CTCF is a barrier for 2C-like reprogramming

Teresa Olbrich1, Maria Vega-Sendino1, Desiree Tillo2, Wei Wu1, Nicholas Zolnerowich1, Raphael Pavani1, Andy D. Tran3, Catherine N. Domingo1, Mariajose Franco1, Marta Markiewicz-Potoczny1, Gianluca Pegoraro4, Peter C. FitzGerald5, Michael J. Kruhlak3, Eros Lazzerini-Denchi1, Elphege P. Nora6,7, André Nussenzweig1 & Sergio Ruiz1

Totipotent cells have the ability to generate embryonic and extra-embryonic tissues. Interestingly, a rare population of cells with totipotent-like potential, known as 2 cell (2C)-like cells, has been identified within ESC cultures. They arise from ESC and display similar features to those found in the 2C embryo. However, the molecular determinants of 2C-like conversion have not been completely elucidated. Here, we show that the CCCTC-binding factor (CTCF) is a barrier for 2C-like reprogramming. Indeed, forced conversion to a 2C-like state by the transcription factor DUX is associated with DNA damage at a subset of CTCF binding sites. Depletion of CTCF in ESC efficiently promotes spontaneous and asynchronous conversion to a 2C-like state and is reversible upon restoration of CTCF levels. This phenotypic reprogramming is specific to pluripotent cells as neural progenitor cells do not show 2C-like conversion upon CTCF-depletion. Furthermore, we show that transcriptional activation of the ZSCAN4 cluster is necessary for successful 2C-like reprogramming. In summary, we reveal an unexpected relationship between CTCF and 2C-like reprogramming.
toipotency is defined as the ability of a single cell to generate all cell types and is found in zygotes and 2-cell (2C) embryos\(^1,2\). As development proceeds, embryonic cells progressively restrict their developmental potential. Embryonic stem cells (ESC) isolated from the inner cell mass (ICM) of blastocysts are defined as pluripotent since they lack the ability to differentiate into extra-embryonic tissues\(^1,2\). Interestingly, a rare (~1–2%) transient population of cells with totipotency-like potential was identified within ESC cultures\(^3–4\). This cell population expresses high levels of transcripts detected in 2C embryos, including a specific gene set regulated by endogenous retroviral promoters of the MERVL subfamily\(^5–6\). At the 2C embryonic stage, these retroviral genetic elements are re-activated and highly expressed when the zygotic genome is first transcribed and quickly silenced after further development. Based on this specific feature, retroviral promoter sequences (LTR) have been used as a reporter system to genetically label 2C-like cells in vitro to study their behavior and properties\(^7–9\). Previous studies have shown the role of different genes and pathways in converting ESC to a 2C-like state in vitro\(^10–11\). Indeed, expression of the transcription factor DUX in ESC is necessary and sufficient to induce a 2C-like conversion characterized by similar transcriptional and chromatin accessibility profiles, including MERVL activation, as observed in 2C-blastomeres\(^8–7\). This reprogramming cell model has been instrumental to study the molecular mechanisms that regulate the acquisition and maintenance of totipotent-like features. DUX belongs to the double homeobox family of transcription factors exclusive to placental mammals\(^8\) and is expressed exclusively in the 2C embryo\(^8–7\). Interestingly, DUX knockout mice revealed that DUX is important but not essential for development, suggesting that additional mechanisms regulate zygotic genome activation (ZGA) and the associated totipotent state in vivo\(^9,10\).

In this study, we demonstrate that the zinc-finger binding protein CTCF, involved in regulating the higher-order organization of chromatin structure, is a barrier in pluripotent cells for 2C-like reprogramming.

**Results**

2C-like conversion correlates with DNA damage and cell death.

To explore new molecular determinants regulating totipotency, we generated ESC carrying a doxycycline (DOX)-inducible DUX cDNA (hereafter, ESC\(^{Dux}\))\(^11\). Upon DOX activation we detected the expected expression of Dux and its downstream ZGA-associated genes (Supplementary Fig. 1a, b). In addition, ESC\(^{Dux}\) containing an LTR-RFP reporter showed reactivation of MERVL sequences after DOX induction (Supplementary Fig. 1c, d). Over-expression of DUX triggers toxicity in myoblasts\(^12\). However, whether sustained expression of DUX leads to cell death in ESC has not been explored thoroughly. We observed that DUX expression induced cell death in a dose and time-dependent manner and correlated with the extent of 2C-like conversion (Fig. 1a). Indeed, live cell imaging of DOX-treated ESC\(^{Dux}\) expressing H2B-eGFP showed efficient cell death in cells asynchronously converting to a 2C-like state (Supplementary Fig. 2a, Supplementary Movie 1). Interestingly, accumulation of DOX-induced ESC\(^{Dux}\) in the G1 and G2 phases of the cell cycle along with a decrease in DNA replication preceded cell death (Fig. 1b, Supplementary Fig. 2b). To exclude that these effects were due to supra-physiological levels of DUX, we analyzed the unperturbed subpopulation of ESC that spontaneously undergoes a 2C-like conversion\(^3\). These endogenous 2C-like ESC were also characterized by G2 accumulation, decreased DNA replication, and overt spontaneous cell death following 2C-like conversion (Supplementary Fig. 2c–e and Supplementary Movie 2). In support of these observations, the activation of the transcriptional 2C program during ZGA following the first cleavage in fertilized zygotes is accompanied by an extremely long G2 phase (around 12–16 h)\(^13,14\).

We next examined whether decreased DNA replication and G2 accumulation in DOX-treated ESC\(^{Dux}\) correlated with elevated levels of replication stress (RS). Indeed, we observed that sustained expression of DUX leads to DNA-damage, revealed by the increased levels of the RS markers KRAB-associated protein 1 (KAP1) and phosphorylated H2AX (γH2AX) in a dose and time-dependent manner (Fig. 1c, d). We also detected higher levels of γH2AX in endogenous 2C-like ESC (Supplementary Fig. 2f). Thus, the decrease in DNA replication and elevated levels of γH2AX observed in 2C-like ESC suggested that RS could underlie the increased levels of DNA damage and reduced cell viability in these cells. Accordingly, increasing RS levels by using an ATR inhibitor showed an additive effect of RS and DUX expression on DNA damage (Supplementary Fig. 3a). We hypothesized that DUX-mediated increased transcription of the 2C-associated genes could be, at least partially, responsible for the RS observed. To test this hypothesis, we performed ChIP-seq analyses to detect chromatin-enrichment of the single-strand DNA (ssDNA) binding protein RPA. RPA accumulates on ssDNA upon transcription-replication conflicts resulting in replication fork stalling and DNA damage during RS\(^15,16\). We examined RPA accumulation in the one hundred most upregulated genes upon DUX expression and observed increased RPA enrichment near the DNA binding site and the transcription start site in DOX-treated ESC\(^{Dux}\) compared to untreated ESC\(^{Dux}\) (Fig. 1e, f). Similarly, we also detected RPA accumulation in re-activated MERVL sequences (Supplementary Fig. 3b, c). These results suggested that transcription-replication conflicts might arise at DUX-induced highly transcribed genes and repeats reinforcing the idea that induction of a 2C-like state by DUX in ESC is associated to RS-mediated DNA-damage. Interestingly, it has been recently shown that cultures of ESC treated with RS agents such as hydroxyurea, UV-light or cisplatin showed elevated expression of specific genes found in 2C embryos and 2C-like cells\(^17,18\).

Combined with our results, these observations suggest the existence of an intertwined relationship between DNA damage and the 2C-like transcriptional program in ESC.

DUX-induced DNA damage localizes to CTCF binding sites.

We next sought to investigate whether DUX-induced RS and DNA damage could result in DNA breakage. To explore this possibility, we performed END-seq, a highly sensitive method to detect recurrent DNA ends genome-wide at base-pair resolution\(^19\). DUX-expressing ESC showed de novo accumulation of END-seq signal at specific genomic locations compared to untreated ESC\(^{Dux}\). Indeed, a total of 1539 END-seq peaks overlapped between two independent ESC\(^{Dux}\) clones (Supplementary Data 1–3). Moreover, the type of lesion (double or single strand DNA break) at each site, showed high correlation when both ESC\(^{Dux}\) clones were compared (Supplementary Fig. 4a–c). More than 25% of the END-seq peaks localized within a 10 kb distance from a DUX binding site (hypergeometric test p-value: 1.62E–32, computed using the average END-seq peak length and the size of the mouse genome). Furthermore, 16% of the 1220 genes associated by proximity to END-seq peaks, including well-known 2C genes, were strongly upregulated by DUX (hypergeometric test p-value: 1.75E–30 using a gene set population size of 27,590 genes\(^2\); Supplementary Fig. 4d and Supplementary Data 4 and 5). These results showed that DUX-induced 2C-like conversion reproducibly generated DNA lesions in specific genomic regions associated with DUX-induced transcription. We next asked
whether these regions shared any feature that could explain the reiterative DNA damage on them. Thus, we performed a transcription factor motif enrichment analysis using our END-seq peak dataset and found the CTCF binding motif as one of the most significant (Fig. 2a). Using published CTCF ChIP-seq datasets in ESC20, we confirmed that around 50% of the END-seq peaks were occupied by CTCF (Fig. 2b, c, Supplementary Fig. 4c, e, f and Supplementary Data 6). Moreover, these sites were also enriched in SMC1 and SMC321, components of the cohesin ring-like protein complex (Fig. 2b, c, Supplementary Fig. 4e).

The transcription factor CTCF is a zinc-finger binding protein involved in chromosome folding and insulation of topologically associated domains (TADs)22. Based on the observed CTCF-associated DNA damage in DOX-induced ESCDux, we speculated that CTCF might represent a barrier for the reprogramming to a 2C-like state. This idea was supported by two observations. First, cohesin depletion in differentiated cells facilitates reprogramming during somatic cell nuclear transfer by activating ZGA23. Second, totipotent zygotes and 2C embryos are characterized by chromatin in a relaxed state associated with weak TADs24,25. Following fertilization, development is accompanied by a progressive maturation of higher-order chromatin architecture24,25. Interestingly, increasing levels of CTCF during human embryonic development are required for the progressive establishment of TADs26. Similarly, we also observed a steady increase in the levels of CTCF during development in mouse embryos (Supplementary Fig. 5). Thus, we first sought to examine the CTCF binding landscape in 2C-like cells by native Cut&Run sequencing. For this, we used LTR-RFP reporter ESCDUX to first induce 2C-like conversion, and then, sort RFP+ and RFP− cells.
24 h after DOX induction (Fig. 2d–f). Interestingly, RFP+ cells are characterized by a decrease in the number of CTCF peaks identified (Fig. 2e, f and Supplementary Data 7). Indeed, a total of 2662 and 657 overlapping CTCF peaks in two independent ESCDux clones were lost and gained, respectively, in RFP+ cells compared to untreated ESCDux (Fig. 2e, f). Of note, although RFP− cells did not reprogram to a 2C-like state, they also showed some level of reorganization in their CTCF binding landscape (Supplementary Fig. 6a). These changes were not due to variations in the total levels of CTCF (Supplementary Fig. 6a). In addition, spontaneously converting 2C-like ESC were also characterized by a similar CTCF binding reorganization (Supplementary Fig. 6b–d and Supplementary Data 8).

CTCF depletion leads to spontaneous 2C-like conversion. We next asked whether CTCF loss influences the acquisition of 2C-like features. Thus, we used an auxin-inducible degron system to deplete CTCF in ESC27. This cell line (ESCCTCF-AID hereafter) harbors both Ctf alleles tagged with an auxin-inducible degron (AID)28 sequence fused to eGFP. Although CTCF-AID protein levels in ESCCTCF-AID are lower compared to untagged CTCF in wild-type cells, ESCCTCF-AID showed negligible transcriptional changes as tagged CTCF retains most functionality27. To test whether CTCF depletion induces conversion to 2C-like cells we first examined in CTCF-depleted cells the expression levels of the zinc finger protein ZSCAN4, a gene cluster that encodes six ZSCAN4 paralogs (ZSCAN4A, ZSCAN4B, ZSCAN4C, ZSCAN4D, ZSCAN4E, and ZSCAN4F) and three pseudogenes (ZSCAN4-PS1, ZSCAN4-PS2 and ZSCAN4-PS3). ZSCAN4 (considered as a cluster unless otherwise noted) is selectively expressed in 2C embryos and 2C-like ESC3,29. Although all ZSCAN4 transcripts are expressed in ESC, ZSCAN4C and ZSCAN4F are the most abundant18. Strikingly, ZSCAN4 levels were elevated two days following CTCF depletion and further increased two days later (Fig. 3a and Supplementary Fig. 7a). Indeed, more than 20% of the cells expressed ZSCAN4 three days following CTCF depletion (Supplementary Fig. 7b). Importantly, similar percentages of RFP+ cells were observed in LTR-RFP reporter ESCCTCF-AID (Fig. 3b). This percentage decreased upon
restoration of the CTCF levels by washing off auxin (Fig. 3b).
Using RNAseq datasets from CTCF-depleted cells at different
timepoints, we observed a progressive increase in the expression
of genes enriched or exclusively expressed in 2C embryos or 2C-
like ESC (Fig. 3c, d). Among these, endogenous MERVL
sequences as well as
Dux were selectively expressed over time
upon CTCF depletion (Fig. 3d). We also observed decreased
expression of the pluripotent gene OCT4 in ZSCAN4+
auxin-treated ESCCTCF-AID as described for 2C-like ESC (Fig. 3e).
Furthermore, CTCF-depleted ESC showed transcriptional similarity
with DUX-overexpressing ESC (Supplementary Fig. 7c).

We next asked whether expression of DUX, known to efficiently
promote 2C-like conversion5–7, could cooperate to further pro-
mote 2C-like reprogramming in auxin-treated ESCCTCF-AID. By
expressing low levels of DUX to limit ESC death and avoid
saturation in 2C-like conversion5–7, we observed increased 2C-like
reprogramming (Supplementary Fig. 7d). We also examined the
effect of HDAC inhibitors in the 2C-like conversion mediated by
CTCF depletion. Indeed, histone acetylation has been shown to
regulate the expression of ZSCAN4 and other 2C genes 30.

**Fig. 3 Spontaneous 2C-like conversion in CTCF-depleted ESC.**

**a** Western blot analysis performed in parental ESC and ESCCTCF-AID treated with auxin for the indicated times. Tubulin levels are shown as a loading control. b HTI quantification of RFP+ cells in untreated or auxin-treated for five days LTR-RFP reporter parental ESC and ESCCTCF-AID. RFP+ cells two days after a wash-off following three days of auxin treatment were also quantified. Center lines indicate mean values. Percentages of RFP+ cells above the threshold are indicated. n = 500; p-values are shown from two-tailed unpaired t-tests. n.s. = non-significant. c Box and whisker plot showing the relative fold change (log2) expression of ten 2C-associated genes/repeats (DUX, ZSCAN4, ZFP352, TCSTV3, SP110, TDP021, DUB1, Elf1ad8, PRAMEL7 and MERVLs) in ESCCTCF-AID treated with auxin for the indicated times. Untreated and wash off ESCCTCF-AID were also included. Data was obtained from RNAseq datasets27. d Genome browser tracks showing RNAseq RPKM read count at the indicated genes in the same samples as in c. e Immunofluorescence analysis of ZSCAN4 and OCT4 in ESCCTCF-AID treated with auxin for 4 days. DAPI was used to visualize nuclei. Scale bar, 100 μm. f Representative bright field images (upper panels) from a time lapse experiment performed in auxin-treated LTR-RFP reporter ESCCTCF-AID. RFP+ cells are shown as they convert over time (lower panels). Time since the addition of auxin is indicated. White arrows indicate 2C-like converted cells undergoing cell death. Scale bars, 100 μm. g Box and whisker plot showing the relative fold change (log2) expression of ten 2C-associated genes (same as in e) in LTR-RFP reporter ESCCTCF-AID untreated or treated with auxin for four days and further incubated with auxin or washed-off for additional 18 h (a total of five days) and sorted based on RFP expression. GAPDH expression was used to normalize. For box plots (c, g), center line indicates the median, box extends from the 25th to 75th percentiles and whiskers show Min to Max values. p values are shown from two-
tailed paired t-tests. In a, b, e–g, one representative experiment is shown but three (a, b) and two (e–g) independent experiments were performed.
showed that 2C-like conversion was further boosted cooperatively with the use of HDAC inhibitors in auxin-treated ESC\textsubscript{CTCF-AID} (Supplementary Fig. 7e). Finally, we validated these observations by generating additional ESC\textsubscript{CTCF-AID} clonal lines (Supplementary Fig. 7f). Collectively, these results demonstrated that CTCF depletion leads to spontaneous 2C-like conversion in ESC.

We next examined the dynamics of the 2C-like conversion by live cell imaging in LTR-RFP reporter ESC\textsubscript{CTCF-AID}. Reprogramming to 2C-like ESC is asynchronous as ESC convert over time after CTCF depletion (Fig. 3f). Interestingly, we observed that spontaneously converted 2C-like ESC undergo similar cell death as shown for endogenous 2C-like ESC while non-converted ESC divide and do not show overt cell death (Fig. 3f, Supplementary Movie 3). Indeed, although CTCF depletion does not lead to increased DNA damage when the population is considered as a whole\textsuperscript{27}, CTCF-depleted 2C-like ESC showed increased \( \gamma \)-H2AX, similar to endogenous 2C-like ESC (Supplementary Fig. 7g). Our data suggested that cell toxicity induced by CTCF-depletion is due to the selective death of the spontaneously converted 2C-like ESC. Firstly, we explored whether restoring CTCF expression facilitates the exit from the 2C-like state. For this, CTCF-depleted LTR-RFP reporter ESC\textsubscript{CTCF-AID} for four days were either further incubated with auxin or washed off for an additional 18 h (5 days total) and sorted based on RFP expression. Gene expression analysis showed that restoration of CTCF levels induced a decrease in the 2C-like transcriptional program in 2C-like cells anticipating the exit from the totipotent-like state (Fig. 3g and Supplementary Fig. 8). Collectively, these results demonstrated that CTCF prevents 2C-like conversion.

Reprogramming roadblocks to 2C-like conversion. We observed that 2C-like conversion mediated by CTCF depletion is not fully penetrant. Indeed, around 15-25% of CTCF-depleted ESC can successfully undergo 2C-reprogramming within 4 days of depletion. We then asked whether intrinsic heterogeneity within ESC cultures could influence the efficiency of the 2C-like reprogramming. Thus, we established a total of 23 single-cell derived clonal ESC lines from the parental ESC\textsubscript{CTCF-AID} and examined reprogramming dynamics. We determined that the endogenous percentage of 2C-like cells within the cultures of these clonal ESC lines varies from 0.17% to 2.87% (Supplementary Fig. 9a). Interestingly, we found a significant correlation between the final percentage of 2C-like cells observed upon CTCF depletion and the starting percentage in the same clonal ESC\textsubscript{CTCF-AID} line (Supplementary Fig. 9a). This observation reinforces the idea that intrinsic transcriptional and/or epigenetic variability is a determinant of successful 2C-like conversion.

To further support this idea, we explored whether lineage committed cells showed the same 2C-like conversion phenotype. For this, we differentiated ESC\textsubscript{CTCF-AID} to proliferative neural stem cells (NSC\textsubscript{CTCF-AID}) (Supplementary Fig. 9b, c). Auxin-treated NSC\textsubscript{CTCF-AID} did not show increased 2C-associated marker expression or undergo 2C-like reprogramming upon CTCF depletion (Fig. 4a–d and Supplementary Fig. 9d). Collectively, these results suggested that the epigenetic and transcriptional changes taking place during lineage commitment and differentiation toward NSC\textsubscript{CTCF-AID} impose additional roadblocks that prevent 2C-like reprogramming in CTCF-depleted cells.

ZSCAN4 expression is required for 2C-like reprogramming. Endogenous emergence of 2C-like cells in ESC cultures is a stepwise process defined by sequential changes in gene expression\textsuperscript{32}. ZSCAN4\textsuperscript{+}MERVL\textsuperscript{−} ESC are detected during this process and represent an intermediate step that precedes the full conversion to a 2C-like state\textsuperscript{32,33}. Levels of ZSCAN4 progressively increase during 2C conversion prior to the activation of MERVL sequences and the expression of chimeric transcripts\textsuperscript{32,33}. Accordingly, we also detected a progressive accumulation of ZSCAN4 in CTCF-depleted ESC starting as early as 24 h after depletion (Fig. 5a and Supplementary Fig. 10a, b). However, upregulation of DUX or MERVL sequences was observed at later timepoints, suggesting that spontaneous conversion upon CTCF depletion followed a similar molecular roadmap as endogenous 2C-like cells. In agreement, we also detected ZSCAN4\textsuperscript{+}MERVL\textsuperscript{−} ESC in early auxin treated LTR-RFP reporter ESC\textsubscript{CTCF-AID} (Supplementary Fig. 10c, d). Importantly, we did not detect changes in the expression level of known regulators of the 2C-like conversion\textsuperscript{32,34,36} 24 or 48 h after CTCF depletion suggesting that CTCF could directly control the expression of the ZSCAN4 cluster in ESC (Supplementary Fig. 11a). Thus, we asked whether early transcriptional activation of ZSCAN4 in ESC precursors is essential for full conversion to 2C-like cells. Therefore, we infected LTR-RFP reporter ESC\textsubscript{CTCF-AID} with lentiviruses expressing shRNAs targeting ZSCAN4 paralogs (see methods for details) and examined transcriptional dynamics during 2C-like conversion upon CTCF removal. We observed that down-regulation of ZSCAN4 in CTCF-depleted cells impaired expression of 2C markers and abrogated reprogramming to 2C-like cells (Fig. 5b, c and Supplementary Fig. 11b). Finally, due to the role of ZSCAN4 in re-activating early embryonic genes and promoting MERVL expression\textsuperscript{37,38}, we examined whether over-expression of ZSCAN4C cooperated with CTCF deletion in promoting 2C-like conversion. Indeed, although ZSCAN4C expression increased the percentage of RFP\textsuperscript{+} parental and untreated ESC\textsubscript{CTCF-AID} cells to a similar extent, over-expression of ZSCAN4C in CTCF-depleted ESC further boosted 2C-like conversion as early as 24 h (Fig. 5d and Supplementary Fig. 11c). These combined results demonstrated that ZSCAN4 proteins are essential for the 2C-like conversion mediated by CTCF depletion.

Discussion

Our study demonstrates that 2C-like ESC are unstable in vitro. We observed increased DNA damage and cell death in endogenous, DUX-induced and CTCF-depleted 2C-like ESC. Similarly, over-expression of DUX in vivo leads to developmental arrest and embryonic death\textsuperscript{39}. We show that the DNA damage observed in DUX-induced 2C-like ESC is, at least partially, associated with R5 mediated by DUX-induced transcription and involves the generation of single or double strand breaks at certain CTCF sites. We propose that DUX-induced transcriptional activity of otherwise silenced genes might induce local de novo transcription/replication conflicts promoting fork stalling and eventual breakage in proximal CTCF binding sites. Importantly, END-seq is a very sensitive technique that allowed us to map precisely DNA lesions recurrently happening in the same genomic location. Thus, it is likely that CTCF-associated DNA damage represents only a fraction of the total DNA damage generated in DUX-expressing ESC. Indeed, non-recurrent random breaks will be indistinguishable from background in END-seq experiments. In support of this idea, additional sources of damage have been associated with the 2C-like state or induced by DUX. In fact, human ortholog DUX4 mediates the accumulation of dsDNA foci and the activation of the dsDNA response contributing to the apoptotic phenotype associated with DUX over-expression\textsuperscript{40}. Further work will be needed to understand the exact origin of these DNA breaks and whether the single DNA ends detected are precursor lesions to double strand breaks or are generated due to specific replication or transcriptional mechanisms.

CTCF depletion triggers spontaneous 2C-like conversion and promotes the acquisition of 2C-like features in ESC (Fig. 3). In
addition, we showed that expression of the ZSCAN4 gene cluster is a necessary early event to successfully promote 2C-like reprogramming upon CTCF-depletion. Similarly, ZSCAN4 down-regulation compromises proper embryo development and efficient somatic cell nuclear transfer performed with cohesin-depleted somatic nuclei. Nevertheless, the precise role of the ZSCAN4 cluster in the 2C-like reprogramming is unclear. The known role of ZSCAN4 in the re-activation of early embryonic genes and promotion of MERVL expression suggests that it might be essential for successful 2C-like conversion. Moreover, ZSCAN4 has also been implicated in the maintenance of telomeres and genome stability of ESCs as well as in protecting the 2C embryo from DNA damage. Thus, ZSCAN4 could participate in limiting the damage associated with the 2C-like conversion. Expression of DUX, which is a later event likely induced by secondary events and not by the direct loss of CTCF, enhances the transcriptional activation of the ZSCAN4 cluster by direct DUX binding to its promoters. In fact, DUX knockout ESC and embryos showed defective ZSCAN4 activation. This positive feedback loop might be required to promote the 2C-like state.

Transition from totipotency to pluripotency during embryonic development is characterized by the progressive accumulation of CTCF and maturation of TADs. Interestingly, CTCF binds to a large number of endogenous RNAs and this interaction seems important for chromatin CTCF deposition. Indeed, CTCF mutants unable to bind RNA showed decreased genome-wide binding. It is tempting to speculate that the progressive strength of TADs during ED correlates with increasing levels of CTCF and RNA transcription after ZGA. Further work will be needed to address how CTCF deposition and TAD insulation take place during early development and if these events play an active role in promoting the exit from totipotency in the early embryo.

Over the past decade, multiple studies have demonstrated that lineage commitment and cell identity are actively reinforced to resist cell fate changes. The best example of these studies is the somatic cell reprogramming into induced pluripotent stem cells (iPSC), which is a very inefficient process. The low reprogramming efficiency is explained by epigenetic roadblocks that need to be overcome to undergo successful reprogramming. Our data also suggest that the 2C-like reprogramming mediated by CTCF-depletion has to overcome similar roadblocks explaining the incomplete reprogramming in every individual CTCF-depleted pluripotent cell.

Finally, the fact that CTCF depletion leads to the reactivation of the 2C transcriptional program is of relevance if we consider...
the possibility that somatic cells with compromised CTCF functionality could re-express genes from this program. Indeed, Ctcf hemizygous mice are prone to spontaneous and induced cancer development in many tissues demonstrating that CTCF is haploinsufficient for tumor suppression. In agreement, somatic missense and non-sense CTCF mutations have been commonly found in human cancers. Interestingly, it has recently been recognized that a broad range of human cancers are characterized by an early zygotic gene signature. Additional studies will be needed to determine whether CTCF functionality is altered in this subtype of human cancers.

In summary, our work revealed the important intertwined relationship between CTCF and 2C-associated features.

**Methods**

**Embryo culture.** C57BL/6J mice were obtained from the Jackson Laboratory. All the animal work included here was performed in compliance with the NIH Animal Care & Use Committee (ACUC) Guideline for Breeding and Weaning. Mice were maintained in a dark/light cycle of 12 h each in a temperature range of 68°–76°F and a range of 30–70% humidity. For embryo isolation, 4-weeks old female mice were injected intraperitoneally with 5 IU Pregnant Mare Serum Gonadotrophin (PMSG, Prospec) followed by 5 IU human Chorionic Gonadotropin (hCG, Sigma-Aldrich) 46–48 h later. Pregnant females were euthanized, and embryos collected in M2 media (MR-015-D, Sigma-Aldrich) 46–48 h later. Embryos were cultured in M2 media (MR-015-D, Sigma-Aldrich) for the indicated time points after hCG injection. Embryos were injected intraperitoneally with 5 IU Pregnant Mare Serum Gonadotrophin (PMSG, Prospec) followed by 5 IU human Chorionic Gonadotropin (hCG, Sigma-Aldrich) 46–48 h later. Pregnant females were euthanized, and embryos collected in M2 media (MR-015-D, Sigma-Aldrich) at indicated time points after hCG injection. Embryos were fixed for 10 min in 4% Paraformaldehyde (Electron Microscopy Sciences), permeabilized for 30 min in 0.3% Triton X-100 and 0.1 M Glycine in PBS 1X and blocked for 1 h (1% BSA, 0.1% Tween in PBS 1X), followed by overnight incubation with primary antibodies against CTCF (1:1000 dilution, ab188408, Abcam). Embryos were washed in 0.1% Tween in PBS 1X and incubated with the appropriate secondary antibody for 1 h at room temperature. Embryos were imaged using a Nikon Ti2-E microscope (Nikon Instruments) equipped with a Yokogawa CSU-W1 spinning disk unit, a Photometrics BSI sCMOS camera and 20x (N.A. 0.75) and 60x (N.A. 1.49) plan-apochromat objective lenses. Confocal z-stacks were acquired and used to generate 3D surfaces were rendered based on nuclear DAPI-staining and the corresponding regions were used to quantify the fluorescence intensity of CTCF. Embryo z-stack images were quantified using Imaris Bitplane (Oxford Instruments).
Cell culture. Wild-type (R1, G4, E14) ESC, ESCDUX and ESCCTCF-AID (ID: EN52.9.1 and EN204.3)27 were grown on a feeder layer of growth-arrested MEFs or on gelatin 0.1% in high-glucose DMEM (Gibco) supplemented with 15% FBS, 1,500 LiF (made in house), 0.1 mM nonessential amino acids, 1% glutamin, 1 mM Sodium Pyruvate, 55 mM β-mercaptoethanol, and 1% penicillin/streptomycin (all from ThermoFisher Scientific) at 37 °C and 5% CO2. Cells were routinely passaged by Trypsin 0.05% (Gibco). Media was changed every other day and passaged every 5–7 days. HEK293T (American Type Culture Collection) cells were grown in DMEM, 10% FBS, and 1% penicillin/streptomycin. Generation of infective lentiviral particles and ESC infections were performed as described21. All experiments were performed using both ESCCTCF-AID (ID: EN52.9.1 and EN204.3)27 with similar results. However, only one set of NSCCTCF-AID (ID: EN52.9.1) is shown in Fig.4 and Extended Fig. 9. To induce differentiation of ESCCTCF-AID (ID: EN52.9.1 and EN204.3)27 with the following primary antibodies overnight at 4 °C: p-KAP1 (dilution 1:1000, 05-636, Millipore), CTCF (1:1000, 07-729, Millipore), Flag (1:1000, F1804, Sigma), anti-H2B-GFP (kind gift from Marcos Malumbres, CNIO, Spain). A total of 40,000 ESC were plated in gelatin-coated μ-Slide 8 wells plates (80826, Ibidi) and imaged untreated or Auxin/Dox-treated for a time period between 43–48 h depending on the experiment. Images were acquired every 15 or 20 min over the course time using either a Nikon spinning disk confocal microscope or a Zeiss LSM780 confocal microscope equipped with 20x plan-apochromat objective lenses (N.A. 0.75 and 0.8, respectively) and stage top incubators to maintain temperature, humidity and CO2 (Tokai Hit STX and Okolab Bold Line, respectively).

Live cell imaging. When indicated, ESC were infected with a lentiviral plasmid encoding H2B-GFP (kind gift from Marcos Malumbres, CNIO, Spain). A total of 40,000 ESC were plated in gelatin-coated μ-Slide 8 wells plates (80826, Ibidi) and imaged untreated or Auxin/Dox-treated for a time period between 43–48 h depending on the experiment. Images were acquired every 15 or 20 min over the course time using either a Nikon spinning disk confocal microscope or a Zeiss LSM780 confocal microscope equipped with 20x plan-apochromat objective lenses (N.A. 0.75 and 0.8, respectively) and stage top incubators to maintain temperature, humidity and CO2 (Tokai Hit STX and Okolab Bold Line, respectively).

Immunoﬂuorescence. Cells were fixed in 4% Paraformaldehyde (PFA, Electron Microscopy Sciences) for 10 min at RT followed by 10 min of permeabilization using the following permeabilization buffer (100 mM Tris-HCl pH 7.4, 50 mM EDTA pH 8.0, 0.5 % Triton X-100). The following primary antibodies were incubated overnight: OCT3/4 (1:100, sc-5279, Santa Cruz Biotechnology), ZSCAN4 (1:2000, AB4340, Millipore Sigma), γH2AX (1:1000, 05-636, Millipore), CTCF (1:1000, sc-32539, Santa Cruz Biotechnology), Flag (1:5000, 32228, Cell Signaling Technology), Alexa Fluor 488 Chicken anti-Rabbit IgG (H+L) (Thermo Fisher Scientific, Cat# 31431), Alexa Fluor 488 Goat anti-Mouse IgG (H+L) (Thermo Fisher Scientific, Cat# A-11008), Alexa Fluor 488 Chicken anti-Rabbit IgG (H+L) (Thermo Fisher Scientific, Cat# A-21413) or Alexa Fluor 647 Chicken anti-Mouse IgG (H+L) (Thermo Fisher Scientific, Cat# A-21663). Secondary antibodies were used at 1:1000. For the experiments using primary antibodies binding (1:1000). For generating the plots shown in Extended Data 2b, image analysis was performed using a custom Python script. In brief, DAPI-stained nuclei were segmented using the StarDist deep-learning image segmentation25. Segmented nuclei ROIs were used to quantify total DAPI intensity and RFP mean intensity.

High throughput imaging (HTI). A total of 10,000–20,000 ESC (depending on the experiment and on the specific ESC line) was plated on gelatinized mCLEAR bottom 96-well plates (Greiner Bio-One, 655087). ESC were treated with DOX (different concentration in the range from 150–600 ng/mL) or 500 μM auxin as indicated or incubate with 10μM EdU (Click Chemistry Tools) for 30 min before fixation with 4% PFA in PBS for 10 min at room temperature. γH2AX and ZSCAN4 staining was performed using standard procedures. EdU incorporation was visualized using Alexa Fluor 488-azide or Alexa Fluor 647-azide (Click Chemistry Tools) Click-it labeling chemistry and DNA was stained using DAPI (4,6-diamidino-2-phenylindole). When indicated, ESCDUX were treated with 1 μM AEC1 (A220, Selleckchem).

Images were automatically acquired using a CellB voyager CV7000 high throughput spinning disk confocal microscope (Yokogawa, Japan). Each condition was performed in triplicate wells and at least 9 different fields of view (FOV) were acquired per well. High-Content Image (HCI) analysis was performed using the Columbus software (PerkinElmer). In brief, nuclei were first segmented using the DAPI channel. Mean fluorescent intensities for γH2AX, ZSCAN4 and RFP signal were calculated over the nuclear masks in their respective channels. Single cell data obtained from the Columbus software was exported as flat tabular. txt files, and then analyzed using RStudio version 1.2.5001, and plotted using Graphpad Prism version 9.0.1. When analyzing HTI data, we considered statistically significant those samples that when compared showed a 1.5-fold difference in the averaged mean and an unpaired two-tail T-test with a p-value of at least 0.05 or lower.

Flow cytometry and cell sorting. For live cell flow cytometry experiments, cells were dissociated into single cell suspensions and analyzed for RFP expression. DAPI was added to detect cells with compromised membrane integrity. For Edu Click-IT experiments, cells were incubated for 20 min with 10 μM Edu, fixed in 4% paraformaldehyde, permeabilized in 0.5% triton X-100, followed by Alexa Fluor 488 (Abcam) or Alexa Fluor 647-azide Click-it labeling chemistry. DNA content was stained using DAPI or Hoechst 33342 (62429, Thermo Fisher Scientific). Analytic flow profiles were recorded on a LSRFortessa (BD Biosciences) or a FACSymphony A1 instrument (BD Biosciences). Data was analyzed using FlowJo version 10.7.1.

Cell sorting experiments were performed on a BD FACSaria Fusion instrument. Post-sort quality control was performed for each sample.

RNA extraction, cDNA synthesis and qPCR. Total RNA was isolated using Isolate II RNA Mini Kit (Bioline). cDNA was synthesized using SensiFAST cDNA Synthesis Kit (Bioline). Quantitative real time PCR was performed with iTaq Universal SYBR Green Supermix (BioRad) in a CFX96 Touch BioRad system. Expression levels were normalized to GAPDH. For a primer list see Supplementary Table 1. When analyzing quantitative real time PCR data, we considered statistically significant (1.633) p-values that when compared showed an averaged of two-fold difference in overall gene expression and an unpaired two-tail T-test with a p-value of at least 0.05 or lower.
**CUT&RUN protocol**: The CUT&RUN protocol was slightly modified as described [28,29]. In brief, trypsinized or cell sorted ESC (between 150,000–500,000 cells demonstrated) were washed with 2X Stop Buffer (20 mM HEPEs-KOH pH 7.5, 150 mM NaCl, 0.5 mM spermidine, Roche complete Protease Inhibitor tablet EDTA free) and bound to activated Concanavalin A beads (Polysciences) for 10 min at room temperature. Cells were then permeabilized in Digitonin Buffer (0.05% Digitonin and 0.1% BSA in Wash Buffer) and incubated with 4 μL of the antibody against CTCF (07-729, Millipore) at 4 °C for 1 h. Samples were washed in Digitonin Buffer, resuspended in 150 μL Digitonin Buffer and equilibrated to 0 °C on ice water for 5 min. To initiate MNase cleavage, 3 μL 100 mM CaCl₂ was added to cells and after 1 h of digestion, reactions were stopped with the addition of 150 μL 2X Stop Buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02% Digitonin, 50 μg/mL RNase A, 30 μg/mL Glycogen). Samples were incubated at 37 °C for 10 min to release DNA fragments and centrifuged at 16,000 g for 5 min. Supernatants were collected and a mix of 1.5 μL 20% SDS/2.25 μL 20 mg/mL Proteinase K was added to each sample and incubated at 65 °C for 35 min. DNA was precipitated with ethanol and sodium acetate and pelleted by high-speed centrifugation at 4 °C, washed, air-dried and resuspended in 10 μL 0.1x TE.

**Library preparation and sequencing**: The entire precipitated DNA obtained from CUT&RUN was used to prepare Illumina compatible sequencing libraries. In brief, end-repair was performed in 50 μL of T4 ligase reaction buffer, 0.4 mM dNTPs, 3 μL of T4 DNA polymerase (NEB), 9 μL of T4 Polynucleotide Kinase (NEB) and 1 U of Klenow fragment (NEB) at 20 °C for 30 min. End-repair reaction was cleaned using AMPure XP beads (Beckman Coulter) and eluted in 16.5 μL of Elution Buffer (10 mM Tris- HCl pH 8.5) following a Tailing reaction in 20 μL of T4-Dll-ligation reaction buffer (NEB) with 2.5 U of Klenow fragment exo- (NEB) at 37 °C for 30 min. The 20 μL of the Tailing reaction were mixed with Quick Ligase buffer 2X (NEB), 3000 μL of Quick Ligase (NEB) and 10 mM of annealed adapter (Illumina truncated adapter) in a volume of 50 μL and incubated at room temperature for 20 min. The adapter was prepared by annealing the following HPLC-puriﬁed oligos: 5′-Phos/GATCCGAGAAGACACAGTCT-3′ and 5′-ACACTCTTCTCCTACTACAGACGAGCTCTCCTTTCCCTA-3′ (phosphorothioate bond). Ligation was stopped by adding 50 μL of EDTA, cleaned using AMPure XP beads and eluted in 14 μL of Elution Buffer. All volume was used for PCR ampliﬁcation in a 50 μL reaction consisting of 1 μL of Ligated Barcoded Adapter primer, 1 μL of 2X Kapa HiFi HotStart Ready mix (Kapa Biosciences). The PCR ampliﬁcation was stopped by adding 50 mM of EDTA, and beads washed 3X WBW, 3X EB, and eluted in 8 μL of EB. PCR products were digested with USER enzyme (NEB, M5050S) at 37 °C for 30 min. PCR ampliﬁcation was performed in 10 μL of PCR reaction mixture containing 1 μL of DNA ligated with USER enzyme, 1 μL of 2X KAPA HiFi HotStart Ready Mix (Kapa Biosciences). Glycogen* was added to the reaction mixture (as internal control) followed by an incubation at 95 °C for 5 min extension at 25 °C for 30 min. Reaction was quenched with 125 mM glycine (Sigma). Cell pellets were washed twice with cold PBS and samples were snap-frozen and stored in −80 °C. Frozen pellets were resuspended in 1 mL RIPA buffer (10 mM Tris-HCl pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.1% sodium dodecyl sulfate, 0.1% sodium deoxycholate, 1% Triton X-100, and 1 Complete Mini EDTA-free protease inhibitor tablet (Roche)). Sonication was performed using Qsonica Q500 sonicator. Supernatants were washed twice with 1 mL Binding and Wash Buffer (1XWB) (10 mM Tris-HCl pH8.0, 1 mM EDTA, 1 NaCl, 0.1% Tween20). After the wash, beads were recovered using a DynaMag-2 magnetic separator (12321D, Invitrogen) and supernatants were discarded. Washed beads were resuspended in 130 μL 2X WBW (10 mM Tris-HCl pH8.0, 2 mM EDTA, 2 NaCl) combined with the 130 μL of sonicated DNA followed by an incubation at 24 °C for 30 min. Next, the supernatant was removed, and the biotinylated DNA bond to the beads was washed thrice with 1 mL WBW, twice with 1 mL EB buffer, once with 1 mL T4 ligase reaction buffer (NEB) and then resuspended in 50 μL of end-repair reaction mix consisting of 9 μL of T4 ligase reaction buffer (NEB), 1 U of Klenow Fragment (NEB) and 1 U of Klenow fragment (NEB) at 25 °C for 30 min. Once again, the supernatant was removed using a magnetic separator and beads were then washed once with 1 mL WBW, twice with 1 mL EB buffer, once with 1 mL T4-NEXT dTailing reaction buffer (NEB) and then resuspended in 50 μL of with NEBNext dTailing reaction buffer (NEB) and 20 U of Klenow fragment exo- (NEB). The A-tailing reaction was incubated at 37 °C for 30 min. The supernatant was removed using a magnetic separator and washed with 1 mL NEBu Buffer 2 and resuspended in 115 mL of Ligation reaction with Quick Ligase buffer (NEB), 6,000 U of Quick Ligase (NEB) and ligated to “END-seq hairpin” adapter 2’ by incubating the reaction at 25 °C for 30 min. Reaction was stopped by adding 50 mM of EDTA, and beads washed 3X WBW, 3X EB, and eluted in 8 μL of EB. Hairpin adapters were digested using USER enzyme (NEB, M5050S) at 37 °C for 30 min. PCR ampliﬁcation was performed in 10 μL of PCR reaction mixture containing 1 μL of DNA ligated with USER enzyme, 1 μL of 2X KAPA HiFi HotStart Ready Mix (Kapa Biosciences). Glycogen* was used as an internal control in the PCR ampliﬁcation. A prewashed Dynabeads Protein A (ThermoFisher) for combination with KAPA Library Quantiﬁcation Kit for Illumina Platforms (Kapa Biosystems) and the sequencing was performed on Illumina NextSeq 500 or 550 (75 bp single end reads).

**Processing of END-seq data**: END-seq reads were aligned to the mouse reference genome mm10 using bowtie (v1.2.2.6) [30] (PMID: 19261174) with parameters -n 3 -l 50 -k 1. Functions “view” and “sort” of samtools (v 1.6) (PMID: 19509543) were used to convert and sort the aligned sam files to sorted bam ﬁles. Bam ﬁles were further converted to bed ﬁles by bedtools bamToBed command (PMID: 20110278). END-seq peaks were called by MACS (v1.4.3.9) [31] with parameters—nomodel—keep-dup=nonlambda (PMID: 18798982) and peaks within blacklisted regions (https://www.ensembl.org/new/ site/anhulkundaje/projects/blacklists) were ﬁltered out (PMID: 31249361). Overlapped peaks from two independent clones were used in this paper. To determine whether the END-seq peak corresponded to a single or double strand lesion (asymmetric versus symmetric) we calculated the ratio between the signal intensity peaks for asymmetric (asymmetric versus symmetric) peaks was lower than 2 and the lesion was considered asymmetric or double stranded. If the ratio was higher than 1 or lower than –1 the lesion was considered asymmetric or single stranded. Gene association was performed by using GREAT [http://great.stanford.edu/public_html/] using “single nearest gene” by default 1000 kb distance.

**Chip-seq**: Twenty million of untreated or DOX-treated ESC [32] cells were ﬁxed using 1% Formaldehyde (Sigma, F1635) at 37 °C for 10 min. Fixation was then quenched with 125 mM glycine (Sigma). Cell pellets were washed twice with cold PBS and samples were snap-frozen and stored in −80 °C. Frozen pellets were resuspended in 1 mL RIPA buffer (10 mM Tris-HCl pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.1% sodium dodecyl sulfate, 0.1% sodium deoxycholate, 1% Triton X-100, and 1 Complete Mini EDTA-free protease inhibitor tablet (Roche)). Sonication power was set to 20% and peak samples were collected in 1 mL of Dynabeads Protein A (ThermoFisher) for...
30 min at 4 °C and then incubated with 4 μL Dynabeads Protein A bound to 10μg of anti-RPA32 antibody (Abcam ab10359) or 10μg of Guinea Pig anti-Rabbit IgG (ABIN101961, Antibodies-online) in 100 μL PBS overnight at 4 °C. Beads were then collected in a magnetic separator (DynaMag-2 Invitrogen), washed twice with cold RIPA buffer, twice with RIPA buffer containing 0.3 M NaCl, twice with LiCl buffer (0.25 M LiCl, 0.5% Igepal-630, 0.5% sodium deoxycholate), once with TE (10 mM Tris pH 8.0, 1 mM EDTA) + 0.2% Triton X-100, and once with TE. Crosslinking was reversed by incubating the chromatin-bound beads at 65 °C for 4 h in the presence of 0.3% SDS and 1 mg/mL of Proteinase K (Qiagen). Chromatin DNA extraction from beads and library preparation were performed as described63.

ChIP-seq data processing. For all ChIP-seq data sets, reads were aligned to the mm10 genome using bowtie27 (paired-end reads) or bwa mem (for single-end reads, in the case of the RPA ChIP-seq data) ref: https://pubmed.ncbi.nlm.nih.gov/19451168/. For paired-end data, duplicate reads were removed using MarkDuplicates from the Picard toolkit (“Picard Toolkit.” 2019. Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/). For single-end data (RPA ChIP-seq), PCR duplicates were removed using the filterdup command from macs2 v2.1.1.7, with the parameter—keep-dup—“auto”. Normalized (RPKM) signal tracks were generated bamCoverage utility from deepTools25, using the parameters bin-size = 25, smooth length = 75, ‘center_reads’ and ‘extend_reads’. For paired-end data, read mates were extended to the fragment size defined by the two read mates. For single-end ChIP-seq data, reads were extended to the estimated fragment length estimated by phantompeaktools66. Gene association was performed by using GREAT (http://great.stanford.edu/public/html) using “single nearest gene” by default 1000 kb distance.

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

Data availability

The sequencing data generated in this study have been deposited in the Gene Expression Omnibus database under accession code GSE165162.Datasets obtained from publicly available sources include GSE5624, GSE5627, GSE85185 and GSE22562 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5624 / GSE85627 / GSE85185 and GSE22562, respectively). Additional data and/or reagents that support the findings of this study are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

Custom code used in this study is provided as Supplementary Software 1 with this paper and it is also available here: https://github.com/desireetillo/CTCF_CutAndRun2167.

Received: 5 January 2021; Accepted: 17 July 2021;

Published online: 11 August 2021

References

1. Lu, F. & Zhang, Y. Cell totipotency: molecular features, induction, and it is also available here: https://github.com/desireetillo/CTCF_CutAndRun2167.

2. Dan, J., Yang, J., Liu, Y., Xiao, A. & Liu, L. Roles for histone acetylation in DNA and protects mouse two-cell embryos from DNA damage.

3. Hirata, T. et al. M. Zscan4 transiently reactivates early embryonic genes during the generation of induced pluripotent stem cells. Cell Rep. 19, 42–50 (2017).

4. Srinivasan, R. et al. and Wysocka. Zscan4 binds nucleosomal microsatellite regions form intranuclear double-stranded RNA foci in human cell models during the generation of induced pluripotent stem cells. Cell 189, 2507–2517 (2019).

5. Zalzman, M. et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 546, 858–863 (2010).
Acknowledgements

We thank Bechara Saykali and Pedro Rocha for critical reading of the manuscript, and to Jacob Pianko for critical discussion. We are grateful to Christian Franke for the continuous technical support on R. We also thank Pedro Rocha, Rafael Casellas and Seol Kyoung Jung for their help on exploring HiC data. We thank Sagartio Ortega and the Transgenic Unit at CNIO for their initial help in this project. David Goldstein and the CCR Genomics Core for sequencing support and Ferenc Livak and the CCR Flow cytometry Core for experimental support. Research in S.R. laboratory is supported by the Intramural Research Program of the NIH. T.O. is supported by a postdoctoral fellowship of the Helen Hay Whitney Foundation.

Author contributions

T.O. and S.R. conceived the study. T.O., M.V-S. designed, performed and analyzed experiments. C.N.D. and M.F. provided technical support. D.T. and P.C.F. analyzed sequencing data. G.P. provided support with high-throughput microscopy imaging. A.D.T. and M.J.K. analyzed confocal microscopy data. E.L.D. and M.M.F. provided critical reagents. N.Z. performed END-seq experiments. W.W. analyzed END-seq data. A.N. supervised END-seq experiments. R.P. performed ChIP-seq experiments. E.P.N. provided critical reagents. S.R. supervised the study and wrote the manuscript with comments and help from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-25072-x.

Correspondence and requests for materials should be addressed to S.R.

Peer review information Nature Communications thanks Didier Trono and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

This is a U.S. Government work and not under copyright protection in the US. Foreign copyright protection may apply 2021