Nucleation and spreading rejuvenate polycomb domains every cell cycle

Giovana M. B. Veronezi and Srinivas Ramachandran*

RNA Bioscience Initiative, and Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine

*Address correspondence to:
Srinivas Ramachandran, Mail Stop: 8101, 12801 East 17th Ave. L-18-9102, Aurora, CO 80045; tel: 303-724-3308; email: srinivas.ramachandran@cuanschutz.edu

Abstract

Gene repression by the polycomb pathway is essential for proper development in metazoans. Polycomb domains, characterized by trimethylation of histone H3 (H3K27me3), carry the memory of repression through successive cell divisions, and hence, need to be maintained to counter dilution of parental H3K27me3 with unmodified nascent H3 during replication. However, how locus-specific H3K27me3 is maintained despite replication is unknown. To address this, we defined nucleation sites within each polycomb domain in mouse embryonic stem cells. Next, to track H3K27me3 dynamics in unperturbed cells during replication, we developed CUT&Flow, which maps H3K27me3 domains as a function of the cell cycle stage. Using CUT&Flow, we show that post-replication rejuvenation of polycomb domains occurs by nucleation and spreading, using the same nucleation sites used during *de novo* domain formation. The timing of nucleation relative to the cell cycle can be explained by a combination of replication timing and H3K27me3 deposition kinetics derived from *de novo* domain formation. Thus, competition between H3K27me3 deposition and nucleosome turnover drives both *de novo* domain formation and maintenance during every cell cycle.

Key words: PRC2, nucleation, Polycomb, epigenome dynamics

Introduction

Many chromatin states are cell-type specific and maintained through successive cell divisions. One such chromatin state is the trimethylation of histone H3 lysine 27 (H3K27me3), catalyzed by Polycomb Repressive Complex 2 (PRC2). The Polycomb and Trithorax groups (PcG and TrxG) of proteins are essential for proper cell fate specification during development (Francis and Kingston, 2001). Regions silenced by PcG proteins form broad domains of H3K27me3 on the genome (Bowman et al., 2014; Kundu et al., 2017). These domains underlie the cell type-specific silencing by PcG, but the mechanisms of their maintenance in spite of dilution during replication remains unclear.

Several factors drive locus-specific deposition of H3K27me3. Core PRC2 consists of EZH2, SUZ12, RBAP46/48, and EED, with EZH2 being the catalytic subunit (Cao et al., 2002; Kuzmichev et al., 2002; Muller et al., 2002). SUZ12 is the central scaffold and has been implicated in directing PRC2 activity genome-wide (Hojfeldt et al., 2019; Hojfeldt et al., 2018). PRC2 spreading is thought to depend on its EED subunit. EED can specifically bind H3K27me3, which allosterically increases the catalytic activity of the EZH2 subunit (Margueron et al., 2009). Thus, EED can bind pre-existing H3K27me3 and simulate EZH2 to modify a spatially adjacent H3 tail. Polycomb Repressive Complex 1 (PRC1) (Shao et al., 1999) is an E3 ubiquitin ligase that mono-ubiquitinates H2A K119 (Wang et al., 2004). H2A ubiquitination can drive PRC2 targeting (Blackledge et al., 2020; Fursova et al., 2019). Finally, H3.3 deposition is important for PRC2 activity (Banaszynski et al., 2013), possibly because PRC2 nucleation sites may feature nucleosome turnover and have DNA accessibility (Deal et al., 2010).

Studies in *Drosophila*, murine, and human cells have shown that two mechanisms lead to H3K27me3 deposition. In the first, specific DNA sequences act as nucleation sites for PRC2 (Schorderet et al., 2013; Sing et al., 2009; Steffen and Ringrose, 2014; Woo et al., 2010). Early work in *Drosophila* identified sequences in Hox clusters that could silence transgenes in a cell type specific manner and these sequences were denoted as Polycomb Response Elements (PREs) (Ringrose and Paro, 2004; Simon et al., 1993). PREs are thought to be direct binding sites of PRC2. Second, via its EED subunit, PRC2 can also use pre-existing H3K27me3 to modify nearby unmodified histones through a second, read-write mechanism similar to DNA methyltransferases (Oksuz et al., 2018). These two mechanisms have been succinctly described as “nucleation” and “spreading” (Kuroda et al., 2020). Nucleation and spreading has been observed in a system where H3K27me3 domains were ablated and then reconstructed *de novo* in an inducible manner (Oksuz et al., 2018).

The extent to which nucleation occurs in cells in absence of perturbation and/or *de novo* domain formation is unknown. DNA replication dilutes global H3K27me3 levels two-fold. Hence, H3K27me3 restoration must occur every cell cycle after DNA is replicated. Tracking H3K27me3 levels through the

cell cycle in HeLa cells has shown that levels of this modification are restored predominantly in the G1 phase following replication (Alabert et al., 2015). However, it is unknown if nucleation is involved in driving the restoration of H3K27me3 post-replication or if the parental H3K27me3 redeposited across the domain (Escobar et al., 2019; Gaydos et al., 2014) predominantly drive methylation of the nascent H3. Mechanism of replenishment of H3K27me3 over the cell cycle is not clear due to the technical challenges in tracking H3K27me3 dynamics over the cell cycle. *De novo* domain formation after PRC2 ablation starts with a clean slate with respect to H3K27me3 enrichment – the H3K27me3 levels over this time course go from almost zero to steady-state levels. However, the change in H3K27me3 post-replication is within a much smaller range during the cell cycle (Reveron-Gomez et al., 2018). Hence, previous studies tracking H3K27me3 temporally post-replication have not tested if the nucleation and spreading model of *de novo* domain formation applies to H3K27me3 maintenance post-replication. Importantly, even if nucleation operated during the cell cycle, it is not known if nucleation sites used during *de novo* domain formation would also be used during the cell cycle.

To ask if nucleation occurs in unperturbed cells during the cell cycle, we developed a framework to identify nucleation sites based just on the intradomain kinetics of H3K27me3. The nucleation and spreading model predicts that within a polycomb domain, its nucleation sites should have the fastest kinetics of increase in H3K27me3 enrichment. We first partitioned chromosomes into H3K27me3 domains and then identified nucleation sites for each of those domains. We then asked if nucleation occurs in unperturbed cells. To do this, we developed a new method, CUT&Flow by coupling Cleavage Under Target and Tagmentation (CUT&Tag) with flow cytometry, to define H3K27me3 domains as a function of the cell cycle stage. With CUT&Flow, we observe nucleation at specific cell cycle stages, at sites that also nucleate during *de novo* domain formation. The cell cycle stage when a site nucleates can be defined by a combination of the site's replication timing and the site's inherent H3K27me3 deposition kinetics. Thus, the opposing forces of PRC2 nucleation and DNA replication can explain the steady-state landscape of Polycomb domains genome-wide. Our results support the idea that a unified nucleation mechanism governs both *de novo* domain formation and rejuvenation of H3K27me3 domains every cell cycle.

Results

Identification of Polycomb domains

To determine nucleation sites for PRC2, we first set out to define Polycomb domains in murine embryonic stem cells (mESCs) using published epigenomic data. Since Polycomb domains are expected to be on the scale of kilobases, traditional peak calling is not appropriate. Instead, we devised a simple domain-calling algorithm, which first performs smoothing of H3K27me3 ChIP-seq datasets by calculating running

averages with a large window size (1000 bp). Then by using a cut-off of 4-fold enrichment of the ChIP signal over input and 2-fold enrichment of the ChIP signal over the genome-wide median, we identified regions with significant levels of H3K27me3. We then coalesced adjacent bins with high H3K27me3 levels to designate domains (see Methods). Reassuringly, our algorithm identified large domains at Hox clusters, which are known to have high enrichment of H3K27me3 in mESCs (**Figure S1**). We identified domains at various scales from 1 kb to greater than 100 kb, which can be observed in heatmaps of enrichment of H3K27me3 plotted relative to domain centers and ordered by size of the domains (**Figure 1A**). We also calculated meta-domain averages at various size classes to observe domain boundaries across many domains (**Figure 1B**). Interestingly, we observe a significant decrease in the number of domains with increasing size (**Figure 1C**). We also observe larger domains to have higher enrichment of H3K27me3 (**Figure 1D**). Due to the stochastic nature of histone partitioning during replication, mechanisms of H3K27me3 deposition that depend on pre-existing H3K27me3 are predicted to be highly sensitive to domain size. Larger domains could partition to both daughter chromosomes more equitably compared to smaller domains (Ramachandran and Henikoff, 2015). This prediction agrees well with our observation that H3K27me3 enrichment increases monotonically with domain size. These observations support our approach of using domains as the unit of epigenomic analyses of Polycomb function.

Large Polycomb domains have diverse intra-domain H3K27me3 kinetics

In a published study (Hojfeldt et al., 2018), mESC were treated with an Ezh2 inhibitor for 10 days, which resulted in a reduction of H3K27me1, H3K27me2, and H3K27me3 to levels undetectable by immunoblots. The inhibitor was then washed out and recovery of H3K27 methylation states was tracked by immunoblot and spike-in normalized ChIP-seq 4, 8, 16, 24, 48, and 96 hours post-washout. We asked if we could use the timed ChIP-seq datasets post-washout to infer H3K27me3 kinetics inside the Polycomb domains we defined.

We reasoned that the absolute enrichment of H3K27me3 at steady state could obscure differences in kinetics between nucleation sites and the rest of the domain for several reasons. First, earlier studies have shown that nucleation sites feature high nucleosome turnover (Deal et al., 2010). Second, H3.3 incorporation (another measure of turnover) is essential for the maintenance of Polycomb domains (Banaszynski et al., 2013), and *in vitro*, PRC2 binding itself might be stimulated by the presence of a nucleosome-free region (Wang et al., 2017). Thus, to remove the effect of absolute H3K27me3 levels on observed kinetics, we plotted all time-points post-drug washout after normalization to the final time-point (96-hour) ChIP signal. We observed a range of H3K27me3 recovery kinetics within the domain once we normalized each segment to the levels at 96 hours post-washout, exemplified by the Polycomb domain covering the *Hoxa* locus (**Figure 2A**). We performed k-means clustering (k=7) of H3K27me3 recovery patterns at 25 bp non-overlapping segments within a domain (**Figure 2B**). To order the clusters based

on the speed of their recovery, we defined t_{40} as the time taken to reach 40% of steady-state H3K27me3 levels. This was chosen because 40% represented a level until which there was a steady increase in H3K27me3 levels across the domain. t_{40} enabled automated ordering of the kinetic groups from fastest to slowest recovery (**Figure 2B, C**).

One model that could explain the variation in kinetics is the nucleation and spreading model (Kuroda et al., 2020). The nucleation and spreading model posits the presence of nucleation sites which regain H3K27me3 first due to direct DNA binding of PRC2, and subsequently the other sites in the domain regain H3K27me3 due to spreading. A consequence of this model is that adjacent sites will have gradual changes in H3K27me3 restoration kinetics: i.e. a slow site will not be adjacent to a fast site. To ask if this was the case, we mapped the kinetic clusters back onto the domain and observed a "peak and valleys" distribution of cluster id that would be expected for nucleation and spreading (**Figure 2D**). Even though the k-means clustering did not use positional information, boundaries between kinetic clusters were strongly enriched for transitions to successive clusters, indicating that there was a stepwise change in the rate of H3K27me3 recovery (**Figure 2E**). The stepwise change could not be reproduced when cluster numbers within the domain were shuffled, indicating that this was due to the non-random distribution of kinetic clusters (**Figure 2E**). This strongly suggests that the spreading of PRC2 from nucleation sites could explain the observed kinetics of H3K27me3 post drug washout. Even though H3K27me3 patterns return to pre-treatment, steady-state distribution ~96 hours after drug washout, our analysis shows that the kinetics of recovery is not uniform.

Polycomb domains are independent units for PRC2 kinetics

We hypothesized that the kinetic group within each domain with the fastest H3K27me3 recovery represents nucleation sites for PRC2. Interestingly, when we compared the H3K27me3 recovery at large domains defined at Hoxa and Hoxd gene clusters, we observed a significant difference in their kinetics: the average t_{40} of the fastest kinetic group of the Hoxa domain was ~7 h 23 m, whereas that of the fastest kinetic group in the Hoxd domain was only 5 h. In other words, even though both the domains feature characteristics of nucleation and spreading, Hoxa nucleation is 32% slower than Hoxd. This result led us to ask how t_{40} varies across domains genome-wide. We chose 1362 domains that were common across different datasets and at least 10 kb in length. For each domain, we performed k-means clustering of H3K27me3 spike-in ChIP across the time points, normalized to the 96-hour time point, then sorted the kinetic groups from slowest to fastest based on each group's average t_{40}. We observed a broad diversity in H3K27me3 recovery genome-wide, with the fastest group within each domain having t_{40} values ranging from 3h 19m to nearly 48 hours (**Figure 3A**). This wide range in recovery times across domains is reflected in the highly overlapping distribution of t_{40} values for each cluster (**Figure 3B**). Thus, each domain has its characteristic nucleation rate, presumably due to differences in PRC2 recruitment strength

at their nucleation sites. This analysis also raises an important point that nucleation and spreading within each domain might not be apparent if recovery kinetics at H3K27me3 peaks were compared genome-wide instead of within individual domains. When compared across the genome, all the fast recovery sites are found in the top few domains. By considering each domain individually, we not only reveal that each domain has its unique nucleation rate, but also identify putative nucleation sites for each of those H3K27me3 domains.

Nucleation site chromatin properties identify distinct classes of Polycomb domains

We hypothesized that kinetic clusters with the fastest H3K27me3 recovery kinetics within each domain represented nucleation sites for PRC2. Shorter t_{40} values would denote stronger nucleation sites. At the Hoxd locus, two of our predicted nucleation sites overlapped with previously characterized nucleation sites: Hoxd11-12 (Woo et al., 2010), and Evx (Oksuz et al., 2018), consistent with our kinetic method being able to identify nucleation sites. Different molecular components are important for PRC2 activity. CpG islands and bivalent promoters have been linked with PRC2 nucleation (Hojfeldt et al., 2018; Mendenhall et al., 2010; Oksuz et al., 2018; Riising et al., 2014). Jarid2 and PCL2 are DNA binding proteins that are auxiliary subunits of PRC2. Based on the well-characterized roles of Jarid2, PCL2, Ezh2, Suz12, H2AK119 ubiquitination (uH2A), chromatin accessibility, and H3.3 incorporation, we asked if these features are enriched at nucleation sites relative to sites with slow kinetics of H3K27me3 recovery. We observed significant enrichment of all these features at putative nucleation sites compared to the slowest sites in each domain, with the PRC2 components showing strongest enrichment (**Figure 3C**).

We then asked if all these components are enriched to the same extent at nucleation sites across domains. We performed k-means clustering of the Z-scores of enrichments of these features at nucleation sites across domains to obtain six groups of domains. The presence/absence of bivalent promoters and CpG islands, and t_{40} at nucleation sites were then added in the same order as the identified groups of domains. Finally, we ordered the groups by time (**Figure 3D, E**). We observed that the PRC2 components, H3K27me3, uH2A, and ATAC clustered together and had the highest enrichment at domains with the fastest nucleation. H3.3 was enriched in groups 2, 3, and 5, with a pattern that was distinct from Polycomb components. Group 5, with the highest levels of H3.3, also featured slower nucleation. As a first approximation, this analysis points to three classes of nucleation sites: fast nucleation sites that feature strong binding of PRC1 and PRC2 at steady state; slower nucleation sites that feature H3.3; and slowest nucleation sites that had low levels of all components surveyed. Interestingly, CpG islands and bivalent promoters overlap with groups 1, 2, and 4, underlining the fact that sequence only partially explains PRC2 kinetics. In summary, our analysis suggests that the components that drive PRC2 activity are enriched at nucleation sites relative to spreading sites and the extent of enrichment correlates with the speed of H3K27me3 deposition post drug washout.

Domain kinetics differentiates role of molecular components in PRC2 nucleation, spreading, and catalytic activity

It is difficult to assign a causal role for different components/features in nucleation with steady-state distributions (as described in Figure 3) for two reasons. First, PRC2 components could cluster together because they exist as a complex. Second, components/features may correlate with nucleation rather than lead to nucleation. For example, PRC1 and PRC2 can form a feedback loop, meaning PRC1 could be enriched at nucleation sites either when it leads to or follows PRC2 nucleation. Hence, to identify drivers of nucleation, we analyzed published H3K27me3 ChIP after various perturbations (Fursova et al., 2019; Hojfeldt et al., 2019; Hojfeldt et al., 2018; Oksuz et al., 2018). We determined the change in the ratio of H3K27me3 enrichment at the nucleation sites to the spreading sites after these perturbations. We term the change in ratio of the H3K27me3 enrichment due to perturbation of a component as the "nucleation strength" of that component. This measure is resilient to changes in overall PRC2 activity: for a given perturbation, even if there is a global decrease in H3K27me3 levels due to reduced overall activity of PRC2, if nucleation is the same, then the nucleation strength wouldn't change significantly. On the other hand, if global H3K27me3 levels were the same post-perturbation, but nucleation was affected, we would see a change in nucleation strength. The loss of components that drive nucleation will lead to a significant loss in nucleation strength. The loss of components that drive spreading will lead to a significant gain in nucleation strength.

The C-terminal region of Suz12 is the central scaffold for the core PRC2 complex, whereas its N-terminal is the binding site for various accessory subunits (Laugesen et al., 2019). Previous studies have deleted the N-terminal region of Suz12 and found the global levels of H3K27me3 unaffected, but the pattern of H3K27me3 changed (Hojfeldt et al., 2018). We observed that Suz12-N-terminal deletion led to the strongest loss in nucleation strength, indicating a central role for full-length Suz12 in nucleation (**Figure 4A**). Although we found the loss of variant PRC1 complexes (vPRC1) to show a significant decrease in nucleation strength, the effect was not as strong as losing Suz12-N-terminal region. This suggested that vPRC1 might be downstream of other components whose signals are integrated by the Suz12-N-terminal region. Loss of canonical PRC1 (cPRC1) complex resulted in minor loss of nucleation strength (**Figure 4A**).

EED binds H3K27me3 through an aromatic cage (Margueron et al., 2009). Mutations of residues in the aromatic cage result in loss of EED binding to H3K7me3. EED's loss of binding to H3K27me3 results in PRC2 unable to spread (Oksuz et al., 2018). Oksuz et al. (Oksuz et al., 2018) introduced these mutations in mESCs and then performed H3K27me3 ChIP. In the cage mutants, the enrichment of H3K27me3 at the nucleation sites we identified based on recovery kinetics was even more exaggerated compared to WT, leading to an apparent increase in nucleation strength with these mutants (**Figure 4A**).

We next asked the extent to which PRC2 accessory subunits drive nucleation. Jarid2, Epop, Aebp2, PCL1, PCL2, and PCL3 bind PRC2 via the Suz12 N-term and have various extents of DNA binding ability. Recent work suggested that these factors collectively maintain H3K27me3 levels at target sites (Healy et al., 2019; Hojfeldt et al., 2019). However, previous analyses were based on H3K27me3 peaks and did not compare nucleation sites to spreading sites. We analyzed the data, where each of these accessory subunits was deleted either individually or in combination with other subunits and H3K27me3 ChIP was performed in mESCs. We calculated the change in H3K27me3 at nucleation sites identified by our analysis of domain kinetics normalized to the spreading sites within the same domains. This analysis showed that no factor by itself had a major impact on nucleation, recapitulating analysis on H3K27me3 peaks (**Figure 4B**). Jarid2, Aebp2, and Epop combined deletions had an effect similar to PCL2. Further deletions in combination with PCL1, 2, 3 resulted in more severe effects on nucleation. The strongest effect was seen when all accessory subunits were deleted. This effect was similar in magnitude to Suz12 N terminal deletion, which supports the idea that the accessory subunits work with PRC2 via the Suz12 N-terminal region.

The accessory subunits have different DNA binding activities and abilities to bind H2A-K119ub. Therefore, even if the global effects of the deletions point to a trend of collective action, at any given polycomb domain, the accessory subunits could be acting individually. To ask if this is the case, we performed k-means clustering of the nucleation strength changes due to each perturbation (**Figure 4C**). Nucleation strength change is the change in the enrichment of H3K27me3 at the nucleation sites we defined normalized to the spreading sites after a perturbation. This analysis identified groups of domains that had differential requirements of accessory subunits for nucleation. Most domains had an increase in nucleation upon loss of Aebp2, Jarid2, and Epop (Groups 2-5, **Figure 4C**), but nucleation was weakened in Group 1 (**Figure 4C, D**). This strongly suggests that nucleation in Group 1 is dependent on Jarid2, Aebp2, and Epop, and the loss of these subunits likely leads to redirection of PRC2 to domains in Groups 2-5. In contrast, PCL proteins were essential for nucleation in Group 2 and loss of PCL proteins led to increased nucleation in Group 1, suggesting redirection of PRC2 to Group 1 when PCL proteins were absent (**Figure 4C, E**). Suz12 and vPRC1 were essential for nucleation of majority of domains, supporting the idea that Suz12 integrates accessory protein binding and that vPRC1 reinforces PRC2 nucleation. Taken together, analysis of nucleation strength reveals that accessory proteins drive domain-specific nucleation, pointing to the possibility of specific cues within each domain that recruit PRC2. Our analyses thus far show that the nucleation sites we identified have chromatin properties known to be important for Polycomb function. Furthermore, the nucleation sites show specific requirements for accessory subunits, indicating a locus-specific ability to recruit PRC2. Next, we wanted to find out if these nucleation sites might also recruit PRC2 to restore H3K27me3 levels every cell cycle.

CUT&Flow profiles Polycomb domains across the cell cycle

If nucleation occurred post-replication, the enrichment of H3K27me3 at nucleation sites relative to spreading sites would be highest at the early stages of rejuvenation of a domain post-replication. To track nucleation in unperturbed cells, we hypothesized that cells in different stages of the cell cycle would feature H3K27me3 domains at different extents of nucleation and spreading. To capture H3K27me3 domains at various stages of the cell cycle, we combined H3K27me3 profiling by Cleavage Under Targets and Tagmentation (CUT&Tag) (Kaya-Okur et al., 2019) with flow-sorting to fractionate cells at different stages of the cell cycle. The technique, which we term CUT&Flow, involves similar steps to CUT&Tag until tagmentation: isolation of nuclei, binding of primary and secondary antibodies, binding of pA-Tn5, and tagmentation. The nuclei are then labeled with propidium iodide and sorted into 5 gates based on DNA content: G1, early S, mid S, late S, and G2/M (**Figure 5A, B**). Finally, the sorted nuclei are lysed and PCR is performed to obtain sequencing libraries, each one representing H3K27me3 profiles of cells at a specific stage of the cell cycle.

To ask if CUT&Flow profiles H3K27me3 dynamics across the cell cycle, we first took advantage of the fact that DNA replication dilutes H3K27me3 levels. We asked if H3K27me3 enrichment was diluted at specific polycomb domains in the different cell cycle gates of CUT&Flow. Notably, just by visual inspection, individual domains showed a significant drop of H3K27me3 enrichment at just one cell cycle gate followed by recovery in the next gates (**Figure 5C**). Extending this analysis genome-wide, we assigned a CUT&Flow cell cycle gate to each domain based on when that domain featured a sudden drop in H3K27me3. We were able to assign a cell cycle gate to most of the domains (**Figure 5D**). The highest level of H3K27me3 for a domain occurred at the third cell cycle gate downstream of dilution, suggesting dilution followed by recovery of H3K27me3 due to replication. We observed the dilution of H3K27me3 to be around one-half fold (**Figure 5D, right**). If the dilution we observe in CUT&Flow gates occurred due to replication, then the replication timing of the domains should correlate with the timing of the dilution. To determine the replication timing of the Polycomb domains, we analyzed published Repli-Seq data for mESC (Zhao et al., 2020), where cells were sorted into 16 S-phase fractions after BrdU labeling. Then, BrdU IP was performed followed by sequencing. The S-phase fraction with the highest Repli-Seq signal for a genomic region represents the predominant replication timing of that region. The domains that underwent dilution in G1 and early S replicated earliest, followed by mid S, late S, and finally G2 (**Figure 5E**). Thus, replication timing strongly correlated with the timing of H3K27me3 dilution mapped by CUT&Flow and leads us to conclude that CUT&Flow captures H3K27me3 dynamics across the cell cycle in unperturbed cells.

CUT&Flow captures nucleation post-replication

To ask if we can observe nucleation in CUT&Flow data, we calculated the ratio of H3K27me3 enrichment at nucleation sites to the enrichment at spreading sites for the different CUT&Flow gates. We first plotted the spreading sites-normalized H3K27me3 profiles from G1, Early S, Mid S, Late S, and G2 at the Hoxd locus and observed robust H3K27me3 domains with similar boundaries to steady-state profiles (**Figure 6A**). Interestingly, we observed an increase in enrichment at sites corresponding to kinetic clusters 6 and 7 (kinetic clusters defined in Figures 2 and 3) in the Mid S dataset compared to the other datasets. We calculated the change in enrichment over the cell cycle for sites in each kinetic cluster normalized by enrichment in spreading sites (kinetic clusters 1 and 2). Clusters 3 and 4 did not fluctuate much over the cell cycle, but clusters 5, 6, and 7 showed a peak in Mid S, with cluster 7 having the strongest increase in Mid S (**Figure 6B**). From published Repli-Seq data for mESC (Zhao et al., 2020), we observed the Hoxd locus to predominantly replicate in Mid S (**Figure 6C**). Thus, the spike in enrichment in nucleation sites (which correspond to kinetic cluster 7) occurred close to when the domain was replicated, pointing to rapid nucleation following replication. Our results suggest that CUT&Flow captures cell cycle-dependent nucleation in the Hoxd polycomb domain and this nucleation appears to be coupled to the replication of the domain.

PRC2 kinetics predict cell cycle-dependent nucleation

To examine nucleation genome-wide, we calculated the ratio of H3K27me3 enrichment at nucleation sites to the enrichment at spreading sites for the different CUT&Flow gates. Based on the nucleation behavior of the Hoxd locus, this ratio is predicted to spike at a specific cell cycle stage when the domain nucleates. We analyzed cell cycle stage-specific enrichment of H3K27me3 at nucleation and spreading sites that were defined based on *de novo* domain formation (in Figures 2 and 3). For most of the domains, we found the nucleation site enrichment increased dramatically in one phase of the cell cycle and then declined (**Figure 7A**). These strong cell cycle stage-specific events were marked by H3K27me3 levels at least 1 Z-score higher than other cell cycle stages, followed by a decrease presumably as H3K27me3 spread through the domain (**Figure 7A**). The spreading sites showed a decrease in Z-scores compared to the nucleation sites, and this decrease correlated with their distance from the nucleation site. Thus, the strongest gain in H3K27me3 was at the nucleation sites at specific cell cycle stages, followed by sites adjacent to the nucleation sites, and so on (**Figure 7B**). This suggests that the post-replication rejuvenation of polycomb domains occurs by nucleation and spreading, via the same nucleation sites used during *de novo* domain formation.

If these nucleation events occur post-replication, the replication timing of the domains that spike in early S is expected to be earlier than those of mid S, and so on, with the domains that spike in late S and

G2 having the latest replication times. For each nucleation site, we determined the peak of the Repli-Seq signal. We observed the cumulative distribution of replication peak to be leftmost for those nucleation sites whose enrichment spikes in early S, followed by mid S, late S, and G2 (**Figure 7C, top**). Nucleation sites that spiked in G1 had replication timing similar to those domains that spiked in late S (**Figure 7C, bottom**). The correlation between the timing of H3K27me3 nucleation events determined by CUT&Flow and replication timing suggests that there is strong coordination between these processes. The fact that nucleation occurs immediately following replication points to rapid PRC2 deposition coupled with replication in contrast to the slower global restoration of total H3K27me3 (Alabert et al., 2015).

We next asked if the cell cycle stage-specific spike in H3K27me3 nucleation sites correlated to the chromatin features of the domain. To do this, we intersected the cell cycle stage of the nucleation spike of a domain with the groups defining chromatin states of the domain (groups that were defined in **Figure 4**). We observed a high overlap that was statistically significant between Group 1 and the nucleation spike in mid S (**Figure 7D**). We observed depletion of Group 1 and enrichment of Group 5 domains with nucleation spike in G1. Finally, we observed an enrichment of Group 6 domains with a nucleation spike in G2. Group 1 has the fastest kinetics in *de novo* domain formation and Group 6 has the slowest kinetics (**Figure 4**). In other words, fast nucleation sites nucleate in Mid S, and slow nucleation sites nucleate in G1 and G2, suggesting a correlation between nucleation kinetics in *de novo* domain formation and nucleation during the cell cycle (**Figure 7D**). This led us to ask if PRC2 nucleation kinetics was coupled to replication timing. We have a measure of *de novo* PRC2 kinetics as the time to 40% recovery after drug washout (t_{40}). If a nucleation site displays a large gap between replication and nucleation during the cell cycle, for example, if a site replicates in early S but nucleates only in G1, that could be due to slow PRC2 nucleation kinetics at that site. To ask if this is the case, we first divided nucleation sites that spike in each phase of the cell cycle into their replication timing bin based on Repli-Seq: early S, mid S, and late S. We then plotted the t_{40} for each replication timing bin. For domains nucleating in early S, we observe a majority of them replicated in early S as expected (**Figure 7E**). The domains that replicated in Early S displayed significantly faster kinetics than those that replicated in mid and late S. The mid and late S replicating domains that nucleate in early S surprisingly have replication timing that would be after nucleation in early S. Thus, these sites seem to nucleate pre-replication, presumably to account for slower nucleation kinetics. We observe a similar phenomenon for domains that nucleate in mid S – the domains that replicate in early and mid S have significantly faster kinetics than those that replicate in late S (**Figure 7F**). Finally, for domains that nucleate in G2, those that replicate in late S have significantly faster kinetics than those that replicate in early and mid S (**Figure 7G**). Thus, H3K27me3 kinetics during *de novo* domain formation correlates with the time between replication and nucleation during the cell cycle. Our results suggest that a balance of locus-specific nucleation strengths and the cell cycle result in the genome-wide H3K27me3 landscape observed at steady state.

Discussion

We identified nucleation sites for every polycomb domain greater than 10 kb in size in mouse embryonic stem cells and have shown that rejuvenation of H3K27me3 every cell cycle occurs by nucleation and spreading. We were able to show this by analyzing a broad histone modification in the context of domains. These analyses have shown that each H3K27me3 domain represents a microcosm of the genome-wide reader/writer/eraser kinetics. Fitting with the microcosm assumption, we find polycomb domains feature a wide range of PRC2 kinetics. This range of PRC2 kinetics across domains has precluded previous attempts to identify nucleation sites genome-wide: any genome-wide criteria to identify nucleation sites would identify only the top enrichments and not segments in each domain that could nucleate.

Genetic studies and reporter assays have been traditionally used to identify developmental enhancers and PREs in *Drosophila* (Steffen and Ringrose, 2014). However, genetic deletions have been notoriously difficult in identifying PREs in mammals presumably due to a high level of redundancy (Schorderet et al., 2013). Here we show that kinetics of H3K27me3 in the timescale of hours, both in the context of drug-washout and replication, can be a viable alternative to identify PRC2 nucleation sites. We defined nucleation sites without relying on any steady-state chromatin properties, nevertheless our analyses tie together steady-state knowledge of Polycomb function with nucleation. These include enrichment of H3K27me3, Suz12, Ezh2, and accessory PRC2 subunits at nucleation sites compared to spreading sites, and significant overlap between nucleation sites and CpG islands and bivalent promoters.

By considering properties important for nucleation domain-by-domain, we find a specific subset of PRC2 components that drive the nucleation of specific domains. Early studies with PCL2 suggested that it directed nucleation at a subset of H3K27me3 peaks, whereas Jarid2 drove nucleation at a different subset (Perino et al., 2020). However, later studies suggested that all accessory subunits collectively drive PRC2 activity (Healy et al., 2019; Hojfeldt et al., 2019). We show a separation of nucleation activity between accessory subunits at different domains by specifically quantifying loss of nucleation strength in each domain rather than absolute loss genome-wide. Notably, we see stronger nucleation in domains that do not rely on a particular subunit when that subunit is lost. This strongly suggests a redistribution of PRC2 to nucleation sites driven by other accessory subunits when a particular subunit is lost. Thus, the intra-domain definition of nucleation enabled us to elucidate how each accessory subunit contributes to PRC2 distribution genome-wide.

We were able to uncover nucleation in drug washout studies because mESCs can both initiate and maintain most of their polycomb domains. A recent study has shown that differentiated cell types cannot regain many polycomb domains upon prolonged treatment with PRC2 inhibitor followed by drug washout (Holoch et al., 2021). Thus, using drug-washout kinetics is not a widely applicable method to identify nucleation sites. Here we show that CUT&Flow can be used to identify PRC2 nucleation in unperturbed cells. CUT&Flow is easily transferrable across cell types and can be used to ask if nucleation sites are

preserved in polycomb domains that are retained upon differentiation of mESC. CUT&Flow also shows that the kinetics of *de novo* nucleation can explain the restoration of polycomb domains post-dilution due to replication.

Surprisingly, we find a subset of domains whose nucleation appears to happen pre-replication. Our findings mirror almost exactly what was observed for BX-C PRE in *Drosophila* cells, where nucleation occurred in the early S-phase, before replication of the PRE in the late S-phase (Lanzuolo et al., 2011). Our observation that the H3K27me3 kinetics at the nucleation site correlate with the timing of nucleation relative to the cell cycle points to a simple model for the generation and maintenance of polycomb domains genome-wide: PRC2, along with its accessory subunits, has different levels of binding affinity and on-rates across the genome, which when balanced by histone turnover due to various processes leads to the observed steady-state distribution of H3K27me3. Such a model places more reliance on starting states during differentiation and ongoing kinetic processes for setting polycomb domains in differentiated cells than pure epigenetic memory. This is exemplified by recent observations that a high burst of transcription can erase polycomb memory in differentiated cells (Holoch et al., 2021). Our observations in this study thus support the model that polycomb memory is a product of cellular history more than just the inheritance via histone modification, even though the histone modification itself is indispensable for the maintenance of the memory.

Author Contributions

S.R. conceptualized this study. G.M.B.V. developed CUT&Flow method and performed the experiments. S.R. performed the data analysis. S.R. wrote the manuscript. G.M.B.V. reviewed and edited the manuscript.

Acknowledgements

We thank Dr. Jay Hesselberth for suggesting the idea that led to CUT&Flow. We thank Dr. Daphne Avgousti, Dr. David Bentley, Dr. Jay Hesselberth, Dr. Aaron Johnson, Dr. Olivia Rissland, Dr. Sujatha Jagannathan, and the members of the Ramachandran lab for critical reading of the manuscript and their insightful comments. We thank Christine Childs and the CU Cancer Center Flow Cytometry Shared Resource for the valuable assistance with the flow sorting procedure. This work was supported by the RNA Bioscience Initiative, University of Colorado School of Medicine, and NIH grant R35GM133434 (S. Ramachandran).

Materials

All materials used in the study are listed in **Table S1**.

Cell lines and culture conditions. Mouse E14 embryonic STEM cells (mESCs) were maintained in complete medium consisting of KnockOut DMEM medium (Gibco) supplemented with 20% ES-qualified fetal bovine serum (FBS, Gibco), 2 mM Glutamax (Gibco), 10 U/ml penicillin-streptomycin (Gibco), 1X non-essential amino acids (Gibco), 1X 2-mercaptoethanol (Gibco) and 1000 U/ml Leukemia Inhibitory Factor (LIF, Millipore-Sigma). mESCs were seeded in 0.1% gelatin-coated culture flasks to a final concentration of 0.4 x 10^6 cells/ml in complete medium and kept in a 37°C humidified incubator with 5% CO_2.

Methods

CUT&Flow. CUT&Flow was adapted from CUT&Tag (Kaya-Okur et al., 2019). About 2,000,000 – 3,000,000 mESCs were collected per sample, in low-retention tubes. Samples were processed by centrifugation at 600 x g for 3 min at RT, in a swinging bucket centrifuge. Cells were lysed and nuclei extracted in nuclear extraction buffer (20 mM HEPES pH 7.9, 10 mM KCl, 0.1% Triton X-100, 20% glycerol, 0.5 mM spermidine and 1x Roche, cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail CAT# 04693159001) for 10 min on ice, and then subdivided into multiple 500,000 nuclei reactions. Each reaction was incubated with primary antibody (H3K27me3 or IgG, 1:100) in antibody 150 buffer (2 mM EDTA in Digitonin 150 buffer) for 1h at RT, followed by incubation with secondary antibody (Anti-Rabbit, 1:100) in Digitonin 150 buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1x protease inhibitor cocktail and 0.01% digitonin) for 30 min at RT. Nuclei were next washed once in Digitonin 150 buffer and incubated with pAG-Tn5 (Epicypher Inc., 1:20) in Digitonin 300 buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM spermidine, 1x protease inhibitor cocktail and 0.01% digitonin) for 1h at RT. After one wash in Digitonin 300 buffer, pAG-Tn5 was activated and tagmentation performed by incubation in tagmentation buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM spermidine, 1x protease inhibitor cocktail and 10 mM $MgCl_2$) for for 1h at 37°C. Tagmented nuclei were combined back into one tube per sample and incubated in TAPS buffer (10 mM TAPS pH 8.5 and 0.2 mM EDTA) for 5 min at RT to stop the tagmentation reaction. Samples were then resuspended in Krishan solution (Krishan, 1975) (3.8 mM sodium citrate; 46 µg/ml propidium iodide (PI); 0.01% NP-40; 10 µg/ml RNase A) and kept at 4°C, protected from light, for up to three days, until sorting.

PI-stained nuclei were sorted based on DNA content. Five fractions were defined: 0-G1, G2-M and three equal subdivisions of the S phase, classified as Early S, Mid S and Late S. A range of 950 to 7000 cells were collected between the gates. Right after sorting, samples were incubated in SDS release buffer (10 mM TAPS pH 8.5 and 0.1% SDS) for nuclei lysis and release of tagmented chromatin for 1h at 58°C. SDS was neutralized by addition of SDS quench buffer (0.67% Triton-X 100 in nuclease-free H2O) and PCR performed with NEBNext High-Fidelity 2x PCR Master Mix (NEB, M0541L) and barcoded primers

for library construction, with the following protocol: 5 min at 58°C, 5 min at 72°C, 45 sec at 98°C, 18 cycles of 15 sec at 98°C and 10 sec at 60°C, followed by 1 min at 72°C for final extension.

Analysis

All publicly available datasets used in this study are listed in **Table S2**.

Defining Polycomb domains in mESCs. For Figure 1, published datasets from (Banaszynski et al., 2013) were aligned to mm10 version of the *Mus musculus* genome. Coverage at 25 bp windows genome-wide was calculated as number of reads that mapped at that window, normalized by the factor N:

N = 2800000000/(Total number of mapped reads)

2800000000 was a number chosen arbitrarily in the range of total mapped base pairs in M. musculus genome. The normalized read density in 25 bp bins were then smoothed with a running average spanning +/- 1000 bp around each bin. Domains were called by linking adjacent windows that satisfied two cutoffs: 4-fold enrichment over input and 2-fold enrichment over the genome-wide median. To account for short disruptions due to mappability issues, jumps of up to 750 bp were allowed while linking windows.

Determining H3K27me3 recovery kinetics from drug washout datasets. Control H3K27me3 ChIP-seq dataset from (Hojfeldt et al., 2018) was used for domain calling as described above. Then domains that overlapped between (Hojfeldt et al., 2018) and (Banaszynski et al., 2013) datasets were used for all further analyses. The 7-day inhibition dataset, and datasets following washout for the following time points were used: 4, 8, 16, 24, 48, 96 hours (Hojfeldt et al., 2018). Kinetics was analyzed at each 25 bp window of a domain. The H3K27me3 enrichment was first normalized at each window by the enrichment at 96 hour time point for all time points, so that the 96 hour time point would be 1 across all windows. After normalizing to final time-point, the time series matrix for each domain (an example shown in **Figure 2A**) was subjected to k-means clustering with k=7 in R (RCoreTeam, 2019). The average time profile of H3K27me3 enrichment was calculated for each kinetic cluster and the time to reach 40% of recovery relative to 96-hour time point was interpolated for each kinetic cluster. This time, t_{40} was used to order the kinetic clusters from slowest to fastest. The sites corresponding to fastest (cluster 7) were used as nucleation sites for that domain for all further analyses.

CUT&Flow analysis. CUT&Flow datasets were aligned to the mm10 version of *Mus musculus* genome using bowtie2, with the following parameters:

--local --very-sensitive-local --no-unal --no-mixed --no-discordant -I 10 -X 2000

Samtools (Li et al., 2009) and bedtools (Quinlan and Hall, 2010) were used for processing aligned reads from sam to bed files. Duplicate reads were discarded for further analysis if the reads had same start and end coordinates. Coverage at 100 bp windows genome-wide was calculated as number of reads that mapped at that window, normalized by the factor N:

N = 2800000000/(Total number of mapped reads)

2800000000 was a number chosen arbitrarily in the range of total mapped base pairs in M. musculus genome. For replication timing, the final data matrix for mESC from (Zhao et al., 2020) was used. The data matrix contains genomic intervals and for each interval, the percentage of replication in 16 S-phase windows. The S-phase window with highest percentage of replication was used for the plots shown in Figure 5E and Figure 7C, E, F, and G. For Figure 5D, the average normalized read count from the whole domain for all gates were calculated. The gate with lowest normalized read count that was lower by at least 0.5 than the next lowest gate was assigned to the domain. For the heatmap, the values for all gates of a domain were normalized by the highest value and log2 transformed.

For Figure 7A, the average normalized read count at each nucleation site was first divided by the average normalized read count at its corresponding spreading sites to get the nucleation site enrichment ratio for each gate. The gate with highest the nucleation site enrichment ratio that was higher by at least 0.5 than the ratio at the lowest gate and by 0.2 than the ratio at the next highest gate was assigned to the nucleation site. If a gate satisfying the above criteria was not found, it went to the unassigned category. For the heatmap, row-wise Z-score of the nucleation site enrichment ratio was plotted for each nucleation site, where the nucleation sites were ordered by the assigned gates. For Figures 7E-G, we assigned replication timing for domains as follows: Early S if %replication peaked in S-phase window <6, Mid S if replication peaked in S-phase window between 6 and 10, and late S if replication peaked in S-phase window > 10.

References

Alabert, C., Barth, T.K., Reveron-Gomez, N., Sidoli, S., Schmidt, A., Jensen, O.N., Imhof, A., and Groth, A. (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev *29*, 585-590.

Banaszynski, L.A., Wen, D., Dewell, S., Whitcomb, S.J., Lin, M., Diaz, N., Elsasser, S.J., Chapgier, A., Goldberg, A.D., Canaani, E.*, et al.* (2013). Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell *155*, 107-120.

Blackledge, N.P., Fursova, N.A., Kelley, J.R., Huseyin, M.K., Feldmann, A., and Klose, R.J. (2020). PRC1 Catalytic Activity Is Central to Polycomb System Function. Mol Cell *77*, 857-874 e859.

Bowman, S.K., Deaton, A.M., Domingues, H., Wang, P.I., Sadreyev, R.I., Kingston, R.E., and Bender, W. (2014). H3K27 modifications define segmental regulatory domains in the Drosophila bithorax complex. Elife *3*, e02833.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science *328*, 1161-1164.

Escobar, T.M., Oksuz, O., Saldana-Meyer, R., Descostes, N., Bonasio, R., and Reinberg, D. (2019). Active and Repressed Chromatin Domains Exhibit Distinct Nucleosome Segregation during DNA Replication. Cell *179*, 953-963 e911.

Francis, N.J., and Kingston, R.E. (2001). Mechanisms of transcriptional memory. Nat Rev Mol Cell Biol *2*, 409-421.

Fursova, N.A., Blackledge, N.P., Nakayama, M., Ito, S., Koseki, Y., Farcas, A.M., King, H.W., Koseki, H., and Klose, R.J. (2019). Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. Mol Cell *74*, 1020-1036 e1028.

Gaydos, L.J., Wang, W., and Strome, S. (2014). Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. Science *345*, 1515-1518.

Healy, E., Mucha, M., Glancy, E., Fitzpatrick, D.J., Conway, E., Neikes, H.K., Monger, C., Van Mierlo, G., Baltissen, M.P., Koseki, Y.*, et al.* (2019). PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. Mol Cell *76*, 437-452 e436.

Hojfeldt, J.W., Hedehus, L., Laugesen, A., Tatar, T., Wiehle, L., and Helin, K. (2019). Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. Mol Cell *76*, 423-436 e423.

Hojfeldt, J.W., Laugesen, A., Willumsen, B.M., Damhofer, H., Hedehus, L., Tvardovskiy, A., Mohammad, F., Jensen, O.N., and Helin, K. (2018). Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. Nat Struct Mol Biol *25*, 225-232.

Holoch, D., Wassef, M., Lovkvist, C., Zielinski, D., Aflaki, S., Lombard, B., Hery, T., Loew, D., Howard, M., and Margueron, R. (2021). A cis-acting mechanism mediates transcriptional memory at Polycomb target genes in mammals. Nat Genet *53*, 1686-1697.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun *10*, 1930.

Krishan, A. (1975). Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. J Cell Biol *66*, 188-193.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. Mol Cell *65*, 432-446 e435.

Kuroda, M.I., Kang, H., De, S., and Kassis, J.A. (2020). Dynamic Competition of Polycomb and Trithorax in Transcriptional Programming. Annu Rev Biochem *89*, 235-253.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev *16*, 2893-2905.

Lanzuolo, C., Lo Sardo, F., Diamantini, A., and Orlando, V. (2011). PcG complexes set the stage for epigenetic inheritance of gene silencing in early S phase before replication. PLoS Genet *7*, e1002370.

Laugesen, A., Hojfeldt, J.W., and Helin, K. (2019). Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. Mol Cell *74*, 8-18.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.
Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLOS Genetics *6*, e1001244.
Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197-208.
Oksuz, O., Narendra, V., Lee, C.H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.P., Garcia, B.A.*, et al.* (2018). Capturing the Onset of PRC2-Mediated Repressive Domain Formation. Mol Cell *70*, 1149-1162 e1145.
Perino, M., van Mierlo, G., Loh, C., Wardle, S.M.T., Zijlmans, D.W., Marks, H., and Veenstra, G.J.C. (2020). Two Functional Axes of Feedback-Enforced PRC2 Recruitment in Mouse Embryonic Stem Cells. Stem Cell Reports.
Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.
Ramachandran, S., and Henikoff, S. (2015). Replicating Nucleosomes. Sci Adv *1*.
RCoreTeam (2019). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
Reveron-Gomez, N., Gonzalez-Aguilera, C., Stewart-Morgan, K.R., Petryk, N., Flury, V., Graziano, S., Johansen, J.V., Jakobsen, J.S., Alabert, C., and Groth, A. (2018). Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication. Mol Cell *72*, 239-249 e235.
Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. Molecular Cell *55*, 347-360.
Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu Rev Genet *38*, 413-443.
Schorderet, P., Lonfat, N., Darbellay, F., Tschopp, P., Gitto, S., Soshnikova, N., and Duboule, D. (2013). A genetic approach to the recruitment of PRC2 at the HoxD locus. PLoS Genet *9*, e1003951.
Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., and Kingston, R.E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell *98*, 37-46.
Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. Dev Biol *158*, 131-144.
Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H.D., and Cordes, S.P. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell *138*, 885-897.
Steffen, P.A., and Ringrose, L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat Rev Mol Cell Biol *15*, 340-356.
Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873-878.
Wang, X., Paucek, R.D., Gooding, A.R., Brown, Z.Z., Ge, E.J., Muir, T.W., and Cech, T.R. (2017). Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. Nat Struct Mol Biol *24*, 1028-1038.
Woo, C.J., Kharchenko, P.V., Daheron, L., Park, P.J., and Kingston, R.E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. Cell *140*, 99-110.
Zhao, P.A., Sasaki, T., and Gilbert, D.M. (2020). High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. Genome Biology *21*, 76.

Figure 1

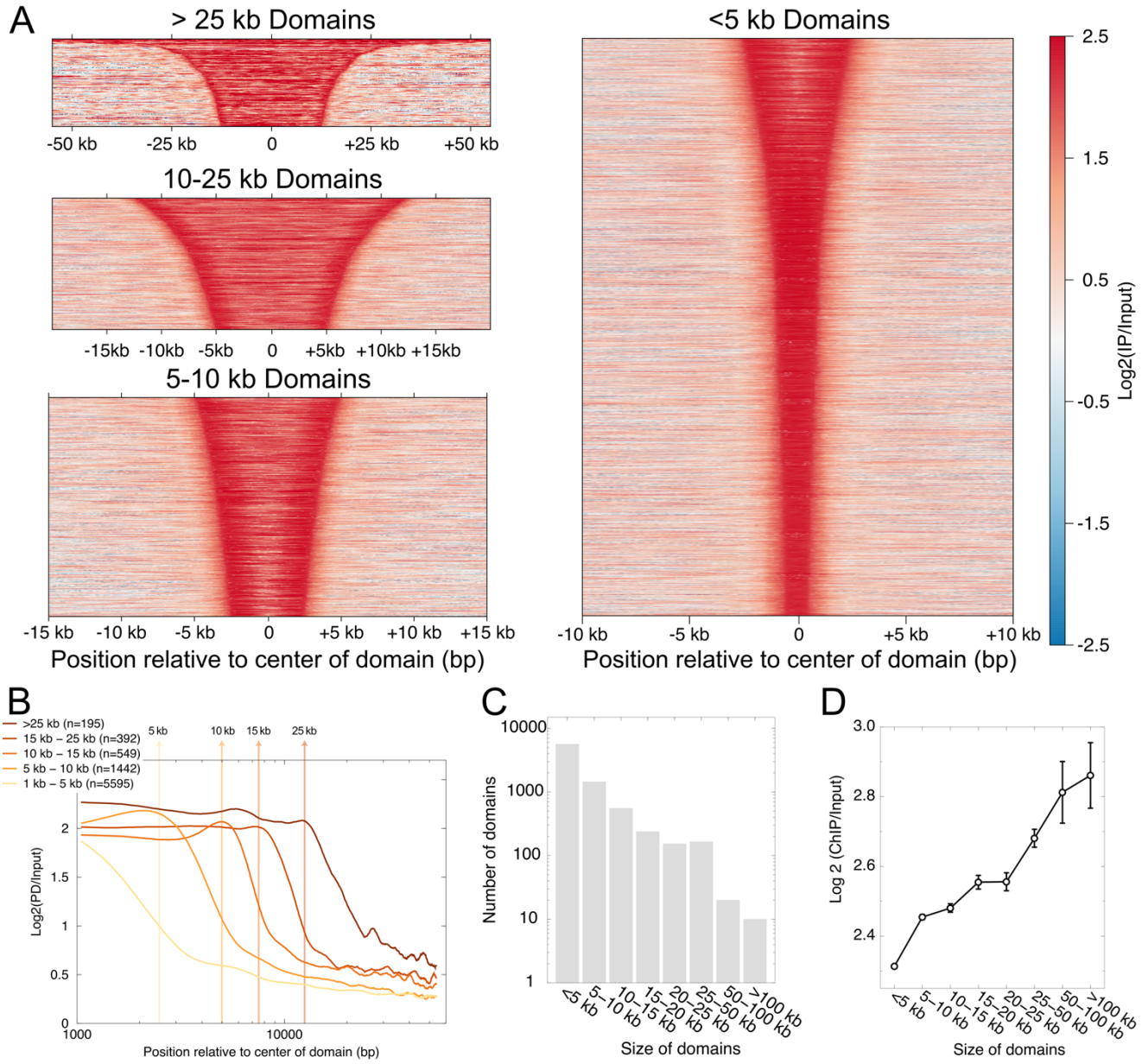

Figure 1. Genome-wide identification of H3K27me3 domains. A) Heatmaps of H3K27me3 ChIP enrichment over input plotted for domains relative to their center, at various scales. **B**) Average profile of H3K27me3 ChIP enrichment over input in the log scale plotted for domains relative to their center, to demonstrate domain edges. The position relative to domain center is plotted on a log scale. **C**) Distribution of the number of domains as a function of their size. **D**) Average (± S.E.M) H3K27me3 ChIP enrichment over input in the log scale plotted for domains as a function of their size.

Figure 2

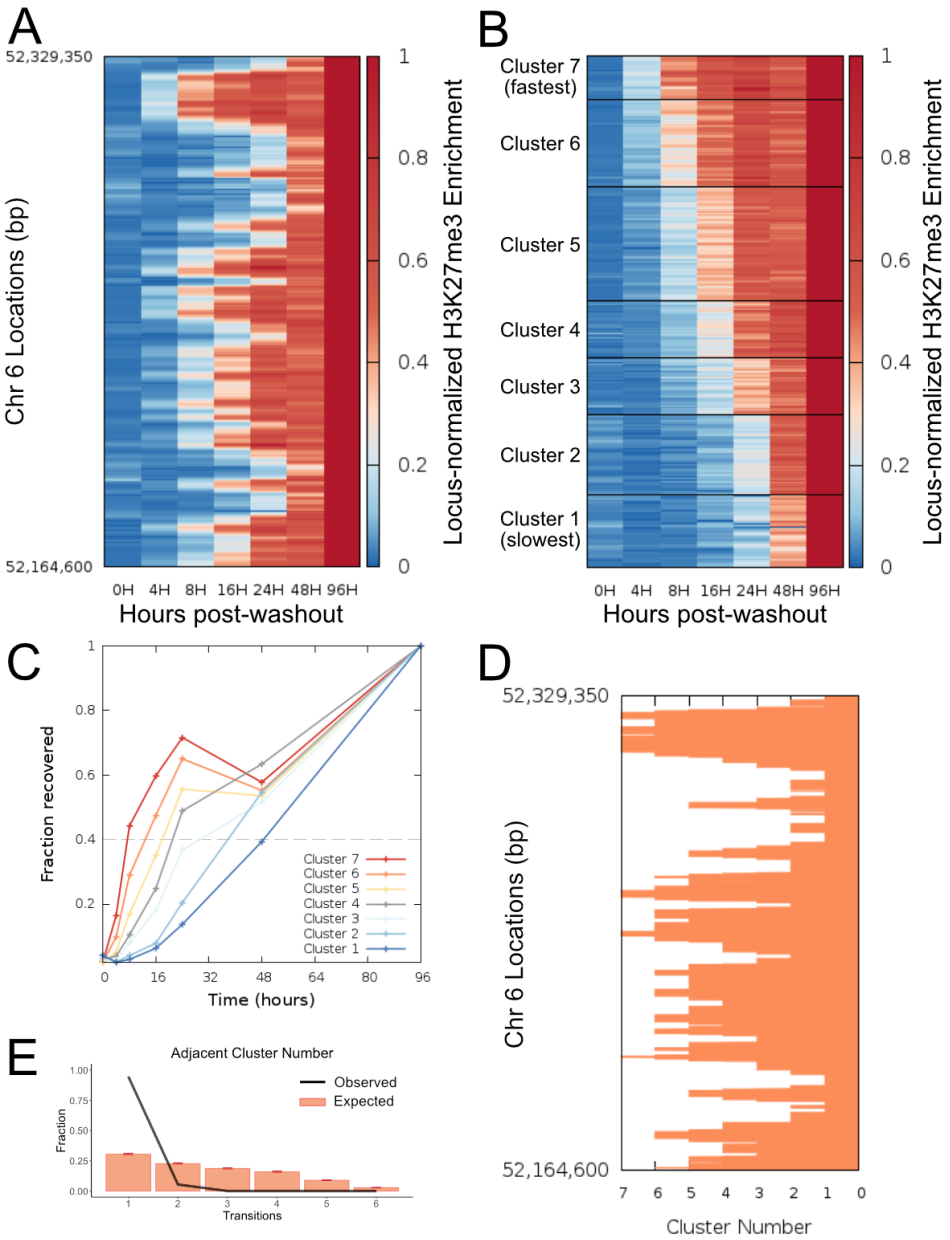

Figure 2. Intradomain kinetics of H3K27me3 recovery at the Hoxa domain. A) Heatmap of H3K27me3 ChIP enrichment at time points post-washout of Ezh2 inhibitor (x axis) normalized to the ChIP enrichment at 96 hours post-washout. Each horizontal line of the heatmap represents 25 bp step of the genomic location on the Hoxa domain. **B**) Heatmap of H3K27me3 ChIP enrichment at time points post-washout of Ezh2 inhibitor (x axis) after performing k-means clustering with 7 clusters. Cluster 7 shows the fastest recovery of H3K27me3 post-washout, and cluster 1 shows the slowest recovery. **C**) Average H3K27me3 ChIP enrichment of each of the 7 clusters obtained by k-means clustering as a function of time post-washout. **D**) The k-means cluster IDs (7 is the fastest recovering cluster post-washout, 1 is the slowest) are plotted as a function of genomic location for the Hoxa domain. **E**) The observed transitions in cluster IDs that occur in adjacent genomic locations (in 25 bp steps) in the Hoxa domain is plotted as a black line. The expected transitions due to random recovery (instead of nucleation and spreading) is modeled by shuffling the cluster ID and the distribution is plotted as orange bars.

Figure 3

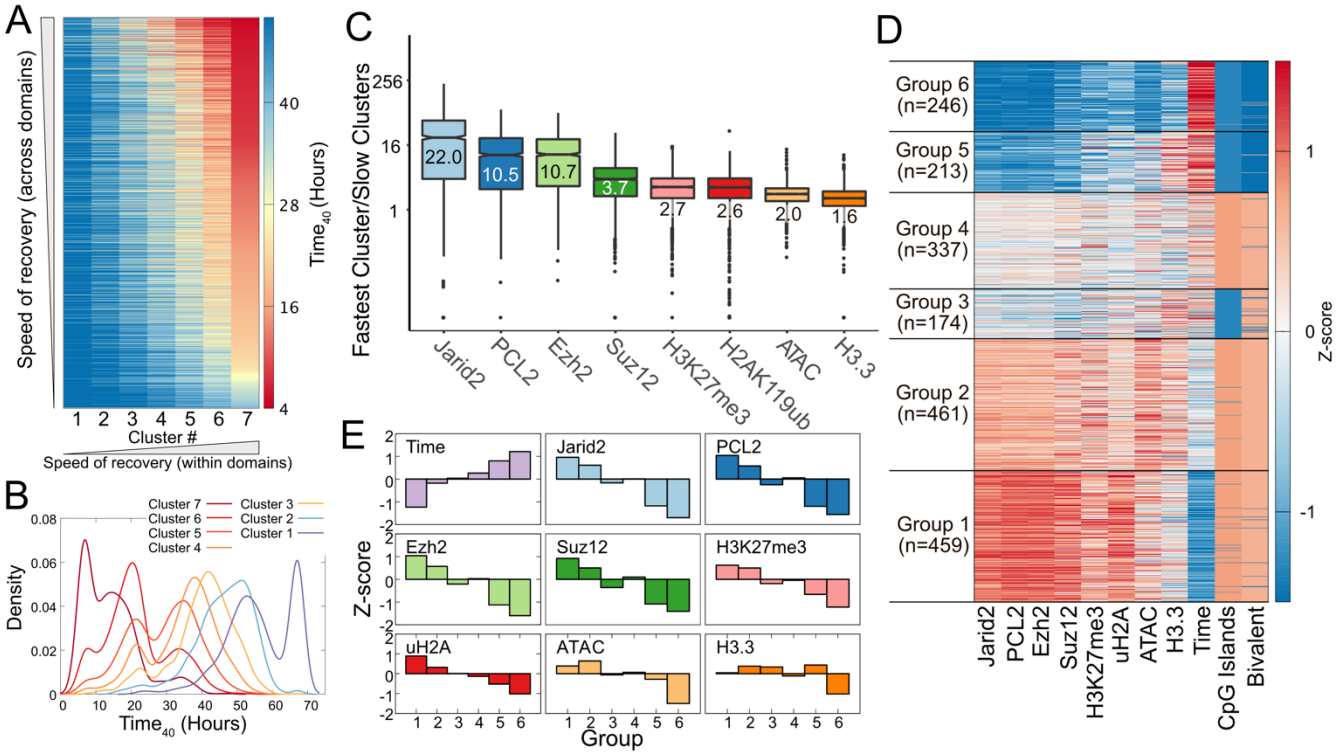

Figure 3. Kinetics of Polycomb domains genome-wide and chromatin properties at nucleation sites. A) Heatmap of t_{40} of each cluster in 1362 domains genome-wide. The clusters are obtained by k-means clustering performed on H3K27me3 ChIP values over time-points of recovery as shown in Figure 2B. For each domain, the clusters are ordered from slowest to fastest. **B**) Histogram of t_{40} for each cluster across domains shown as a heatmap. **C**) Ratio of enrichment of different components/chromatin features at nucleation sites to the enrichment at spreading sites for all domains shown as boxplots. **D**) Groups obtained by k-means clustering of domains based on the z-score of features shown as a heatmap. **E**) Average of each property of each group shown as a bar graph.

Figure 4

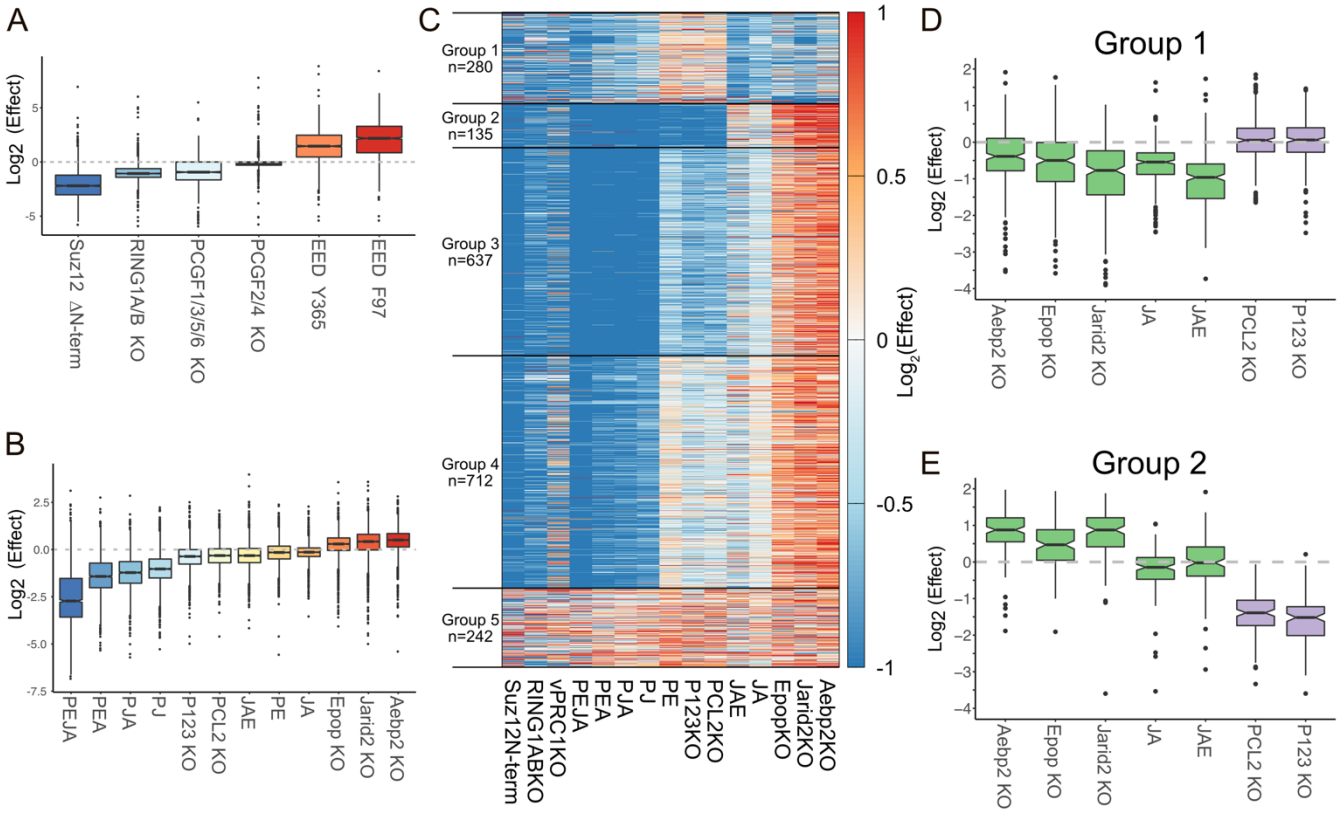

Figure 4. Domain kinetics differentiates role of molecular components in PRC2 nucleation, spreading, and catalytic activity. A) Nucleation strength, the change in enrichment of H3K27me3 at nucleation sites relative to spreading sites upon perturbations indicated in the X-axis, shown as a boxplot across domains. **B**) same as (A) for different combinations of PRC2 accessory components: P – KO of PCL1, PCL2, PCL3, E – KO of Epop, J – KO of Jarid2, and A – KO of Aebp2. **C**) 5 groups defined by k-means clustering of nucleation strength of perturbations shown in (A) and (B), shown as a heatmap. Each row represents one domain **D**) Distribution of nucleation strengths for selected perturbations in Group 1 shown as a boxplot. **E**) Same as (D) for Group 2.

Figure 5

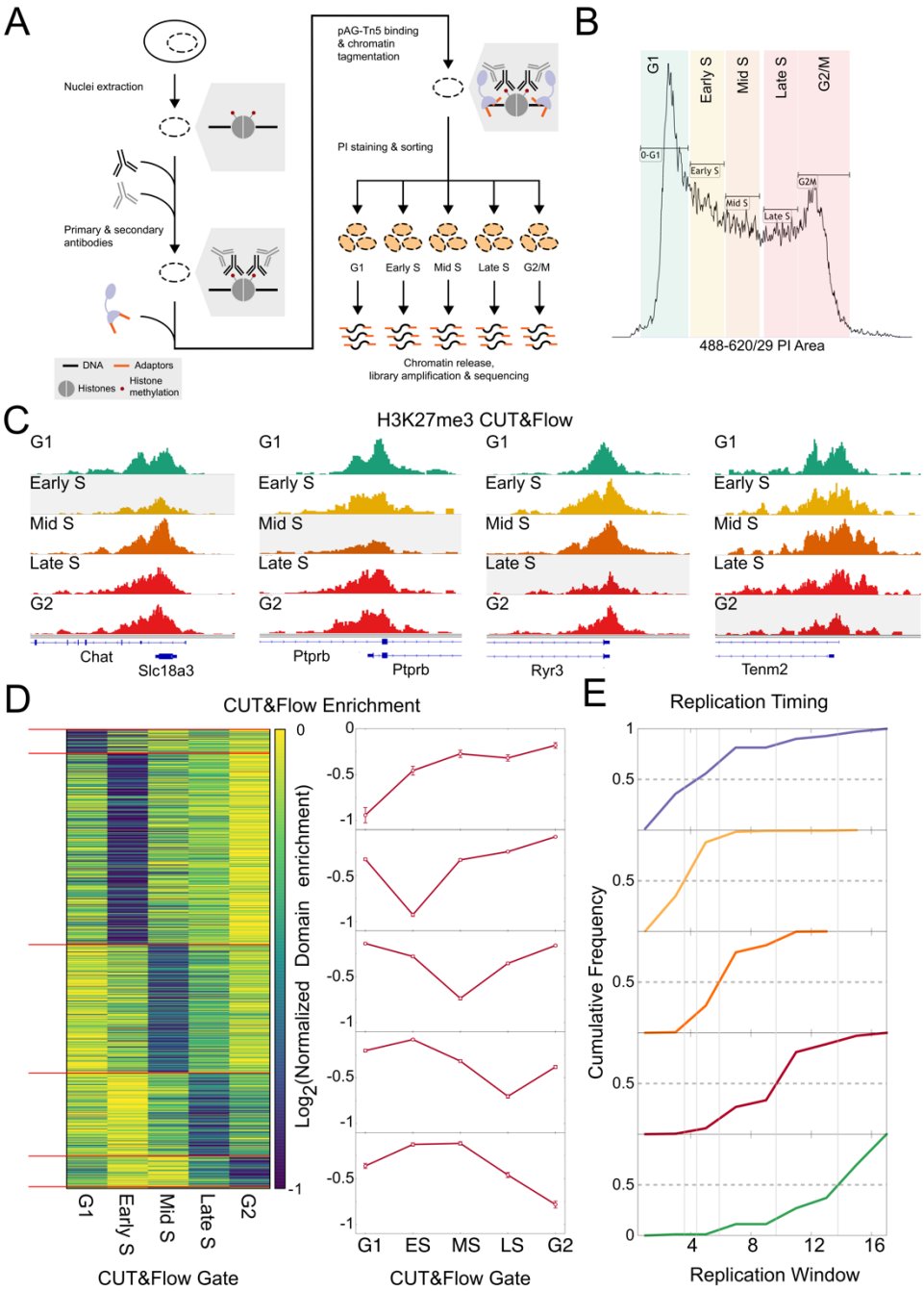

Figure 5. CUT&Flow captures cell cycle stage-specific chromatin profiles. A) Schematic of the CUT&Flow protocol. **B**) Distribution of Propidium Iodide fluorescence of nuclei. The 5 gates to segregate nuclei based on stage of cell cycle are shown. **C**) Examples of domains where the normalized enrichment of H3K27me3 drops significantly in only one of the CUT&Flow gates. The gate with significant drop is shaded gray. The genomic regions displayed are chr14:32,447,403-32,471,747 (first from left), chr10:116,295,946-116,306,843 (second from left), chr2:113,204,914-113,225,200 (second from right), and chr11:37,224,873-37,244,919 (first from right). **D**) Heatmap (left) showing classification of domains based on the CUT&Flow gate which shows the lowest H3K27me3 enrichment. The enrichments for each domain are normalized to the gate with highest H3K27me3. Mean and S.E.M. for each classification from the heatmap are shown as a line plot (right). **E**) Cumulative histogram of the S-phase window (1-16) in which each group of domains sites (defined in (**D**)) replicate most often. Domains that have reduced enrichment in G1 and Early S have earliest replication timing, followed by Mid S, Late S, and then G2.

Figure 6

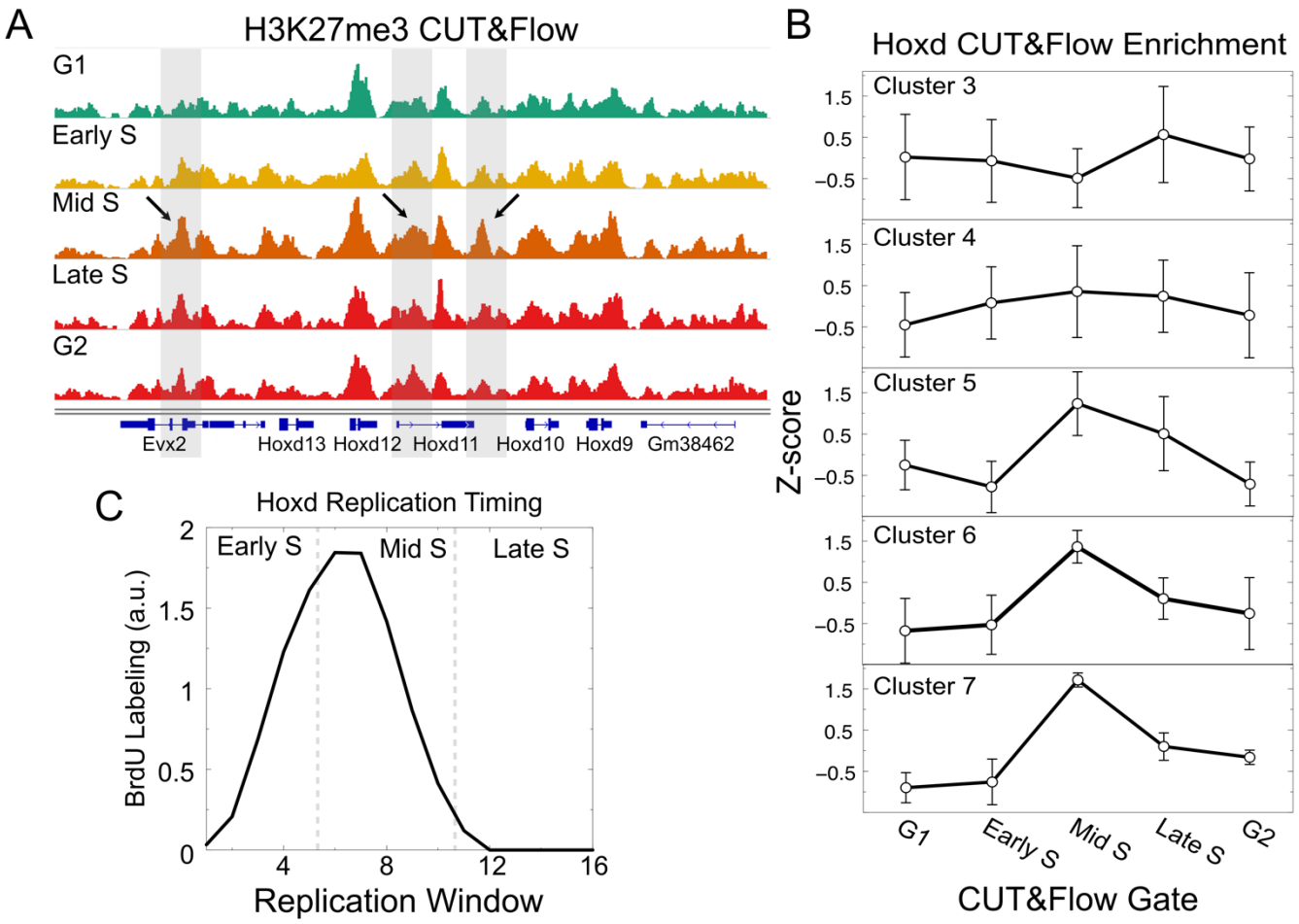

Figure 6. CUT&Flow captures cell cycle stage-specific nucleation site dynamics for Hoxd locus. A) H3K27me3 profiles at the Hoxd locus from each CUT&Flow gate. The regions shaded in gray are nucleation sites. The arrows point to peaks with increased enrichment in the mid S. **B**) Quantification of enrichment of regions in each kinetic cluster (defined in Figures 2 and 3) normalized to enrichment of Clusters 1 and 2 shows a significant increase for Clusters 5, 6, and 7 at Mid S. **E**) Repli-seq enrichment of the Hoxd locus from published data indicates that the locus replicates in Mid S.

Figure 7.

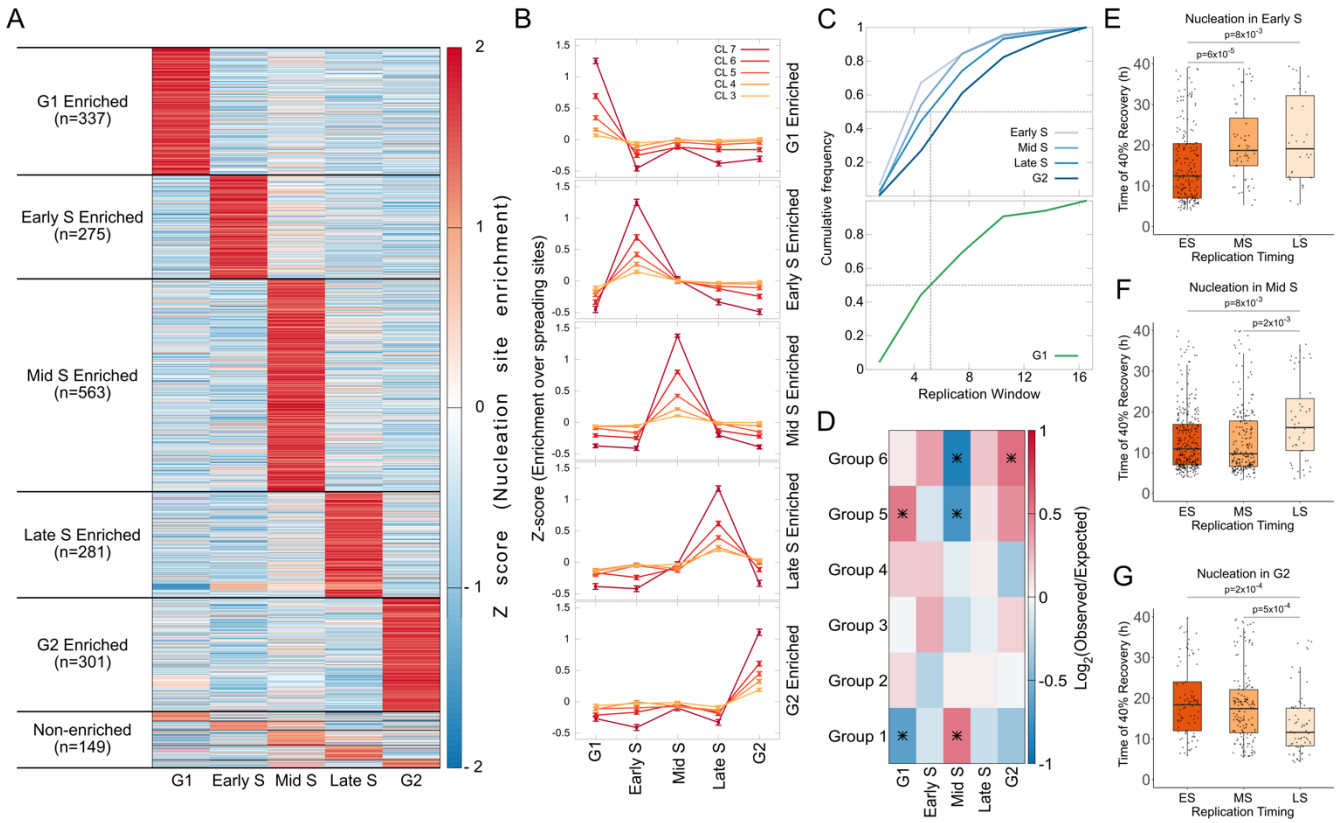

Figure 7. CUT&Flow shows H3K27me3 nucleation post-replication. A) Z-score of nucleation site enrichment over spreading sites across the cell cycle plotted as a heatmap. Nucleation sites are arranged based on the stage of the cell cycle that they spike in. **B**) For domains that feature spike in nucleation at each stage of cell cycle, the enrichment of H3K27me3 at clusters 3-7 over the enrichment at spreading sites (clusters 1,2) plotted as Z-scores. **C**) Cumulative distribution of the S-phase window (1-16) in which each group of nucleation sites (defined in (**A**)) replicate most often. Nucleation sites that spike in Early S have earliest replication timing, followed by Mid S, Late S, and G2 (top). G1 distribution is at the middle of the spread (bottom). **D**) Overlap of domains based on cell cycle stage-specific nucleation and chromatin features from Figure 4. Overlaps that are significant by hypergeometric test ($p<0.05$, $fdr<0.1$) are denoted by asterisk. **E**) Distribution of the recovery kinetics of nucleation sites (measured by t_{40}) plotted for domains separated by the window of replication, for domains that feature H3K27me3 nucleation spike in Early S. **F**) Same as (**E**) for domains that feature H3K27me3 nucleation spike in Mid S. **G**) Same as (**E**) for domains that feature H3K27me3 nucleation spike in G2.