μM, respectively (Fig. 4C). Notably, shifts observed on the naked DNA were even more pronounced than those observed on the nucleosomal DNA, likely because this 147-bp sequence contains 11 Zscan4 binding sites, some of which may not be available for binding when the DNA is wrapped around the nucleosome. Nonetheless, our results indicated that the Zscan4 zinc fingers confer sequence-dependent DNA binding and also have robust nucleosome-binding activity. We confirmed that Zscan4(ZnF) and the histone octamer are part of the same complex by performing Western-EMSA and probing for histone H3 antibody or Cy5-labeled octamer (fig. S4); the shift in H3 or Cy5 signal appears only upon addition of Zscan4 protein. Furthermore, because EMSA assays were performed under

conditions favoring B-DNA conformation, and Z-DNA structure is incompatible with nucleosome assembly, we conclude that Zscan4 must recognize Z-DNA susceptible (TG)<sub>n</sub>/(CA)<sub>n</sub> sequences in their B-DNA form.

### Zscan4 favors a stable nucleosomal state

Transcription-coupled DNA supercoiling may predispose the genome to topological stress leading to nucleosomal destabilization and/or facilitate formation of non-B-DNA structures (20). We hypothesized that Zscan4 binds microsatellite repeats in their nucleosomal form and protects them from nucleosomal disassembly under conditions of torsional stress. In cancer cells, nucleosome disassembly has been reported to occur upon treatment with the anticancer drug curaxin (cb10137), which induces torsional stress (12). Curaxin intercalates into DNA, destabilizes nucleosomes, and the resulting negative supercoiling promotes B-DNA-to-Z-DNA transition at susceptible regions (12). Recent reports showed that in curaxin-treated cancer cells, nucleosome disassembly and Z-DNA formation are sensed by the SSRP1 (structure specific recognition protein 1) subunit of the facilitates chromatin transcription (FACT) complex (12). Upon curaxin treatment, SSRP1 relocalizes from actively transcribed genes to these nucleosome-destabilized regions and becomes especially enriched at the (TG)<sub>n</sub>/(CA)<sub>n</sub> repeats, suggesting that this type of repeat may be particularly sensitive to nucleosome disassembly sensed by the FACT complex (12). Our reanalysis of the SSRP1 ChIP-seq from human fibrosarcoma HT1080 cells revealed that upon curaxin treatment, SSRP1 relocalized to (TG)<sub>n</sub>/(CA)<sub>n</sub> repeats, but not to other Z-DNA prone sequences such as (GC)<sub>n</sub> (fig. S5A). Given that previous experiments were performed in human cancer lines, we asked whether curaxin treatment also induces SSRP1 relocalization to (TG)<sub>n</sub>/(CA)<sub>n</sub> repeats in mESCs and whether these sensitive sites correspond to those occupied by Zscan4 in 2C-like cells. To this end, we conducted SSRP1 ChIP-seq in mESCs treated with two different concentrations of curaxin (Fig. 5A). In response to the treatment, SSRP1 binding relocalized from TSSs to Zscan4 binding sites in a drug dose-dependent manner, suggesting that



**Fig. 5. Zscan4 stabilizes nucleosomal DNA under torsional strain.** (A) Top: Schematic illustrating the relationship between curaxin and SSRP1. Curaxin intercalation into DNA results in nucleosome destabilization, negative supercoiling, and conversion of susceptible microsatellite regions to Z-DNA, which is recognized by SSRP1 (12). Bottom: Heat maps of SSRP1 ChIP-seq in mESCs treated with DMSO or indicated concentrations of curaxin. ChIP-seq signal enrichments were sorted and centered as in Fig. 1B. (B and C) EMSAs monitoring interaction between NCP and Zscan4(ZnF) in the presence of curaxin. Schematic above each panel summarizes the reaction performed. Gels were visualized using the fluorescent Cy5 signal present on the octamer (bottom panels) and SYBR Gold for visualization of nucleic acids (top panels). (D) Titration curve displaying the relationship between increasing concentrations of curaxin and percentage of free DNA released upon nucleosome destabilization, in the presence and absence of Zscan4(ZnF).

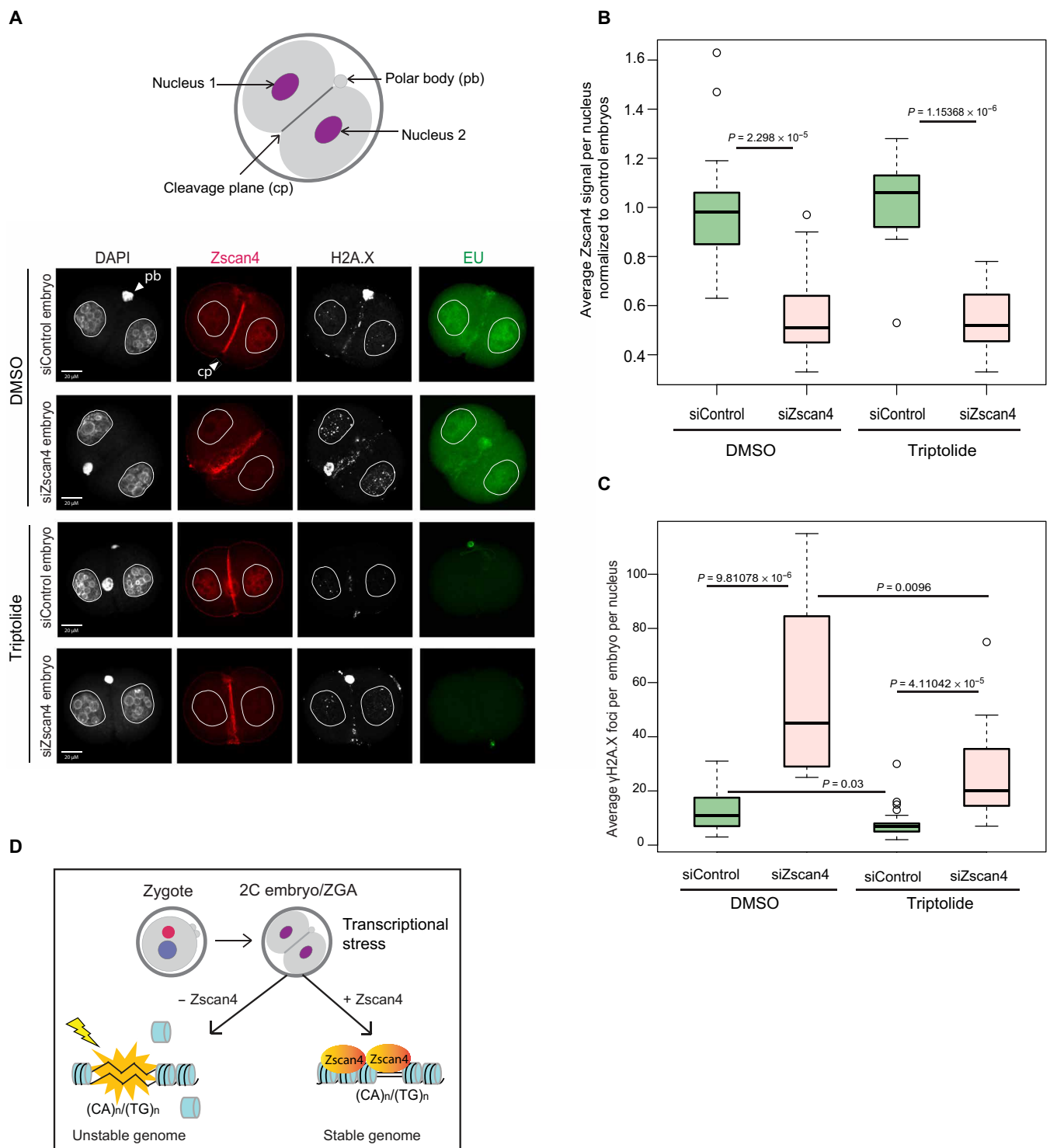
these sites are susceptible to nucleosomal instability and potentially also to B-form-to-Z-form transition in mESCs. Consistently, *in vitro* analysis of B-to-Z conversion by circular dichroism (CD) spectroscopy demonstrated that a (TG)<sub>n</sub>/(CA)<sub>n</sub> repeat corresponding to a consensus Zscan4 binding site has a higher propensity for conversion upon curaxin treatment than (GC)<sub>n</sub> repeats (fig. S5B).

To directly examine whether Zscan4 can protect consensus (CA)<sub>n</sub>/(TG)<sub>n</sub> sequences from nucleosome disassembly under conditions of torsional stress, we used an *in vitro* reconstitution system to detect the interaction between recombinantly assembled NCPs and the purified Zscan4(ZnF) in the presence of curaxin. First, NCP alone is destabilized by the addition of curaxin in a dose-dependent manner, releasing the octamer and free DNA at concentrations of curaxin as low as 1 μM (Fig. 5B). Second, upon addition of saturating amounts of Zscan4(ZnF) protein to the NCP and allowing for complex formation, curaxin disrupts NCP or the NCP-Zscan4(ZnF) complex only at much higher concentrations, with 25 μM drug concentra-

tion required to destabilize the nucleosome and release free DNA (Fig. 5, C and D). Thus, Zscan4 appears to prevent disassembly of the nucleosome, suggesting that under torsional stress, it protects repetitive DNA by maintaining a nucleosome-rich state, thereby preventing potential formation of fragile alternative DNA structures, such as Z-DNA.

### Zscan4 loss induces DNA damage in the embryo

Microsatellite and Z-DNA prone regions are hotspots for genetic instability (21–25). We hypothesized that the massive transcriptional up-regulation associated with ZGA in mouse 2C embryos, combined with the overall decreased nucleosomal density, may subject the genome of an early embryo to an unusual level of negative supercoiling associated with Pol II passage and in turn lead to DNA damage, especially at fragile sites at which transcription occurs. To examine whether Zscan4-bound microsatellites are transcribed during ZGA, we analyzed previously reported mouse embryo expression



**Fig. 6. Zscan4 depletion leads to transcriptionally dependent elevation of DNA damage in 2C embryos.** (A) Representative immunofluorescence images of  $\gamma$ H2A.X and Zscan4 staining in mouse 2C embryos in G<sub>2</sub> stage (31 hpf), derived from zygotes injected with control or Zscan4 siRNA and treated with DMSO or triptolide (2 hours prior fixation in the presence of EU) as indicated. (B and C) Quantification of Zscan4 staining (B) or  $\gamma$ H2A.X foci (C) in siControl (green) and siZscan4 (pink) 2C embryos (31 hpf) treated with DMSO (23 siControl and 12 siZscan4 embryos) or triptolide (21 siControl and 15 siZscan4 embryos) as shown. P values were determined by Wilcoxon test. (D) Proposed model of transcriptionally dependent regulation of genome stability by Zscan4 in early development. See the main text for details.

data (26). We found that over a thousand Zscan4 sites have appreciable levels of transcription in late 2C stage embryos, following the major wave of ZGA (fig. S6).

To test whether Zscan4 counteracts DNA damage at this developmentally relevant time point, we monitored DNA damage response by staining for  $\gamma$ H2A.X in Zscan4-depleted mouse 2C embryos undergoing ZGA. We generated small interfering RNA (siRNA) diced pools against Zscan4 or luciferase control and injected in vitro-fertilized zygotes at 5 hours post-fertilization (hpf). At 31 hpf, when the embryo is at the 2C stage and in G<sub>2</sub> phase, we measured levels of Zscan4 and  $\gamma$ H2A.X. Depletion of Zscan4 led to significantly elevated DNA damage as indicated by an increase in the number of  $\gamma$ H2A.X foci in the nuclei of the 2C embryos (Fig. 6, A to C). Because the 2C embryo is torsionally burdened by substantial levels of transcription, we tested whether suppression of DNA damage is transcription dependent. Control or Zscan4 siRNA-injected embryos were treated with triptolide (an inhibitor of Pol II transcription) between 29 and 31 hpf during a time window, when major ZGA was ongoing, but after the induction of Zscan4 expression had already occurred at minor ZGA. Triptolide treatment not only resulted in expected inhibition of transcription [no detectable 5-ethynyl-uridine (EU) signal] but also diminished  $\gamma$ H2A.X foci observed upon Zscan4 knockdown (Fig. 6, A to C). Of note, triptolide treatment also reduced the baseline level of  $\gamma$ H2A.X observed in control siRNA-injected embryos (Fig. 6C). These results indicate that ZGA is associated with transcription-dependent DNA damage and that Zscan4 acts to counteract it.

## DISCUSSION

A number of cellular processes including transcription, replication, and chromatin remodeling are associated with DNA supercoiling and torsional strain (20, 27). As development switches from maternal control to the zygote, the overwhelming changes in transcription, replication, and chromatin state that occur at this time of embryogenesis may delay timely alleviation of torsional stress related to one or more of these events, increasing the propensity for genome instability at susceptible regions. Here, we uncovered a new mechanism, by which transcription-associated genomic instability is counteracted in a developmentally regulated manner.

We found that 2C-specific zinc finger protein Zscan4 recognizes the nucleosome-rich form of simple repeat sequences characterized by (CA)<sub>n</sub>/(TG)<sub>n</sub> dinucleotide tracts, which are prone to recombination. Under torsional strain, these sequences are predicted to form a high-energy alternative structure called Z-DNA. We further demonstrated that, in vitro, Zscan4 directly binds nucleosomes and protects them from disassembly upon torsional stress induced by a DNA intercalator curaxin. Last, loss of Zscan4 in mouse 2C embryos increases DNA damage in a manner dependent on the transcriptional status of the embryo. On the basis of our results, we propose that transcription and the associated negative supercoiling behind the polymerase, coupled to overall high histone dynamics, global DNA demethylation, and low nucleosomal density that characterize 2C embryos, generate conditions that may cause at least sporadic transition to Z-DNA and fragility of the genome. Our results are consistent with a model that shows that nucleosomal binding of Zscan4 at the Z-DNA prone sites stabilizes the B-DNA form of microsatellite repeat DNA and protects these regions from damage (Fig. 6D). Given that Zscan4 expression also coincides with ZGA in

human embryos (28), this function is likely conserved across species and important for protecting the genome integrity of germ line and soma, both of which originate from the totipotent 2C embryo.

Directly establishing prevalence of Z-DNA in cells has been problematic owing to questions regarding specificity of existing Z-DNA antibodies and a lack of reliable orthogonal reagents to monitor and detect this transient DNA form. However, regardless of the challenges of the in vivo DNA structure detection, accumulating evidence links simple repeats, alternative DNA structures, and transcription with genetic instability in higher eukaryotes (3, 21–23). Plasmids carrying Z-DNA prone sequences introduced into mammalian cells accumulate large deletions and rearrangements, which are further exacerbated by inducing transcription through the repeat region (22, 29). Transgenic mice carrying Z-DNA-forming sequences on a reporter have higher chromosomal abnormalities at these regions (23). Comprehensive analyses comparing non-B-DNA-forming sequences within cancer genomes found Z-DNA prone repeats to be in close proximity to translocation breakpoints (24). Specifically, stretches of TG repeats at 10q24 and 11p13 loci are within 1 kb of the breakpoint in B cell tumors (25). Clusters of TG stretches were also found to occur near a human myeloma translocation breakpoint at 11q13 (25). If Zscan4 can protect the most fragile DNA sequences in the genome from damage during early embryogenesis, it is interesting why somatic cells do not use this mechanism more broadly to prevent carcinogenesis. Similarly, it remains an open question whether Zscan4 overexpression could at least partially suppress instability of microsatellite sequences in cancer cells. Intriguingly, during reprogramming, Zscan4 overexpression does not have a major effect on reprogramming efficiency per se, but it markedly improves the quality, developmental competency, and genomic stability of the resultant iPSCs (11). The mechanism described in our study likely contributes to this effect (in addition to the previously described role of Zscan4 in promoting telomere maintenance) and suggests a promise for Zscan4 gain-of-function approach in stabilizing fragile sites.

## MATERIALS AND METHODS

### Cell lines and plasmids

E14Tg2a (E14) male mESCs were cultured in a defined, LIF (leukemia inhibitory factor) containing medium (2i + LIF) as previously described (30). Briefly, cells were grown on poly-L-ornithine-coated (7.5  $\mu$ g/ml; Sigma-Aldrich) and laminin-coated (5  $\mu$ g/ml; Life Technologies) tissue culture dishes, in a serum-free N2B27-based Dulbecco's modified Eagle's medium (DMEM)/F12 medium supplemented with bovine serum albumin (BSA) AlbuMAX II (5 mg/ml), sodium pyruvate, L-glutamine, N2 Neuroplex, B27 without retinoic acid, non-essential amino acids,  $\beta$ -mercaptoethanol, penicillin/streptomycin, 0.8  $\mu$ M mitogen-activated protein kinase (MAPK) kinase inhibitor (PD0325901, Selleck Chemicals), and 3.3  $\mu$ M glycogen synthase kinase 3 $\beta$  inhibitor (CHIR99021, Selleck Chemicals).

*Zprom::GFP* and *Zprom::GFP-Zscan4* plasmids and reporter lines: The *Zscan4c* open reading frame was amplified from E14 mouse complementary DNA (cDNA) and cloned in-frame into a piggyBac vector containing an enhanced GFP sequence to create an N-terminal GFP-Zscan4c fusion construct. A 3566-bp putative promoter region upstream from the *Zscan4c* start codon was amplified from *pZscan4-Emerald* vector (a gift from M. S. H. Ko, Keio University School of Medicine) and cloned upstream of either GFP or the

GFP-Zscan4c fusion in the same vector. Constructs were independently transfected into E14 with a piggyBac transposase for stable integration, using Lipofectamine 2000 (Invitrogen). Clonal lines were selected and expanded with blasticidin (5 µg/ml; Invitrogen). GFP fluorescence was quantified by passing trypsinized cells through a 35-µm cell strainer and analyzed using an LSRFortessa Analyzer (BD Biosciences) and further analyzed with FlowJo software (TreeStar).

### RNA sequencing

2C-like cells or pluripotent cells were isolated from *Zprom::GFP* and *Zprom::GFP-Zscan4* reporter lines on the basis of GFP fluorescence using a fluorescence-activated cell sorter (Aria II, BD Biosciences) and collected in Trizol (Thermo Fisher Scientific). Total RNA was extracted according to the manufacturer's recommendations, and mRNA was purified using an Oligo-dT Dynabeads mRNA Purification Kit (Invitrogen). mRNA was fragmented using 10× Fragmentation reagent (Thermo Fisher Scientific) and subjected to first-strand synthesis with random hexamers and SuperScript II (Invitrogen). Second-strand synthesis was carried out with ribonuclease H (RNaseH; Invitrogen) and *Escherichia coli* DNA polymerase I (NEB, New England Biolabs). Libraries were prepared as described in (31) and were multiplexed and sequenced on a HiSeq 2500 Illumina platform (Elim Biopharmaceuticals) using 50-bp single reads. Reads were aligned using Tophat with mm10 genomic index and analyzed as previously described (32). Aligned reads were converted to counts for each gene using HTSeq, and RPKMs (Reads Per Kilobase of transcript, per Million mapped reads) were calculated in R.

### Western immunoblotting

Whole-cell extracts were prepared by lysing cells and extracting proteins for 30 min at 4°C in a buffer containing 300 mM NaCl, 100 mM tris (pH 8), 0.2 mM EDTA, 0.1% NP-40, 10% glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF), and protease inhibitor cocktail (Roche). Supernatant was collected after centrifugation and protein concentrations were determined using Bradford reagent (Bio-Rad). Proteins were separated on an SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to nitrocellulose membranes followed by immunoblotting with the following antibodies as indicated: Zscan4 (AB4340, Millipore) and GFP (ab290, Abcam).

### Chromatin immunoprecipitation

ChIP in FACS-sorted cells: Cells were trypsinized, centrifuged to remove trypsin, and cross-linked in media containing 1% formaldehyde at room temperature and then quenched with 0.125 M glycine. Cross-linked cells were collected by centrifugation, washed and resuspended in phosphate-buffered saline (PBS), strained through a 35-µm cell strainer, and subjected to two rounds of FACS to enrich for GFP<sup>+</sup> or GFP<sup>-</sup> cells as required. A post-sort analysis was conducted to confirm purity of the sorted populations. Sorted cells were lysed and sonicated using either a bioruptor sonicator (Diagenode) or a Covaris ultra sonicator followed by ChIP as previously described (30, 31). Sonicated chromatin was recovered by centrifugation and combined with the antibody of interest overnight at 4°C followed by a further 4-hour incubation with magnetic Dynabeads Protein G (Invitrogen). After washes and reversal of cross-links, the ChIP and input DNA were purified.

For SSRP1 ChIP, cells were grown on fibronectin (5 µg/ml)-coated tissue culture dishes. After treatment with curaxin (CBL0137,

Cayman Chemical) for 1 hour, cells were immediately cross-linked in 1% formaldehyde for 10 min and quenched with 0.125 M glycine for 5 min. Cells were lysed and sonicated with a bioruptor sonicator followed by ChIP as above.

The following antibodies were used in ChIP: Zscan4 (AB4340, Millipore), GFP (ab290, Abcam), pan-H3 (ab1791, Abcam), and SSRP1 (609702, 10D1, BioLegend). For sequencing, libraries were prepared as described in (31) and sequenced with either single-end 50-bp reads or paired-end 75-bp reads on either HiSeq 2500 or NextSeq Illumina platform, respectively.

### Quantitative polymerase chain reaction

Primers used in ChIP-qPCR are listed in table S1. qPCR was performed on a LightCycler 480 II machine (Roche). ChIP-qPCR signals were calculated as percentage recovery:  $100\% \times \text{primer efficiency}^{(\text{Ct Input [adjusted]} - \text{Ct ChIP})}$ .

### ATAC-seq

Fifty thousand GFP<sup>+</sup> and GFP<sup>-</sup> cells were isolated from the *Zscan4::GFP* reporter line by FACS. Tagmentation with a Tn5 transposase and library generation were conducted using the Nextera DNA Library Preparation Kit (Illumina) according to published protocols (19) and as described in (31). Libraries were multiplexed and sequenced with 50-bp paired-end reads on a HiSeq 2500 Illumina platform.

### Data analysis for ChIP-seq and ATAC-seq

ChIP-seq and ATAC-seq reads were trimmed with cutadapt and aligned with bowtie2 to mm10 reference genome. The read coverage was normalized to total aligned reads in a given library and average signals from two biological replicates were plotted for Dux and H3K4me3 ChIP-seq data. Peak calling was conducted using MACS2 callpeak function. Endogenous Zscan4 ChIP signals were normalized to reads from the GFP<sup>-</sup> population to account for non-specific background. GFP ChIP-seq from the GFP<sup>+</sup> population was used as the background model for the GFP-Zscan4 ChIP-seq.

Heat maps were generated by intersecting bam alignment files with intervals of interest (bedtools v2.25.0), followed by tabulation of the distances of the reads relative to the center of the interval and scaling to account for total aligned read numbers (10<sup>6</sup> per number aligned). Heat maps were plotted using a custom R function. Aggregate plots were generated by averaging rows of the heat map matrix. For nucleosomal positioning in ATAC-seq, only reads corresponding to library inserts >147 bp were used.

For visualization, tracks were generated with bedtools genomecov (-bg -scale) with scaling factor being 10<sup>6</sup> per number aligned reads and converted to BigWig with bedGraphToBigWig (Kent Tools). BigWigs were plotted with the IGV (integrative genomics viewer) browser.

For analysis of the relationship between ChIP-seq and repetitive sequences, bed files for various classes of repetitive elements were obtained from repeat masker and intersected with ChIP-seq peaks. Enriched families of repeats were identified with R fisher.test() followed by FDR (false discovery rate) correction with *q* value().

### Protein purification

The cDNA-encoding amino acids 369 to 506 of the mouse Zscan4c protein were cloned into a modified pGOOD bacterial expression vector encoding an N-terminal glutathione *S*-transferase (GST) and

a C-terminal hexahistidine fusion protein with a PreScission protease cleavage site after the GST. The protein was expressed in *E. coli* BL21(DE3) grown in LB in the presence of ampicillin (100 µg/ml) and 50 µM ZnCl<sub>2</sub>. The culture was induced at an OD<sub>600</sub> (optical density at 600 nm) of 0.8 to 1.0 with 1 mM isopropyl-1-thio-D-galactopyranoside and grown at room temperature (RT) for 5 hours. The cell pellet from a 1-liter culture was resuspended in 40 ml of lysis buffer consisting of 20 mM tris-HCl (pH 7.5), 0.25 M NaCl, 15% glycerol, 1% Triton X-100, 20 µM ZnCl<sub>2</sub>, 5 mM benzamidine, 1 mM β-mercaptoethanol, 2 mM PMSF, and deoxyribonuclease (DNase). Cells were lysed and sonicated. After centrifugation at 40,000g for 30 min, the clarified cell lysate was applied on a column packed with 3-ml TALON metal affinity resin (Clontech). The column was consequently washed with 20 ml of Wash Buffer A [20 mM tris-HCl (pH 7.5), 0.25 M NaCl, 5 mM imidazole, 1 mM β-mercaptoethanol, 2 mM PMSF, and 5 mM benzamidine] and 20 ml of Wash Buffer B [20 mM tris-HCl (pH 7.5), 0.25 M NaCl, and 10 mM imidazole]. The protein was eluted with 5 ml of 20 mM tris-HCl (pH 7.5), 0.25 M NaCl, and 500 mM imidazole. The N-terminal GST-tag was removed by overnight incubation with PreScission protease (1:2500 v/v) at 4°C. Protein was concentrated to 2 mg/ml in the buffer with 20 mM tris-HCl (pH 7.5), 0.25 M NaCl, 10 µM ZnCl<sub>2</sub>, 5 mM benzamidine, 5 mM β-mercaptoethanol, and 2 mM PMSF; because its function is dependent on the presence of Zn<sup>2+</sup>, all buffers containing ZnCl<sub>2</sub> and protein purity were assessed by SDS-PAGE. The concentrated purified protein was stored at 4°C for 1 to 2 weeks and at -80°C for longer periods.

### Nucleosome labeling and reconstitution

Recombinant *Xenopus laevis* histones were expressed and purified from *E. coli* as previously described (33). Histone octamer was reconstituted as previously described (33, 34). Labeled NCPs were generated using Cy5-labeled histone H3. The Cy5 on histone H3 was generated via a cysteine introduced at position 33, while native H3 C110 was mutated to alanine. The labeling reaction was done before histone octamer assembly via cysteine-maleimide chemistry. The DNA that contained 147 bp around one of the strongest Zscan4 binding site as determined by ChIP-seq was generated by PCR with high-performance liquid chromatography-purified primers and purified by PAGE. This DNA was assembled with labeled octamer by salt gradient dialysis, purified by glycerol gradient centrifugation, and quantified by native gel (34). DNA sequence used in this work is given below (mm9, chr1: 194238860–194239006).



### Electrophoretic mobility shift assays

All binding reactions were performed under conditions where the protein was in excess of oligos or nucleosomes. Saturation was determined by increasing the concentration of protein by twofold until no observable shift. Various concentrations of Zscan4(ZnF) protein were incubated with 20.6 nM Cy5-labeled oligo or Cy5-labeled NCP. Binding reactions with oligos in Fig. 2 were performed in reaction buffer containing 20 mM tris (pH 7.5), 150 mM KCl, 1 mM ethylenediaminetetraacetic acid, 1 mM dithiothreitol, 1 mM MgCl<sub>2</sub>,

and 50 ng/uL polydI:dC double-stranded (poly(deoxyinosine-deoxycytidine) poly(dI-dC)). The samples with oligos were incubated at room temperature for 30 min and resolved by native pre-electrophoresed PAGE [8%, 19:1 acrylamide/bis-acrylamide, 0.5× TBE (Tris-Borate-EDTA)] at 4°C for 1 hour at 150 V. The buffer conditions used for running EMSAs in Figs. 4 and 5 contained 20 mM tris (pH 7.5), 50 mM NaCl, 0.1% NP-40, BSA (1 µg/µl), and no polydI:dC, as described previously (35) and were resolved on 59:1 5% gels in 6-µl volume reactions.

The gels were scanned on a Typhoon variable mode imager (GE Life Sciences, Pittsburgh, PA) by scanning for Cy5 and then quantified by densitometry using ImageJ, and  $K_d$  was calculated using Prism (36). Binding with NCP was performed in reaction buffer containing 10 mM tris (pH 7.5), 50 mM NaCl, and 0.02% NP-40. In conditions where both Zscan4(ZnF) and curaxin were present, indicated concentrations of Zscan4 were added first and incubated for 30 min and then curaxin was added, and the reaction was incubated for an additional 30 min. The samples were resolved by native pre-electrophoresed PAGE (5%, 59:1 acrylamide/bis-acrylamide, 0.2× TBE) at 4°C for 1 hour at 150 V. The gels were first scanned on a Typhoon variable mode imager (GE Life Sciences, Pittsburgh, PA) by scanning for Cy5 and then stained with SYBR Gold to image DNA. The gels were quantified by densitometry using ImageJ, and  $K_d$  was calculated using Prism (36). The following oligos were used to generate figures 2D, E, F:

Oligo DNA Oct4: 5'-cy5-ACTTTTTAGTCATGTATTCTAT-3'  
 Oligo DNA (GC)<sub>n</sub>: 5'-cy5-TCGCGCGTCGCGCGTCGCGCGC-3'  
 Oligo DNA (TG)<sub>n</sub>: 5'-cy5-TGTGTGCATCTGTGTGCATCAT-3'

### Western blotting–electrophoretic mobility shift assay

WEMSA (Western blotting–electrophoretic mobility shift assay) was performed essentially as previously described (37). All binding reactions were performed as described above with fivefold more protein and nucleosomes containing Cy5-labeled histone H3. The samples were incubated at room temperature for 30 min and resolved by native pre-electrophoresed PAGE (5%, 29:1 acrylamide/bis-acrylamide, 0.5× TBE). EMSA was scanned for Cy5 fluorescence as described above. Western blot was performed using a wet transfer setup (4°C) onto nitrocellulose membrane in transfer buffer [25 mM tris, 192 mM glycine, and 20% methanol (v/v) (pH 8.3)] for 80 min at 100 V. Primary antibody incubations were performed with rabbit anti-histone H3 antibody (ab1791, Abcam) overnight at 4°C. Secondary incubations were performed using donkey anti-rabbit immunoglobulin G horseradish peroxidase (HRP) (ab16284, Abcam) for 1 hour at room temperature. Chemiluminescence was detected using Amersham ECL prime Western blotting detection reagent (GE Life Sciences) on Amersham Imager 680. Membrane was subsequently blotted for each primary antibody after inactivation of respective secondary HRP-conjugated antibody using 0.05% sodium azide in blocking buffer for 8 hours.

### Circular dichroism

The synthetic polynucleotide oligos investigated in this study, as indicated below, were annealed into double-stranded oligos and dialyzed against 5 mM cacodylate buffer (pH 7.0). The CD spectra were recorded on a Jasco J-810 (Tokyo, Japan) spectropolarimeter. The oligonucleotide concentration was fixed to 40 µM, and aliquots of curaxin at various concentrations were added to the oligonucleotide solution. The reaction was incubated overnight at room temperature,

and spectra were collected after 16 hours. The path length for all CD measurement was 0.5 cm. All measurements were carried out at 25°C. The following oligos were used to generate supplementary figure S5B:

Oligo DNA (GC)<sub>n</sub>: 5'-cy5-TCGCGCGTCGCGCGTCGCGCGC-3'

Oligo DNA (TG)<sub>n</sub>: 5'-cy5-TGTGTGCATCTGTGTGCATCAT-3'

### In vitro fertilization of mouse oocytes

All animal experiments were carried out in accordance with the Stanford University Administrative Panel on Laboratory Animal Care and the authorizing committee of the Medical University Vienna. Spermatozoa collection and in vitro fertilization procedures were carried out as previously described (38). In short, sperm was isolated from the cauda epididymis of adult F1 (C57BL6 × DBA) male mice and capacitated by preincubation for 1.5 hours in pre-gassed modified KSOM medium (Millipore) supplemented with BSA (30 mg/ml). Mature oocytes were collected 14 hours after human chorionic gonadotropin injection of adult F1 (C57BL6 × DBA) female mice according to standard procedures (38). Cumulus-oocyte complexes were placed into a 400- $\mu$ l drop of KSOM medium with capacitated sperm and incubated at 37°C in a humidified atmosphere of 5% CO<sub>2</sub> and 95% air.

### Zscan4 knockdown experiments

The siRNA diced pools against Zscan4 and luciferase (control) were generated using recombinant *Giardia lamblia* Dicer. Experimental setup was done as previously described (39). Briefly, 4  $\mu$ M siRNAs against Zscan4 or luciferase control siRNA were coinjected with dextran-tetramethyl-rhodamine (Invitrogen, 3000 molecular weight 100  $\mu$ g/ml) in late G<sub>1</sub> stage (4 to 5 hpf) zygotes derived from in vitro-fertilized mouse oocytes. After incubation at 37°C in a humidified atmosphere of 5% CO<sub>2</sub> and 95% air, Zscan4 or control siRNA-injected 2C embryos at 30 hpf were briefly washed in M2 medium, treated with acidic tyrodes, and fixed in 3.7% paraformaldehyde in PBS at 4°C. After permeabilization with 0.2% Triton X-100 in PBS for 10 min at RT, embryos were blocked overnight at 4°C in 1% BSA and 0.1% Triton X-100 in PBS. For DNase/MNase (micrococcal nuclease) treatment, fixed and permeabilized embryos were incubated with TURBO DNase 1 (2 U/ $\mu$ l, Ambion) and MNase (80 gel units, NEB) for 2 hours at 37°C under mineral oil. Embryos were analyzed for Zscan4  $\gamma$ H2A.X by immunostaining. For triptolide experiments, 2C embryos at 29 hpf were incubated with triptolide (10  $\mu$ M; Sigma-Aldrich) or dimethyl sulfoxide (DMSO) for 2 hours in the presence of 5-ethynyl-uridine (5 mM; EU, Thermo Fisher Scientific) before fixation at 31 hpf and analyzed for Zscan4,  $\gamma$ H2A.X-foci, and EU signals.

### Embryo staining and immunofluorescence microscopy

For Zscan4 and  $\gamma$ H2A.X immunostaining, fixed 2C embryos were incubated with anti-Zscan4 (1:200; AB4340, Millipore) and anti- $\gamma$ H2A.X (1:250; ab22551, Abcam) for 3 to 4 hours at RT. Followed by several washes in blocking solution, embryos were incubated at RT with anti-mouse and anti-rabbit secondary antibodies for 1.5 hours coupled with Alexa Fluor 488, 594, or 568 (1:500; Invitrogen), respectively. EU staining was performed using the Click-It RNA Alexa Fluor 488 Imaging Kit (Thermo Fisher Scientific). Embryos were washed and mounted on slides with a small drop of VECTASHIELD (VectorLab) mounting medium. Mounted embryos were analyzed on a Zeiss LSM510 Meta inverted laser scanning confocal microscope and computations of z-stack images were processed as described previously (40). ImageJ software was used to quantify Zscan4 antibody signals

and  $\gamma$ H2A.X foci of z-stack computed (~16 inner stacks with 0.3  $\mu$ m per sample) Immunofluorescence images as previously described (39).

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/12/eaaz9115/DC1>

Fig. S1. Zscan4 reporter validation.

Fig. S2. Enrichment of Zscan4 at repetitive elements.

Fig. S3. Nucleosomal binding of Zscan4 in 2C-like cells.

Fig. S4. Nucleosomal binding of Zscan4 in vitro.

Fig. S5. SSRP1 binding at repetitive elements.

Fig. S6. Transcription at Zscan4 sites in 2C embryo.

Table S1. List of primers for ChIP-qPCR in Fig. 3C.

Reference (42)

[View/request a protocol for this paper from Bio-protocol.](#)

### REFERENCES AND NOTES

1. D. Jukam, S. A. M. Shariati, J. M. Skotheim, Zygotic genome activation in vertebrates. *Dev. Cell* **42**, 316–332 (2017).
2. M. T. Lee, A. R. Bonneau, A. J. Giraldez, Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* **30**, 581–613 (2014).
3. M. Butučić, A. B. Williams, M. M. Wong, B. Kramer, W. M. Michael, Zygotic genome activation triggers chromosome damage and checkpoint signaling in *C. elegans* primordial germ cells. *Dev. Cell* **34**, 85–95 (2015).
4. S. A. Blythe, E. F. Wieschaus, Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. *Cell* **160**, 1169–1181 (2015).
5. T. S. Macfarlan, W. D. Gifford, S. Driscoll, K. Lettieri, H. M. Rowe, D. Bonanomi, A. Firth, O. Singer, D. Trono, S. L. Pfaff, Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
6. A. De Iaco, E. Planet, A. Coluccio, S. Verp, J. Duc, D. Trono, DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
7. G. Falco, S.-L. Lee, I. Stanghellini, U. C. Basse, T. Hamatani, M. S. H. Ko, Zscan4: A novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev. Biol.* **307**, 539–550 (2007).
8. D. Rodríguez-Terrones, X. Gaume, T. Ishiuchi, A. Weiss, A. Kopp, K. Kruse, A. Penning, J. M. Vaquerizas, L. Brino, M.-E. Torres-Padilla, A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* **50**, 106–119 (2018).
9. M. Zalzman, G. Falco, L. V. Sharova, A. Nishiyama, M. Thomas, S.-L. Lee, C. A. Stagg, H. G. Hoang, H.-T. Yang, F. E. Indig, R. P. Wersto, M. S. H. Ko, Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858–863 (2010).
10. J. Dan, P. Rousseau, S. Hardikar, N. Veland, J. Wong, C. Autexier, T. Chen, Zscan4 inhibits maintenance DNA methylation to facilitate telomere elongation in mouse embryonic stem cells. *Cell Rep.* **20**, 1936–1949 (2017).
11. J. Jiang, W. Lv, X. Ye, L. Wang, M. Zhang, H. Yang, M. Okuka, C. Zhou, X. Zhang, L. Liu, J. Li, Zscan4 promotes genomic stability during reprogramming and dramatically improves the quality of iPSC cells as demonstrated by tetraploid complementation. *Cell Res.* **23**, 92–106 (2013).
12. A. Safina, P. Cheney, M. Pal, L. Brodsky, A. Ivanov, K. Kirsanov, E. Lesovaya, D. Naberezhnov, E. Neshler, I. Koman, D. Wang, J. Wang, M. Yakubovskaya, D. Winkler, K. Gurova, FACT is a sensor of DNA torsional stress in eukaryotic cells. *Nucleic Acids Res.* **45**, 1925–1945 (2017).
13. P. G. Hendrickson, J. A. Doráis, E. J. Grow, J. L. Whiddon, J.-W. Lim, C. L. Wike, B. D. Weaver, C. Pflueger, B. R. Emery, A. L. Wilcox, D. A. Nix, C. M. Peterson, S. J. Tapscott, D. T. Carrell, B. R. Cairns, Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
14. J. A. Dahl, I. Jung, H. Aanes, G. D. Greggains, A. Manaf, M. Lerdrup, G. Li, S. Kuan, B. Li, A. Y. Lee, S. Preissl, I. Jermstad, M. H. Haugen, R. Suganthan, M. Bjørås, K. Hansen, K. T. Dalen, P. Fedorcsak, B. Ren, A. Klungland, Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).
15. A. Jolma, J. Yan, T. Whittington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J. M. Vaquerizas, R. Vincentelli, N. M. Luscombe, T. R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, J. Taipale, DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
16. D. B. Haniford, D. E. Pulleyblank, Facile transition of poly[d(TG) x d(CA)] into a left-handed helix in physiological conditions. *Nature* **302**, 632–634 (1983).
17. M. M. Garner, G. Felsenfeld, Effect of Z-DNA on nucleosome placement. *J. Mol. Biol.* **196**, 581–590 (1987).

18. M. J. Guertin, J. T. Lis, Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr. Opin. Genet. Dev.* **23**, 116–123 (2013).
19. J. D. Buenostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* **109**, 21.29.1–21.29.9 (2015).
20. J. Ma, M. Wang, Interplay between DNA supercoiling and transcription elongation. *Transcription*. **5**, e28636 (2014).
21. P. Zeller, S. M. Gasser, The importance of satellite sequence repression for genome stability. *Cold Spring Harb. Symp. Quant. Biol.* **82**, 15–24 (2017).
22. G. Wang, L. A. Christensen, K. M. Vasquez, Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2677–2682 (2006).
23. G. Wang, S. Carbajal, J. Vijg, J. DiGiovanni, K. M. Vasquez, DNA structure-induced genomic instability in vivo. *J. Natl. Cancer Inst.* **100**, 1815–1817 (2008).
24. A. Bacolla, J. A. Tainer, K. M. Vasquez, D. N. Cooper, Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. *Nucleic Acids Res.* **44**, 5673–5688 (2016).
25. T. Boehm, L. Mengle-Gaw, U. R. Kees, N. Spurr, I. Lavenir, A. Forster, T. H. Rabbitts, Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumours. *EMBO J.* **8**, 2621–2631 (1989).
26. B. Zhang, H. Zheng, B. Huang, W. Li, Y. Xiang, X. Peng, J. Ming, X. Wu, Y. Zhang, Q. Xu, W. Liu, X. Kou, Y. Zhao, W. He, C. Li, B. Chen, Y. Li, Q. Wang, J. Ma, Q. Yin, K. Kee, A. Meng, S. Gao, F. Xu, J. Na, W. Xie, Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).
27. J. B. Schwartzman, P. Hernández, D. B. Krimer, J. Dorier, A. Stasiak, Closing the DNA replication cycle: From simple circular molecules to supercoiled and knotted DNA catenanes. *Nucleic Acids Res.* **47**, 7182–7198 (2019).
28. R. Vassena, S. Boué, E. González-Roca, B. Aran, H. Auer, A. Veiga, J. C. Izpisua Belmonte, Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development* **138**, 3699–3709 (2011).
29. A. Jaworski, J. A. Blahó, J. E. Larson, M. Shimizu, R. D. Wells, Tetracycline promoter mutations decrease non-B DNA structural transitions, negative linking differences and deletions in recombinant plasmids in *Escherichia coli*. *J. Mol. Biol.* **207**, 513–526 (1989).
30. C. Buecker, R. Srinivasan, Z. Wu, E. Calo, D. Acampora, T. Faial, A. Simeone, M. Tan, T. Swigut, J. Wysocka, Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).
31. K. M. Dorigi, T. Swigut, T. Henriques, N. V. Bhanu, B. S. Scruggs, N. Nady, C. D. Still, B. A. Garcia, K. Adelman, J. Wysocka, MII3 and MII4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation. *Mol. Cell* **66**, 568–576.e4 (2017).
32. N. Nady, A. Gupta, Z. Ma, T. Swigut, A. Koide, S. Koide, J. Wysocka, ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *Elife* **4**, e10150 (2015).
33. K. Luger, T. J. Rechsteiner, T. J. Richmond, Preparation of nucleosome core particle from recombinant histones. *Meth. Enzymol.* **304**, 3–19 (1999).
34. C. Y. Zhou, G. J. Narlikar, Analysis of nucleosome sliding by ATP-dependent chromatin remodeling enzymes. *Meth. Enzymol.* **573**, 119–135 (2016).
35. L. A. Cirillo, F. R. Lin, I. Cuesta, D. Friedman, M. Jarnik, K. S. Zaret, Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
36. S. P. Ryder, M. I. Recht, J. R. Williamson, Quantitative analysis of protein-RNA interactions by gel mobility shift. *Methods Mol. Biol.* **488**, 99–115 (2008).
37. M. Fernandez Garcia, C. D. Moore, K. N. Schulz, O. Alberto, G. Donague, M. M. Harrison, H. Zhu, K. S. Zaret, Structural features of transcription factors associating with nucleosome binding. *Mol. Cell* **75**, 921–932.e6 (2019).
38. R. Behringer, *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, ed. 4, 2014).
39. M. Wossidlo, T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, J. Walter, 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* **2**, 241 (2011).
40. M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Schöler, J. Walter, Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J.* **29**, 1877–1888 (2010).
41. R. Z. Cer, K. H. Bruce, U. S. Mudunuri, M. Yi, N. Volfovsky, B. T. Luke, A. Bacolla, J. R. Collins, R. M. Stephens, Non-B DB: A database of predicted non-B DNA-forming motifs in mammalian genomes. *Nucleic Acids Res.* **39**, D383–D391 (2010).
42. R. Hubley, R. D. Finn, J. Clements, S. R. Eddy, T. A. Jones, W. Bao, A. F. A. Smit, T. J. Wheeler, The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44**, D81–D89 (2016).

**Acknowledgments:** We thank K. Cimprich and Wysocka laboratory members for comments; M. Ko for pZscan4-Emerald vector; B. Gu for custom Matlab script; J. Mohammed, S. Prescott, H. Long, and E. Calo-Velazquez for discussions and ideas; Stanford Functional Genomics Facility for sequencing support; and the Stanford Stem Cell Institute FACS Core. **Funding:** This work was supported by the Howard Hughes Medical Institute, NIH R35 GM131757, the Virginia and D.K. Ludwig Fund for Cancer Research (J.W.), NIH K99CA212204 (N.N.), and a Siebel Stem Cells Scholarship (M.W.). **Author contributions:** R.S., N.N., and J.W. conceived and designed the study, and wrote the manuscript with input from all coauthors. R.S., N.N., and N.A. performed experiments. M.W. conducted the mouse embryo studies. T.S. performed bioinformatical analyses and advised on experimental design and data analysis interpretations. L.J.H. and G.J.N. provided reagents and guidance for nucleosome binding experiments. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper, Supplementary Materials, and/or the Gene Expression Omnibus repository under accession number GSE140621.

Submitted 18 October 2019

Accepted 2 January 2020

Published 20 March 2020

10.1126/sciadv.aaz9115

**Citation:** R. Srinivasan, N. Nady, N. Arora, L. J. Hsieh, T. Swigut, G. J. Narlikar, M. Wossidlo, J. Wysocka, Zscan4 binds nucleosomal microsatellite DNA and protects mouse two-cell embryos from DNA damage. *Sci. Adv.* **6**, eaaz9115 (2020).

## Zscan4 binds nucleosomal microsatellite DNA and protects mouse two-cell embryos from DNA damage

Rajini SrinivasanNataliya NadyNeha AroraLaura J. HsiehTomek SwigutGeeta J. NarlikarMark WossidloJoanna Wysocka

*Sci. Adv.*, 6 (12), eaaz9115. • DOI: 10.1126/sciadv.aaz9115

### View the article online

<https://www.science.org/doi/10.1126/sciadv.aaz9115>

### Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).