bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nucleosome-directed replication origin licensing independent of a consensus DNA sequence

Sai Li[1], Michael R. Wasserman[1], Olga Yurieva[2], Lu Bai[3], Michael E. O'Donnell[2]*, Shixin Liu[1]*

[1]Laboratory of Nanoscale Biophysics and Biochemistry, The Rockefeller University, New York, NY, USA
[2]Howard Hughes Medical Institute, Laboratory of DNA Replication, The Rockefeller University, New York, NY, USA
[3]Center for Eukaryotic Gene Regulation, Department of Biochemistry & Molecular Biology, Department of Physics, The Pennsylvania State University, University Park, PA, USA
*Correspondence: odonnel@rockefeller.edu; shixinliu@rockefeller.edu

ABSTRACT

The numerous enzymes and cofactors involved in eukaryotic DNA replication are conserved from yeast to human, and the budding yeast *Saccharomyces cerevisiae* (S.c.) has been a useful model organism for these studies. However, there is a gap in our knowledge of why replication origins in higher eukaryotes do not use a consensus DNA sequence as found in S.c.. Using in vitro reconstitution and single-molecule visualization, we show here that S.c. origin recognition complex (ORC) stably binds nucleosomes and that ORC-nucleosome complexes have the intrinsic ability to load the replicative helicase MCM double hexamers onto adjacent nucleosome-free DNA regardless of sequence. Furthermore, we find that *Xenopus laevis* nucleosomes can substitute for yeast ones in engaging with ORC. Combined with new analysis of genome-wide ORC binding, our results lead us to propose that the yeast origin recognition machinery contains the cryptic capacity to bind nucleosomes near a nucleosome-free region and license origins, and that this nucleosome-directed origin licensing paradigm generalizes to all eukaryotes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Complete and accurate genome duplication prior to cell division is a critical process for essentially all living organisms. The basic mechanism for the initiation of DNA replication is shared across domains of life [1,2]. An "initiator" first binds to genomic sites referred to as origins of replication, then recruits the replicative helicase responsible for unwinding parental duplex DNA, creating two complementary templates for DNA polymerases to form two identical copies of daughter chromosomes. In eukaryotes, multiple origins across the genome are licensed for firing once, and only once, per cell cycle and their firing follows a temporally controlled program [3-5]. The eukaryotic initiator is known as the origin recognition complex (ORC), which consists of six conserved subunits, Orc1-6 [6,7]. ORC works in concert with Cdc6 and Cdt1 to coordinate the loading of two Mcm2-7 helicase complexes (MCM) onto the origin DNA [8,9], forming the MCM double hexamer (DH), or "pre-replication complex" (pre-RC). Only a fraction of MCMs loaded on chromatin become activated during the subsequent S phase to produce bidirectional replication forks. The excess of chromatin-associated MCM over those used for replication is referred to as the "MCM paradox" [10].

In the budding yeast *Saccharomyces cerevisiae* (S.c.), origins are located at a set of "replicator" positions known as Autonomously Replicating Sequences (ARS), each containing a 17-bp AT-rich ARS consensus sequence (ACS) element and other less conserved "B elements" [1]. The ARS sequence is AT rich which is not conducive to nucleosome assembly, forming a "nucleosome free region" (NFR). An NFR at an ARS is important for origin function [11], and when nucleosomes are permitted to encroach upon the ARS they are demonstrated to be inhibitory [12-14]. While S.c. has been an instrumental model organism to study eukaryotic replication initiation, the use of a consensus origin DNA sequence is only found in some budding yeasts [15], and therefore a consensus origin sequence is the exception rather than the norm in eukaryotes [15,16]. For example, defined origin elements are not present in the yeast, *Schizosaccharomyces pombe* [17]. The specificity of S.c. ORC for ARS is due to a unique insertion helix (IH) in the Orc4 subunit, which specifically interacts with the ACS sequence [18-21]. S.c. strains with Orc4 IH mutations alter genome-wide firing patterns [19,20] and may enable initiation in NFRs that are different from ARS, such as at transcriptional promotors. It is also worth noting that ARS and its internal ACS and B elements are not strictly required in vitro for S.c. ORC binding or MCM loading [9,22]. These findings, combined with the lack of ACS-like elements in higher eukaryotes, indicate that there exist other chromosomal features that enable the licensing and firing of eukaryotic replication origins [15,23].

One distinct challenge for ORC is the need to navigate through chromosomes prevalently packaged into nucleosomes [24], which influence multiple aspects of eukaryotic replication [25-27]. The nucleosome is often viewed as a barrier to ORC binding and origin activity that must be overcome by certain DNA sequences or chromatin-remodeling enzymes [12-14]. ORC is known to directly interact with nucleosomes by multiple connections [28,29]. Given the continuing uncertainties of the various contacts of ORC subunits to nucleosomes, the interplay between ORC and nucleosomes and its role in origin selection and function remain an area of active research.

In the current study we unexpectedly discovered that budding yeast contain a cryptic nucleosome-dependent mechanism for origin licensing that is independent of an ARS consensus, implying its capability to license origins independent of DNA sequence, and suggesting this process may provide a general mechanism for origin selection that applies to all eukaryotes. On hindsight, this is supported by an extensive ChIP-seq study that determined about one-third of S.c. origins lack a recognizable consensus ACS sequence [11]. Based on the facts that (1) ORC binds nucleosomes [28,29], and (2) pre-RCs are predominantly licensed at nucleosome-free regions (NFRs) [11,12], we hypothesize that

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nucleosomes recruit ORC and that this can lead to assembly of the MCM DH onto DNA provided, at a minimum, that there is an adjacent NFR that can accommodate ORC binding and MCM loading (**Figure 1**). We presume that there may be additional requirements for origin activation in vivo at these "non-ARS" sites, such as histone modifications or nucleosome remodelers.

Why would S.c. have developed sequence-specific origins and also utilize a general nucleosome-directed but non-sequence-specific method for origin selection? We presume S.c. evolved ARS sequences, located at intergenic regions, for preferred ORC binding due to its small genome size and therefore urgent need to avoid transcription-replication conflicts, as proposed earlier [20]. To test whether origin licensing in S.c. can occur at random DNA sequence that is adjacent to a nucleosome, we develop herein a single-molecule platform to visualize the dynamic behavior of ORC and MCM on nucleosomal DNA. We find that S.c. ORC stably associates with nucleosomes and loads MCM DHs at adjacent nucleosome-free regions. This finding of ARS sequence-independent yet nucleosome-directed ORC binding and subsequent MCM DH loading reveals that the S.c. system has an inherent sequence-independent origin licensing activity that is akin to that in higher eukaryotes, thereby providing insight and a unifying mechanism for eukaryotic origin selection.

RESULTS

A single-molecule platform to directly visualize eukaryotic origin licensing

To perform single-molecule studies on pre-RC formation, we used the bacteriophage λ genomic DNA containing an engineered S.c. ARS1 sequence [30] placed 14 kb from one end (**Figure 2A & Supplementary Figure 1**). The DNA construct (termed λ_{ARS1}) was biotinylated at both ends. We expressed and purified the S.c. Orc1-6 and Mcm2-7 complexes (referred to as ORC and MCM hereafter) as well as Cdc6 and Cdt1 (**Supplementary Figure 2**). To generate fluorescently labeled ORC for direct visualization, we site-specifically attached a Cy3 fluorophore to the N terminus of the Orc1 subunit via a 12-residue S6 peptide tag. A single λ_{ARS1} DNA molecule was tethered between a pair of streptavidin-coated beads in the microfluidic chamber of a dual-trap optical tweezers instrument combined with multicolor confocal fluorescence microscopy [31,32] (**Figure 2B**). Upon moving the tethered DNA into a channel containing Cy3-ORC (+/- Cdc6) and ATP, we observed ORC binding to DNA in real time. In the presence of Cdc6, ORC displayed short-lived and diffusive binding to non-ARS1 DNA, while remaining stably bound at the ARS1 site (**Figure 2C**). Cdc6 enhances the overall binding of ORC to DNA and is required for its ARS specificity (**Figure 2D & 2E**). These results are consistent with previous single-molecule studies [33-35], thus confirming the normal function of proteins used in this study. The distinctive behavior of ORC at the engineered ARS1 site versus all other DNA sites indicates that the λ_{ARS1} template does not contain a second strong ACS motif, which is corroborated by sequence analysis of the λ genomic DNA (**Supplementary Figure 3**).

ORC mediates MCM DH loading onto both ARS and non-ARS DNA

Next we generated a S.c. strain expressing Mcm2-7 with S6-tagged Mcm3, enabling us to site-specifically label MCM complexes with a fluorophore. We used LD650-labeled MCM, along with unlabeled ORC, Cdc6 and Cdt1, to examine pre-RC formation on λ_{ARS1} DNA (**Figure 3A**). Surprisingly, we observed that the majority (86%) of MCM that stably bound to DNA were at positions distant from ARS1 (**Figure 3B & 3C**). Considering that non-ARS DNA sites vastly outnumber the sole ARS1 site in our template, ORC/MCM may still have a higher affinity to ARSs than to other DNA sequences. We then conducted photobleaching analysis to examine the stoichiometry of MCM bound to DNA, and observed that a significant fraction (36%) of MCMs underwent two-step photobleaching (**Figure 3D & 3E**). Given

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the estimated labeling efficiency for MCM (~60%), this result suggests that the majority of MCMs on DNA observed in our experiments were double hexamers. The presence of ORC is strictly required for MCM binding to DNA (**Figure 3F**). Notably, we have shown that ORC displays diffusive behavior at non-ARS sites (**Figure 2C & 2E**). Therefore, it appears that MCM stabilizes ORC binding to non-ARS DNA sequences.

To obtain further evidence for MCM DH formation, we moved the DNA tether into a separate channel containing a high-salt buffer (0.5 M NaCl) after ORC-mediated MCM binding (**Figure 3G**). We found that a significant portion of the MCM complexes remained associated with DNA upon high-salt wash—either stably residing at the initial position or diffusing along the DNA (**Figure 3H & 3I**), consistent with the behavior of properly loaded MCM DHs [9,34,35]. The rest of the MCMs dissociated from DNA at high salt, likely representing those not topologically encircling the DNA duplex. The fraction of loaded MCMs at the ARS1 site (62%) is much higher than that at non-ARS sites (25%). Therefore, our results confirm that ARS1 represents a preferred position for ORC-mediated MCM loading, but also unambiguously show that this process can still occur on non-ARS sequences. In other words, our data suggest that ARS enhances the likelihood, but is not strictly required, for ORC to license an origin.

ORC preferentially binds nucleosomes over nucleosome-free DNA

Next, we set out to examine the behavior of ORC on chromatinized DNA. We reconstituted both S.c. and *Xenopus laevis* (X.l.) histone octamers for comparative studies (**Supplementary Figure 2 & Supplementary Figure 4**). In both cases, a unique cysteine residue is placed on the histone H2A, enabling site-specific labeling of the nucleosome. We used the histone chaperone Nap1 for in situ nucleosome assembly [26,36]. By titrating the protein concentrations, we found a condition that yields sparsely populated nucleosomes (usually between 1 and 7) within the tethered λ_{ARS1} DNA, such that most if not all of the nucleosomes are flanked by substantial NFRs (**Figure 4A & Supplementary Figure 5**). When incubating the DNA tether harboring Cy3-labeled S.c. nucleosomes with LD650-ORC and Cdc6, we made the striking observation that ORC predominantly colocalized with the nucleosomes, rather than residing within the long stretches of bare DNA between nucleosomes (**Figure 4B**). Importantly, stable association of ORC with the nucleosome is independent of whether the nucleosome is located at the ARS1 position or non-ARS sites (**Figure 4C**). In fact, most nucleosomes were located at non-ARS positions and yet the vast majority (>90%) had a stably bound ORC using only a low concentration (2 nM) of ORC (**Figure 4D**). Sometimes we observed ORC at the nucleosome-free ARS1 DNA position where it remained stably bound in the presence of Cdc6 (e.g. second kymograph in **Supplementary Figure 6A**), again supporting the expected and normal ORC function. In contrast, when ORC binds to non-ARS nucleosome-free regions of the nucleosomal DNA tether, it displays diffusive behavior and the diffusion was confined between adjacent nucleosomes that are occupied by ORC (**Figure 4B & Supplementary Figure 6**).

Comparing **Figure 2E** (ORC behavior on non-nucleosomal DNA) and **Figure 4C** (ORC behavior on nucleosomal DNA), it becomes apparent that the presence of nucleosomes on DNA abrogates the requirement of ARS sequences for stable ORC binding. In other words, ORC has the ability to stably engage with a nucleosome regardless of its adjacent DNA sequence. Interestingly, stable association of ORC with nucleosomes appears to be a conserved phenomenon as we obtained similar results using either S.c. or X.l. histone octamers (**Figures 4C & 4D**). Nucleosome targeting by ORC can conceivably be achieved by either a three-dimensional (3D) search (direct binding from solution) or a one-dimensional (1D) search (sliding along the DNA from a non-nucleosomal site). We observed both modes in our data,

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with 3D search being the more dominant mode, especially for S.c. nucleosomes (**Supplementary Figure 6**).

ORC mediates MCM loading to nucleosomes

Next we asked whether ORC binding to nucleosomes can lead to MCM helicase loading and pre-RC formation. We used fluorescently labeled MCM complexes (with unlabeled ORC, Cdc6, and Cdt1) to examine their recruitment to nucleosomal DNA and its dependence on ORC. We found that MCM frequently colocalized with nucleosomes that were sparsely distributed across the DNA tether (**Figure 5A & 5B**). This result was observed for both S.c. and X.l. nucleosomes. MCM binding to the nucleosome requires the presence of ORC, as omitting ORC completely eliminated MCM-nucleosome colocalization (**Figure 5C**). Considering that the free DNA sites vastly outnumber the nucleosome sites in our assay (a few nucleosomes within 48-kbp DNA), the observed frequency of MCM-nucleosome colocalization indicates that nucleosomes are preferred sites for ORC-mediated MCM binding (**Figure 5D**). Again, ARS sequences are not required for MCM-nucleosome colocalization, as these events were mostly observed at non-ARS sites (**Figure 5E**). To examine whether MCM can form double hexamers at the nucleosome, we performed three-color fluorescence experiments using A488-labeled histone octamers and a mixture of Cy3- and LD650-labeled MCMs (**Figure 5F**). Indeed, we detected colocalization of dual-color MCMs (green and red) with nucleosomes (blue), suggesting MCM DH formation at the nucleosome site.

To test whether MCM DHs are truly loaded (i.e. topologically encircling DNA), we moved the tethered nucleosomal DNA with bound ORC and MCM into a high-salt buffer channel containing 0.5 M NaCl. As explained earlier, recruited MCMs that do not encircle DNA are expected to dissociate at high salt, whereas loaded MCM DHs are expected to stay on DNA and can slide on it if ORC is dislodged by high salt [8,9]. Notably, we found that, unlike ORC removal from free DNA (ARS1 or non-ARS1) by high salt in the absence of nucleosomes, MCM-ORC complexes appeared more resistant to salt when bound to nucleosomes. Thus we use labeled ORC to examine the ability to dissociate ORC from nucleosomes at 0.5 M NaCl. Upon comparing the two types of nucleosomes (S.c. vs. X.l.), we found that about half of S.c. nucleosome-bound ORC was dissociated by 0.5 M NaCl, and about 20% of X.l. nucleosome-bound ORC was removed (**Supplementary Figure 7**). Then we analyzed the behavior of fluorescent MCM (red) bound to S.c. ORC-S.c. fluorescent nucleosomes (blue) upon high-salt wash. The example kymograph in **Figure 6A** shows multiple MCMs on tethered DNA. Upon high-salt wash, the MCM colocalized with a non-ARS positioned nucleosome underwent sliding on DNA, as did another MCM located at the ARS1 DNA site (without a nucleosome). More kymographs are shown in **Supplementary Figure 8**, including an example of dual-color MCM DH sliding at high salt. Overall we observed ~30% of MCMs formed at ORC-nucleosome sites to diffuse on DNA and another ~30% to remain associated with the nucleosome (**Figure 6B**). The immobile MCM fraction can be attributed to the strong engagement of ORC with the nucleosome (**Supplementary Figure 7**), which in turn holds MCM next to the nucleosome. The remainder of the MCMs dissociated into solution upon high-salt wash, which presumably represent MCMs that were not fully loaded onto DNA.

Genome-wide analysis of ORC/MCM localization

The single-molecule results presented in this study suggest that ORC binding and MCM loading may occur over genomic regions that lack an ARS consensus sequence (ACS) element. In support of this observation, it has been previously shown that only 67% of well documented ARS sites in the S.c. genome contain a recognizable ACS sequence [11]; while the remaining 33% were presumed to contain “novel ACS

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequences". We propose, based on the work presented in the current report, that the novel ACS sequences may have been random DNA sequences adjacent to a nucleosome at an NFR.

To test this hypothesis, we re-analyzed published S.c. Orc1 and Mcm2-7 ChIP-seq data [11,37]. Consistent with previous analysis [11], we identified 295 Orc1 ChIP peaks genome-wide, and most of these sites also bind to Mcm2-7 (**Figure 7A**). Most of the Orc1 peaks (225 out of 295) overlap with annotated origins of replication [38] (**Supplementary Table 1**). Examples of ORC/MCM peaks at known origin sites are shown in **Figure 7B**. We extracted DNA sequences +/- 500 bp flanking the center of the ChIP peaks, scanned each 17-bp window with the ACS position weight matrix (PWM) [39], and recorded all the windows with a PWM score of > 9 (previous studies recommended a cutoff score of 11.9). We found that, even with this less stringent criterion, 145 out of the 295 Orc1 peaks contain no qualified window, 105 with one qualified window, and the remaining 45 with multiple qualified windows. The PWM scores are not correlated with the strength of the ChIP peaks (**Figure 7C**). For example, among the six examples shown in **Figure 7B**, ARS609, ARS523, and ARS416 contain an ACS with high-to-medium scores of 13.2, 14.0, and 10.1, respectively, while the rest contain no window with a score above 9. Yet all of these sites show comparable Orc1 and Mcm2-7 ChIP signals. This unexpected data suggests that ORC/MCM binding to established S.c. origins does not require a highly consensus ACS.

To rule out the possibility that the above conclusion depends on the choice of the cutoff score, we calculated the fraction of origin sequences that contain at least one qualified window using different cutoff values (**Figure 7D**). As expected, fewer qualified sequences were identified with a higher cutoff ACS PWM score. Using the conventional cutoff of 11.9, the ACS consensus is only identified in ~20% of Orc1 ChIP peaks. Among the 225 Orc1 peaks that overlap with annotated origins, we observed a similar number (~20%) (**Figure 7E**). This number only mildly increases to ~25% when we analyzed the subset of Orc1 ChIP peaks that also have high Mcm2-7 ChIP signals (**Supplementary Figure 9A&B**). To test the above conclusion further, we also examined the recent Orc1 ChIP-exo data [40]. Although only 81 Orc1 peaks were identified in this dataset, the probability of finding a consensus ACS motif in these peaks is essentially the same (~20% with a cutoff score of 11.9) (**Supplementary Figure 9C**). As a quality control of our method, we applied the same analysis to sequence-specific transcription factors Abf1 and Reb1, and found that their cognate DNA motifs occur at a much higher frequency near their respective ChIP peaks (~75% and ~55% at recommended cutoff of 8.2 and 8.7, respectively) (**Figure 7F**). Overall, these results suggest that a significant fraction of established replication origins in the yeast genome do not contain a highly consensus ACS.

We next investigated the location of Orc1 ChIP peaks relative to the neighboring nucleosomes. Nucleosome occupancy measured by MNase assay is also shown in **Figure 7B** [41]. For the first three origins with a well-defined ACS, the consensus sequence all locates inside NFRs (indicated by the red lines). Indeed, when we collected all 62 Orc1 peaks with a consensus ACS (score > 11.9), aligned at their ACS locations, and plotted the nearby nucleosome occupancy, we observed strong nucleosome depletion over most of the sites (**Figure 7G**). This is consistent with the previous finding that a consensus ACS engineered into a well-positioned nucleosome leads to nucleosome displacement [42]. For some Orc1 binding sites without a strong ACS (score < 9), the Orc1/Mcm2-7 peaks seem to locate near a nucleosome-NFR junction (**Figure 7B**).

Taken together, these findings are consistent with the model that ORC either binds to consensus ACS elements in the genome—which may lead to NFR formation—or binds to nucleosomes near an NFR, indicating a level of flexibility in yeast origin architecture that has also been reported before [43].

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DISCUSSION

Nucleosomes play a major role in directing ORC function in origin licensing

The prevailing model for eukaryotic origin selection takes a DNA-centric view, which argues that ARS—especially the ACS element within ARS—largely dictates where ORC binds in the genome [1]. Nonetheless, in vitro studies demonstrate that these elements are not strictly required for pre-RC formation or replication initiation even for the yeast system where ARSs were initially identified [6,9,22]. For example, the in vitro MCM loading efficiency was similar between wildtype ARS1 and ACS-deleted ARS1, and the specificity of MCM DH loading at ARS1 required addition of competitor DNA [9]. Moreover, the yeast genome contains far more ACS motifs than functional origins [44]. Therefore, the determinants of eukaryotic origin selection must include other chromosomal features, the identity of which are still under study. By imaging ORC's behavior on nucleosomal DNA in real time, our work provides clear evidence that ORC preferentially binds nucleosomes—independent of a nearby ACS motif—over non-nucleosomal DNA, and importantly, that this interaction is functionally relevant to MCM DH loading at nucleosomal sites. These findings support the hypothesis illustrated in **Figure 1** that nucleosomes are the dominant directive of origin function in all cells. This demonstration of ARS-sequence independent, but nucleosome-directed origin licensing, is applicable to all eukaryotes including those where a consensus origin sequence has not been found. In this hypothesis, a nucleosome and an adjacent NFR sufficiently wide to accommodate ORC and an MCM DH are the two main prerequisites for pre-RC formation.

The influence of nucleosomes on origin function has been previously investigated [45], but mainly reported as a secondary mechanism to reinforce the DNA sequence-encoded origin specificity by targeting ORC to ARS instead of other DNA sequences occupied by nucleosomes [26,27]. Here we show that nucleosomes—in and of themselves—represent a primary instructive code for replication origins. Given the highly conserved nature of replication initiation factors, we presume that this principle will generalize to higher eukaryotes, just as the many earlier findings in the S.c. system.

NFR adjacent to a nucleosome enables pre-RC assembly independently of ARS

The second prerequisite for pre-RC formation in our model is the presence of a sufficiently wide NFR that can accommodate the pre-RC complex. It is known that an NFR flanked by regularly positioned nucleosomes is a pronounced feature of S.c. ARS origin sites [11,12]. Our single-molecule data suggest that an MCM DH can be loaded onto DNA lacking an ARS sequence as long as it is adjacent to a nucleosome. Our genomic analyses show that S.c. indeed utilizes many origin sites without a strong ACS element. ARS may be involved in creating some of these NFRs [42], but is not a necessity. Indeed, there exist numerous non-ARS NFRs in yeast chromosomes located at 5' and 3' ends of genes. These NFRs are 80-300 bp long and flanked by well-positioned nucleosomes [46-48]. Importantly, recent studies have shown that "humanizing mutations" of S.c. ORC that abrogate its ability to bind ACS change the genome-wide origin pattern such that initiation occurs at NFRs lacking ACS, such as at promotors [19,20], supporting the notion that some other NFRs that border a nucleosome can in principle be utilized as an origin site.

On the other hand, the requirement for an NFR permissive to helicase loading excludes many genomic regions from becoming origin sites, such as the tightly packaged heterochromatin prevalent in eukaryotic chromosomes. This view explains why replication initiation cannot occur at just any nucleosome. The hierarchical chromatin organization in vivo also influences how the timing of different

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

origins is controlled [17]. There is evidence that origins located within heterochromatin fire later in S phase compared to those in the more open euchromatin [49], perhaps stemming from the time needed for factors such as chromatin remodelers to create wide enough NFRs for ORC binding and MCM DH assembly.

Why have some yeast species evolved DNA sequence-specific origins?

Our model (**Figure 1**) infers that there is no fundamental difference in the origin selection mechanism between yeast and higher eukaryotes, but simply that some yeast species (for example, S.c.) have evolved a dependence on ARS sequences, perhaps to increase fitness. Yet S.c. cells contain some origins that simply lack a recognizable ACS for ORC [11]. This evolutionary pressure may be related to the high gene density in organisms such as *S. cerevisiae* with a small genome size. The slight advantage that the ACS element confers on yeast ORC binding to non-nucleosomal DNA, along with the strong binding of ORC to a nucleosome adjacent to an ARS sequence, may serve to place replication origins within intergenic regions to help avoid replication-transcription conflicts and genome instability [20]. As such, ARSs add an additional layer of "security" to prevent spurious origin firing. Nevertheless, our results here show that the S.c. system still retains the nucleosome-directed origin licensing capability that is likely also used by higher eukaryotes. Metazoans, which have much larger genomes, may have evolved other mechanisms to circumvent or tolerate transcriptional interference, and thus did not need to evolve sequence-specific origins [50].

A potential explanation for the MCM paradox

Our proposed model that any NFR might enable MCM DH formation mediated by ORC-nucleosome interaction in G1 phase suggests a rather "sloppy" process of origin licensing. It follows that many more sites may be licensed than can be used in S phase for origin firing. While we do not expect replication initiation to take place in every NFR in each cell cycle, because of additional unknown requirements (e.g. histone modifications, MCM requirements to mature to CMG helicases), it is still possible that many NFRs that occur within the genome allow MCM DHs to be assembled but not normally used. Thus, our findings may offer a mechanistic basis for the "MCM paradox" that refers to the excess MCMs distributed in the genome compared to actively used origins [10]. One may question whether the extra MCMs can be those binding nonspecifically to chromatin or are actually in the form of pre-RC MCM DHs. The first cryo-EM study of the MCM DH provides evidence for the latter scenario [51]. In that study, sufficient MCM DHs were obtained by DNase treatment of yeast chromosomal DNA without protein overexpression for high-resolution structural analysis of the MCM DH. Hence, one may infer a natural abundance of MCM DH on normal yeast chromosomes. These unused "dormant" origins may become activated in the event of DNA damage or other cellular stress that limits the ability of replication forks to progress [52].

There are further steps in S phase that prompt the maturation of MCM DH to dual CMG replicative helicases [26], some of which may involve elements of ARS sequences that can recruit nucleosome remodeling enzymes. Moreover, epigenetic features of the chromatin, such as posttranslational modifications of histones—which were missing in our current study and thus not required for functional binding of nucleosomes to ORC to facilitate pre-RC formation—may nonetheless contribute to fine-tuning the "nucleosome origin code" by making particular nucleosomes a better or worse binding partner for ORC [53,54].

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In conclusion, our work provides a model for how ORC specifies eukaryotic replication origins in general and an explanation for how an excess of MCM DHs can be loaded onto DNA in G1 phase. Future studies, conducted in diverse eukaryotic systems including human, are needed to fully elucidate the regulatory mechanisms for nucleosome-ORC-MCM interaction and function.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACKNOWLEDGMENTS

We thank X. Zhao (Sloan Kettering Institute) for critical reading of the manuscript, B. Stillman (Cold Spring Harbor Laboratory) for feedback on a preprint version of this manuscript, L. Langston, R. Mayle, G. Schauer, N. Yao, and D. Zhang in the O'Donnell Laboratory for reagents, J. Watters in the Liu Laboratory for data analysis codes, and other members of the Liu and O'Donnell groups for discussions. L.B. is supported by National Institutes of Health (R01GM118682 & R35GM139654). M.E.O. is supported by NIH (R01GM115809) and the Howard Hughes Medical Institute. S.Liu is supported by the Robertson Foundation, the Alfred P. Sloan Foundation, the Pershing Square Sohn Cancer Research Alliance, and an NIH Director's New Innovator Award (DP2HG010510).

AUTHOR CONTRIBUTIONS

S.Liu and M.E.O. oversaw the project. M.R.W. prepared the DNA templates. O.Y. prepared the yeast expression constructs for protein production, and purified the factors. M.R.W., O.Y. and S.Li prepared and labeled the replication proteins. S.Li prepared the nucleosome samples and performed the biochemical assays. S.Li and M.R.W. performed the single-molecule experiments. L.B. performed the genomic data analysis. S.Liu, M.E.O., S.Li and L.B. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MATERIALS AND METHODS

Protein expression and purification

S6-ORC

A recombinant strain of *S. cerevisiae* co-expressing the six subunits of ORC, having a 12-aa S6 tag for fluorescent labeling at the N terminus of Orc1, was constructed as follows. We used either an Orc1 gene with an N-terminal 3× Flag (wt ORC) or further modified the Orc1 subunit gene by insertion of DNA encoding the "S6" peptide (GDSLSWLLRLLN) at the N terminus (S6-ORC) [55]. The six subunits of ORC complex were cloned into integration vectors having the galactose inducible Gal1/10 bidirectional promotor for induction by galactose and were cloned and integrated into the genome of strain OY001 (*ade2-1 ura31 his311,15 trp11 leu23,112 can1100 bar1*Δ, *MAT*a *pep4øKANMX6*), a strain constructed from W303 [56]. The order of integration was: (1) genes encoding 3×Flag-Orc1 or S6-3×Flag-Orc1, and Orc3 (both cloned into pRS404/GAL); (2) genes encoding Orc4 and Orc5 (both cloned into pRS405/GAL); (3) the gene encoding Orc2 (cloned into pRS403/GAL); and (4) the gene encoding Orc6 (cloned into pRS402/GAL). The wt and S6-ORC overexpression strains were constructed by integrating the expression plasmids described above in the yeast genome. Both wt ORC and S6-ORC were purified by the same procedure below.

One liter of S6-ORC cells were grown under selection at 30 °C in SC glucose, then split into twelve 2-L fluted flasks, each containing 1 L of YP-glycerol media and grown to an optical density at 600 nm (OD_{600}) of 0.4 at 30 °C, arrested with α-factor (50 µg/L) for 2 h, and then induced for 3 h upon addition of 20 g/L of galactose. Cells were harvested by centrifugation, resuspended in a minimal volume of 20 mM HEPES pH 7.6, 1.2% polyvinyl pyrrolidone, and protease inhibitors (Sigma-Aldrich #5056489001) and 0.5 mM PMSF, then snap frozen by dripping into liquid nitrogen. Purification of the S6-ORC complex was performed by lysis of 12 L of frozen cells using two SPEX cryogenic grinding mills (6970 EFM). Ground cells were thawed and debris removed by centrifugation (19,000 rpm in an SS34 rotor for 2 h at 4 °C); then the supernatant was applied to a 1-mL anti-Flag M2 affinity column (Sigma) equilibrated in buffer H (50 mM HEPES pH 7.5, 250 mM potassium glutamate, 1 mM EDTA, 10% glycerol). Elution was in buffer H containing 0.15 mg/mL 3× Flag peptide (EZBiolab). Peak fractions were diluted with 2 volumes of buffer H and loaded onto a 1-mL SP HP column (GE Healthcare) and washed with buffer C (50 mM HEPES pH 7.5, 100 mM KOAc). Elution was with an 8-mL linear gradient of 100–600 mM KOAc in 50 mM HEPES pH 7.5. The S6-ORC complex eluted at approximately 400 mM KOAc. Protein concentration of column fractions was determined by Bradford reagent (Bio-Rad), and stored at -80 °C.

S6-MCM

Recombinant S.c. strains that co-expressed the six subunits of Mcm2-7 complex having an N-terminal S6-3×Flag tag for fluorescent labeling (S6-MCM), or only a 3×Flag tag on Mcm3 (wt MCM) were constructed as follows. The six subunits of Mcm2-7 were cloned into integration vectors, each having the galactose inducible Gal1/10 bidirectional promotor for induction by galactose. The vectors were then integrated into the genome of strain OY001. Genes encoding Mcm4 and 3×Flag-S6-Mcm3 (or 3×Flag-Mcm3) were cloned into pRS404/GAL (Trp); genes encoding Mcm6 and Mcm7 were cloned into pRS405/GAL (Leu); the gene encoding Mcm2 was cloned into pRS403/GAL (His); and the Mcm5 gene was cloned into pRS406/GAL (Ura). The vectors were integrated in OY001 in the order described above to yield S6-MCM (or wt MCM).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S6-MCM cells were grown and induced as described for S6-ORC. Purification of S6-MCM complex was performed by lysis of 12 L of frozen cells with a SPEX cryogenic grinding mill (6970 EFM). Ground cells were thawed and debris removed by centrifugation (12,500 rpm in a SLC-1500 rotor at 4 °C). The clarified extract was applied to a 3-mL anti-Flag M2 affinity column (Sigma) equilibrated in buffer M [50 mM HEPES pH 7.5, 100 mM potassium glutamate, 2 mM DTT, 1 mM EDTA, 10 mM $Mg(OAc)_2$, 0.5 mM ATP, 10% glycerol]. The column was washed with 25 mL buffer M, then eluted with buffer M containing 0.15 mg/mL 3× Flag peptide (EZBiolab). Protein concentration of column fractions was determined by Bradford reagent (Bio-Rad), then aliquoted, snap frozen in liquid N_2 and stored at -80 °C.

Cdc6

An *E. coli* optimized gene sequence for expression of S.c. Cdc6 in *E. coli* (a generous gift of Dr. Megan Davey, Western University, Canada) was cloned into pET11 (Novagen). The Cdc6 gene was then subcloned into pGEX-6P-1 (GeneScript) to provide an N-terminal GST tag with a PreScission protease site which we refer to as pGST-Cdc6 PST/P. *E. coli* BL21(DE3) cells were transformed with the pGST-Cdc6 PST/P plasmid and cells were grown in LB + 100 µg/mL ampicillin at 37 °C with shaking until reaching an OD_{600} of 0.49, at which time the culture temperature was quickly reduced to 14 °C by shaking in an ice bath. Then IPTG (1 mM) and 0.2% ethanol were added to induce GST-Cdc6 expression for 24 h at 15 °C. Cells were harvested by low speed centrifugation at 4 °C, and the cell pellet was resuspended in buffer G (50 mM Tris-HCl pH 7.5, 1mM EDTA, 10% glycerol) plus 30 µM spermidine and 500 mM NaCl. Cells were lysed by French Press and the lysate was clarified by centrifugation at 14,000 rpm for 1 h at 4 °C. The supernatant was loaded onto a 5-mL glutathione column (GE Healthcare), followed by washing with 20 column volumes of buffer G plus 300 mM NaCl. Elution was performed using 25 mL of 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 10% glycerol, 47 mM glutathione and 300 mM NaCl. Fractions of 1 mL were collected and analyzed by SDS PAGE. Fractions containing GST-Cdc6 were treated with 200 U PreScission protease (ThermoFisher) for 2 h on ice, then were diluted with buffer A (20 mM HEPES pH 7.5, 1 mM EDTA, 10% glycerol) to a conductivity of 85 uS/CM and loaded onto an 8-mL SP-Sepharose column (Sigma). Cdc6 was eluted in 20-mL steps of buffer A containing either 0.2 M, 0.3 M, 0.4 M, 0.5 M NaCl. Fractions were analyzed for Cdc6 by SDS PAGE, and fractions containing Cdc6 were pooled and passed through a 2-mL GST column to remove remaining GST contaminants. Protein concentration was determined by Bradford reagent (Bio-Rad), then aliquoted, snap frozen in liquid N_2 and stored at -80 °C.

Cdt1

Cdt1 was cloned into a pET16b vector followed by replacement of the NcoI/NdeI region with insertion of a DNA segment encoding a 3× Flag tag. *E. coli* was transformed with the pFlag-Cdt1-pET plasmid and cells were grown in LB + 100 µg/mL ampicillin at 37 °C with shaking until reaching an OD_{600} of 0.6, at which time the culture temperature was quickly reduced to 15 °C by shaking in an ice bath. Then IPTG (1 mM) was added to induce Cdt1 expression for 10 h at 15 °C. Cells were then harvested by low speed centrifugation at 4 °C, and the cell pellet was resuspended in 50 mL of buffer B (50 mM HEPES pH 7.5, 1 mM EDTA, 2 mM DTT, 2 mM $MgCl_2$, 10% glycerol) plus 800 mM NaCl, and lysed by French Press. Cell lysate was clarified by centrifugation at 14,000 rpm for 1 h at 4 °C. The supernatant was treated with 1.5 mL Flag beads (Sigma) for 1 h with

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

gentle agitation, then packed into a 5-mL GE C column equilibrated in buffer B at 4 °C. Cdt1 was eluted with buffer B containing 175 mM NaCl and 0.2 mg/mL 3× Flag peptide. The preparation was then diluted 2-fold with buffer B to reduce conductivity, applied to a 1-mL Heparin agarose column (Sigma), and eluted with a 10-mL linear gradient of 100 mM NaCl to 500 mM NaCl in buffer B. Protein concentration of column fractions was determined by Bradford reagent (Bio-Rad), then aliquoted, snap frozen in liquid N_2 and stored at -80 °C.

Nap1

S.c. Nap1 was expressed in *E. coli* from the gene inserted into pGEX-6P-1, a kind gift of Dr. Aaron Johnson (University of Colorado, Denver). The GST-Nap1 was purified essentially as described [57,58], except for passage of the final prep through a GST column (ThermoFisher). Briefly, the pGEX-GST-Nap1 expression plasmid was transformed into *E. coli* BL21 (DE3) cells, and 6 L were grown in LB plus 100 μg/mL ampicillin to an OD_{600} of 0.5 at which time the culture temperature was quickly reduced to 15 °C by shaking in an ice bath. Then IPTG (1 mM) was added to induce Nap1 expression for 10 h at 15 °C. Cells were harvested by low speed centrifugation at 4 °C, and the cell pellet was resuspended in 100 mL of PBS (ThermoFisher) plus 500 mM NaCl, 1.5 mM DTT, 1 mM EDTA, 30 mM spermidine and 0.5% Triton X-100. Cells were lysed by French Press, and the cell lysate was clarified by centrifugation. The supernatant was loaded onto a 4-mL GST column (GE Healthcare) equilibrated in PBS containing 1 mM EDTA, 0.1 mM DTT, 0.1 mM PMSF, 0.5% Triton X-100, 10% glycerol, and 350 mM NaCl. Elution was with 50 mM Tris-HCl pH 8.0, 1 mM EDTA, 5 mM DTT, 300 mM NaCl, 10% glycerol and 47 mM glutathione. Fractions of 2 mL were collected and analyzed by SDS PAGE for presence of Nap1. Fractions containing Nap1 were pooled and 100 U of PreScission protease (ThermoFisher) was added prior to dialysis overnight against buffer C (20 mM Tris pH 7.5, 1 mM EDTA, 1mM DTT, 150 mM NaCl, 10% glycerol). Twenty eight mL of the dialysate was then loaded onto a 1-mL MonoQ column and eluted with a 20-mL gradient from 150 mM to 1 M NaCl in buffer C. Peak fractions containing Nap1 were pooled and then passed over a GST column to remove any contaminating GST tag and GST-PreScission protease. Protein concentration of column fractions was determined by Bradford reagent (Bio-Rad), then aliquoted, snap frozen in liquid N_2 and stored at -80 °C.

Histone octamer

Recombinant *Xenopus laevis* and S.c. histones and their mutants were purified as previously described [59]. Briefly, histones were expressed in BL21 (DE3) cells. Histones were extracted from inclusion bodies under denaturing conditions and purified through Q FF and SP FF columns (GE Healthcare). Histone octamers were then refolded by dialysis and purified by gel filtration using a Superdex 200 10/300 GL column.

Fluorescent labeling of proteins

To label S6-ORC and S6-MCM, the protein, Sfp synthase and dye-CoA (dye = Cy3 or LD650) were incubated at a 1:2:5 molar ratio for 1 h at room temperature in the presence of 10 mM $MgCl_2$. Excess dyes and Sfp were removed by a 100-kDa Amicon spin filter. The sample was then buffer-exchanged into 50 mM HEPES pH 7.5, 100 mM KOAc, 250 mM potassium glutamate, 1 mM EDTA, 10% glycerol, and 10 mM $MgSO_4$ for S6-ORC; or into 50 mM HEPES KOH pH 7.5, 100 mM KOAc, 2

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mM DTT, 10 mM $Mg(OAc)_2$, 0.5 mM ATP and 10% glycerol for S6-MCM. The final products were aliquoted, flash frozen, and stored at -80 °C.

To label histones, the single-cysteine construct $H2A^{K120C}$ was generated by site-directed mutagenesis. All histones were purified and labeled as previously described [59]. Briefly, they were incubated with Cy3 maleimide (GE Healthcare) or A488 C_5 maleimide (ThermoFisher) at 1:5 molar ratio in a labeling buffer (20 mM Tris-HCl pH 7.0, 7 M Guanidine-HCl, 5 mM EDTA, and 1.25 mM TCEP) for 4 h at room temperature. The labeling reaction was quenched with 80 mM β-mercaptoethanol. Excess dyes were removed by dialysis with a 10-kDa Amicon spin filter. The labeling efficiency varies among batches, ranging from ~50% to >90%.

Preparation of DNA template for single molecule experiments

To generate λ_{ARS1} DNA, a 501-bp DNA fragment containing the 185-bp Stillman minimum ARS1 sequence [60] (5'-TCGAGAAGCAGGTGGGACAGGTGAACTTTTGGATTGGAACTCGATTTCTGACTGGGTTGGAAGGCAAG AGAGCCCCGAAAGCTTACATTTTATGTTAGCTGGTGGACTGACGCCAGAAAATGTTGGTGATGCGCTTA GATTAAATGGCGTTATTGGTGTTGATGTAAGCGGAGGTGTGGAGACAAATGGTGTAAAAGACTCTAAC AAAATAGCA**AATTTCGTCAAAAATGCT**AAGAAATAGGTTATTACTGAGTAGTATTT**ATTTAAGTATT**GTT TGTGCACTTGCCTGCAGGCCTTT**TGAAAAGCAAGCA**TAAAAG**ATCTAAACATAAAATCT**GTAAAATAAC AAGATGTAAAGATAATGCTAAATCATTTGGCTTTTTGATTGATTGTACAGGAAAATATACATCGCAGGG GGTTGACTTTTACCATTTCACCGCAATGGAATCAAACTTGTTGAAGAGAATGTTCACAGGCGCATACGCT ACAATGACCCGATTCTTG-3'; the 185-bp ARS1 sequence is underlined; bold sequences represent B3, B2, B1 and ACS elements) was amplified from the yeast chromosome (S288C_ChrIV BK0069382: 462,279–462,787) and inserted into λ DNA (Roche, Cat# 11558706910) with XhoI and NheI restriction enzymes (New England BioLabs). The product was then packaged into phage particles using phage extract (MaxPlax, Epicentre). Plaques were generated on LE392 *E. coli* bacterial lawns (Epicentre) and screened for the ARS1 insert. A screened plaque was used as a phage source to purify λ_{ARS1} DNA by lytic growth [61]. The final λ_{ARS1} DNA is 47,822 bp in length and the ARS1 site is located 33,499–33,999 bp from the left end of the the phage genome. To create a terminally biotinylated λ_{ARS1} DNA, the 12-base 5' overhang on each end was filled in with a mixture of unmodified and biotinylated nucleotides by the exonuclease-deficient DNA polymerase I Klenow fragment (New England BioLabs). The reaction was conducted by incubating 10 nM λ_{ARS1} DNA, 33 μM each of dGTP/dATP/biotin-11-dUTP/biotin-14-dCTP (Thermo Fisher), and 5 U Klenow in 1× NEB2 buffer at 37 °C for 45 min, followed by heat inactivation at 75 °C for 20 min. The DNA was then ethanol precipitated overnight at -20 °C in 2.5× volume cold ethanol and 300 mM NaOAc pH 5.2. Precipitated DNA was recovered by centrifugation at 20,000× g at 4 °C for 15 min. After removing the supernatant, the pellet was air-dried, resuspended in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and stored at 4 °C.

Single-molecule experiments

Data acquisition

Single-molecule experiments were performed at room temperature on a LUMICKS C-Trap instrument as previously described [32]. Laminar-flow-separated channels 1-3 were used to form DNA tethers between two 4.35-μm streptavidin-coated polystyrene beads (Spherotech). Channels 4 and 5 served as protein loading and imaging channels. A488, Cy3, and Cy5/LD650

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fluorophores were excited by 488 nm, 532 nm and 638 nm laser lines, respectively. Kymographs were generated via confocal line scanning through the center of the two beads.

Nucleosome assembly in situ

Optical traps tethering a single DNA were moved to a channel containing 1 nM of fluorescently labeled S.c. or X.l. histone octamers and 2 nM Nap1 in HR buffer [30 mM Tris-OAc pH 7.5, 20 mM $Mg(OAc)_2$, 50 mM KCl, 1 mM DTT, 40 µg /mL BSA], and incubated under a tension below 1 pN until a few fluorescent spots were seen. The tether was then moved to Channel 3 containing 0.5 mg/mL salmon sperm DNA (Thermo Fisher) in HR buffer except for MCM/ORC salt stability assays, in which Channel 3 contained a high-salt buffer [50 mM HEPES pH 7.5, 40 µg/mL BSA, 2 mM DTT, 10 mM $Mg(OAc)_2$, 500 mM NaCl, 2.5 mM ATP]. A microfluidic flow was turned on for 1 min to gently remove free histones and free Nap1.

Visualization of ORC

Optical traps tethering a single bare DNA or nucleosomal DNA were moved to a separate channel containing 2 nM Cy3- or LD650-ORC and 5 nM Cdc6 in an imaging buffer containing 25 mM Tris-OAc pH 7.5, 5% glycerol, 40 µg/mL BSA, 3 mM DTT, 2 mM TCEP, 0.1 mM EDTA, 10 mM $Mg(OAc)_2$, 50 mM potassium glutamate, and 2.5 mM ATP. The imaging buffer was supplemented with an ATP-regeneration system [60 mg/mL creatine phosphokinase (Sigma) and 20 mM phosphocreatine (Sigma)], a triplet-state quenching cocktail [1 mM cyclooctatetraene (Sigma), 1 mM 4-nitrobenzyl alcohol (Sigma) and 1 mM Trolox (Sigma)], as well as an oxygen scavenging system [10 nM protocatechuate-3,4-dioxygenase (Sigma) and 2.5 mM protocatechuic acid (Sigma)]. The kymograph was typically recorded for 10 min.

Visualization of MCM

For two-color MCM/nucleosome experiments, a tethered nucleosomal DNA loaded with Cy3 or A488-labeled octamers was moved to a channel containing 10 nM LD650-MCM and 14 nM Cdt1, with or without 2 nM ORC and 5 nM Cdc6, in the imaging buffer described above in the presence of 5 mM ATP. For MCM DH loading experiments, a tethered nucleosomal DNA was moved to a channel containing 10 nM Cy3-MCM, 10 nM LD650-MCM, 4 nM unlabeled ORC, 10 nM Cdc6, 28 nM Cdt1, and 5 mM ATP. For MCM/ORC salt stability experiments, the tether was moved to a channel containing the high-salt buffer described above. The oxygen scavenging system was omitted from the imaging buffer for the MCM photobleaching experiments.

Data analysis

Single-molecule force and fluorescence data from the .h5 files generated from C-Trap experiments were analyzed using tools in the lumicks.pylake Python library supplemented with other Python modules in a custom GUI Python script titled “C-Trap .h5 File Visualization GUI” (https://harbor.lumicks.com/single-script/c5b103a4-0804-4b06-95d3-20a08d65768f), which was written to extract confocal images and traces.

Genomic analysis

We analyzed the following published datasets: SRR034475 and SRR034476 (Orc1 ChIP-seq) [11]; SRR1261333 (Mcm2-7 ChIP-seq) [37]; GSE147927 (ChIP-exo data for Orc1, Abf1, and Reb1) [40];

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

GSM2589911 (MNase data) [41]. The ChIP-seq data were aligned to yeast genome (version Scer3) with bowtie2, and peaks were called using MACS2 (threshold: effective p value 0.01). Orc1 and Mcm2-7 ChIP peaks tend to be broader than typical transcription factors, probably due to the large size of the complex. As a result, multiple peaks tend to appear as clusters within a few hundred bps. We wrote a MATLAB algorithm to select a single peak with the largest area underneath within each cluster. We found 295 Orc1 peaks in total (sorted based on the total intensity of the Orc1 ChIP-seq signal in the +/- 1kb region in **Figure 7A**). ARS annotation was downloaded from the Saccharomyces Genome Database. All Orc1 peak locations (peak center +/- 100bp) and the corresponding ARSs are shown in **Supplementary Table 1**.

To get the Orc1 motif information, we extracted the genomic sequences within peak center +/- 500bp. Using the PWM in [39], we used a MATLAB program to scan through the sequence and calculated the PWM score for each 17-bp elements on both strands. At each base, PWM score calculated the $\log_2$ of probability of the appearance of a base divided by the probability for that base to appear in the genome. For example, if for Position X, "A" appears 90% of time, while in the genome it is 25%, then the score of Position X is $\log_2(0.9/0.25)$. The final score is the sum of scores for all positions in the element. Abf1 and Reb1 motifs were analyzed the same way, except that we scanned through shorter sequences (peak center +/- 150bp) because the ChIP peaks of these factors are sharper.

QUANTIFICATION AND STATISTICAL ANALYSIS

Errors reported in this study represent the standard deviation (SD). *P* values were determined from unpaired two-tailed *t* tests using GraphPad Prism 9 (ns, not significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. Bell, S.P. & Labib, K. Chromosome Duplication in Saccharomyces cerevisiae. *Genetics* **203**, 1027-67 (2016).
2. Bleichert, F., Botchan, M.R. & Berger, J.M. Mechanisms for initiating cellular DNA replication. *Science* **355**(2017).
3. Rhind, N. & Gilbert, D.M. DNA replication timing. *Cold Spring Harb Perspect Biol* **5**, a010132 (2013).
4. Kelly, T. Historical Perspective of Eukaryotic DNA Replication. *Adv Exp Med Biol* **1042**, 1-41 (2017).
5. Siddiqui, K., On, K.F. & Diffley, J.F. Regulating DNA replication in eukarya. *Cold Spring Harb Perspect Biol* **5**(2013).
6. Bell, S.P. & Stillman, B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**, 128-34 (1992).
7. Duncker, B.P., Chesnokov, I.N. & McConkey, B.J. The origin recognition complex protein family. *Genome Biol* **10**, 214 (2009).
8. Evrin, C. et al. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc Natl Acad Sci U S A* **106**, 20240-5 (2009).
9. Remus, D. et al. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139**, 719-30 (2009).
10. Das, M., Singh, S., Pradhan, S. & Narayan, G. MCM Paradox: Abundance of Eukaryotic Replicative Helicases and Genomic Integrity. *Mol Biol Int* **2014**, 574850 (2014).
11. Eaton, M.L., Galani, K., Kang, S., Bell, S.P. & MacAlpine, D.M. Conserved nucleosome positioning defines replication origins. *Genes Dev* **24**, 748-53 (2010).
12. Simpson, R.T. Nucleosome positioning can affect the function of a cis-acting DNA element in vivo. *Nature* **343**, 387-9 (1990).
13. Venditti, P., Costanzo, G., Negri, R. & Camilloni, G. ABFI contributes to the chromatin organization of Saccharomyces cerevisiae ARS1 B-domain. *Biochim Biophys Acta* **1219**, 677-89 (1994).
14. Alexiadis, V., Varga-Weisz, P.D., Bonte, E., Becker, P.B. & Gruss, C. In vitro chromatin remodelling by chromatin accessibility complex (CHRAC) at the SV40 origin of DNA replication. *EMBO J* **17**, 3428-38 (1998).
15. Prioleau, M.N. & MacAlpine, D.M. DNA replication origins-where do we begin? *Genes Dev* **30**, 1683-97 (2016).
16. Gilbert, D.M. In search of the holy replicator. *Nat Rev Mol Cell Biol* **5**, 848-55 (2004).
17. Kelly, T. & Callegari, A.J. Dynamics of DNA replication in a eukaryotic cell. *Proc Natl Acad Sci U S A* **116**, 4973-4982 (2019).
18. Li, N. et al. Structure of the origin recognition complex bound to DNA replication origin. *Nature* **559**, 217-222 (2018).
19. Lee, C.S.K. et al. Humanizing the yeast origin recognition complex. *Nat Commun* **12**, 33 (2021).
20. Hu, Y. et al. Evolution of DNA replication origin specification and gene silencing mechanisms. *Nat Commun* **11**, 5175 (2020).
21. Yuan, Z. et al. Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat Struct Mol Biol* **24**, 316-324 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

644 22. Yeeles, J.T., Deegan, T.D., Janska, A., Early, A. & Diffley, J.F. Regulated eukaryotic DNA
645 replication origin firing with purified proteins. *Nature* **519**, 431-5 (2015).
646 23. Smith, O.K. & Aladjem, M.I. Chromatin structure and replication origins:
647 determinants of chromosome replication and nuclear organization. *J Mol Biol* **426**,
648 3330-41 (2014).
649 24. Bellush, J.M. & Whitehouse, I. DNA replication through a chromatin environment.
650 *Philos Trans R Soc Lond B Biol Sci* **372**(2017).
651 25. Devbhandari, S., Jiang, J., Kumar, C., Whitehouse, I. & Remus, D. Chromatin
652 Constrains the Initiation and Elongation of DNA Replication. *Mol Cell* **65**, 131-141
653 (2017).
654 26. Azmi, I.F. et al. Nucleosomes influence multiple steps during replication initiation.
655 *Elife* **6**(2017).
656 27. Kurat, C.F., Yeeles, J.T.P., Patel, H., Early, A. & Diffley, J.F.X. Chromatin Controls DNA
657 Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates.
658 *Mol Cell* **65**, 117-130 (2017).
659 28. De Ioannes, P. et al. Structure and function of the Orc1 BAH-nucleosome complex.
660 *Nat Commun* **10**, 2894 (2019).
661 29. Hizume, K., Yagura, M. & Araki, H. Concerted interaction between origin recognition
662 complex (ORC), nucleosomes and replication origin DNA ensures stable ORC-origin
663 binding. *Genes Cells* **18**, 764-79 (2013).
664 30. Stinchcomb, D.T., Struhl, K. & Davis, R.W. Isolation and characterisation of a yeast
665 chromosomal replicator. *Nature* **282**, 39-43 (1979).
666 31. Candelli, A., Wuite, G.J. & Peterman, E.J. Combining optical trapping, fluorescence
667 microscopy and micro-fluidics for single molecule studies of DNA-protein
668 interactions. *Phys Chem Chem Phys* **13**, 7263-72 (2011).
669 32. Wasserman, M.R., Schauer, G.D., O'Donnell, M.E. & Liu, S. Replication Fork Activation
670 Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. *Cell* **178**, 600-611 e16
671 (2019).
672 33. Duzdevich, D. et al. The dynamics of eukaryotic replication initiation: origin
673 specificity, licensing, and firing at the single-molecule level. *Mol Cell* **58**, 483-94
674 (2015).
675 34. Sanchez, H. et al. DNA replication origins retain mobile licensing proteins. *Nat*
676 *Commun* **12**, 1908 (2021).
677 35. Ticau, S., Friedman, L.J., Ivica, N.A., Gelles, J. & Bell, S.P. Single-molecule studies of
678 origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* **161**,
679 513-525 (2015).
680 36. Crickard, J.B., Moevus, C.J., Kwon, Y., Sung, P. & Greene, E.C. Rad54 Drives ATP
681 Hydrolysis-Dependent DNA Sequence Alignment during Homologous
682 Recombination. *Cell* **181**, 1380-1394 e18 (2020).
683 37. Belsky, J.A., MacAlpine, H.K., Lubelsky, Y., Hartemink, A.J. & MacAlpine, D.M. Genome-
684 wide chromatin footprinting reveals changes in replication origin architecture
685 induced by pre-RC assembly. *Genes Dev* **29**, 212-24 (2015).
686 38. Cherry, J.M. et al. Saccharomyces Genome Database: the genomics resource of
687 budding yeast. *Nucleic Acids Res* **40**, D700-5 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

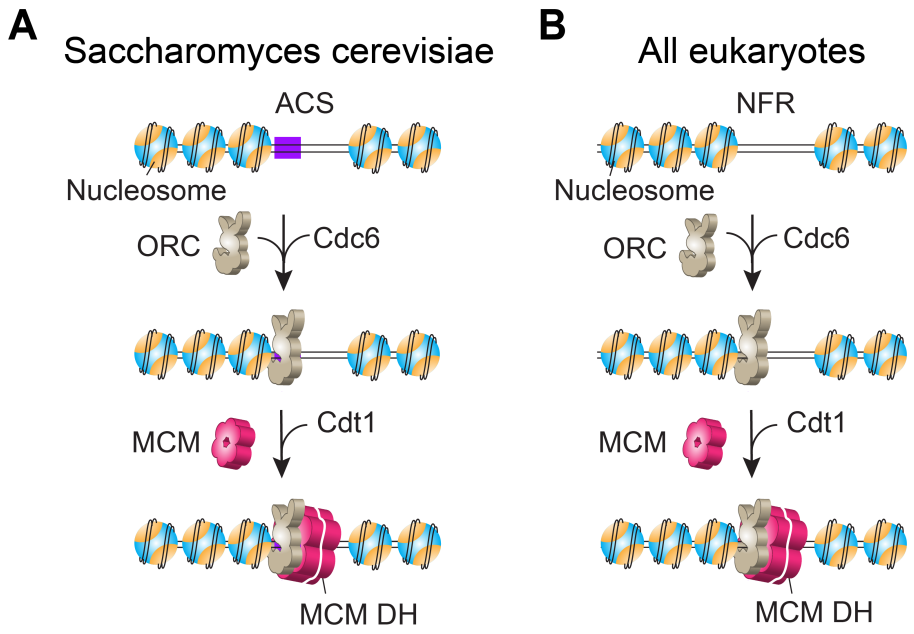
39. Spivak, A.T. & Stormo, G.D. ScerTF: a comprehensive database of benchmarked position weight matrices for Saccharomyces species. *Nucleic Acids Res* **40**, D162-8 (2012).
40. Rossi, M.J. et al. A high-resolution protein architecture of the budding yeast genome. *Nature* **592**, 309-314 (2021).
41. Kubik, S. et al. Sequence-Directed Action of RSC Remodeler and General Regulatory Factors Modulates +1 Nucleosome Position to Facilitate Transcription. *Mol Cell* **71**, 89-102 e5 (2018).
42. Yan, C., Chen, H. & Bai, L. Systematic Study of Nucleosome-Displacing Factors in Budding Yeast. *Mol Cell* **71**, 294-305 e4 (2018).
43. Chang, F. et al. High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. *Nucleic Acids Res* **39**, 6523-35 (2011).
44. Nieduszynski, C.A., Knox, Y. & Donaldson, A.D. Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev* **20**, 1874-9 (2006).
45. MacAlpine, D.M. & Almouzni, G. Chromatin and DNA replication. *Cold Spring Harb Perspect Biol* **5**, a010207 (2013).
46. Raisner, R.M. et al. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123**, 233-48 (2005).
47. Albert, I. et al. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature* **446**, 572-6 (2007).
48. Yadon, A.N. et al. Chromatin remodeling around nucleosome-free regions leads to repression of noncoding RNA transcription. *Mol Cell Biol* **30**, 5110-22 (2010).
49. Fragkos, M., Ganier, O., Coulombe, P. & Mechali, M. DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol* **16**, 360-74 (2015).
50. Aladjem, M.I. & Redon, C.E. Order from clutter: selective interactions at mammalian replication origins. *Nat Rev Genet* **18**, 101-116 (2017).
51. Li, N. et al. Structure of the eukaryotic MCM complex at 3.8 A. *Nature* **524**, 186-91 (2015).
52. Ge, X.Q., Jackson, D.A. & Blow, J.J. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev* **21**, 3331-41 (2007).
53. Long, H. et al. H2A.Z facilitates licensing and activation of early replication origins. *Nature* **577**, 576-581 (2020).
54. Kuo, A.J. et al. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* **484**, 115-9 (2012).
55. Zhou, Z. et al. Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem Biol* **2**, 337-46 (2007).
56. Georgescu, R.E. et al. Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat Struct Mol Biol* **21**, 664-70 (2014).
57. Speck, C., Chen, Z., Li, H. & Stillman, B. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nat Struct Mol Biol* **12**, 965-71 (2005).
58. Johnson, A. et al. Reconstitution of heterochromatin-dependent transcriptional gene silencing. *Mol Cell* **35**, 769-81 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

59. Li, S., Zheng, E.B., Zhao, L. & Liu, S. Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. *Cell Rep* **28**, 2689-2703 e4 (2019).
60. Marahrens, Y. & Stillman, B. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* **255**, 817-23 (1992).
61. Lockett, T.J. A bacteriophage lambda DNA purification procedure suitable for the analysis of DNA from either large or multiple small lysates. *Anal Biochem* **185**, 230-4 (1990).
62. Liachko, I., Youngblood, R.A., Keich, U. & Dunham, M.J. High-resolution mapping, characterization, and optimization of autonomously replicating sequences in yeast. *Genome Res* **23**, 698-704 (2013).
63. Coster, G. & Diffley, J.F.X. Bidirectional eukaryotic DNA replication is established by quasi-symmetrical helicase loading. *Science* **357**, 314-318 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

746 **FIGURES**
747



748
749 **Figure 1. Model for origin licensing in *Saccharomyces cerevisiae* vs. in most other eukaryotes**
750 (**A**) In a chromatinized genome, ORC searches for nucleosomes and stably associates with them
751 regardless of whether a nearby ARS consensus sequence (ACS) exists. ORC then loads MCM
752 double hexamers in conjunction with Cdc6 and Cdt1 at nucleosomal sites in G1 phase. Due to its
753 small genome, *S. cerevisiae* may have evolved ARS-dependent origins in order to limit ORC
754 binding to specific sequences and thus avoid replication-transcription conflicts [20], which does not
755 generalize to higher eukaryotes. (**B**) Origins in most eukaryotes lack a consensus ACS motif, but
756 still utilize ORC which is known to bind nucleosomes. We demonstrate in this study that S.c. ORC
757 harbors a cryptic ability to bind nucleosomes and direct MCM DH formation within nucleosome-
758 free regions (NFRs) that lack ARS consensus sequences. We propose that this mechanism is the
759 normal process for most eukaryotes, whose origin sites are chiefly defined by the nucleosomal
760 architecture. In this model, ACS confers origin specificity in S.c. by facilitating ORC binding to
761 nucleosomes proximal to a nucleosome-free ARS sequence.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

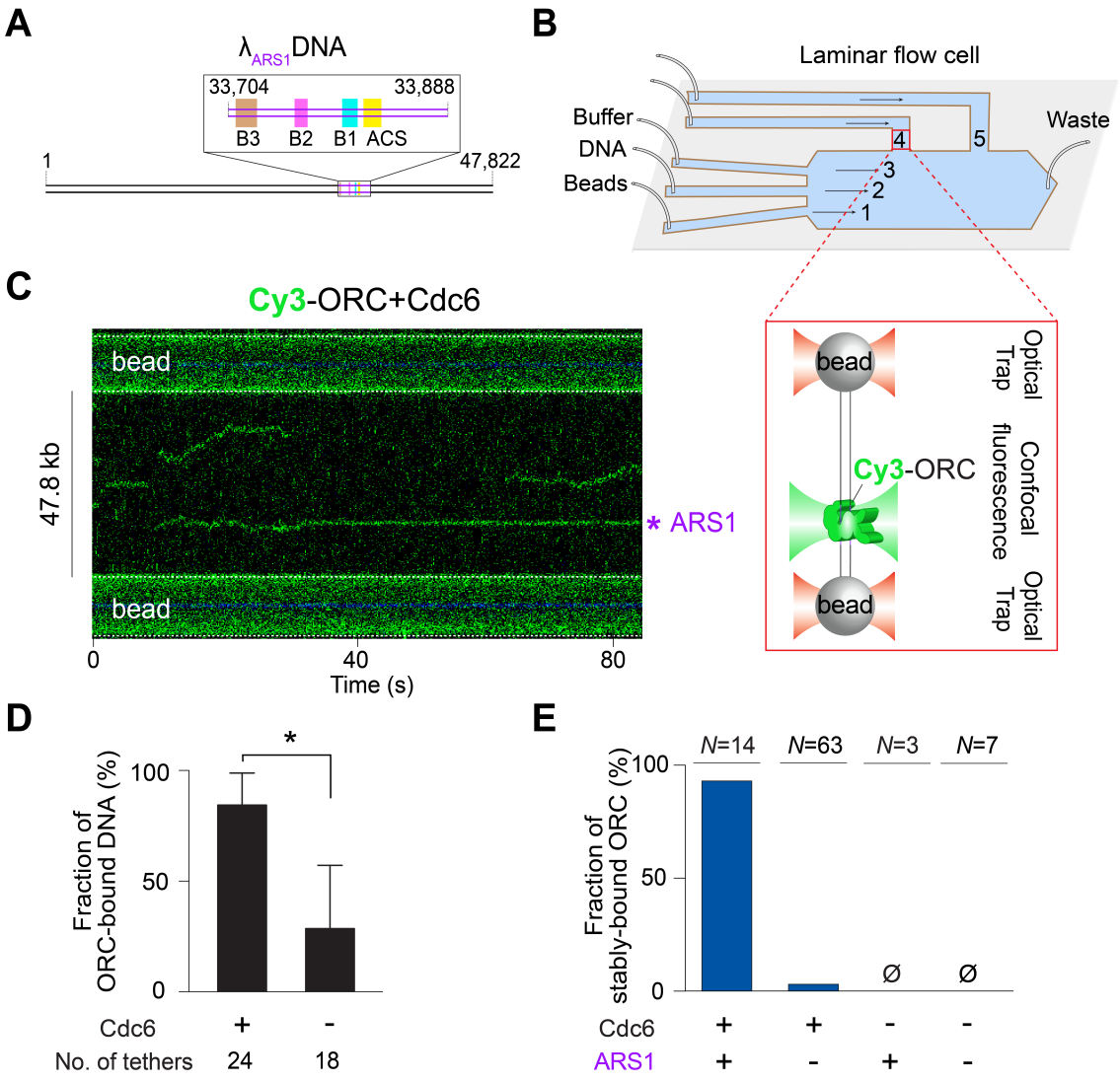


Figure 2. A single-molecule platform to study eukaryotic replication initiation

(**A**) Cartoon of the λ_{ARS1} DNA template. The inserted ARS1 element is illustrated in the inset box. (**B**) Schematic of the single-molecule experimental setup. Channels 1-3 are separated by laminar flow. Beads are optically trapped in channel 1, moved to channel 2 to tether DNA, then moved to channel 3 to characterize the force-extension curve of the tether. Once a correct tether is confirmed, the beads-DNA is moved to channel 4 or 5 containing the proteins. The zoom-in box illustrates the final assembly in the imaging channel (not drawn to scale). (**C**) A representative kymograph showing the behavior of Cy3-labeled ORC on λ_{ARS1} DNA in the presence of Cdc6. The engineered ARS1 position is indicated. (**D**) Fraction of λ_{ARS1} DNA tethers that were observed to have at least one ORC bound in the presence or absence of Cdc6. The protein concentrations used in this experiment are: 2 nM for ORC, and 5 nM for Cdc6. The number of tethers analyzed for each condition is indicated. Error bars represent SD. (**E**) Fraction of ORC molecules that stably reside at the ARS1 site versus non-ARS1 sites in the presence or absence of Cdc6. *N* incidates the number of ORC molecules analyzed for each condition.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

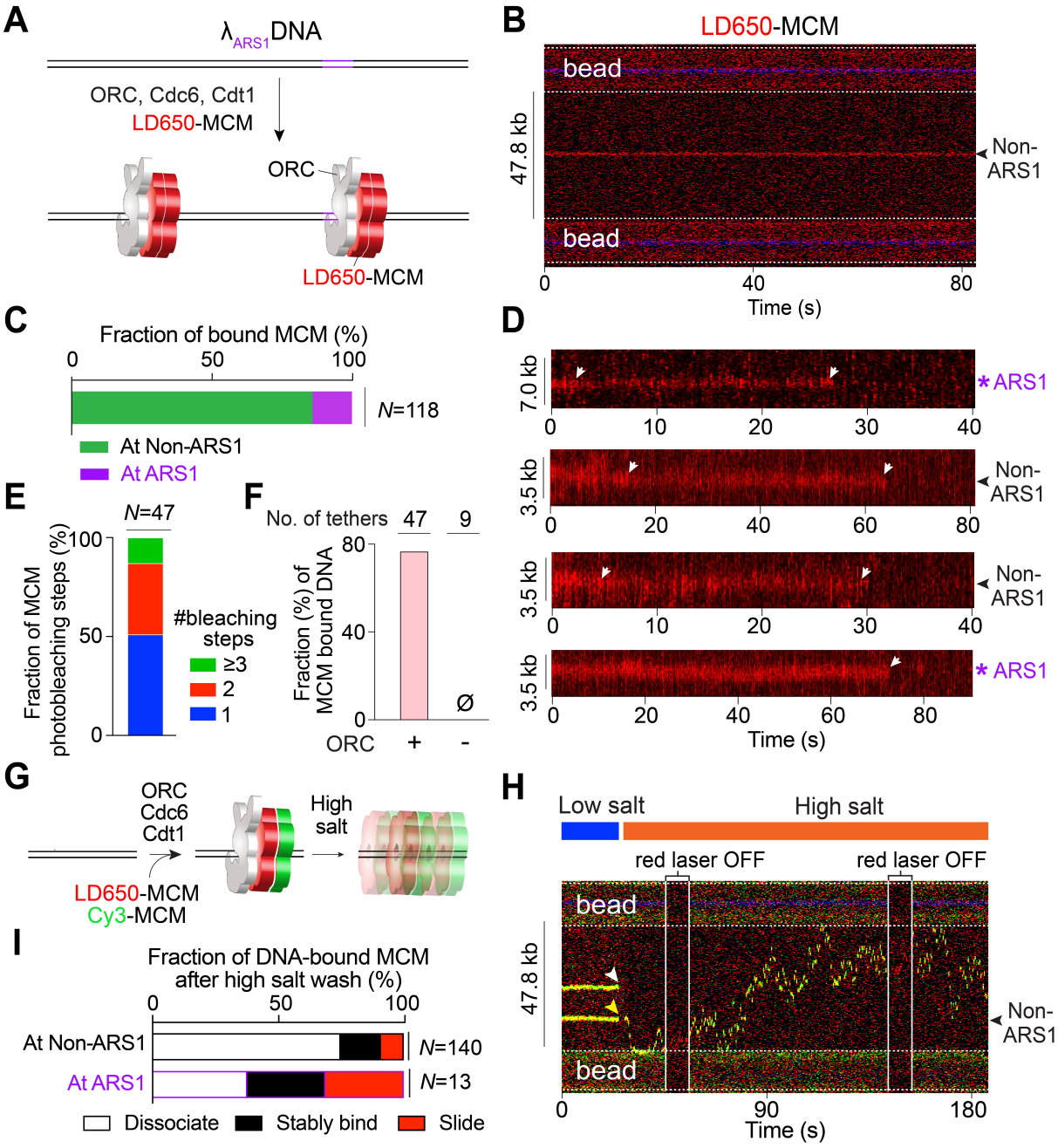


Figure 3. ORC-dependent MCM loading occurs frequently at non-ARS DNA sites

(**A**) Cartoon of the single-molecule pre-RC assembly experiment using λ_{ARS1} DNA, unlabeled ORC, Cdc6, Cdt1, and LD650-labeled MCM (red). (**B**) An example kymograph showing that the MCM fluorescence signal appeared at a non-ARS1 position on DNA. (**C**) Fraction of stably bound MCM complexes that were observed at ARS1 versus non-ARS1 positions. *N* indicates the number of MCM complexes analyzed. (**D**) Example kymographs showing the photobleaching steps (white arrows) of MCM fluorescence at ARS1 and non-ARS1 positions on the DNA tether. (**E**) Distribution of the number of photobleaching steps observed in each MCM fluorescence trajectory. *N*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

indicates the number of trajectories analyzed. (**F**) Fraction of DNA tethers that were observed to harbor at least one fluorescent MCM complex in the presence or absence of ORC. The protein concentrations used in this experiment are: 10 nM for MCM, 2 nM for ORC, and 5 nM for Cdc6. The number of tethers analyzed for each condition is indicated. (**G**) Cartoon of the high-salt wash experiment to demonstrate MCM loading on DNA using a mixture of LD650-MCM and Cy3-MCM, unlabeled ORC, Cdc6 and Cdt1. (**H**) A representative kymograph showing large-scale mobility of an MCM DH (indicated by the dual-color complex which appeared as yellow) loaded at a non-ARS1 position traversing the entire length of the tethered DNA upon high-salt wash (yellow arrowhead). The other MCM complex dissociated at high salt (white arrowhead). (**I**) Fraction of MCM complexes on nucleosome-free DNA that underwent diffusion without dissociation (red), remained stably bound to the DNA position (black), or dissociated into solution (white) upon high-salt wash. *N* indicates the number of MCM complexes analyzed.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

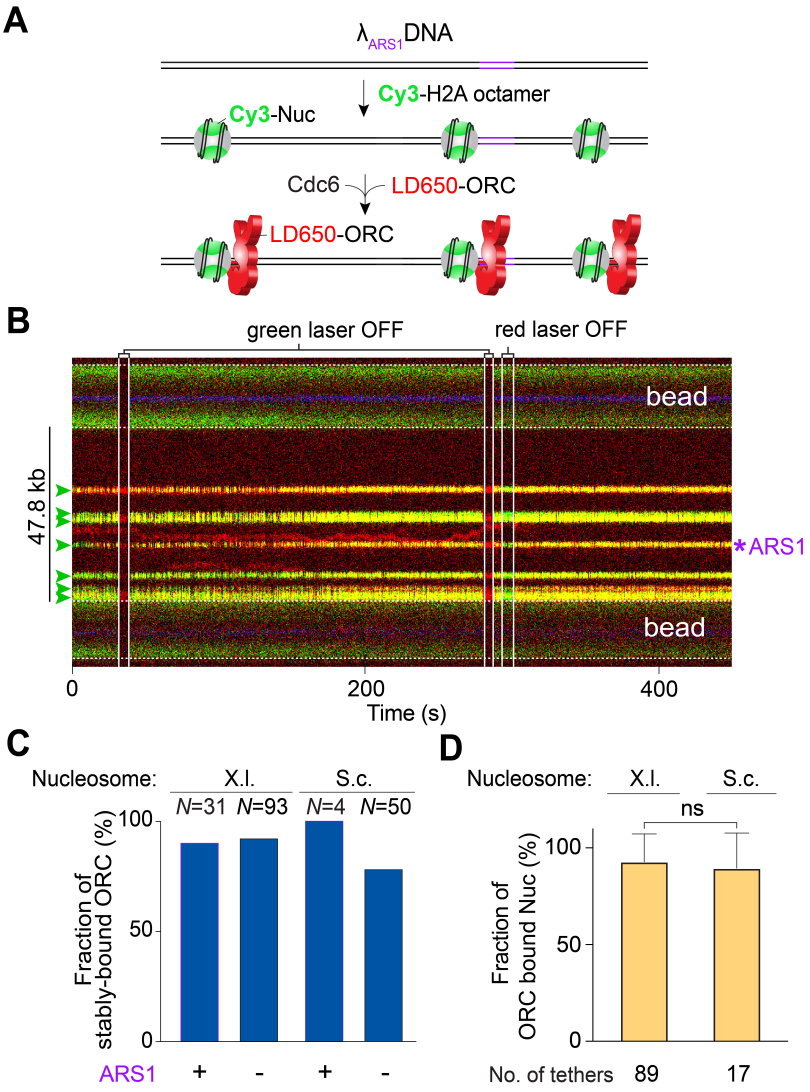


Figure 4. ORC predominantly binds to nucleosomes over ARS DNA

(**A**) Cartoon of the λ_{ARS1} DNA sparsely loaded with Cy3-labeled S.c. nucleosomes (green) and incubated with LD650-labeled ORC (red) and Cdc6. (**B**) A representative kymograph showing a λ_{ARS1} DNA tether loaded with multiple nucleosomes (positions indicated by green arrowheads), all of which were located at non-ARS1 sites except one. Each nucleosome was observed to be stably bound by ORC (red). The presence of ORC on the nucleosomes is confirmed by turning off the green laser, which showed only the red fluorescence from ORC; alternatively, turning off the red laser showed only the green fluorescence from the nucleosomes. (**C**) Fraction of ORC stably bound to nucleosomes (X.l. or S.c.) located at either the ARS1 site or non-ARS1 sites. *N* indicates the number of ORC molecules analyzed for each condition. (**D**) Fraction of nucleosomes (X.l. or S.c.) within a given DNA tether that were observed to be ORC-bound in the presence of 2 nM ORC and 5 nM Cdc6. The number of tethers analyzed for each condition is indicated. Error bars represent SD.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

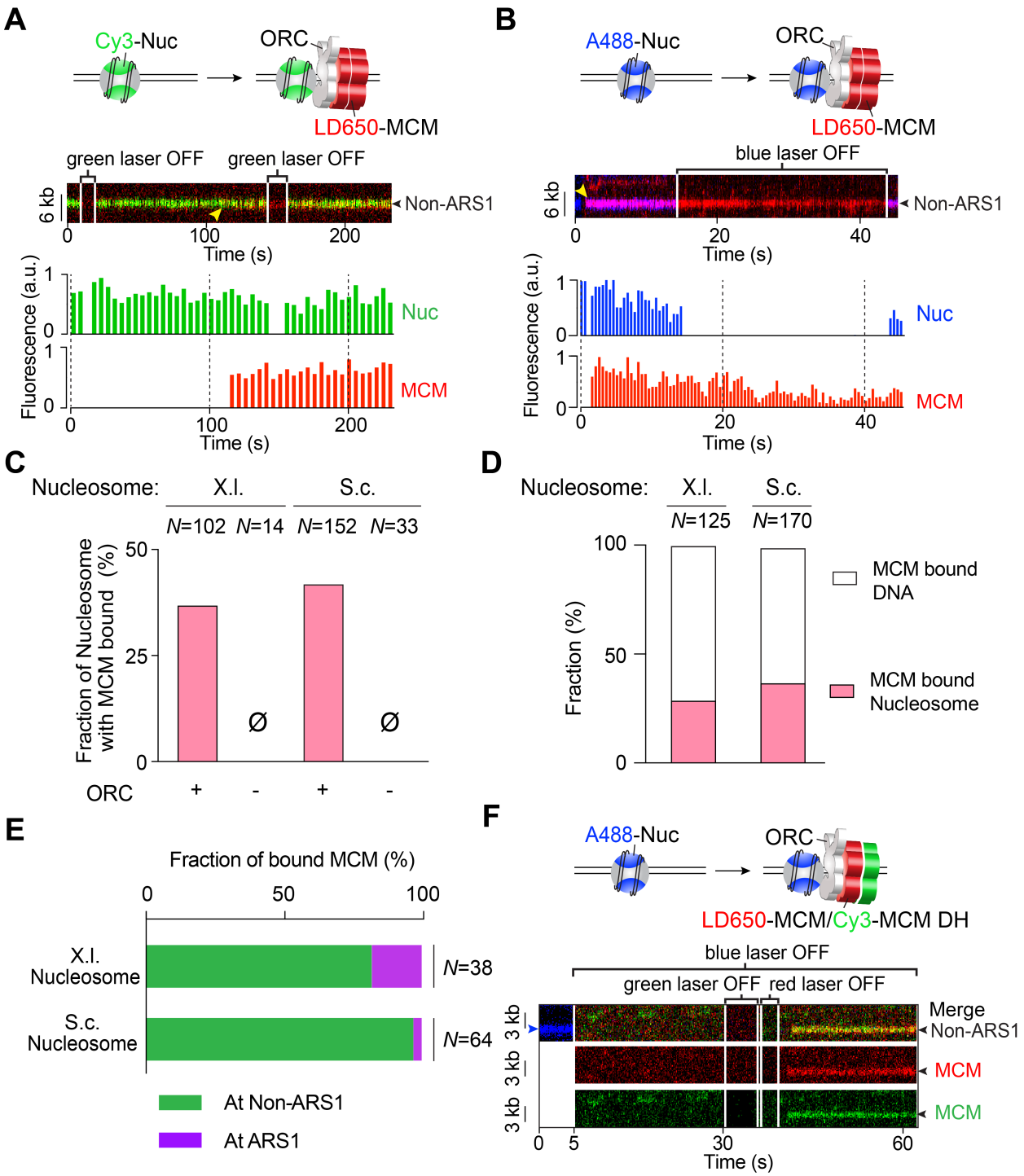


Figure 5. MCMs are recruited by ORC to nucleosomes independently of ARS DNA

(**A**) Cartoon (top), an example kymograph (middle) and the corresponding fluorescence intensities (bottom) of the pre-RC assembly experiment using Cy3-labeled X.l. nucleosomes (green), LD650-labeled MCM (red), unlabeled ORC, Cdc6 and Cdt1. Yellow arrowhead in the kymograph indicates the time when the MCM fluorescence signal appeared at the nucleosomal site. (**B**) Cartoon (top), an example kymograph (middle) and the corresponding fluorescence intensities (bottom) of the pre-RC assembly experiment using A488-labeled S.c. nucleosomes

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(blue), LD650-labeled MCM (red), unlabeled ORC, Cdc6 and Cdt1. In both examples in **A** and **B** the nucleosomes were at non-ARS1 positions on the DNA. (**C**) Fraction of nucleosomes (X.l. or S.c.) that were observed to have colocalized MCM signals in the presence or absence of ORC. *N* indicates the number of nucleosomes analyzed for each condition. (**D**) Fraction of MCM complexes on a nucleosome-loaded (X.l. or S.c.) tether that colocalized with a nucleosome versus with nucleosome-free DNA. *N* indicates the number of MCM complexes analyzed. (**E**) Fraction of MCM-nucleosome (X.l. or S.c.) colocalization events observed at ARS1 versus non-ARS1 positions. *N* indicates the number of events analyzed. (**F**) Cartoon (top) and an example kymograph (bottom) of the three-color experiments using A488-labeled S.c. nucleosomes (blue), both LD650-labeled MCM (red) and Cy3-labeled MCM (green), unlabeled ORC, Cdc6 and Cdt1. The colocalization of a dual-color MCM with a nucleosome indicates MCM DH recruitment to the nucleosomal site. Individual lasers were occasionally turned off to confirm the fluorescence signals from the other channels.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

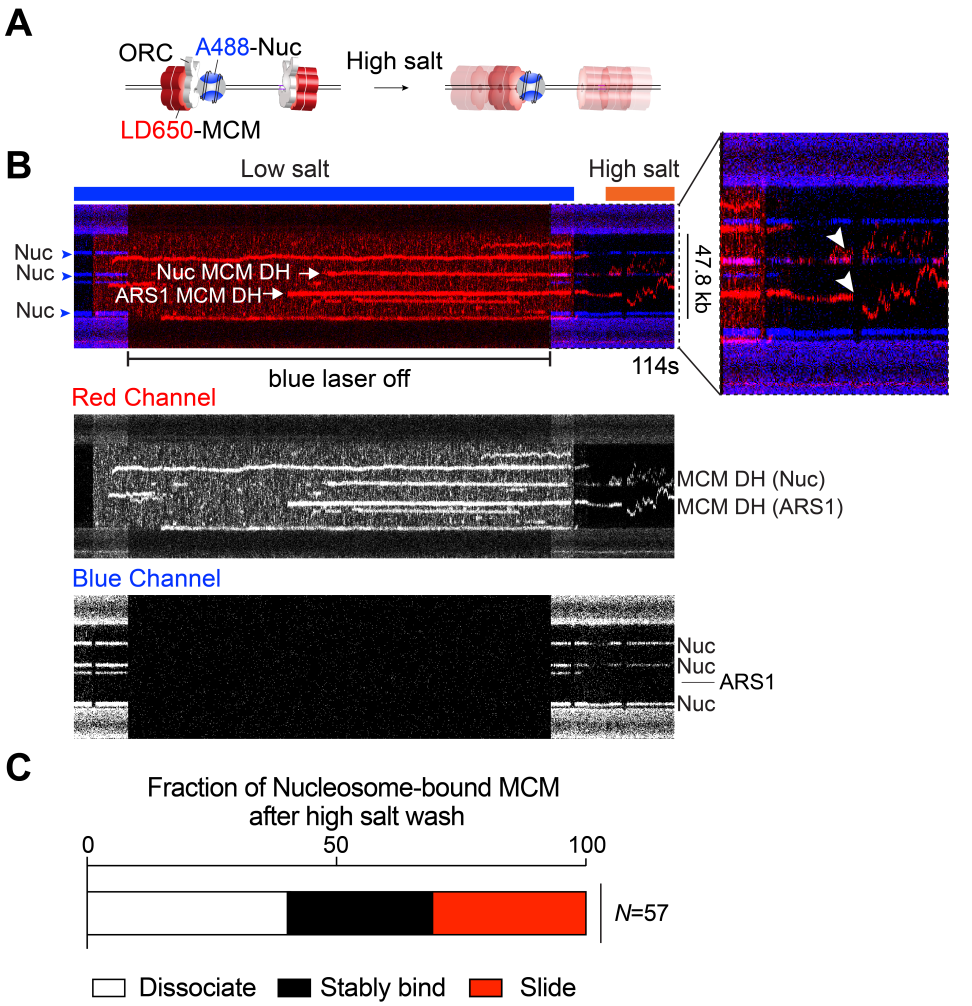


Figure 6. ORC-mediated MCM loading occurs at nucleosomal sites

(**A**) Cartoon illustrating the experimental assay that evaluates MCM loading at nucleosomal sites via high salt wash. (**B**) A representative kymograph showing that MCM complexes (red) formed on a section of bare DNA in the presence of unlabeled ORC, Cdc6 and Cdt1. Some of the MCMs colocalized with a S.c. nucleosome (blue). Upon moving to a high-salt buffer (0.5 M NaCl), a fraction of the MCMs displayed diffusive behavior from both bare DNA sites and nucleosomal sites (white arrowheads in the zoomed-in view) without dissociation, demonstrating their successful loading onto DNA. Blue arrowheads indicate nucleosome positions, all of which were at non-ARS1 sites on the tether shown in this example. MCM and nucleosome fluorescence signals are also separately shown in grey scale at the bottom. (**C**) Fraction of nucleosome-colocalized MCM complexes that underwent diffusion without dissociation (red), remained stably bound to the nucleosome (black), or dissociated into solution (white) upon high-salt wash. *N* indicates the number of MCM complexes analyzed.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

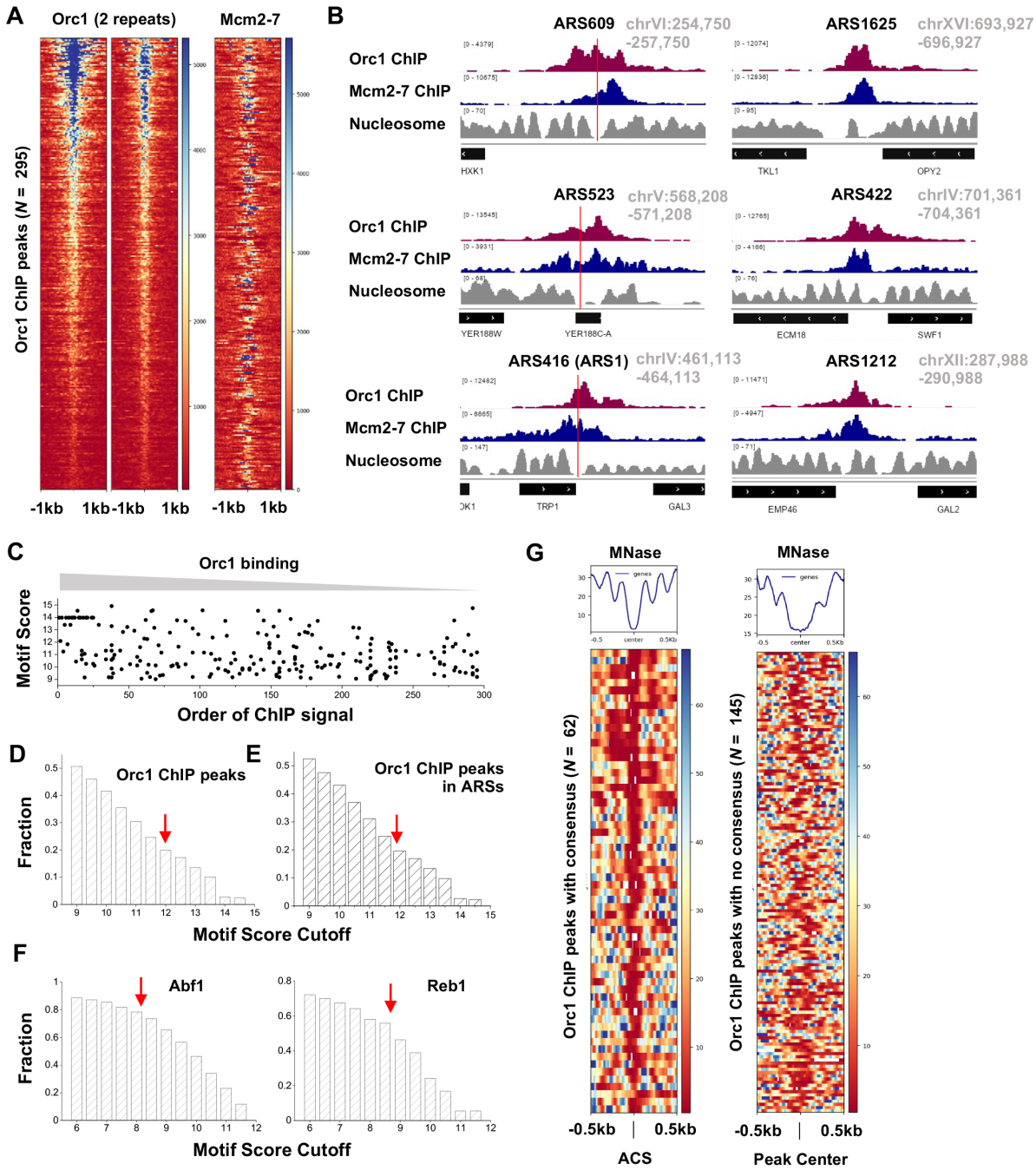


Figure 7. Genome-wide analysis of ORC/MCM localization

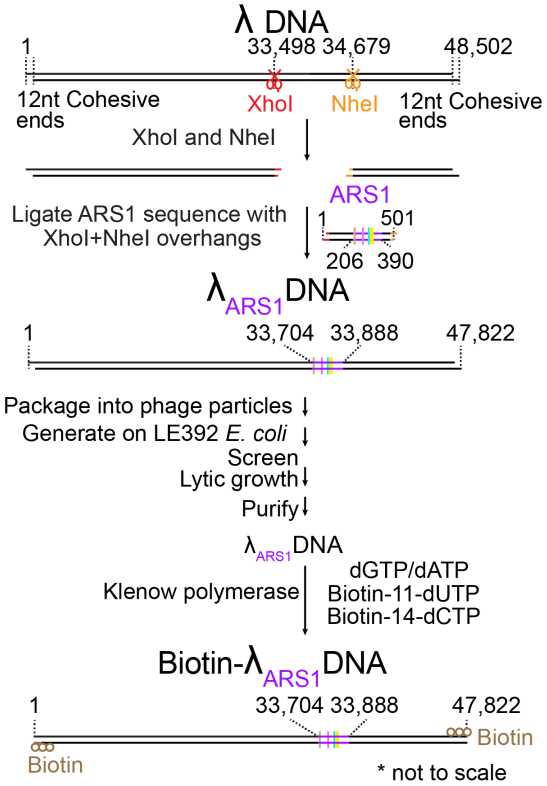
(**A**) Heatmap of Orc1 ChIP peaks (*N* = 295) and Mcm2-7 ChIP signal at the corresponding sites. (**B**) Examples of Orc1 and Mcm2-7 ChIP data and nucleosome occupancy at six ARSs. Red vertical lines represent the location of ACS consensuses in these regions. The three ARSs on the right contain no consensus with score above 9. (**C**) ACS motif score vs. the Orc1 ChIP peak strength. The motifs were identified in a 1kb region near Orc1 ChIP peaks (peak center +/- 500 bp). All elements with a PWM score >9 are shown here. For the *x* axis, "1" represents the largest Orc1 ChIP peak, and "295" is the weakest. (**D**) Fraction of sequences underlying all Orc1 peaks that

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contain motifs above a certain threshold (varying from 9 to 15). The red arrow represents the recommended cutoff of 11.9. (**E**) Same as in **D** except using a subset of Orc1 peaks that overlap with previously annotated ARSs. (**F**) Same as in **D** except for Abf1 and Reb1. The red arrows represent recommended cutoffs for these two factors. (**G**) Heatmap of nucleosome occupancy near Orc1 ChIP peaks. The left panel includes 62 peaks containing ACS consensus (score > 11.9), with each row aligned at the consensus site. The right panel includes 145 peaks with no consensus above 9, and it was aligned at the center of the ChIP peaks.

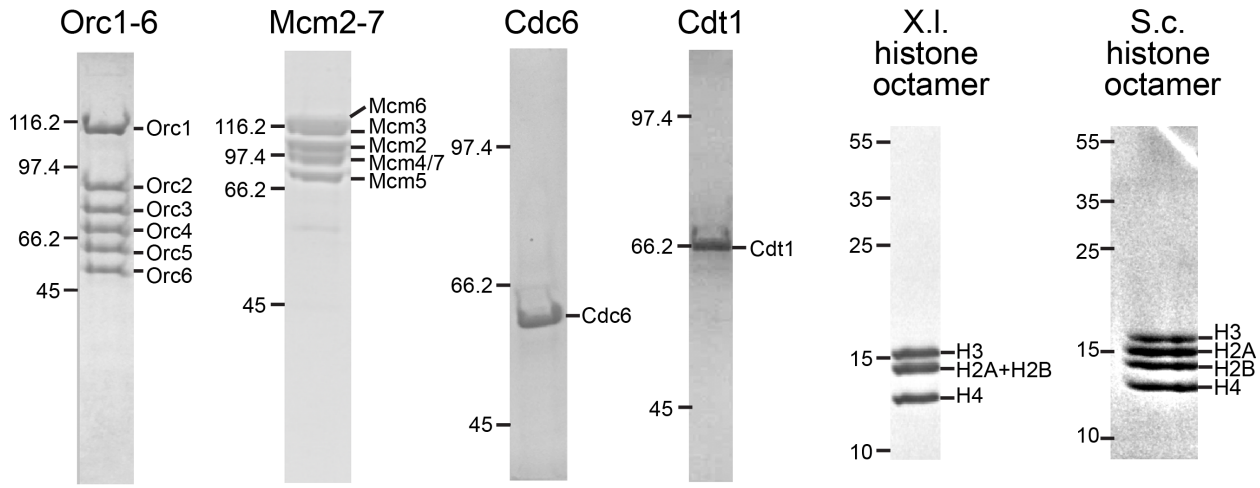
bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SUPPLEMENTARY FIGURES



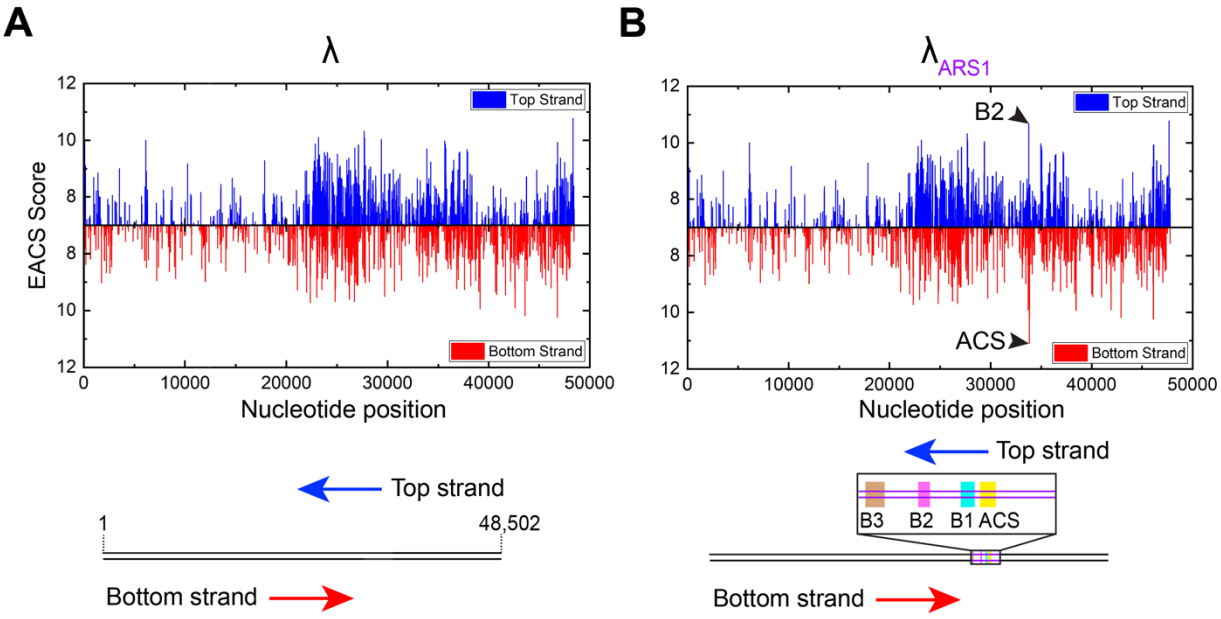
Supplementary Figure 1. Workflow for generating the biotinylated λ_{ARS1} DNA

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 2. SDS-PAGE gels showing the proteins used in this work

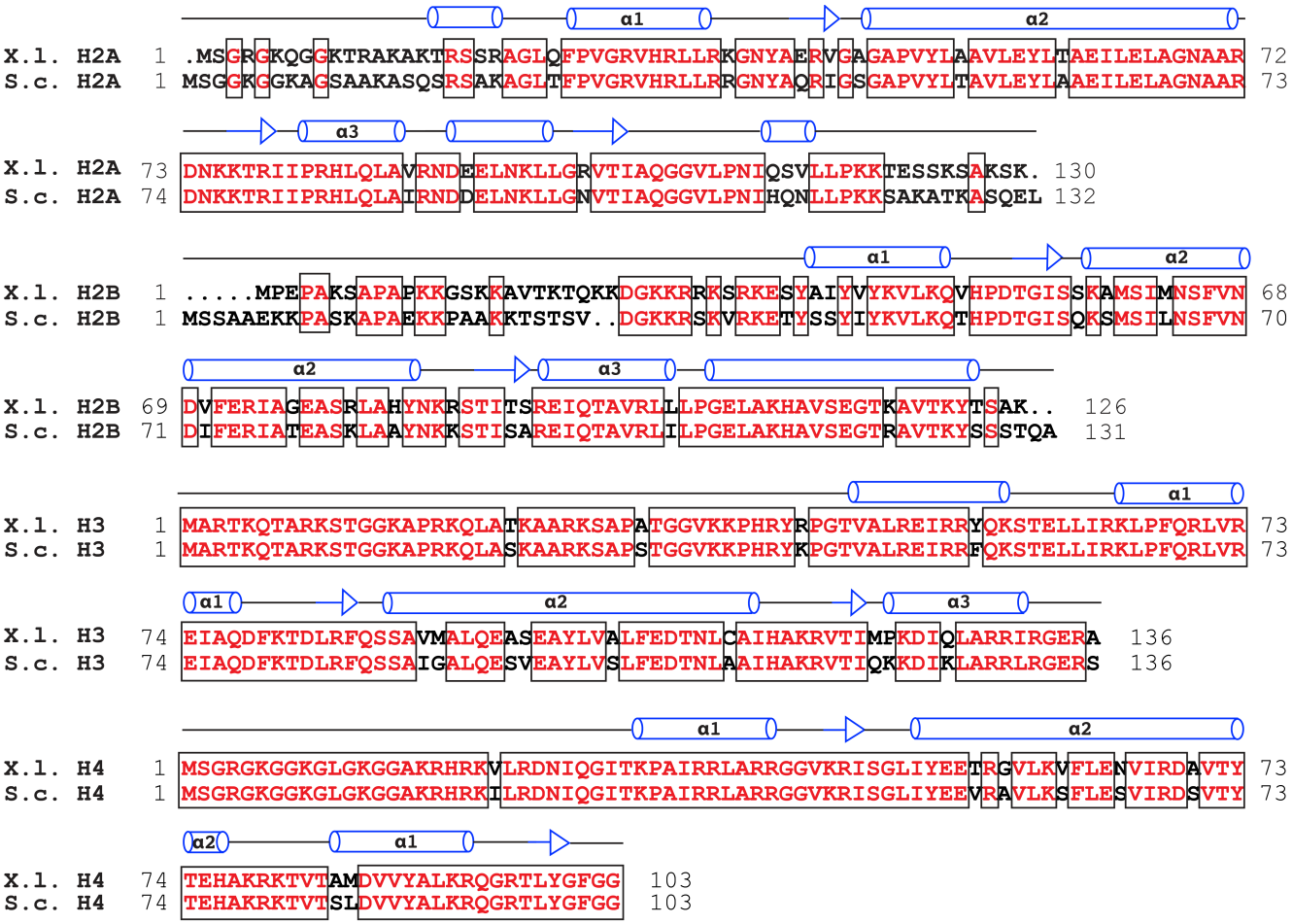
bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 3. Analysis of potential ACS motifs within λ DNA

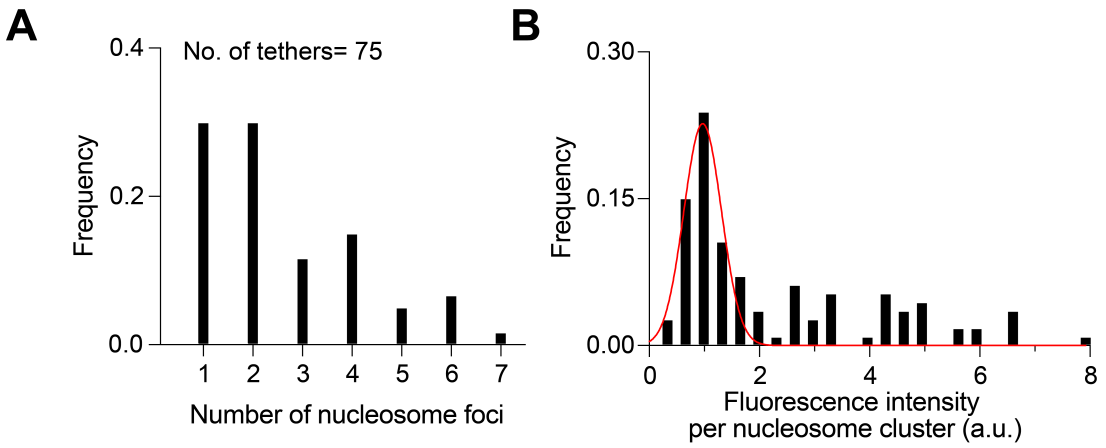
(**A**) Extended ACS (EACS) scores for the native λ genomic DNA using a 17-bp position weight matrix (PWM) generated from multiple sequence alignment of functional ARS elements identified in *S. cerevisiae* and previously used to score potential ORC binding sites in a given DNA sequence [62,63]. (**B**) EACS scores for the engineered λ_{ARS1} DNA. The highest scores correspond to the ARS1 elements (black arrowheads) inserted in the λ template.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 4. Sequence alignment of *Xenopus laevis* and *Saccharomyces cerevisiae* histones. Identical residues are colored in red.

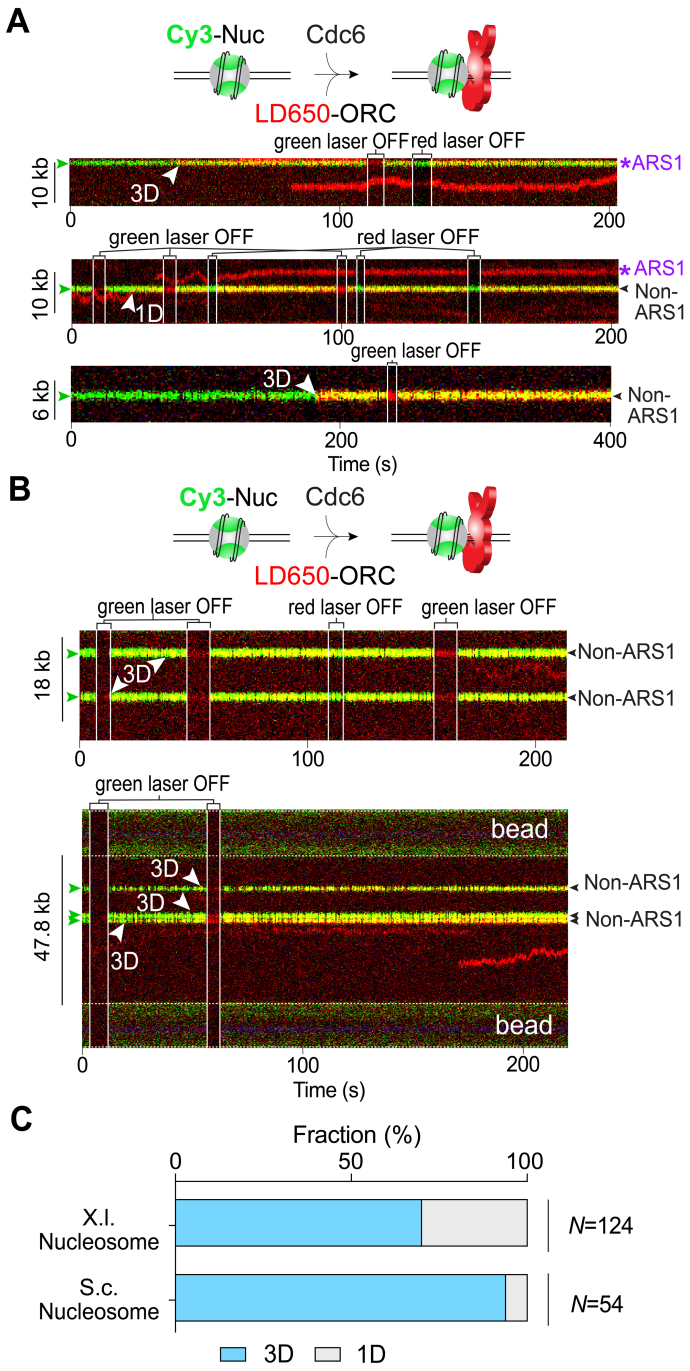
bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 5. Evaluation of nucleosome loading on tethered DNA

(**A**) Distribution of the number of fluorescently labeled nucleosome foci per DNA tether. (**B**) Distribution of the fluorescence intensity per nucleosome cluster. The main peak (red curve) represents the mononucleosome population.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



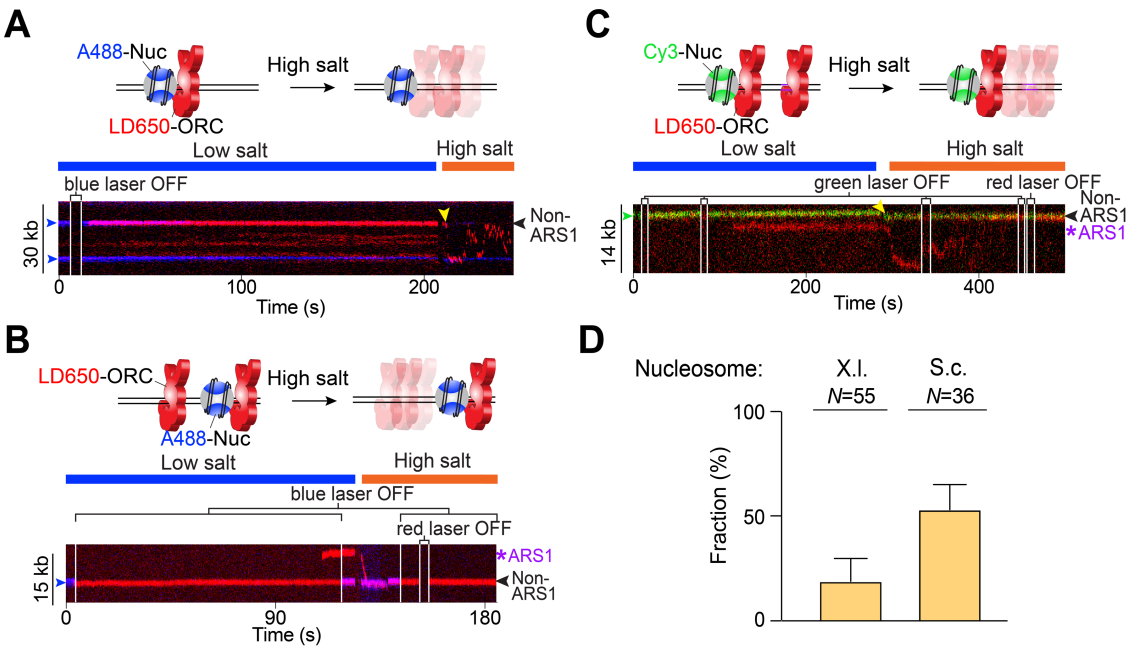
Supplementary Figure 6. Analysis of the search mode used by ORC to target nucleosomes

(**A**) Three example kymographs showing LD650-labeled ORC (red) stably associated with Cy3-labeled X.l. nucleosomes (green). White arrowheads indicate the time when ORC arrived at the nucleosomal site (ARS1 or non-ARS1) via three-dimensional (3D) or one-dimensional (1D) search. Green arrowheads indicate the positions of nucleosomes. (**B**) Two example kymographs showing LD650-labeled ORC (red) binding to Cy3-labeled S.c. nucleosomes (green). (**C**) Fraction of ORC molecules that were observed to

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

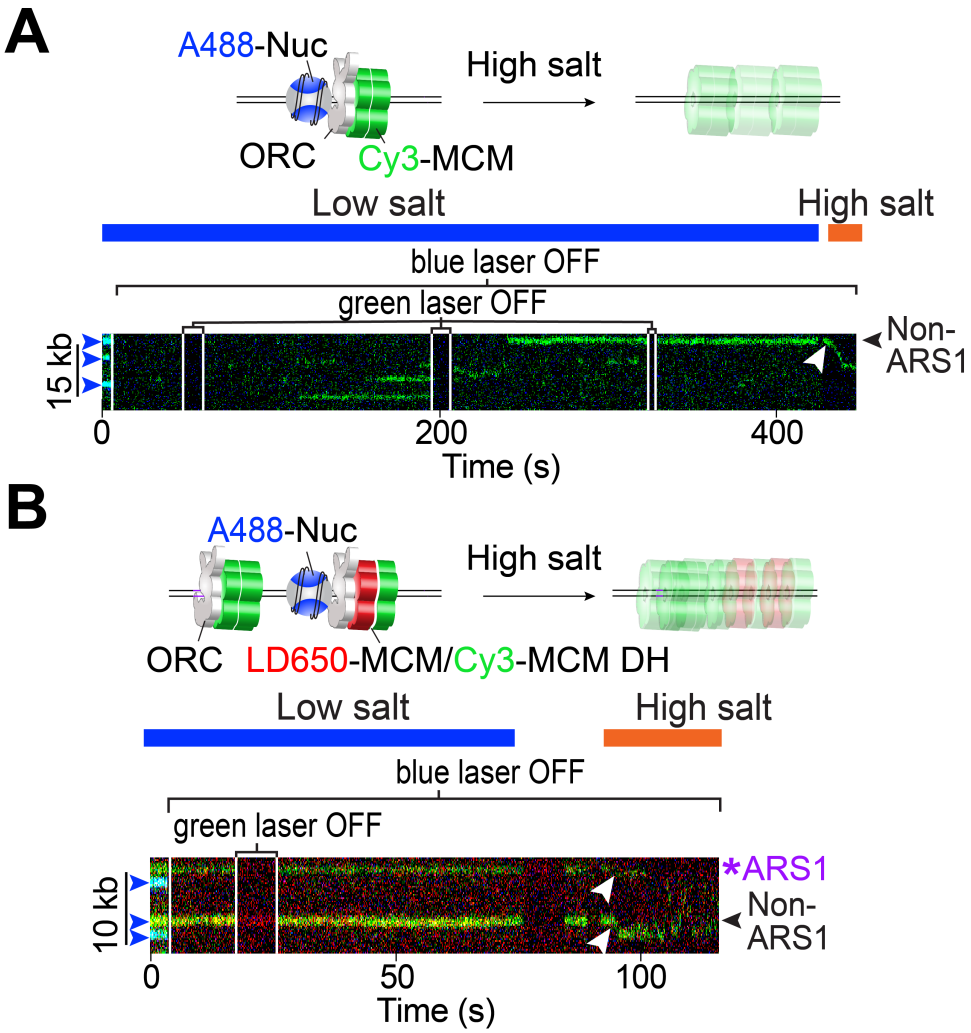
907 target a nucleosome (X.l. or S.c.) by 3D or 1D search. *N* indicates the number of events analyzed for each
908 condition.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 7. The fraction of ORC that remained associated with nucleosomes at high salt (**A**) Cartoon and an example kymograph showing ORC (red) colocalized with a S.c. nucleosome (blue) in a low-salt buffer but departed from the nucleosome and underwent diffusive motions on DNA upon high-salt wash (yellow arrowhead). (**B**) Cartoon and an example kymograph showing an ORC molecule (red) colocalized with a S.c. nucleosome (blue) and remained bound to the nucleosome at high salt. Another ORC bound at the ARS1 DNA position underwent diffusion upon high-salt wash. (**C**) Cartoon and an example kymograph showing an ORC molecule (red) colocalized with a Cy3-labeled X.l. nucleosome (green) and remained bound to it at high salt. Another ORC bound at the ARS1 DNA position underwent diffusion upon high-salt wash (yellow arrowhead). (**D**) Fraction of ORC molecules that were observed to depart from a nucleosomal site (X.l. vs. S.c.) upon high-salt wash. *N* indicates the number of ORC molecules analyzed for each condition. In all experiments, ORC was first incubated with tethered nucleosomal DNA at low salt in the presence of Cdc6; then the whole assembly was moved to another channel containing high salt.

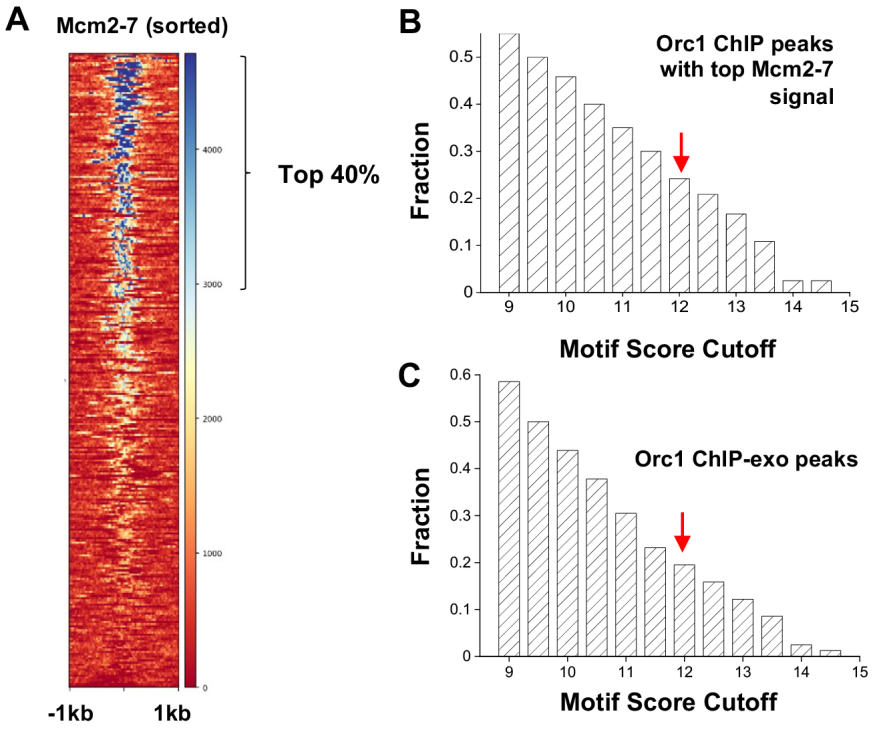
bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 8. Other examples of nucleosome-bound MCM undergoing diffusive motions upon high-salt wash

(**A**) Cartoon and kymograph showing MCM (green) colocalized with a S.c. nucleosome (blue) in the presence of unlabeled ORC, Cdc6 and Cdt1. Upon moving to a high-salt buffer, the MCM diffused away from the nucleosome (white arrowhead) as expected for a loaded MCM DH that encircles DNA. Blue arrowheads indicate the positions of nucleosomes. (**B**) Cartoon and kymograph showing two MCM complexes—one formed on the ARS1 DNA (top) and the other at a non-ARS1 nucleosome (bottom)—both of which displayed diffusive behavior upon high-salt wash (white arrowheads).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456647; this version posted March 26, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 9. Additional genome-wide analysis of ORC/MCM localization
(**A**) Heatmap of sorted Mcm2-7 ChIP signals among the 295 Orc1 ChIP peaks. The top 40% of these sites are included in the analysis in **B**. (**B**) Fraction of sequences that contain motifs above a certain threshold (varying from 9 to 15). These sequences are extracted from the Orc1 peaks with high Mcm2-7 ChIP signals. The red arrow represents the recommended cutoff of 11.9. (**C**) Same as in **B** except that the sequences are extracted from Orc1 ChIP-exo peaks (*N* = 81).