Master transcription factor binding sites are necessary for early replication control element activity

Jesse L. Turner[1,2], Laura Hinojosa-Gonzalez[3], Takayo Sasaki[2], Athanasios Vouzas[1,2], Mariella S. Soto[2], Abhijit Chakraborty[3], Karen E. Alexander[4], Cheryl A. Fitch[1], Amber N. Brown[1], Ferhat Ay[3,5], David M. Gilbert[2,6]

[1] Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA

[2] San Diego Biomedical Research Institute, San Diego, CA 92121, USA

[3] La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037, USA

[4] Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, FL 32306, USA

[5] Department of Pediatrics, University of California San Diego, La Jolla, CA 92093, USA

[6] Corresponding author: gilbert@sdbri.org

ABSTRACT

Eukaryotic genomes are replicated in a defined temporal order called the replication timing (RT) program. RT is developmentally regulated, and has potential to drive cell fate transitions, but mechanisms controlling RT remain elusive. We previously identified “Early Replication Control Elements” (ERCEs) necessary for chromosome domain early replication, transcription, 3D chromatin architecture and compartmentalization in mouse embryonic stem cells (mESCs). Deletions of these elements were too large to identify functional elements. Here, we performed deletion analyses of three ERCEs within one domain that is early replicating uniquely in pluripotent mammalian cells. All three of these ERCEs required one or more sites for co-binding of the pluripotency transcription factors Oct4, Sox2, Nanog (OSN) for early replication. Deletion of TSSs throughout the domain eliminated transcription without affecting RT, indicating that early replication activity of ERCEs is independent of transcription. However, in the absence of all ERCE OSN sites, a TSS was sufficient to maintain mid-late replication. ERCE OSN sites do not align with initiation sites, suggesting that they function as replication enhancers. We conclude that pluripotency transcription factor binding sites ensure early replication independent of transcription, suggesting a means for co-regulation of RT with cell fate transitions during development.

INTRODUCTION

DNA replication is an essential process that facilitates the faithful duplication of genetic information. In eukaryotes, chromosomes are replicated in a defined temporal order called the replication timing (RT) program. In mammals, changes in RT occur during development in units of 400-800 kilobases termed replication domains (RD), and changes in RT during development correlate with changes in transcriptional programs (Rivera-Mulia et al. 2019; Vouzas and Gilbert 2023). Defects in RT are correlated with dysregulation of genes and genome instability, observed in several diseases and cancers, and associated with defects in chromosome condensation and sister chromatid cohesion (Blumenfeld et al. 2017; Rivera-Mulia et al. 2017; Sasaki et al. 2017; Platt et al. 2018; Rivera-Mulia et al. 2019; Heskett et al. 2020). Recently was it revealed that RT perturbations result in immediate epigenetic disturbances (Klein et al. 2021), demonstrating that RT is essential to maintain the epigenome.

We previously demonstrated (Sima et al. 2019) that early RT of several RDs in mouse embryonic stem cells (mESCs) is controlled by discrete, cooperative, and redundant chromosome segments termed early replication control elements (ERCEs). ERCEs are required for early RT, transcription, euchromatic compartmentalization, and Topologically Associating Domain (TAD) organization of their entire domain. However, the deletions identifying ERCEs spanned tens of kilobases and multiple discrete regions with active chromatin signatures, leaving open the possibility that they may contain several different types of control elements. ERCEs contain enhancer histone marks, the histone acetyltransferase CBP/P300, multiple Oct4/Sox2/Nanog co-bound (OSN) sites, transcription start sites (TSS), and they interact in a CTCF-independent manner. Changes in cooperative binding of core transcriptional regulatory network (TRN) transcription factors (TFs) have been shown to strongly correlate with changes in RT during differentiation (Rivera-Mulia et al. 2019). In budding yeast, TFs Fkh1/Fkh2 mediate early RT through spatial clustering of early origins, independent of their transcriptional activity (Knott et al. 2012; Fang et al. 2017; Ostrow et al. 2017). Together, these results suggest the hypothesis that ERCE activity may reside within their resident OSN co-binding sites.

To test this hypothesis, we performed a series of finer deletions to delineate the locations of ERCE activity. Indeed, Oct4/Sox2/Nanog co-bound (OSN) sites accounted for early RT activity in all 3 regions. Deletion of all 3 TSSs that resided within the prior deletions eliminated nearly all transcription in the domain with very little impact on RT. Only when all the ERCE-containing OSN sites were deleted, did an independent role for transcription unveil itself. Thus, the co-binding of lineage specific master TFs can control replication independent of transcription. Moreover, OSN sites do not overlap efficient replication origins and deletion of most of the initiation activity in the locus, leaving behind multiple OSN site, did not affect RT, demonstrating that ERCEs are enhancers of initiation that function to advance RT, not origins themselves. Together with prior knowledge of their 3D interactions, we propose that ERCEs create a sub-nuclear environment that is favorable for early firing of potential replication origins located throughout the domain.

RESULTS & DISCUSSION

Anatomy of the Dppa2/4 Replication Domain

The murine "Dppa domain" contains three active genes in mESCs, *Dppa2*, *Dppa4*, and *Morc1* (**Fig. 1A**). *Dppa2* and *Dppa4* are markers of pluripotency (Maldonado-Saldivia et al. 2007) and *Morc1* is essential for spermatogenesis (Inoue et al. 1999), although it is

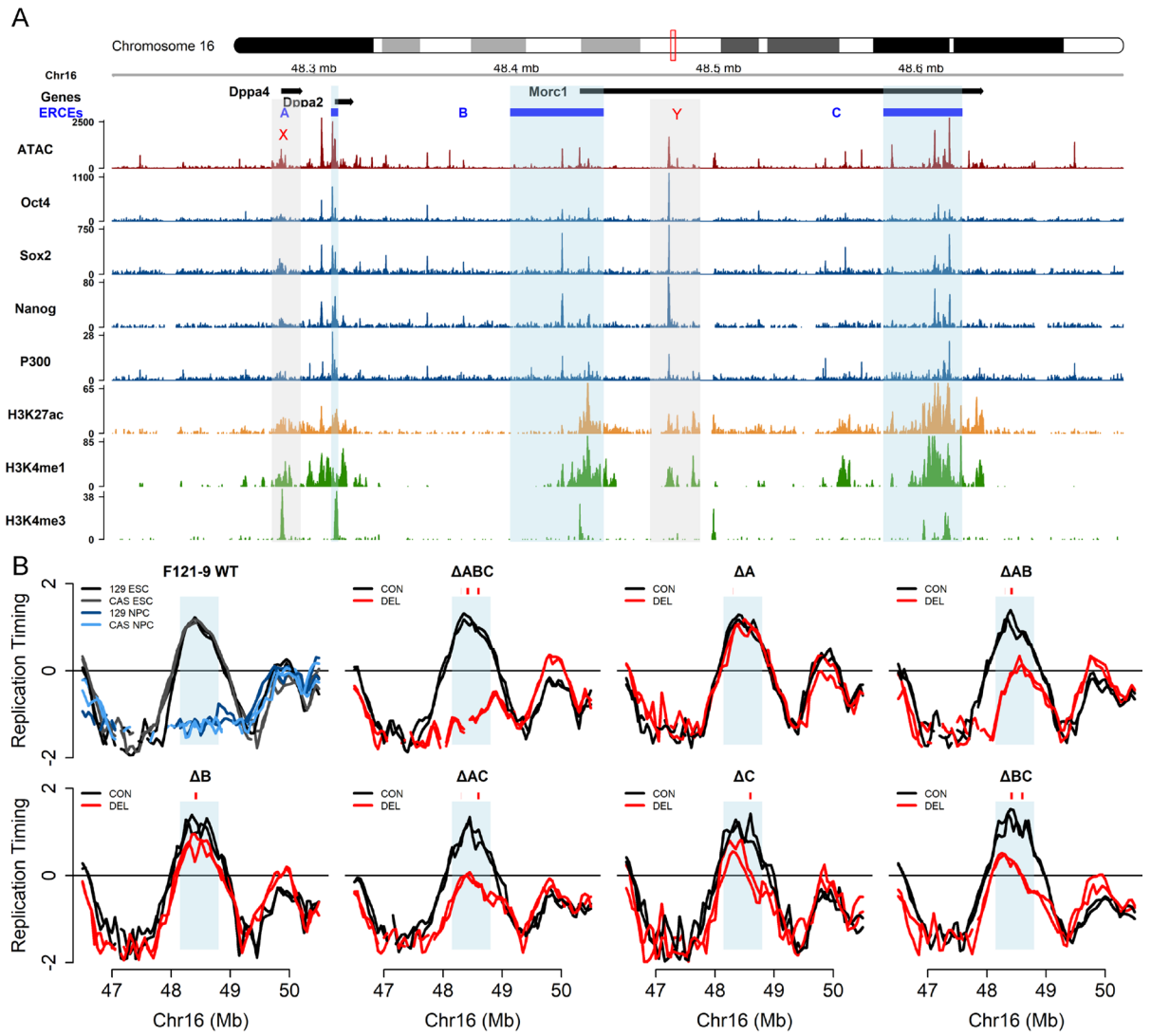


Figure 1. Epigenetic signature of the Dppa2/4 replication domain ERCEs. A) GVIZ browser tracks showing the locations of activity in the Dppa2/4 domain. ERCEs, defined by deletions in Sima et. al., 2019, are highlighted in blue. X and Y (grey highlight) represent regions that display epigenetic features of ERCEs but did not display ERCE activity in deletion analyses (Sima et. al., 2019). B) Log2(E/L) Repli-seq (Marchal et al. 2018) from Sima et. al., 2019. The first panel shows RT before and after mESC (black) to neural precursor cell (NPCs; blue) differentiation. The remaining panels show replicate (independent CRISPR clones) deletions of one or more ERCEs on one allele (DEL; red) vs. the homologous unmodified allele (CON; black). Light blue highlight indicates the Dppa2/4 replication domain boundaries, which align with its TAD boundaries.

also highly expressed in ESCs (Sima et al. 2019). The Dppa domain corresponds to a TAD whose boundaries align with Hi-C A/B compartment boundaries and are demarcated by convergent CTCF/cohesin binding sites (Sima et al. 2019). Our previous study identified three segments with ERCE activity in the domain, designated A, B and C (**Fig. 1**). These ERCEs interact and contribute to TAD strength in a CTCF-independent manner (Sima et al. 2019). ERCEs, but not CTCF protein or the CTCF-containing TAD boundaries, were necessary to maintain interactions with the A compartment and early replication. The Dppa domain resides in the nuclear interior but is flanked on both sides by late-replicating lamina-associated domains (Takebayashi et al. 2012a) and deletion of the ERCE containing segments causes a movement of the Dppa domain toward the nuclear lamina (Brueckner et al. 2020). Our studies of *cis*-acting elements have been greatly facilitated by the use of mESCs derived from F1 *castaneusXmusculus* hybrid mice (F121-9). These two subspecies are separated by 500,000 years and harbor single nucleotide polymorphisms (SNPs) at, on average, every 150bp, facilitating distinction between paternal and maternal homologues (Rivera-Mulia et al. 2018). This has allowed us to make heterozygous manipulations in one allele (red plots in all figures) while retaining the WT allele in the same datasets (black plots in all figures). During differentiation, the Dppa domain undergoes a domain-wide shift from early to late RT (**Fig. 1B, first panel**), concomitant with a switch from the A compartment to the B compartment and silencing of genes throughout the domain (Hiratani et al. 2008; Hiratani et al. 2010; Takebayashi et al. 2012b; Sima et al. 2019). These changes occur in all three germ layers, thus the Dppa domain is a pluripotency-specific early replicating domain (Hiratani et al. 2008; Hiratani et al. 2010). The remaining panels in Fig. 1B show the results of single, double and triple ERCE deletions. These results illustrate the redundant nature of ERCEs and serve as comparators for the deletion series in **Figures 2-4**.

Cooperative Contributions of 3 OSN Sites Maintain RT Activity of ERCE C

We first performed a series of 8 deletions in the 39kb ERCE C, the only ERCE to demonstrate necessity on its own (**Fig.1B**). ERCE C contains 3 OSN sites (COSN3, COSN2, COSN1, 5' to 3') and one annotated alternative intragenic TSS for Morc1 ~135kb downstream from the start of the gene (CTSS) (**Fig. 2A**). C also contains an annotated mESC super-enhancer (Dowen et al. 2014; Khan and Zhang 2016). The left half of C contains the only replication origin activity detected by the small nascent strand (SNS) assay (Prorok et al. 2019), and is within the initiation zone (IZ) (**Supplemental Fig. 1**), while the right half of C contains all the enhancer activity (**Figs. 1A and 2A**) but lacks SNS activity and is outside of the IZ. While deletion of the left half (ΔCLeft20k) had no detectable effect, deletion of the right half (ΔCRight14.6k) caused a delay similar to the original C deletion (**Fig 2B)**. Thus, the strong SNS site, including 20kb of the IZ, is dispensable for ERCE early RT activity. In fact, one of the deletions in our prior report removed 245kb between ERCEs A and C, included a large portion of the IZ, with no effect on RT (**Fig. S1**). These observations underscore the flexibility of replication initiation and demonstrate that ERCEs regulate RT independently of origin choice.

Next, we deleted the TSS (ΔCTSS), the OSN site downstream of the TSS (ΔCOSN1), as well as a deletion of the entire right half of C, leaving COSN1 intact but removing COSN2, CTSS and COSN3 (ΔCRight11.8k). ΔCTSS and ΔCOSN1 had no effect on RT, while ΔCRight11.8k resulted in a partial delay of RT. Next, we deleted COSN2 or both COSN2

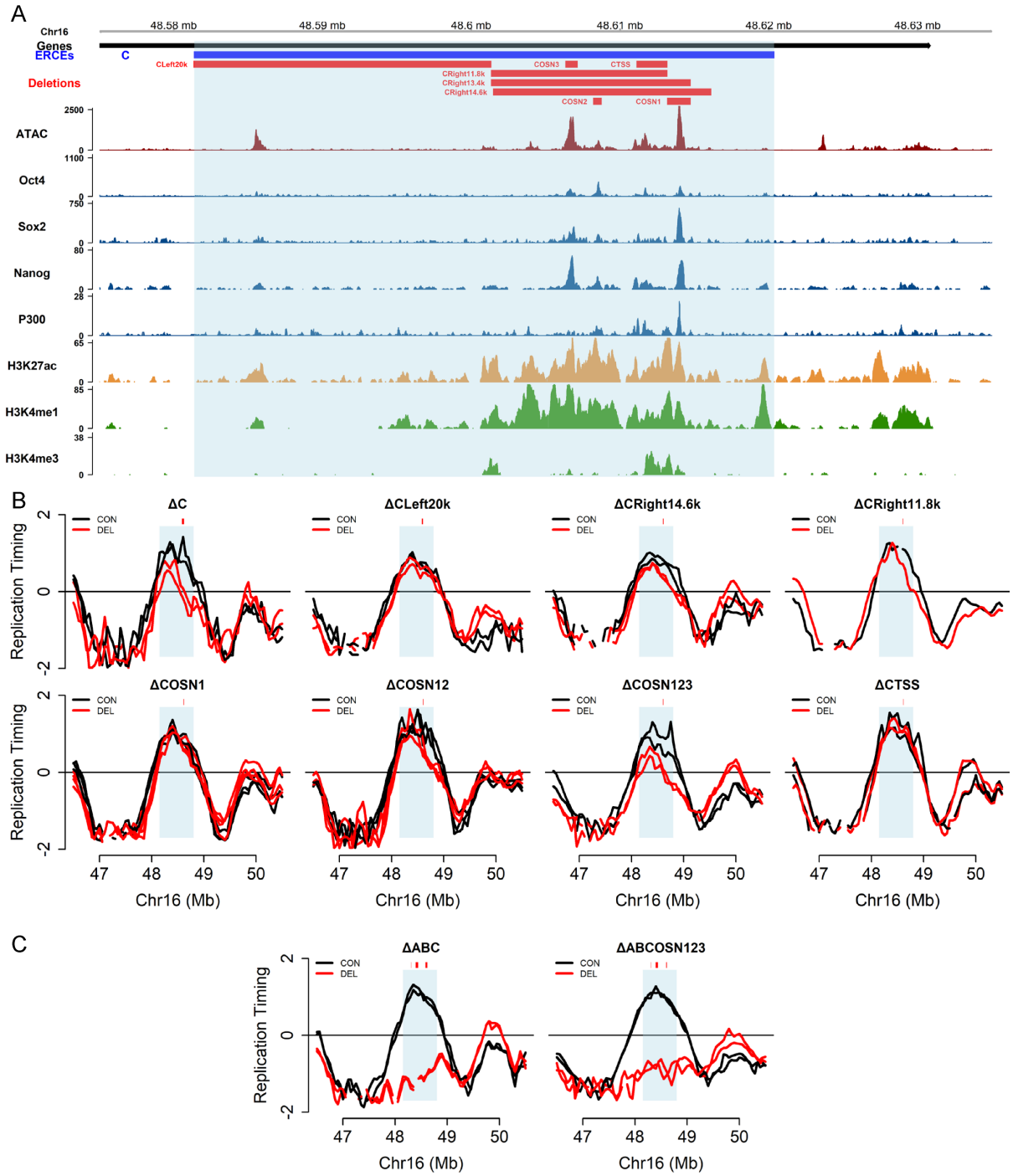


Figure 2. Genetic dissection of the C element implicates OSN as having RT control. A) GVIZ browser tracks as in Figure 1, zooming into the C element (blue horizontal line). Positions of the deletions are demarcated at the top in red. B) RT profiles of cell lines harboring the deletions in C that are shown in panel A, displayed as in Figure 1B. C) RT profiles of ABC from Sima et al, 2019 and ABCOSN123, demonstrating that three OSN sites at ERCE C fully account for the RT activity of element C.

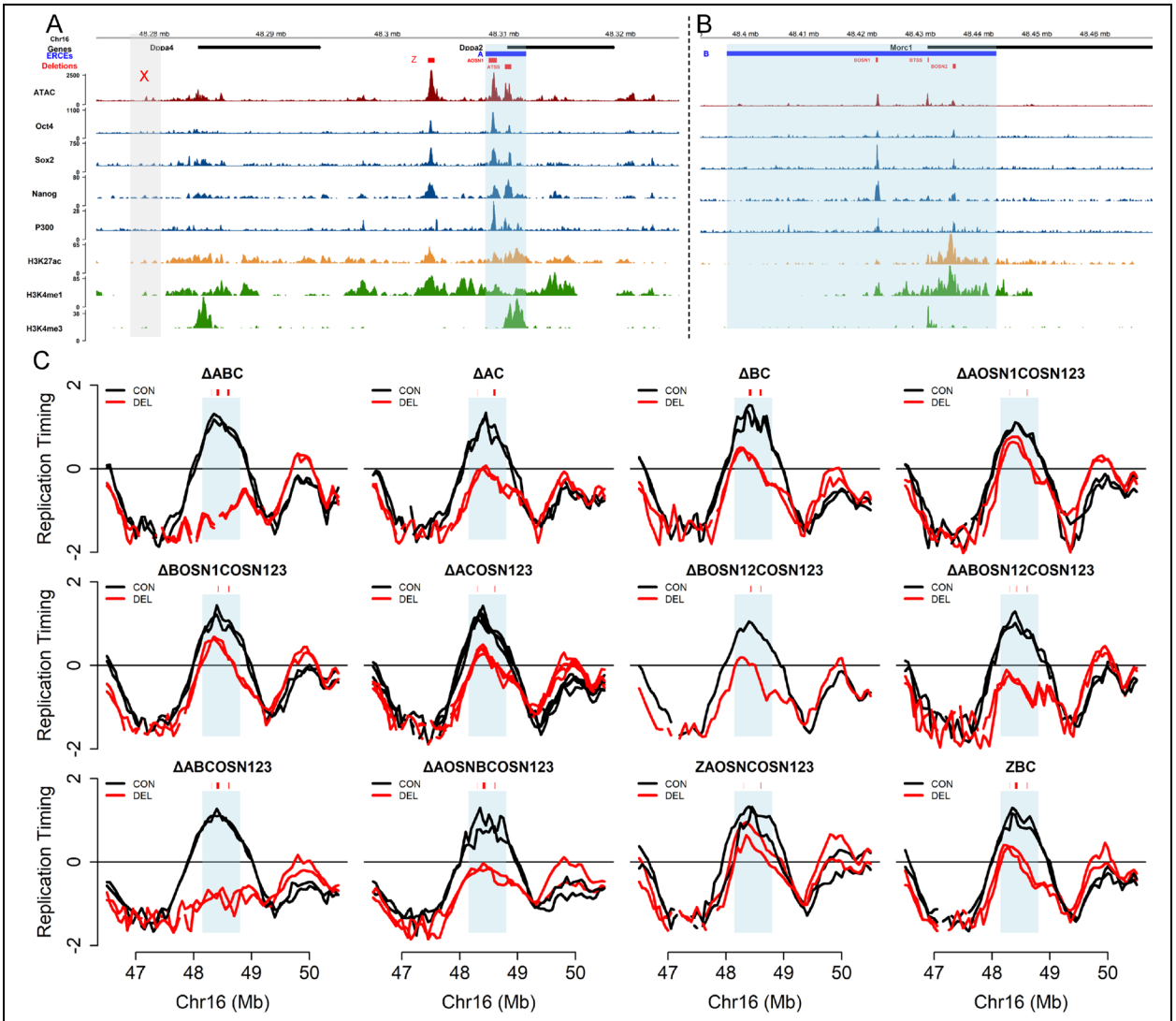


Figure 3. OSNs account for RT activity of ERCEs A and B. A) GVIZ browser tracks as in Figures 1 and 2, zooming into the A element (blue horizontal line). B) GVIZ browser as in Figures 1 and 2, zooming into the B element (blue horizontal line). In A and B, deletions are demarcated at the top in red as in Figure 2B. C) RT profiles of cell lines harboring the deletions in A and B, displayed as in Figures 1B and 2B. Z represents an OSN site near A that, like X and Y shown in Fig.1, does not harbor ERCE activity in our deletion analyses.

COSN3 in the ΔCOSN1 cell line, generating ΔCOSN12 and ΔCOSN123, which give a partial "C effect" and a full "C effect", respectively (**Fig. 2B**). The CRight11.8k deletion, which deleted not only COSN2 and 3 but also the CTSS, did not delay RT any more than ΔCOSN12. We also generated a larger deletion that spans all 3 OSNs (CRight13.4k), which also gave a full "C effect", similar to ΔCOSN123 (**Fig. 2B**). generated a 2kb To fully validate that the three OSN sites in C are responsible for the RT activity of C, we re-generated the A and B deletions from Sima et al 2019 in the ΔCOSN123 background, giving rise to ΔABCOSN123, which fully delays RT reminiscent of ΔABC from Sima et al 2019 (**Fig. 2C**). We thus conclude that ERCE activity within the "C" element can be

accounted for by the three OSN co-bound sites and that both the SNS origin and the TSS are dispensable for ERCE activity.

OSNs also Account for the Activity of ERCEs A and B

ERCE A contains one OSN site (AOSN1) and one active TSS for the Dppa2 gene that also overlaps an OSN. Since we cannot distinguish whether this latter element's activity is due to the TSS or the OSN, we will call it ATSS/AOSN2 (**Fig. 3A**). ERCE B contains two OSN sites, upstream (BOSN1) and downstream (BOSN2) of the TSS for Morc1 (BTSS) (**Fig. 3B**). Because deletion of ERCEs A or B on their own do not significantly delay RT (**Fig 1B**), and because we were able to recreate the full "C" effect by deleting the 3 OSN sites at the C element, we dissected the "A" and "B" elements in the ΔCOSN123 background expecting loss of the full activity of these elements to delay RT similar to ΔAC, ΔBC and ΔABC from Sima et al 2019 (**Fig 1B**). First, we introduced either the original A (removing both AOSN1 and ATSS/AOSN2) or the AOSN1 deletions into the ΔCOSN123 cell line. We found that targeting AOSN1 alone in the COSN123 background did not give a substantial shift in RT compared to COSN123 alone (**Fig. 3C**). Deleting all of A, however, which includes ATSS/AOSN2, yielded similar changes in RT to ΔAC (**Figs 3A and C**). Thus, the two OSN sites in ERCE A account for its RT activity, although due to their coincidence, we could not confirm that the ATSS is dispensable. Note that upstream of the A element is another OSN (OSN Z) site that is outside of the original A deletion. Deletion of this additional OSN site in the context of either ΔBC or ΔAOSN1COSN123, producing ΔZAOSNCOSN123 and ΔZBC, respectively, did not delay RT relative to their parent cell lines (**Fig 2C**). We conclude that, within ERCE A, the OSN sites contribute to ERCE RT activity while an upstream OSN (Z) outside of A does not. OSN "Z" is reminiscent of the regions we previously designated "X" and "Y" (Sima et al. 2019), that also had no ERCE activity (**Fig. 1A**). It will be important to understand what essential component of ERCEs is missing from these OSN binding sites.

To dissect the B element, we deleted the two OSN sites, one ~8kb upstream (BOSN1) and one ~4kb downstream (BOSN2) of the Morc1 TSS (BTSS) (**Fig 2B**). Deletion of BOSN1 in the context of either ΔCOSN123 or ΔAC3OSN resulted in only minor delays, but after additionally deleting BOSN2, we found significant delays in ΔBOSN12COSN123 (reminiscent of ΔBC), while ΔABOSN12COSN123 was more delayed than ΔAC and ΔBC (**Fig 3C**). However, ΔABOSN12COSN123 is less delayed than either ΔABC or ΔABCOSN123, indicating that something else within the B element is contributing to the promotion of early replication in the absence of all OSN sites with ERCE activity. Since the BTSS remains in this configuration, we speculate that transcription of the large Morc1 gene may be sufficient to mediate late-middle RT by other mechanisms (further discussed below). We conclude that OSN sites are contributing to most or possibly all of the ERCE activity in the Dppa2/4.

Transcription is not Necessary for Early RT of the Dppa Domain

In our previous study, we reported that a triple deletion of A, B and C led to silencing of transcription throughout but confined to the Dppa domain. However, it remained to be determined whether ERCE activity could be uncoupled from transcription activity. To address this question, we generated a series of deletions targeting the three TSSs at A, B, and C followed by E/L Repli-seq to measure RT and Bru-seq to measure nascent

transcription (**Figs. 4, S2**). We began by deleting BTSS (TSS for the Morc1 gene) in the WT background and in the ΔCTSS background to generate ΔBTSS and ΔBTSSCTSS. ΔBTSS and ΔBTSSCTSS both result in a near-complete loss of detectable Bru-seq signal from the Morc1 gene (**Figs. 4A, S2**), with a modest impact on RT (**Fig. 4B**). Next, to eliminate transcription from the Dppa2 gene, we deleted the region corresponding to the ATSS. ΔATSS alone resulted in a near-complete loss of Bru-seq signal from the Dppa2 gene (**Figs. 4A, S2**), with no detectable effect on RT (**Fig. 4B**), while deletion of all three TSSs (ΔATSSBTSSCTSS) resulted in almost no detectable Bru-seq signal from Dppa2 or Morc1 (**Figs. 4A, S2**) with a modest change in RT similar to ΔBTSS and ΔBTSSCTSS (**Fig. 4B**). Thus, we conclude that transcription is not necessary for early RT at the Dppa domain.

Because ΔBTSS resulted in a modest RT delay, we also generated a BTSS deletion in the ΔAC background, and found that ΔABTSSC resulted in a modestly greater delay than

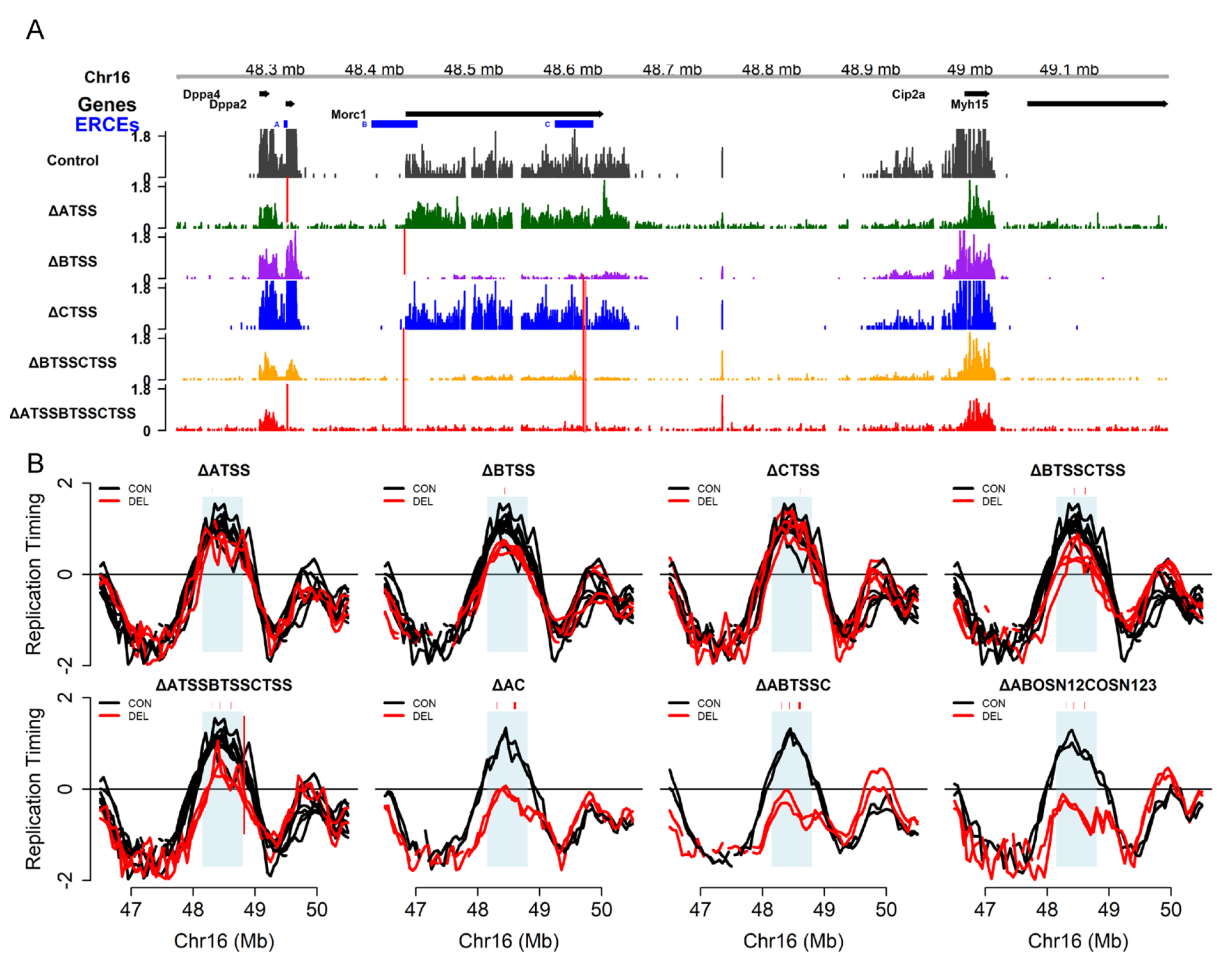


Figure 4. Transcription is not required for early RT of the *Dppa2/4* domain. A) Representative GVIZ Bru-seq tracks demonstrating that deletions (red vertical lines) of TSSs wipes out transcription of the genes they regulate. Vertical pink lines in each track represent the positions of the deletions. All of the data tracks generated in this study are in **Fig. S2**. B) Replication timing profiles of alleles harboring different TSS deletions. The first five plots (from left to right) all use the same control profiles (black lines). Red lines represent individual alleles harboring the indicated deletion.

ΔAC (**Fig. 4B**), but less than ΔABC, similar to ΔABOSN12COSN123 (**Fig. 3C**). At present we do not know whether the modest delay of RT after deleting the BTSS is a direct effect on mechanisms regulating RT activity. The fact that ΔABOSN12COSN123 is less delayed than either ΔABC or ΔABCOSN123 (**Fig. 3C**) suggests that in the presence of ERCE activity, transcription is not necessary for early replication but can have a modest contribution, while in the absence of ERCEs, transcription can be sufficient on its own to maintain mid-late replication at the Dppa2/4 locus.

Our results strongly implicate master transcription factors in developmental control of RT. ERCEs were originally discovered by virtue of their CTCF-independent interactions. We proposed that the interactions between ERCEs create a hub of hyperacetylation and open chromatin that promotes early RT. OSN sites with ERCE activity is reminiscent of binding sites for the Fkh1/2 transcription factors in budding yeast, which reside near a set of early firing origins and promote early RT by dimerization domain-mediated interactions, independently of their role in transcriptional activation (Ostrow et al. 2017). In fact, Oct4 and Nanog, as well as Yamanaka factor Klf4, have been shown to mediate chromatin looping (De Wit et al. 2013; Di Giammartino et al. 2019; Choi et al. 2022). FoxP1, the mammalian homolog that possesses the same dimerization motif as Fkh1/2 (Ostrow et al. 2017), is highly expressed in ESCs and required for pluripotency and works closely with OSN (Gabut et al. 2011). Moreover, both Sox2 and Klf5 interact with the pluripotency-specific isoform of FoxP1 (Göös et al. 2022), which contains a self-dimerization motif homologous to that of budding yeast Fkh1/2. Independent transcription and RT activities for master TFs suggests one possible reason why RT and transcriptional regulation have been so hard to untangle, and a partial explanation for why there are so many more cell type specific TF binding sites than there are essential transcriptional enhancers (Lo et al. 2022; Rothenberg 2022). It also suggests that developmental regulation of RT may be mediated by cell type specific ERCEs. It is now of high interest to determine whether OSN (and master TFs in other cell types) mediates interactions between ERCEs, and if so, whether these interactions function to regulate RT independently of their role in cell type specific transcription. RT is correlated with transcription, sub-nuclear localization, chromatin compartmentalization and epigenomic composition but we lack knowledge of the biological systems linking large-scale chromatin structure and function. Elucidating the mechanism by which ERCEs regulate RT and which correlated structural and functional features can be uncoupled is one potential means to reveal such systems.

Materials and Methods

Cell Culture

mESCs were cultured, maintained, and passaged as previously described (Sima et al. 2019). Briefly, cells were cultured on 0.1% gelatin-coated dishes in serum-free 2i/LIF media.

Cell Line Engineering

Cells were transfected using P3 Primary Cell 4D Nucleofector X Kit (Lonza, V4XP3024) or XFect Transfection Reagent (Takara, 631318) with two px458 (Addgene, 48138) plasmids containing gRNA sequences targeting the deletion breakpoints. Three days after transfection, GFP-positive cells were single-cell sorted by FACS into 96-well plates.

Ten to fourteen days after sorting, colonies were picked and expanded into duplicate 96-well plates and PCR screened. Primers upstream and downstream of the breakpoints are used to identify deletion mutants, and for large (>2kb) deletions, primers within the deleted sequences are used to identify heterozygous mutants. Deletion PCR fragments were then sanger-sequenced to determine the haplotype of the deletion allele which was then verified by NGS.

Repli-seq Library Preparation and Data Analysis

Repli-seq was performed as previously described (Marchal et al. 2018). Briefly, cells were labeled with 100uM BrdU (Sigma Aldrich, B5002) for 2 hours and then fixed in 75% ethanol. Fixed cells were then sorted by FACS into early-S and late-S fractions based on propidium iodide staining of DNA. DNA was then purified from each fraction, 46 sheared using Covaris ME220, and used to construct Illumina sequencing libraries with NEBNext Ultra Library DNA Prep Kit for Illumina (NEB, E7370). BrdU-labeled nascent DNA library fragments were then enriched by immunoprecipitation with anti-BrdU antibody (BD, 555627). IP products are then PCR amplified and indexed. A minimum of 30 millions reads/read pairs per library were requested. Raw sequencing data were first quality and adaptor trimmed using Trim Galore (https://github.com/FelixKrueger/TrimGalore) (Martin 2011) and then aligned to the reference genome with bowtie2 (Langmead and Salzberg 2012). For F121-9, reads were aligned to a custom N-masked mm10 genome for subsequent haplotype phasing with SNPsplit (Krueger and Andrews 2016). Aligned reads were then quality filtered and duplicate reads removed using Samtools (Li et al. 2009). Processed aligned reads were then counted in 50kb windows across the entire genome for both early- and late-fraction using bedtools (Quinlan and Hall 2010) and then a log2 ratio of early-to-late read counts for each window was calculated to generate raw timing files. Raw E/L data was then scaled with R, quantile normalized with R package preprocessCore (Bolstad 2023) and loess-smoothed in 300kb windows to generate the final replication timing profiles used for plotting and analysis.

Bru-seq Library Preparation and Data Analysis

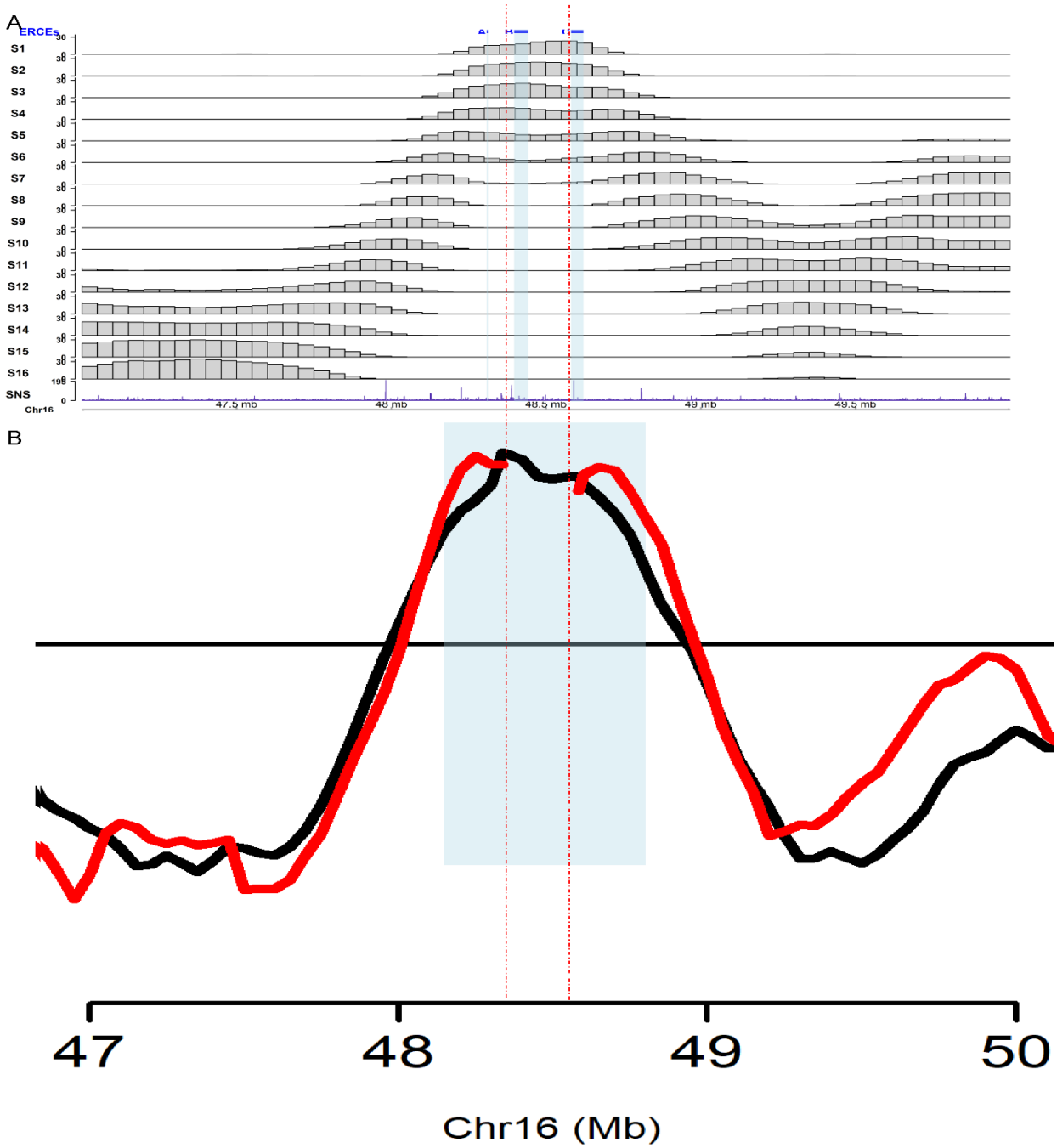
Bru-seq was performed as previously described (Paulsen et al. 2014; Bedi et al. 2020). Briefly, cells were labeled with 2mM Bru (Sigma Aldrich, 850187) for 30 min and then total RNA collected using Direct-Zol RNA Miniprep Plus Kit (Zymo, R2070). Bru-labeled nascent RNA was then enriched 47 using anti-BrdU antibodies conjugated to magnetic beads. Nascent RNA was then fragmented at 85C for 10 min and converted to cDNA using NEBNext Ultra II RNA First Strand Synthesis Module (NEB, E7771) and NEBNext Ultra II Directional Second Strand Synthesis Kit (NEB, E7550) according to manufacturer's instructions. cDNA fragments were then constructed into libraries using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645). 100M reads/read pairs per library were requested. Raw sequencing data were processed and aligned like in E/L repli-seq, except duplicate reads are retained (Paulsen et al. 2014). RPM-normalized basepair coverage tracks were generated from the processed alignment files with bedtools and visualized in bigwig file format using R package GVIZ (Hahne and Ivanek 2016).

CUT&RUN Library Preparation and Data Analysis

CUT&RUN was performed using the Epicypher CUTANA CUT&RUN kit according to manufacturer's instructions. 50M read pairs per library were requested. Raw sequencing data were processed and aligned like in E/L repli-seq. Read pileups and peak calls were generated from the processed alignment files using MACS3 (Zhang et al. 2008) and visualized in bigwig file format using R package GVIZ.

Author Contributions

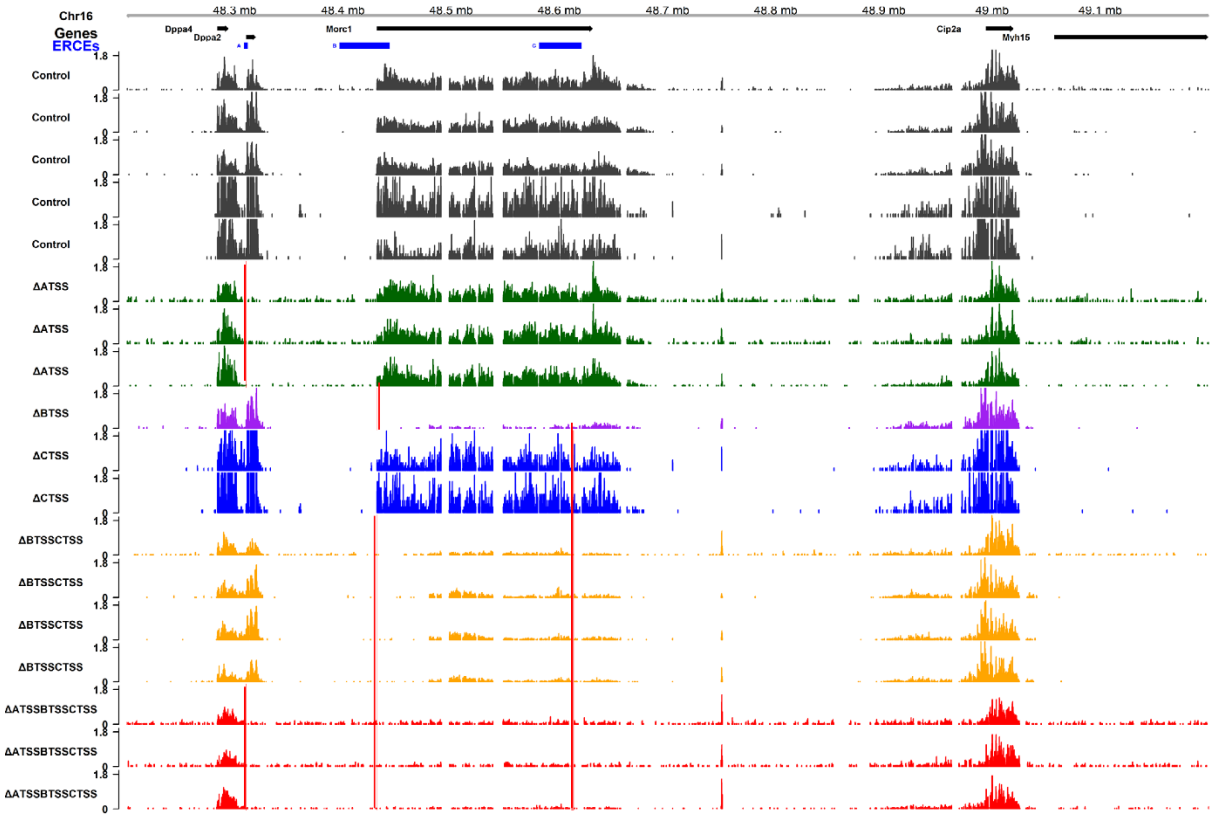
Conceptualization & Study Design, J.L.T & D.M.G; CRISPR Cloning, J.L.T & C.A.F; Cell Culture, J.L.T, M.S.S, & T.S; Cell Line Engineering, J.L.T & M.S.S; PCR Screening, J.L.T; Repli-seq & Analysis, J.L.T & T.S; Bru-seq & Analysis, J.L.T, A.V, & A.N.B; Flow Cytometry, J.L.T & K.E.A; CUT&RUN & Analysis, J.L.T & A.V; Bioinformatics Analyses, J.L.T, L.H.G, A.C, & F.A; J.L.T, D.M.G, L.H.G & F.A. wrote the manuscript with input from all authors.



Supplemental Figure 1. Initiation activity in the Dppa2/4 domain is not required for RT. A) GVIZ tracks showing read coverage in 50kb genomic windows from high-resolution repli-seq (Zhao et al. 2020), which accurately captures regions of initiation (IZs). Bottom-most track shows signal from the small nascent strand (SNS) assay which is an alternative method for mapping initiation sites (Prorok et al. 2019). ERCEs are represented by blue boxes beneath the tracks and the blue vertical highlight. B) E/L Repli-seq of the 245k deletion line from Sima et. al., 2019. Blue box represents the Dppa2/4 replication domain. Both A) and B) have been lined up to demonstrate that the 245k deletion, which removes a large portion of the IZ, has no effect on RT.

SUPPLEMENTARY FIGURE 2

A



Supplemental Figure 2. All Bru-seq data sets generated in this study. All of the Bru-seq tracks represented by the tracks in figure 4 of the main text, demonstrating that deletions (red vertical lines) of TSSs wipes out transcription of the genes they regulate. Vertical pink lines in each track represent the positions of the deletions.

REFERENCES

Bedi K, Paulsen MT, Wilson TE, Ljungman M. 2020. Characterization of novel primary miRNA transcription units in human cells using Bru-seq nascent RNA sequencing. *NAR Genom Bioinform* **2**: lqz014.

Blumenfeld B, Ben-Zimra M, Simon I. 2017. Perturbations in the Replication Program Contribute to Genomic Instability in Cancer. *International Journal of Molecular Sciences* **18**: 1138.

Bolstad B. 2023. preprocessCore: A collection of pre-processing functions. p. Version 1.62.61, Bioconductor.

Brueckner L, Zhao PA, Van Schaik T, Leemans C, Sima J, Peric-Hupkes D, Gilbert DM, Van Steensel B. 2020. Local rewiring of genome–nuclear lamina interactions by transcription. *The EMBO Journal* **39**.

Choi K-J, Quan MD, Qi C, Lee J-H, Tsoi PS, Zahabiyon M, Bajic A, Hu L, Prasad BVV, Liao S-CJ et al. 2022. NANOG prion-like assembly mediates DNA bridging to facilitate chromatin reorganization and activation of pluripotency. *Nature Cell Biology* **24**: 737-747.

De Wit E, Bouwman BAM, Zhu Y, Klous P, Splinter E, Verstegen MJAM, Krijger PHL, Festuccia N, Nora EP, Welling M et al. 2013. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* **501**: 227-231.

Di Giammartino DC, Kloetgen A, Polyzos A, Liu Y, Kim D, Murphy D, Abuhashem A, Cavaliere P, Aronson B, Shah V et al. 2019. KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nature Cell Biology* **21**: 1179-1190.

Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schujiers J, Lee TI, Zhao K, Young RA. 2014. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**: 374-387.

Fang D, Lengronne A, Shi D, Forey R, Skrzypczak M, Ginalski K, Yan C, Wang X, Cao Q, Pasero P, Lou H. 2017. Dbf4 recruitment by forkhead transcription factors defines an upstream rate-limiting step in determining origin firing timing. *Genes & Development* **31**: 2405-2415.

Gabut M, Samavarchi-Tehrani P, Wang X, Slobodeniuc V, O'Hanlon D, Sung H-K, Alvarez M, Talukder S, Pan Q, Mazzoni O, Esteban et al. 2011. An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming. *Cell* **147**: 132-146.

Göös H, Kinnunen M, Salokas K, Tan Z, Liu X, Yadav L, Zhang Q, Wei G-H, Varjosalo M. 2022. Human transcription factor protein interaction networks. *Nature Communications* **13**.

Hahne F, Ivanek R. 2016. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol* **1418**: 335-351.

Heskett MB, Smith LG, Spellman P, Thayer MJ. 2020. Reciprocal monoallelic expression of ASAR lncRNA genes controls replication timing of human chromosome 6. *RNA* **26**: 724-738.

Hiratani I, Ryba T, Itoh M, Rathjen J, Kulik M, Papp B, Fussner E, Bazett-Jones DP, Plath K, Dalton S et al. 2010. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome Research* **20**: 155-169.

Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang C-W, Lyou Y, Townes TM, Schübeler D, Gilbert DM. 2008. Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation. *PLoS Biology* **6**: e245.

Inoue N, Hess KD, Moreadith RW, Richardson LL, Handel MA, Watson ML, Zinn AR. 1999. New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum Mol Genet* **8**: 1201-1207.

Khan A, Zhang X. 2016. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res* **44**: D164-171.

Klein KN, Zhao PA, Lyu X, Sasaki T, Bartlett DA, Singh AM, Tasan I, Zhang M, Watts LP, Hiraga SI et al. 2021. Replication timing maintains the global epigenetic state in human cells. *Science* **372**: 371-378.

Knott RV, Simon, Peace M, Jared, Ostrow Z, A., Gan Y, Rex E, Alexandra, Viggiani J, Christopher, Tavaré S, Aparicio M, Oscar. 2012. Forkhead Transcription Factors Establish Origin Timing and Long-Range Clustering in S. cerevisiae. *Cell* **148**: 99-111.

Krueger F, Andrews SR. 2016. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Res* **5**: 1479.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357-359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Lo JH-H, Edwards M, Langerman J, Sridharan R, Plath K, Smale ST. 2022. Oct4:Sox2 binding is essential for establishing but not maintaining active and silent states of dynamically regulated genes in pluripotent cells. *Genes & Development* **36**: 1079-1095.

Maldonado-Saldivia J, Van Den Bergen J, Krouskos M, Gilchrist M, Lee C, Li R, Sinclair AH, Surani MA, Western PS. 2007. Dppa2 and Dppa4 Are Closely Linked SAP Motif Genes Restricted to Pluripotent Cells and the Germ Line. *Stem Cells* **25**: 19-28.

Marchal C, Sasaki T, Vera D, Wilson K, Sima J, Rivera-Mulia JC, Trevilla-García C, Nogues C, Nafie E, Gilbert DM. 2018. Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. *Nat Protoc* **13**: 819-839.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**: 10.

Ostrow AZ, Kalhor R, Gan Y, Villwock SK, Linke C, Barberis M, Chen L, Aparicio OM. 2017. Conserved forkhead dimerization motif controls DNA replication timing and spatial organization of chromosomes in S. cerevisiae. *Proceedings of the National Academy of Sciences* **114**: E2411-E2419.

Paulsen MT, Veloso A, Prasad J, Bedi K, Ljungman EA, Magnuson B, Wilson TE, Ljungman M. 2014. Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods* **67**: 45-54.

Platt EJ, Smith L, Thayer MJ. 2018. L1 retrotransposon antisense RNA within ASAR lncRNAs controls chromosome-wide replication timing. *Journal of Cell Biology* **217**: 541-553.

Prorok P, Artufel M, Aze A, Coulombe P, Peiffer I, Lacroix L, Guédin A, Mergny J-L, Damaschke J, Schepers A et al. 2019. Involvement of G-quadruplex regions in mammalian replication origin activity. *Nature Communications* **10**.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Rivera-Mulia JC, Desprat R, Trevilla-Garcia C, Cornacchia D, Schwerer H, Sasaki T, Sima J, Fells T, Studer L, Lemaitre J-M, Gilbert DM. 2017. DNA replication timing alterations identify common markers between distinct progeroid diseases. *Proceedings of the National Academy of Sciences* **114**: E10972-E10980.

Rivera-Mulia JC, Dimond A, Vera D, Trevilla-Garcia C, Sasaki T, Zimmerman J, Dupont C, Gribnau J, Fraser P, Gilbert DM. 2018. Allele-specific control of replication timing and genome organization during development. *Genome Research* **28**: 800-811.

Rivera-Mulia JC, Kim S, Gabr H, Chakraborty A, Ay F, Kahveci T, Gilbert DM. 2019. Replication timing networks reveal a link between transcription regulatory circuits and replication timing control. *Genome Research* **29**: 1415-1428.

Rothenberg EV. 2022. Transcription factors specifically control change. *Genes & Development* **36**: 1097-1099.

Sasaki T, Rivera-Mulia JC, Vera D, Zimmerman J, Das S, Padget M, Nakamichi N, Chang BH, Tyner J, Druker BJ et al. 2017. Stability of patient-specific features of altered DNA replication timing in xenografts of primary human acute lymphoblastic leukemia. *Exp Hematol* **51**: 71-82.e73.

Sima J, Chakraborty A, Dileep V, Michalski M, Klein KN, Holcomb NP, Turner JL, Paulsen MT, Rivera-Mulia JC, Trevilla-Garcia C et al. 2019. Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication. *Cell* **176**: 816-830.e818.

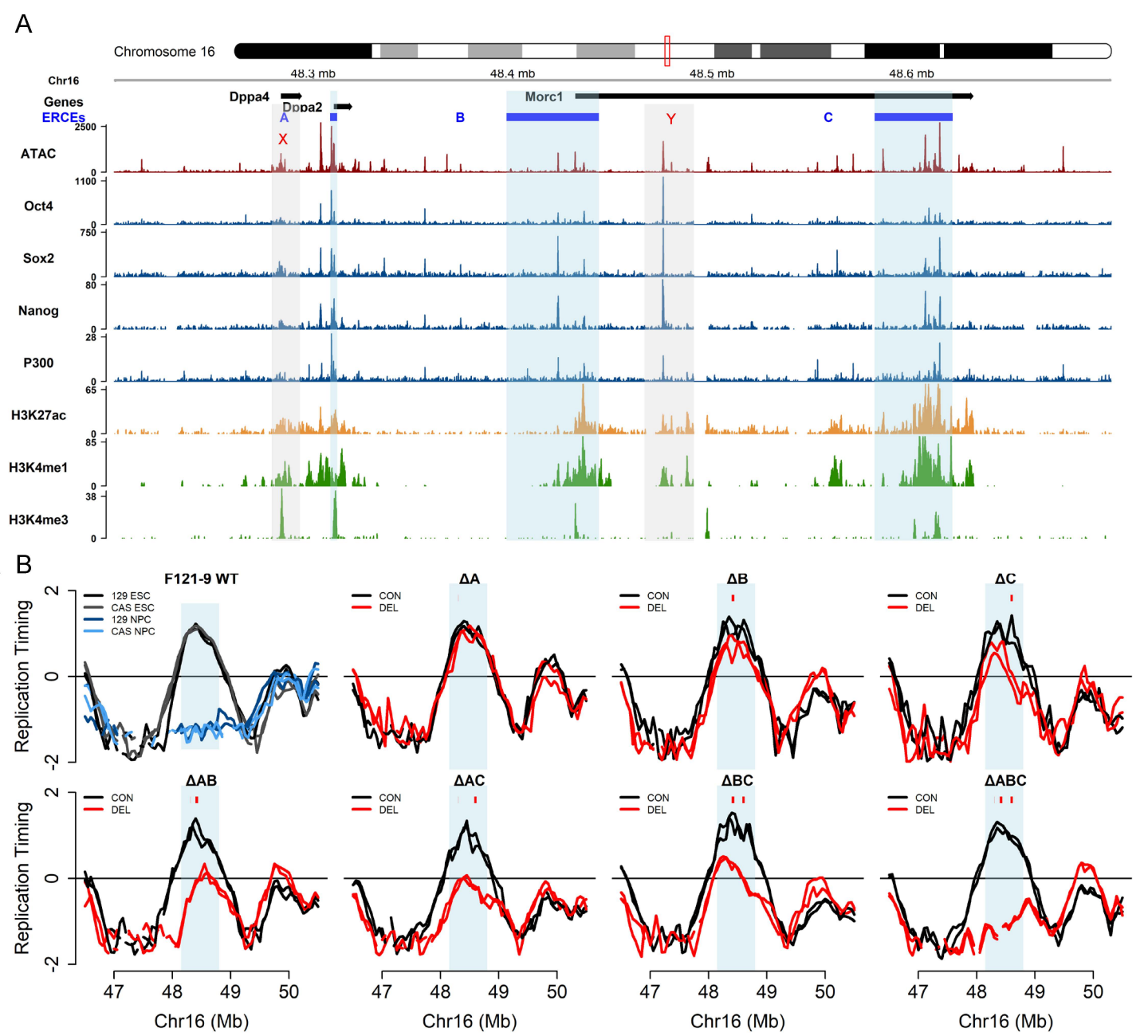
Takebayashi S-I, Dileep V, Ryba T, Dennis JH, Gilbert DM. 2012a. Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding. *Proceedings of the National Academy of Sciences* **109**: 12574-12579.

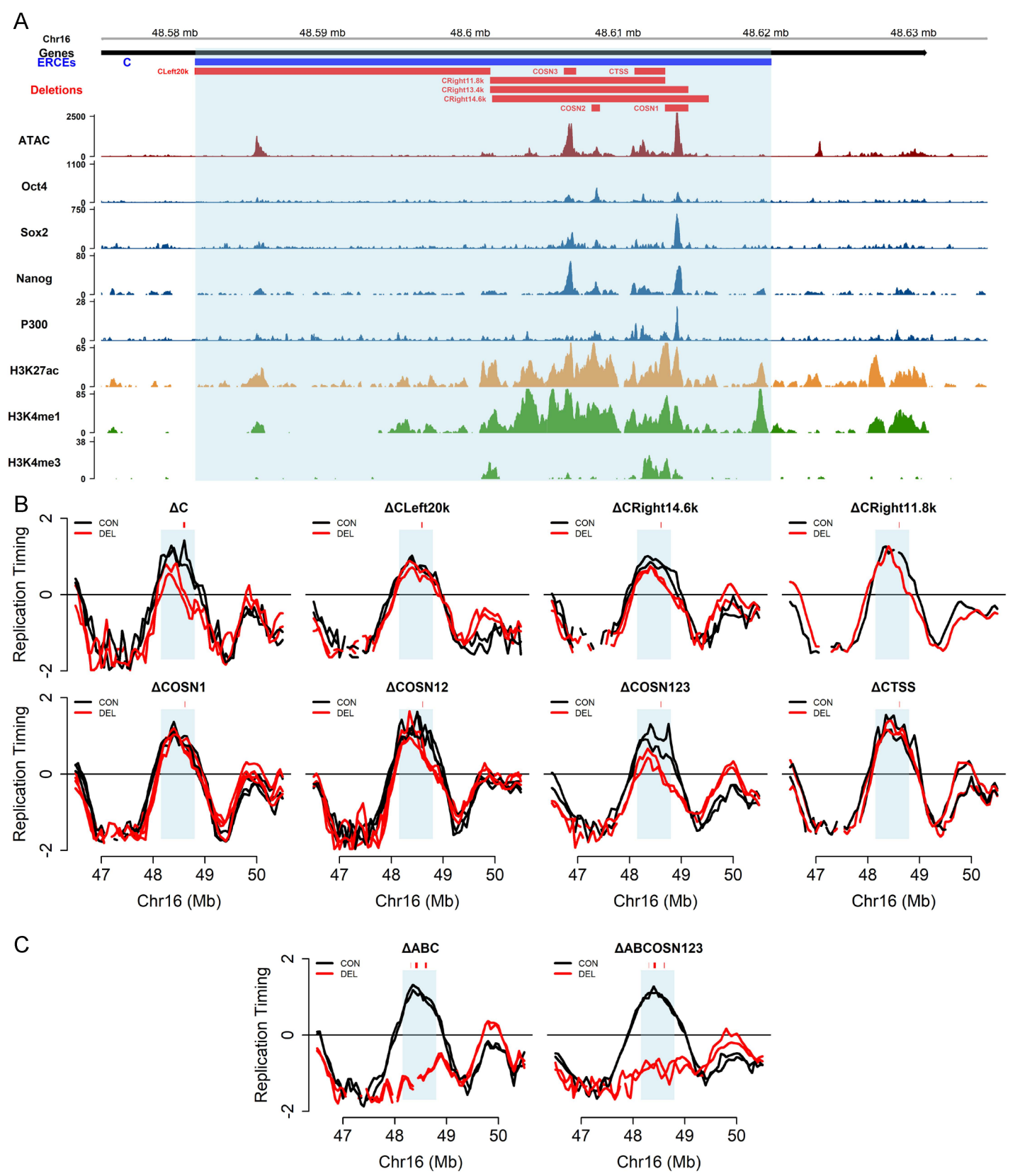
Takebayashi S-I, Ryba T, Gilbert DM. 2012b. Developmental control of replication timing defines a new breed of chromosomal domains with a novel mechanism of chromatin unfolding. *Nucleus* **3**: 500-507.

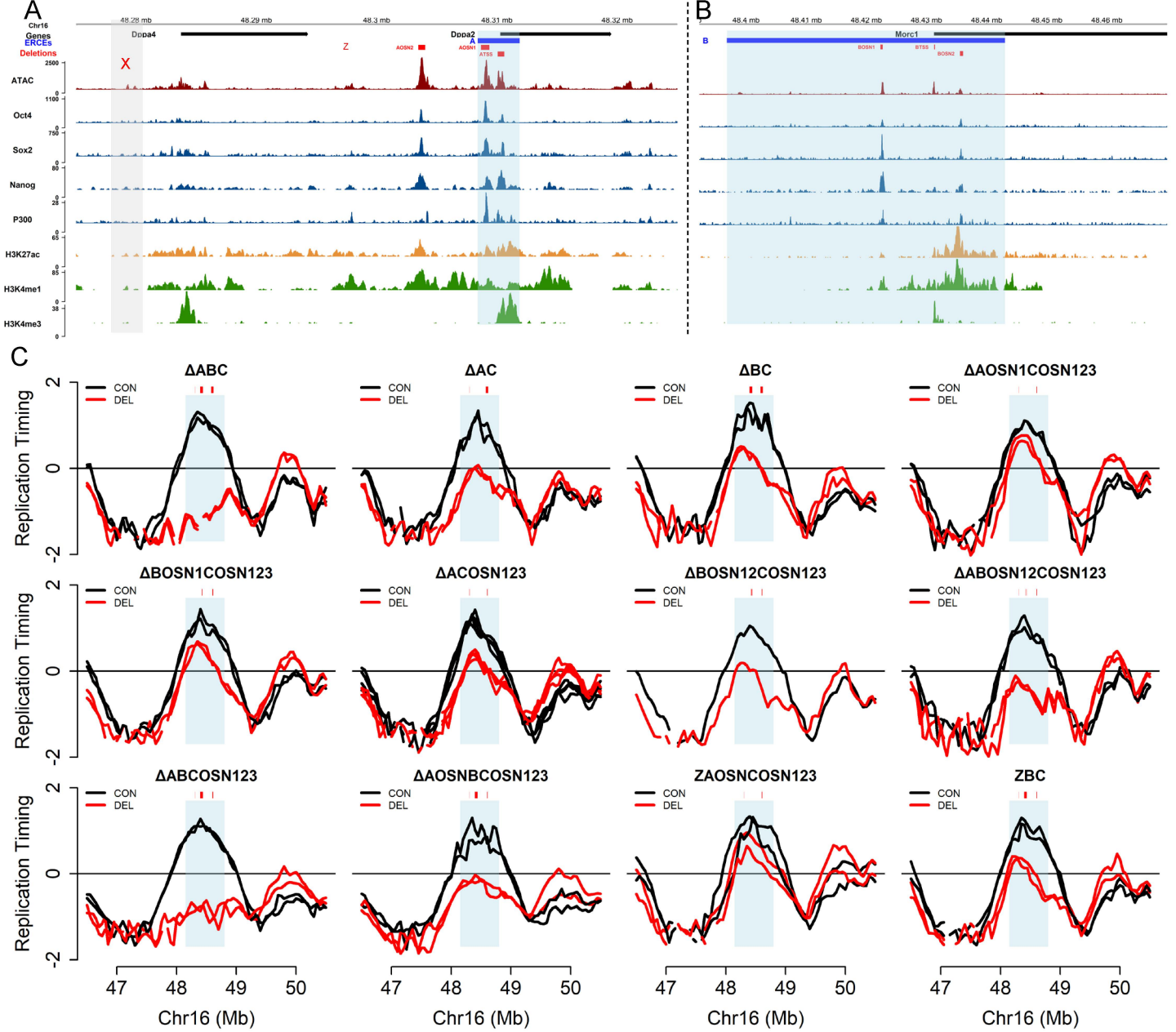
Vouzas AE, Gilbert DM. 2023. Replication timing and transcriptional control: beyond cause and effect - part IV. *Curr Opin Genet Dev* **79**: 102031.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**: R137.

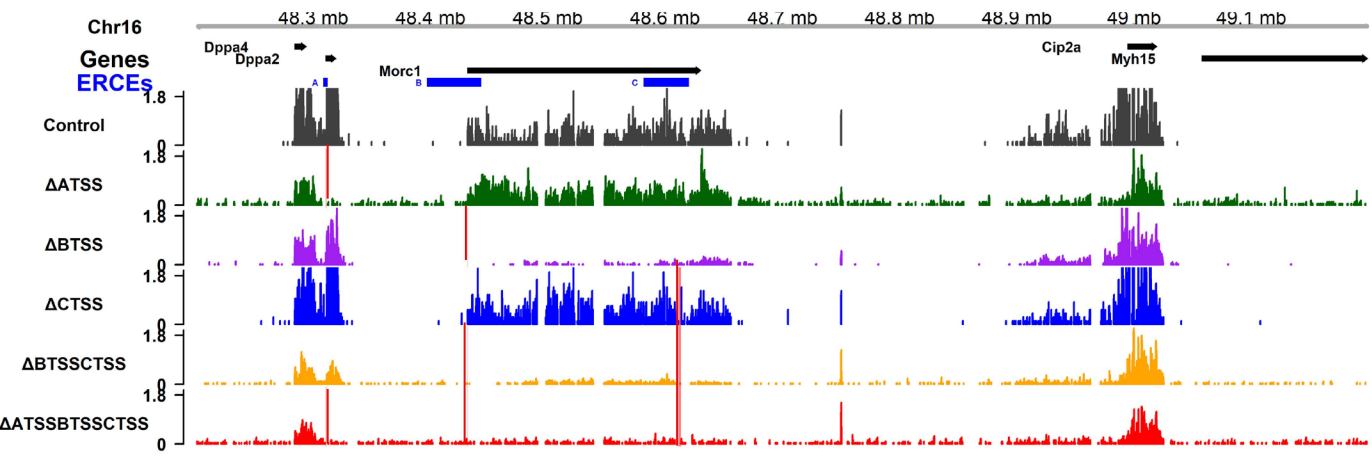
Zhao PA, Sasaki T, Gilbert DM. 2020. High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. *Genome Biology* **21**.


A
Chromosome 16
Chr16
48.3 mb
48.4 mb
48.5 mb
48.6 mb
Genes
ERCEs
Dppa4
Dppa2
Morc1
A
B
C
X
Y
ATAC
Oct4
Sox2
Nanog
P300
H3K27ac
H3K4me1
H3K4me3
B
F121-9 WT
129 ESC
CAS ESC
129 NPC
CAS NPC
ΔA
ΔB
ΔC
ΔAB
ΔAC
ΔBC
ΔABC
CON
DEL
Replication Timing
Chr16 (Mb)

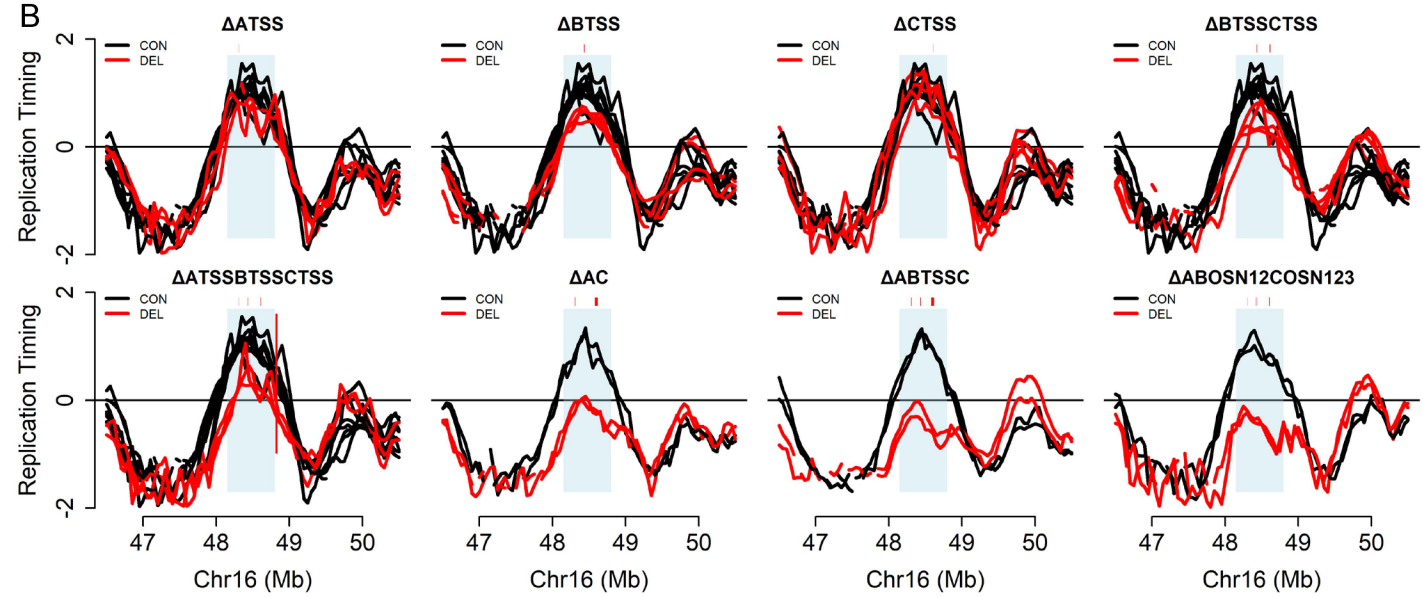

A
Chr16
48.58 mb
48.59 mb
48.6 mb
48.61 mb
48.62 mb
48.63 mb
Genes
ERCEs
C
Deletions
CLeft20k
COSN3
CTSS
CRight11.8k
CRight13.4k
CRight14.6k
COSN2
COSN1
ATAC
Oct4
Sox2
Nanog
P300
H3K27ac
H3K4me1
H3K4me3
B
ΔC
ΔCLeft20k
ΔCRight14.6k
ΔCRight11.8k
ΔCOSN1
ΔCOSN12
ΔCOSN123
ΔCTSS
CON
DEL
Replication Timing
Chr16 (Mb)
C
ΔABC
ΔABCOSN123

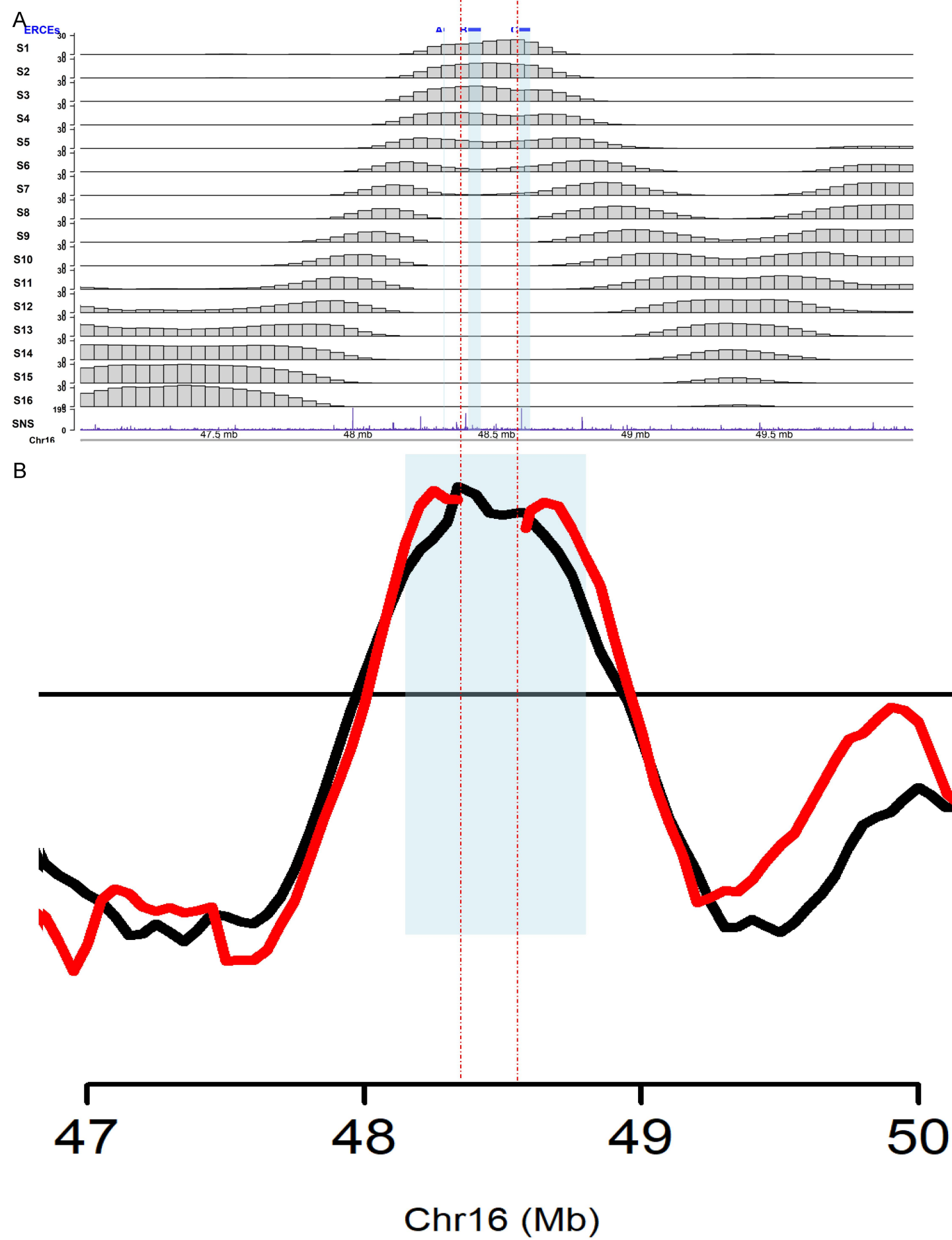

A
Chr16
Genes
ERCEs
Deletions
48.28 mb
48.29 mb
48.3 mb
48.31 mb
48.32 mb
Dppa4
Dppa2
X
Z
A
AOSN2
AOSN1
ATS5
ATAC
Oct4
Sox2
Nanog
P300
H3K27ac
H3K4me1
H3K4me3
B
48.4 mb
48.41 mb
48.42 mb
48.43 mb
48.44 mb
48.45 mb
48.46 mb
Morc1
BOSN1
BTS5
BOSN2
C
ΔABC
ΔAC
ΔBC
ΔAOSN1COSN123
ΔBOSN1COSN123
ΔACOSN123
ΔBOSN12COSN123
ΔABOSN12COSN123
ΔABCOSN123
ΔAOSNBCOSN123
ZAOSNCOSN123
ZBC
CON
DEL
Replication Timing
Chr16 (Mb)

A

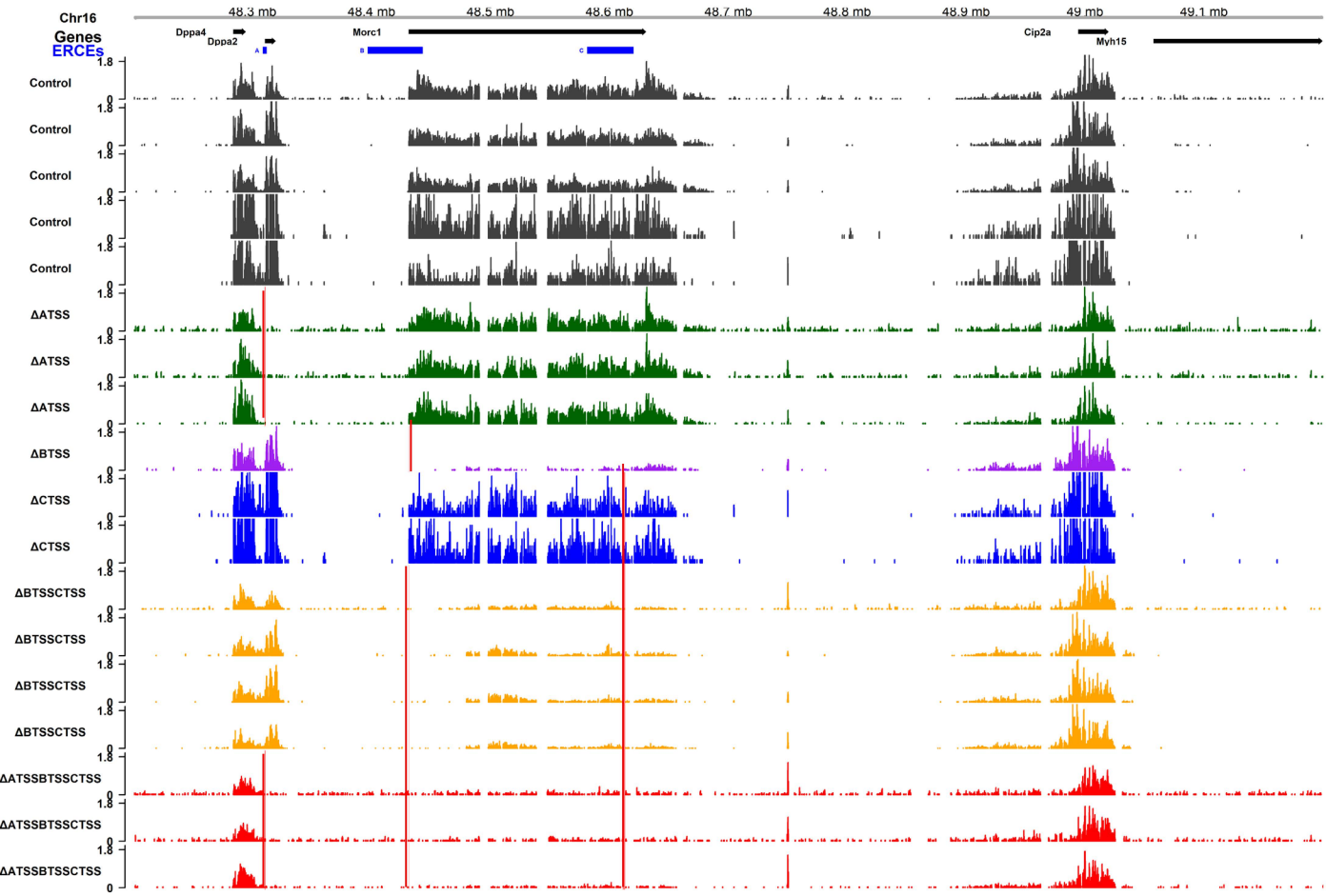

Chr16
48.3 mb
48.4 mb
48.5 mb
48.6 mb
48.7 mb
48.8 mb
48.9 mb
49 mb
49.1 mb
Genes
Dppa4
Dppa2
Morc1
Cip2a
Myh15
ERCEs
A
B
C
Control
ΔATSS
ΔBTSS
ΔCTSS
ΔBTSSCTSS
ΔATSSBTSSCTSS
1.8
0

B


ΔATSS
ΔBTSS
ΔCTSS
ΔBTSSCTSS
ΔATSSBTSSCTSS
ΔAC
ΔABTSSC
ΔABOSN12COSN123
CON
DEL
Replication Timing
2
0
-2
47
48
49
50
Chr16 (Mb)


A
ERCEs
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15
S16
SNS
Chr16
47.5 mb
48 mb
48.5 mb
49 mb
49.5 mb
B
47
48
49
50
Chr16 (Mb)

A


Chr16
48.3 mb
48.4 mb
48.5 mb
48.6 mb
48.7 mb
48.8 mb
48.9 mb
49 mb
49.1 mb
Genes
Dppa4
Dppa2
Morc1
Cip2a
Myh15
ERCEs
A
B
C
Control
Control
Control
Control
Control
ΔATSS
ΔATSS
ΔATSS
ΔBTSS
ΔCTSS
ΔCTSS
ΔBTSSCTSS
ΔBTSSCTSS
ΔBTSSCTSS
ΔBTSSCTSS
ΔATSSBTSSCTSS
ΔATSSBTSSCTSS
ΔATSSBTSSCTSS
1.8
0