

# Origin Replication Complex Binding, Nucleosome Depletion Patterns, and a Primary Sequence Motif Can Predict Origins of Replication in a Genome with Epigenetic Centromeres

Hung-Ji Tsai,<sup>a</sup> Joshua A. Baller,<sup>a\*</sup> Ivan Liachko,<sup>b</sup> Amnon Koren,<sup>a\*</sup> Laura S. Burrack,<sup>a\*</sup> Meleah A. Hickman,<sup>a\*</sup> Mathuravani A. Thevandavakkam,<sup>c</sup> Laura N. Rusche,<sup>d</sup> Judith Berman<sup>a,c</sup>

Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, Minnesota, USA<sup>a</sup>; Department of Genome Science, University of Washington, Seattle, Washington, USA<sup>b</sup>; Department of Molecular Microbiology and Biotechnology, George Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Israel<sup>c</sup>; Department of Biological Sciences, State University of New York, University at Buffalo, New York, USA<sup>d</sup>

\* Present address: Joshua A. Baller, Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota, USA; Amnon Koren, Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA, and Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; Laura S. Burrack, Department of Biology, Grinnell College, Grinnell, Iowa, USA; Meleah A. Hickman, Department of Biology, Emory University, Atlanta, Georgia, USA.

**ABSTRACT** Origins of DNA replication are key genetic elements, yet their identification remains elusive in most organisms. In previous work, we found that centromeres contain origins of replication (ORIs) that are determined epigenetically in the pathogenic yeast *Candida albicans*. In this study, we used origin recognition complex (ORC) binding and nucleosome occupancy patterns in *Saccharomyces cerevisiae* and *Kluyveromyces lactis* to train a machine learning algorithm to predict the position of active arm (noncentromeric) origins in the *C. albicans* genome. The model identified bona fide active origins as determined by the presence of replication intermediates on nondenaturing two-dimensional (2D) gels. Importantly, these origins function at their native chromosomal loci and also as autonomously replicating sequences (ARSs) on a linear plasmid. A “mini-ARS screen” identified at least one and often two ARS regions of  $\geq 100$  bp within each bona fide origin. Furthermore, a 15-bp AC-rich consensus motif was associated with the predicted origins and conferred autonomous replicating activity to the mini-ARSs. Thus, while centromeres and the origins associated with them are epigenetic, arm origins are dependent upon critical DNA features, such as a binding site for ORC and a propensity for nucleosome exclusion.

**IMPORTANCE** DNA replication machinery is highly conserved, yet the definition of exactly what specifies a replication origin differs in different species. Here, we utilized computational genomics to predict origin locations in *Candida albicans* by combining locations of binding sites for the conserved origin replication complex, necessary for replication initiation, together with chromatin organization patterns. We identified predicted sequences that exhibited bona fide origin function and developed a linear plasmid assay to delimit the DNA fragments necessary for origin function. Additionally, we found that a short AC-rich motif, which is enriched in predicted origins, is required for origin function. Thus, we demonstrated a new machine learning paradigm for identification of potential origins from a genome with no prior information. Furthermore, this work suggests that *C. albicans* has two different types of origins: “hard-wired” arm origins that rely upon specific sequence motifs and “epigenetic” centromeric origins that are recruited to kinetochores in a sequence-independent manner.

Received 4 August 2014 Accepted 5 August 2014 Published 2 September 2014

**Citation** Tsai H-J, Baller JA, Liachko I, Koren A, Burrack LS, Hickman MA, Thevandavakkam MA, Rusche LN, Berman J. 2014. Origin replication complex binding, nucleosome depletion patterns, and a primary sequence motif can predict origins of replication in a genome with epigenetic centromeres. *mBio* 5(5):e01703-14. doi:10.1128/mBio.01703-14.

**Editor** Arturo Casadevall, Albert Einstein College of Medicine

**Copyright** © 2014 Tsai et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Judith Berman, [jberman@umn.edu](mailto:jberman@umn.edu) or [jberman@post.tau.ac.il](mailto:jberman@post.tau.ac.il).

This article is a direct contribution from a Fellow of the American Academy of Microbiology.

Proper inheritance of genetic information from each cell to its progeny requires faithful DNA replication. In *Saccharomyces cerevisiae*, origins of replication (ORIs) are determined by specific DNA binding motifs that are necessary for origin firing (1). *S. cerevisiae* also has point centromeres that are specified by a DNA sequence necessary for kinetochore assembly on the centromere region. In contrast, in most other eukaryotes, centromeres are determined by epigenetic factors, and the precise genomic properties that specify origins of replication remain elusive.

The replication machinery is highly conserved among eukaryotes and includes a heterohexameric origin recognition complex (ORC) that binds DNA and recruits the prereplication complex to initiate DNA replication from replication origin sites (2, 3). ORC binding alone is not sufficient to specify origin activity, as ORC binds to dormant origins (those that are licensed but inactive) (4–6). In the fission yeasts, origins are associated with regions that exclude nucleosomes and that are associated with the ORC but do not have a specific consensus motif (7–10). In *Drosophila*

and mammals, the number of licensed origins is far greater than the number that fire in a given cell. Furthermore, while origins in *S. cerevisiae* are defined by a specific DNA sequence (11–14) and a GC-rich DNA sequence may be associated with human origins (15), it is not known whether specific DNA sequences are necessary or sufficient for origin function or how such DNA might direct origin firing (16–19).

In *Saccharomyces cerevisiae*, autonomously replicating sequences (ARSs) direct replication on plasmids and include the 11- to 17-bp ARS consensus sequence (ScACS) that is necessary but not sufficient for origin firing (1). A T-rich B-element also contributes to efficient ARS function (20–22). In related budding yeast species, genome-wide mapping revealed other origin motifs (23–26). In *Kluyveromyces lactis*, a 50-bp ARS consensus sequence (KlACS) is highly diverged from the ScACS. Thus, it appears that evolutionary constraints on ACSs are weak (23).

Another challenge in identifying active origins is that ARSs from plasmid screens do not necessarily function in the chromosomal context. However, in *S. cerevisiae*, ACS motifs that are tightly bound by the ORC are often associated with an asymmetric pattern of positioned nucleosomes that is not found at nonreplicative ACSs (27–29). A similar asymmetric nucleosome occupancy pattern is also seen at origins in *Lachancea waltii* (24), suggesting that nucleosomes positioned around a nucleosome-depleted region are a feature of origins in the chromosomal context. Nucleosome exclusion is also observed at *Schizosaccharomyces pombe* origins, due to either specific sequences that exclude nucleosomes or *trans*-acting factors (8). While nucleosome depletion patterns are associated with active origins in *S. cerevisiae* and in *Drosophila melanogaster* (27, 30), they have not been used previously to predict sequences that may be active origins.

The most efficient or earliest-firing origins on each *Candida albicans* chromosome are associated with the centromeres (31). Importantly, centromere function and origin activity are epigenetically inherited together: deletion of a centromere is accompanied by a major shift from early/efficient to late/inefficient replication timing/efficiency and the appearance of new and highly efficient origins at neocentromeres (31). The limited resolution of replication timing profiles (~20 kb) precluded precise localization of arm origin positions. Early studies in *C. albicans* using circular plasmids identified putative ARSs (CaARSs) based on transformation efficiency (32–34), but origin firing was never demonstrated for these CaARSs. Furthermore, some CaARSs were chimeric fragments derived from two different genomic loci (34). These observations, together with those from many other eukaryotes, suggested that *C. albicans* origins might not be associated with a specific DNA motif.

Here, we identified chromosome arm origins in the *C. albicans* genome. We first used a machine learning approach to predict origins (proposed ORIs) based on the distribution of ORC binding sites and nucleosome occupancy. While strong ORC sites alone were not necessarily origins, those ORC sites associated with specific nucleosome depletion patterns were more likely to include bona fide origins, as determined using two-dimensional (2D) gel analysis of replication intermediates. Furthermore, these bona fide origins conferred increased transformation efficiency to linear plasmids, allowing us to narrow down the sequence regions required for origin function. Taken together, these results predicted origins across the *C. albicans* genome, identified minimal DNA fragments that were capable of directing autonomous repli-

cation on a linear plasmid, and identified a motif common to the predicted origins. Thus, specific genomic DNA regions from chromosome arms can direct replication on chromosome arms and coexist in the same genome with epigenetically inherited centromeric origins of replication.

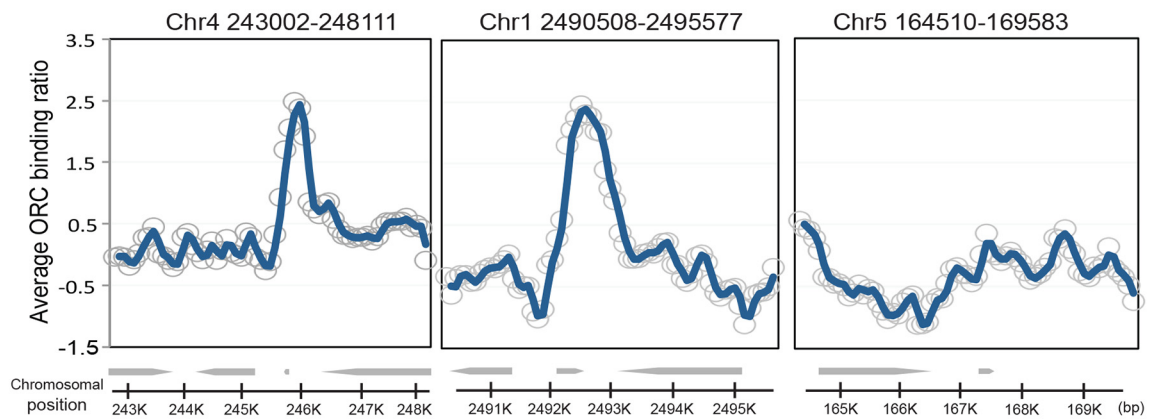
## RESULTS

**Using ORC binding sites and conserved nucleosome patterns to predict origin locations.** To identify regions of the genome that are potential replication origins in *C. albicans*, we mapped ORC binding sites using chromatin immunoprecipitation (ChIP) with a polyclonal antibody that recognizes the ORC complex, followed by whole-genome microarray hybridization (ChIP-chip) (Fig. 1A; see Fig. S1A in the supplemental material) on a microarray with 60-bp probes every 80 bp tiled across the genome. Consistent with previous work, all eight *C. albicans* centromeres exhibited strong ORC binding affinity (see Fig. S1A). The large number of local maxima in the ORC binding data is similar to data reported for *S. cerevisiae* ORC binding (5). We tested eight different fragments that included strong ORC binding peaks (see Table S2 and Fig. S1B in the supplemental material) using 2D gels, and we were unable to detect replication intermediates for any of these fragments. This suggests that ORC binding sites may not be sufficient to identify efficient origins of replication in the *C. albicans* genome.

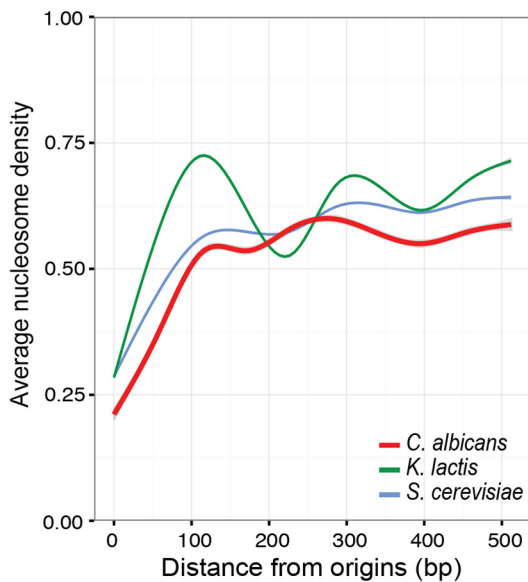
From *S. cerevisiae* to metazoans, a specific nucleosome pattern is associated with origin function (27, 29, 35). We first asked if this is the case in *K. lactis* as well, given that separate studies had determined ARS positions and nucleosome positions for this organism (23, 36, 37). A nucleosome-depleted region flanked by positioned nucleosomes was evident in *K. lactis* origins as well as in *S. cerevisiae* origins (Fig. 1B). We then used *K. lactis* ORC binding data together with nucleosome occupancy information (37) to train a logistic regression model to discriminate between known origins from *K. lactis* (23) and other, randomly chosen, sites. To determine the predictive strength of the model and an appropriate cutoff, we used a cross-validation approach in which one *K. lactis* chromosome was held out for testing, while the others were used for training the model. This process was repeated for each chromosome, and the prediction performance was evaluated based on predictions for the held-out data. This determination also allowed for the selection of an appropriate cutoff (Materials and Methods). We validated the capacity for cross-species prediction by predicting origins in *S. cerevisiae* with similar power. The model retained features corresponding to low nucleosome occupancy in the central region and high nucleosome occupancy to either side, along with higher than average ORC occupancy closer to the edges of the 1,024-bp window. Based on the assumption that similar chromatin context had been seen in *Schizosaccharomyces* species (7–10) and in *Drosophila* (27, 30), we then applied the model to the ORC binding data and nucleosome occupancy information (37) in *C. albicans*. This yielded 386 proposed origins of replication (proORIs) on chromosome arms throughout the genome as well as the previously identified origins at all eight centromeres (see Fig. S1 and Table S1 in the supplemental material).

The number of proORIs per chromosome was roughly proportional to chromosomal length (see Fig. S2A in the supplemental material), and proORIs were biased toward intergenic regions (see Fig. S2B). In general, proORIs were associated with early replicating regions ( $P < 6.5e-17$ ) (Fig. 1C); 30% (115/386) of the

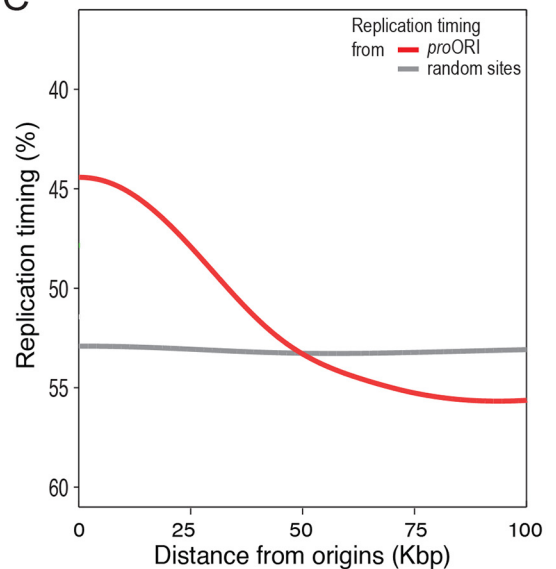
A



B



C

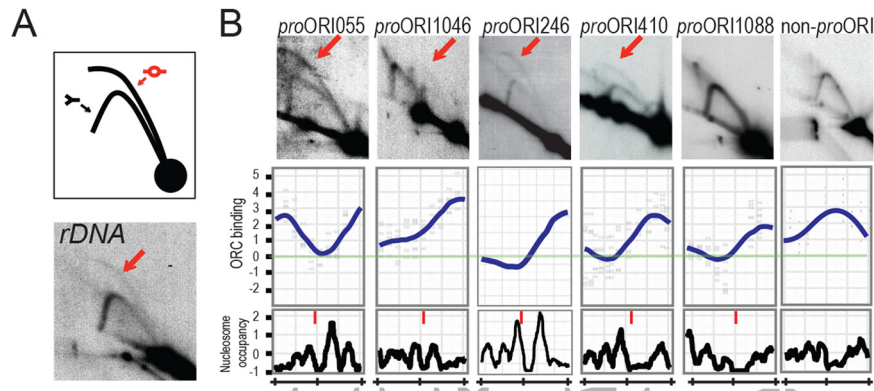


**FIG 1** Combined ORC binding profile and a conserved nucleosome depletion pattern at origins to predict origin positions in *C. albicans*. (A) ORC binding profile (blue) at two selected strong ORC sites (left and middle panels) from ChIP-chip experiments and a region that does not bind ORC (right panel). Each gray circle represents the average ORC binding ratio from seven experiments. The corresponding chromosomal coordinates and ORFs (gray bars) are shown at the bottom of each panel. (B) Average nucleosome density ( $y$  axis) as a function of distance from origins of replication ( $x$  axis) in *S. cerevisiae* (blue) and *K. lactis* (green). Nucleosome occupancy at the positions of proposed origins (proORIs) in *C. albicans* (red) was predicted based on a model trained on the *K. lactis* nucleosome occupancy pattern combined with ORC binding sites. (C) Average replication timing ( $y$  axis, % of S phase) of all proORIs (red) is earlier/more efficient than the replication timing of 500 random genomic loci ( $P < 6.5e-17$ , Kolmogorov-Smirnov test). Values between 0 and 50 indicate replication during the first half of S phase.

proORIs were found within 20 kb of a replication timing peak, consistent with the resolution of the timing profile data. A very strong association between tRNA loci and proORIs was also detected (see Fig. S2C), a trend seen to a lesser degree in other yeasts (6, 23, 38).

Two-dimensional (2D) nondenaturing gels directly detect the bidirectional replication fork of actively firing origins as “bubble arcs” (39) and are usually performed on synchronous cultures of cells. In *C. albicans*, full synchrony is difficult to achieve (40), yet a bubble arc was evident for the *CEN5* origin (31) and for an anticipated origin within *C. albicans* ribosomal DNA (rDNA) (Fig. 2A), both detected in asynchronous log-phase cultures. Furthermore,

four of the eight proORIs tested in the same manner produced bubble arcs (Fig. 2B; see Table S2 in the supplemental material), indicating that they fire efficiently in the genome context. We cannot distinguish between the possibility that the other four proORIs are inefficient, inactive (dormant) origins or that they are simply not origins. Therefore, we use the term “bona fide ORIs” for those origins that clearly had bubble arc replication intermediates on 2D gels. In contrast, no bubble arcs were evident from ORC sites that were not proORIs (e.g., Fig. 2B, right), presumably because they did not have well-positioned nucleosomes. The identification of four bubble arcs from eight proposed ORIs relative to zero bubble arcs from eight non-proORI ORC sites is significant



**FIG 2** Detection of bubble arc replication intermediates within some proORIs. (A) DNA 2D gels detect the migration pattern of replication intermediates. A bubble arc (red arrow) is indicative of a DNA fragment containing an active replication initiation site (upper panel), while a Y arc (black arrow) is indicative of passive replication from an origin outside the fragment being analyzed. The rDNA repeats include a conserved active origin that fires within a subset of the repeats to generate a bubble arc (bottom panel, red arrow). (B) Four proORIs formed bubble arcs (red arrows) and thus are bona fide origins of replication (proORI055, ChrR 1787352; proORI1046, Chr1 1168803; proORI246, Chr2 1619861; proORI410, Chr4 245520), while four (e.g., proORI1088) did not. In addition, eight ORC binding fragments that are not proORIs did not form bubble arcs (see Table S2 in the supplemental material [e.g., in the far-right panel, the fragment is centered on ChrR 1881144]). ORC binding (middle panels) and nucleosome occupancy patterns (lower panels) are shown for each tested genomic locus (with open reading frames [ORFs] in the window shown at the bottom corresponding to DNA 2D gels on the top) shown across the 1,024-bp window of proORIs. The y axis is the ratio of ORC binding (top) and nucleosome occupancy (bottom). ORC binding scores for individual probes on the microarrays are shown as gray dots, and smoothed curves were calculated across the entire chromosome. ORC binding scores were calculated across the entire window and are reported as the central proORI nucleotide positions, as labeled with red bars.

(see Table S2 [ $P < 0.04$ , Fisher's exact test]). Importantly, while the four bona fide ORIs all included a strong ORC binding score (calculated across the 1,024-bp window), strong ORC binding sites that were not proORIs did not necessarily form bubble arcs (Fig. 2B, right). Thus, the nucleosome positioning information learned from *K. lactis* ARSs significantly increased the likelihood of identifying those *C. albicans* ORC sites that fired efficiently.

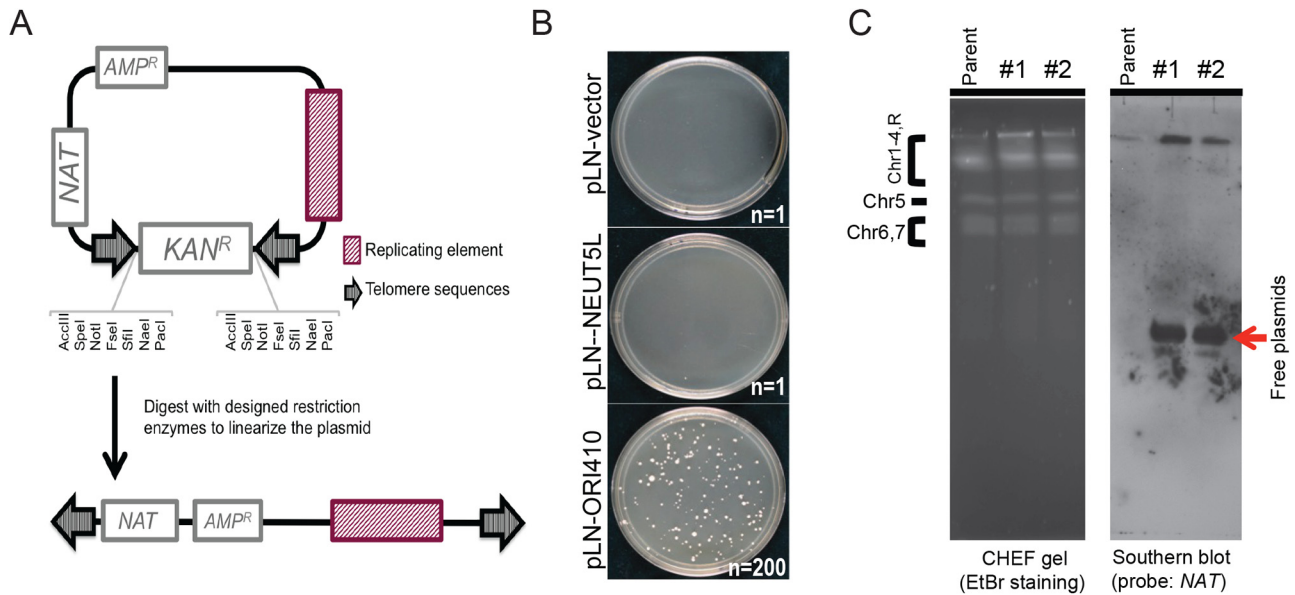
**Bona fide origins (ORIs) can direct the autonomous replication of linear plasmids.** *C. albicans* circular plasmids integrate into the genome (34, 41), which limits their usefulness for detecting active ARSs. This is especially the case when circular plasmids carry  $>1$  kb of DNA with high homology to genomic sequences (see Fig. S3A in the supplemental material) (H. J. Tsai, unpublished data). We postulated that linear plasmid vectors may solve this problem, as integration events would result in chromosome truncation, and cells with truncated chromosomes would be expected to grow poorly or to die. We tested this idea using a linearizable vector (pLN) with inverted telomere repeats separated by a “spacer” cassette (42) that, when excised, yields a linear plasmid capped with telomere sequences (Fig. 3A). Control plasmids with no insert (pLN-empty) or with a non-ORC/proORI intergenic region (pLN-NEUT5L) yielded few transformants, which we assume arose via integration. Transformation with pLN-ORI410, carrying a bona fide origin, resulted in significantly higher transformation frequency ( $\sim 50$  transformants/ $\mu\text{g}$  of DNA) (Fig. 3B). Contour-clamped homogeneous electric field (CHEF) gel separation of undigested chromosomes of several independent pLN-ORI410 transformants detected a band that hybridized to *NAT1*, the selectable marker on the pLN-ORI410 (Fig. 3C; see Fig. S3B). We interpret this as maintenance of an autonomous plasmid, as the band was far smaller than the undigested chromosomes. Together, these results demonstrate that DNA that functions as a bona fide origin in its chromosomal context also confers the ability to maintain an autonomous plasmid. This implies that the primary DNA sequence of the bona fide origins is sufficient to be

assembled into a functional prereplication complex and to initiate replication on a plasmid.

**Delineation of minimal functional ARS regions.** To ask if ARS function within *C. albicans* ORIs (1 to 3 kb) can be delimited to a smaller functional region, we next performed a minimal ARS (mini-ARS) screen (Fig. 4A) (43) by amplifying 1- to 3-kb regions of bona fide origin DNA (ORI055, ORI1046, and ORI410), digesting the DNA into small fragments and subsequently inserting the fragments into the pLN vector to produce “mini-ARS libraries” from each bona fide origin (43). Individual plasmid clones carrying a “mini-ARS” from *Escherichia coli* libraries were isolated and screened for high transformation efficiency in *C. albicans*.

For ORI410 and ORI055, insert sequences recovered from selected transformants ranged in size from 65 to  $\sim 200$  bp, and multiple nonoverlapping fragments conferred high transformation efficiency (Fig. 4B). This is reminiscent of the situation in *S. cerevisiae*, where multiple ARS consensus sequences are often found within efficient origins (44, 45). Additionally, while genes encoding tRNAs were frequently found in proORIs (see Fig. S2C in the supplemental material), the tRNA near ORI410 was not included within the ORI410 minimal ARS fragment. Thus, minimal ARS function does not require the presence of adjacent tRNAs.

**proORIs share a primary sequence motif.** With a large collection of proORIs in hand, we used MEME SUITE (46) to look for DNA consensus motifs. Interestingly, an AC-rich 15-bp motif was identified within proORI sequences (E value,  $1.3\text{e}-132$ ) (Fig. 5A, top; see Table S1B in the supplemental material). Furthermore, this proORI motif can be detected within the overlapping mini-ARS fragments (Fig. 4B), supporting the idea that the consensus motif may confer ARS function. Moreover, a similar motif is also seen within strong ORC sites (see Fig. S4 in the supplemental material [E value,  $1.6\text{e}-154$ ]), suggesting that this motif may recruit ORC, at least in the context of the genome. Consistent with this, the 97-bp mini-ARS fragment derived from ORI410 directed



**FIG 3** Origin DNA confers ARS activity to a linear plasmid. (A) Linearization of pLN plasmids. pLN plasmids include a nourseothricin resistance marker (NAT1) in *C. albicans*, an ampicillin resistance marker ( $AMP^R$ ) in *E. coli*, and a pair of inverted *C. albicans* telomere repeats separated by a kanamycin resistance marker ( $KAN^R$ ). Potential replicating elements were inserted (red hatched box). After the excision of the  $KAN^R$  gene (lower half), the plasmid was linearized with telomere sequences exposed at the termini. (B) Transformation of pLN plasmids into *C. albicans*. pLN-ORI410 yielded high yeast transformation efficiency (bottom plate, ~200 transformants), while the pLN vector (pLN, no insert) and pLN-NEUT5L (DNA from a locus without ORC binding, intergenic region between *orf19.1963* and *orf19.1961*) did not (top and middle plates, respectively; 1 transformant). (C) CHEF gel analysis of transformants detects autonomous plasmid DNA. Two independent *C. albicans* transformants with pLN-ORI410 (no. 1 and 2) and the parental strain (parent) were subjected to CHEF gel separation using conditions that maintain intact chromosomes in the top portion of the gel (left), subjected to Southern blotting, and analyzed with a plasmid-specific probe (NAT1 gene) (right). pLN-ORI410 is detected as a band (red arrow) with mobility much faster than that of chromosomal DNA, indicating that it is maintained as an autonomous plasmid.

ORC binding when inserted in the *NEUT5L* non-origin region of the genome (H. J. Tsai, unpublished preliminary result).

In addition to the AC-rich motif, a 21-bp-long T-rich motif resides within the proORIs (Fig. 5A, bottom). The T-rich sequences usually aligned with the centers of the proORI positions, overlapping the central nucleosome exclusion region surrounded by phased nucleosome patterns that was a prominent feature of the model that defined proORIs (Fig. 5B). This is consistent with the observation that T-rich sequences generally provide strong nucleosome exclusion signals (47).

To test the hypothesis that the conserved T-rich motif in proORI sequences may serve as an auxiliary element for origin activity, analogous to the B-elements in *S. cerevisiae* origins (20), we analyzed the relative positions of the AC-rich and the T-rich sequences in proORIs. Intriguingly, the AC-rich proORI motifs were usually located upstream relative to the position of the proORIs (Fig. 5B, blue, upper left panel) and the T-rich sequences (putative B-elements) were located approximately 100 bp downstream of the proORI motif (Fig. 5B, green, bottom right panel). This A:T asymmetry at *C. albicans* origins may be a feature of origins conserved between *C. albicans* and *S. cerevisiae* (27).

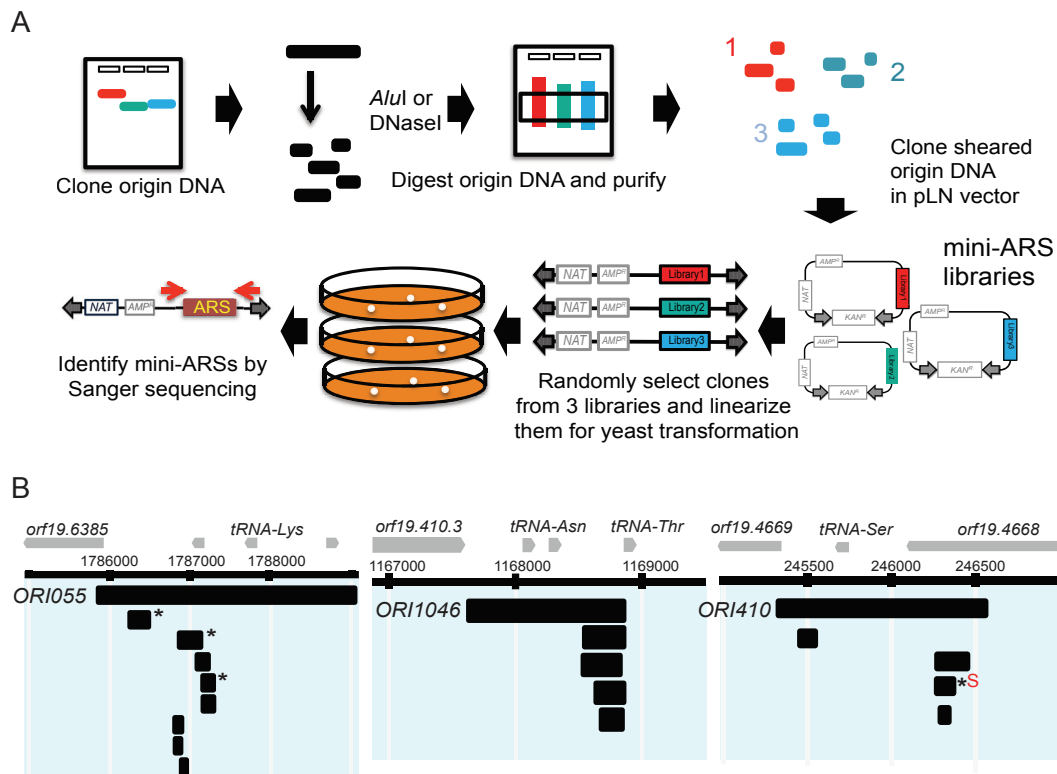
**Mutation of the 15-bp consensus motif disrupts ARS function.** To ask if the proORI motif is directly associated with origin activity, we mutated the consensus sequences, changing the two ACC sequences to GGG and TTT, respectively (Fig. 5C, left panel). Plasmids carrying the mutated motif exhibited a reduced transformation frequency relative to that of plasmids with the intact motif (Fig. 5C, right panel). This supports the idea that the proORI motif is important for ARS activity. Linear plasmids with

the mutated ARS or without an ARS insert yielded few transformants, and these spurious transformants had integrated into the genome, as determined by CHEF gel analysis (data not shown). Together, these results support the idea that the proORI motif in a bona fide origin is required for autonomous plasmid replication. Whether ARSs that are not bona fide origins in the genome context behave similarly remains to be tested.

## DISCUSSION

In this study, we predicted the position of *C. albicans* replication origins using a machine learning logistic regression approach based on ORC binding ChIP-chip data together with the nucleosome depletion patterns. We identified several bona fide origins that efficiently produce replication intermediates and that function on linear plasmids. We then identified ~100-bp mini-ARS fragments that were sufficient to direct plasmid replication. Importantly, this is the first example in which origin locations were predicted based on a conserved nucleosome depletion pattern together with ORC binding at known origins and then confirmed in an organism for which there was no *a priori* knowledge of an associated primary DNA sequence or an ARS assay. In most eukaryotes, an understanding of the DNA requirements for replication origin function is limited; this sequence-independent strategy provides a paradigm for identifying origins in model or nonmodel organisms that have no known origin features and/or no available plasmid assays.

Of note, the proORI prediction model was trained by identifying ORC and nucleosome distributions associated with *K. lactis* origins but not with random sites in the *K. lactis* genome. Cross-



**FIG 4** Minimal-ARS (mini-ARS) screen to identify proORI DNA sequences with ARS function. (A) Minimal-ARS screen strategy. Bona fide origin DNA regions from proORI055 (3.2 kb), proORI1046 (1.2 kb), and proORI410 (1.2 kb) were linearized, subcloned into plasmid pLN, and then used to transform yeast strain RM1000. Subfragments of the bona fide origins that yielded high transformation efficiency were individually analyzed by Sanger sequencing. (B) Mini-ARS fragments identified from bona fide origins. Each black bar represents an isolated DNA fragment that yielded relatively high transformation efficiency. A 97-bp sub-ARS fragment from ORI410 (red "S" [right panel]) was used in Fig. 5.

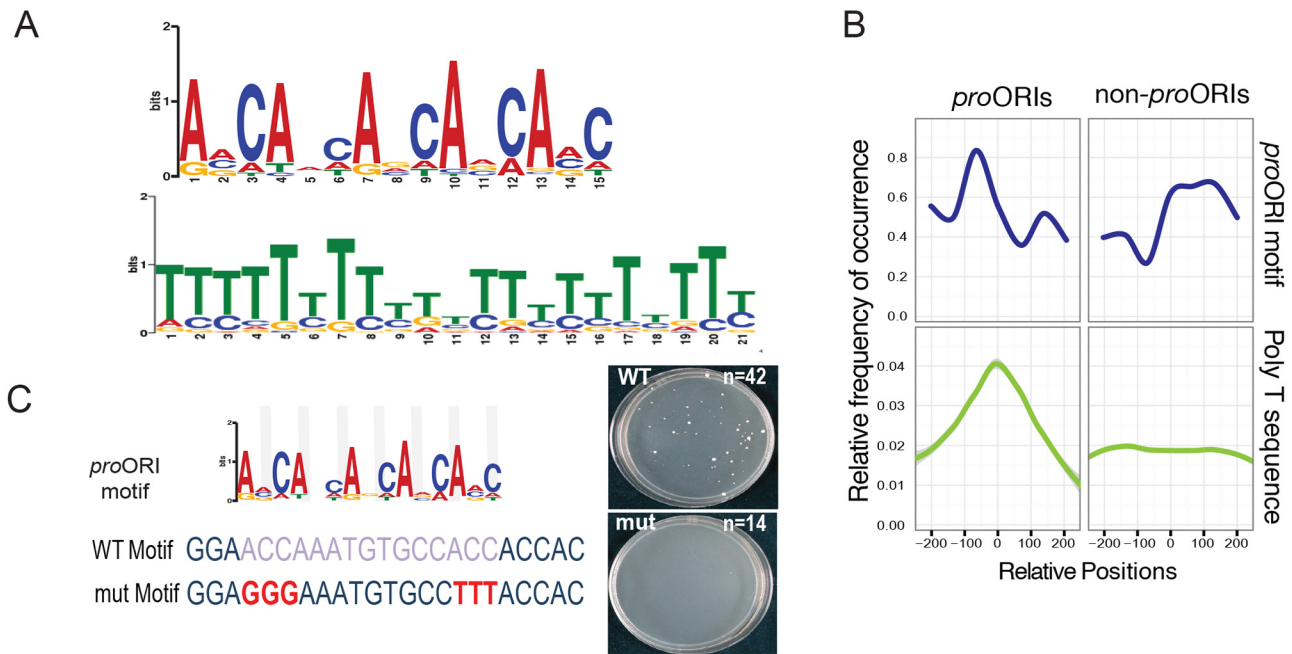
species predictive ability was validated on *S. cerevisiae* origins. The model coefficients indicate that the primary features allowing the model to discriminate between *K. lactis* origins and random *K. lactis* regions are a central nucleosome depletion trough flanked by phased nucleosome occupancy and more distal local ORC maxima. Four of eight proORIs tested were demonstrated to be bona fide origins of replication (Fig. 2B). Several genomic regions, including several fragments that bind ORC but lack the conserved nucleosome pattern, did not produce detectable bubble arcs. This is consistent with the idea that the conserved nucleosome pattern helps to distinguish regions that are origins in the chromosome context. The results suggest that, in the context of an appropriate nucleosome occupancy pattern, strong ORC binding is a common feature of bona fide origins.

In *Saccharomyces*, the origins flanking centromeres are sequence dependent and are activated earlier than many other origins on the chromosome arms (24, 48). In *C. albicans*, the epigenetic centromeres are associated with the earliest-firing origins and the inheritance of their function is independent of the DNA sequence (31). Importantly, when a centromere is deleted and a neocentromere forms at a new region of the chromosome where an origin was not found previously, the neocentromere region becomes the region firing earliest, and the region flanking the deleted centromere replicates far later (31). Furthermore, an evolutionary vestige of constitutive replication origin activity, a GC-skew pattern characteristic of constitutive origins in bacterial

chromosomes was evident at all eight *C. albicans* centromeres (31).

In contrast, here we found that the arm origins appear to be sequence dependent, requiring ORC binding sites, a nucleosome depletion domain, and a specific DNA binding motif. Furthermore, chromosome arm origins are not associated with a GC-skew pattern (<2.5% association,  $P = 0.78$ ), consistent with the idea that they do not fire as consistently at the centromeric origins. Together with the finding of a proORI consensus motif, which is necessary for ARS function, this provides additional support to the idea that *C. albicans* harbors at least two types of origins: centromere origins that are specified epigenetically and chromosome arm origins that are specified by primary sequence motifs. Moreover, the recent discovery of two distinct types of ACS motifs that specify ARS function in *Pichia pastoris* (49) supports the idea that some organisms may have more than one mechanism for specifying replication initiation. Future studies will need to explore how an organism utilizes two apparently separate mechanisms—one epigenetic and one sequence dependent—for the specification of replication origins.

As in higher eukaryotes and *S. pombe*, circular plasmids in *C. albicans* tend to integrate into the genome, irrespective of whether they carry replicating elements; linear plasmids carrying bona fide origins appear to have reduced this problem. Importantly, the *URA3* gene, which was used as a marker for the identification of ARSs in earlier studies, is a predicted proORI and has



**FIG 5** Consensus sequences from proORIs confer ARS function. (A) The 15-bp AC-rich and 21-bp T-rich sequence motifs enriched within proORI sequences. (B) Distribution between the proORI motif (blue) and poly(T) sequences (green) in proORIs (left) and non-proORIs (right panels). A poly(T) tract is found in the center with the proORI motifs, oriented by the Watson strand (AC-rich strand), localized to left of the poly(T) tract. (C) The 15-bp proORI wild-type (WT) motif (ORI410S [Fig. 4B]) and the derived insert carrying the illustrated mutated (mut) motif were transformed into yeast. The mutated plasmid exhibited reduced replication function, measured as transformation frequency.

weak ARS function on a circular plasmid (see Fig. S3A in the supplemental material) (H. J. Tsai, unpublished data). Thus, negative controls are critically important for studies using *URA3* to measure transformation frequencies. For example, in a recent report, origins flanking *CEN7* yielded a high level of transformants but without any negative controls/empty plasmid controls (50).

Several properties of origins are shared by different yeast species that have point centromeres (see Table S4 in the supplemental material). In contrast, origin properties are more enigmatic in most organisms that have regional centromeres. In particular, the sequence features, if any, that specify genomic origins remain to be elucidated. Because it has small regional, epigenetic centromeres, more akin to regional centromeres in other organisms with less-well-characterized origins, *C. albicans* is a useful model organism for studying genome organization. We propose that insights gained from studies of *C. albicans* centromeres have the potential to inform studies of centromeres of higher eukaryotes. Similarly, we suggest that the origin prediction algorithms used here for *C. albicans* have the potential to provide new insights into the elusive nature of replication origins in metazoans.

## MATERIALS AND METHODS

**Yeast strains and growth conditions.** The yeast strains used in this study were derived from *C. albicans* laboratory strain SC5314 (wild type, diploid). Yeast cultures were grown in yeast extract-peptone-dextrose plus adenine (YPAD) medium at 30°C. The lithium acetate transformation protocol (51) was used for the transformation of plasmids and PCR products into *C. albicans*. For *NAT1* marker selection, cells were recovered on nonselective YPAD medium for 6 h prior to replica plating to selective medium containing 400 µg/ml nourseothricin (52).

**ChIP-chip analysis.** ChIPs were performed as described previously (53, 54). Polyclonal antibodies against the *S. cerevisiae* ORC were kindly

provided by Stephen P. Bell. ChIP-chip (28), with seven replicates, was performed according to Agilent protocols, and ChIP DNA was labeled and hybridized to custom-designed Agilent microarrays (template number available upon request) containing 60-bp probes, one every 80 bp, spanning the genome. ORC binding positions were identified by LOESS smoothing of the data set (55), calling local maxima (no larger peaks in the 1,024-bp window) with strong peaks being local maxima at least 2 standard deviations (SD) above the mean of the population of local maxima (see Tables S3 and S5 in the supplemental material).

**proORI prediction from nucleosome-ORC association pattern analyses.** All statistical analyses and computational modeling were performed in the R environment using custom scripts and standard packages. proORIs were predicted using a logistic regression model trained on ORC binding as determined by ChIP microarrays (see Table S5 in the supplemental material) combined with published nucleosome occupancy data (37). Model training and prediction were implemented using the glmnet package of the R statistical software with a binomial kernel and least absolute shrinkage and selection operator (LASSO) regularization (56). For training, established origin regions in *K. lactis* served as a positive set, while random *K. lactis* regions served as a negative set with 6-fold cross-validation. The likelihood of proORI presence was evaluated at each position in the *C. albicans* genome, using the *K. lactis* trained model, based on a sliding 1,024-bp window. ORC and nucleosome data were used identically and concurrently when training the model. The model was given 2,048 total features per site on which to train (1,024 ORC values plus 1,024 nucleosome values). The LASSO regularization has the effect of removing from the model features with minimal predictive contribution. The reduced feature set provides a compact model and minimizes overfitting, thus increasing predictive power. Often multiple local maxima would be found in a close area in sync with arrayed nucleosomes. The maximum local peak was found by selecting the non-endpoint maximum based on a 1,024-bp window. Only those maxima over 2 SD were retained. This cut-off corresponded to a 33% precision over 40% recall in the *K. lactis* data set. For determination of precision recall, a cross-validation was used,

leaving out a single chromosome at a time in order to avoid fragmenting chromosomes. The predictive accuracy based on the area under the curve representing true-positive rate (AUC) for testing in *K. lactis* was 0.925. When tested in *S. cerevisiae*, the same model had an AUC of 0.905 (Fig. S5).

**Two-dimensional nondenaturing DNA gel electrophoresis.** Two-dimensional DNA gels were performed as previously described (39) following protocols provided on the Brewer laboratory website (<http://fangman-brewer.genetics.washington.edu/2Dgel.html>). *C. albicans* cells were harvested from asynchronous log-phase cultures, and replication intermediates were further isolated.

**Plasmid library constructions.** Linear plasmid vector pLN was constructed in a pGEM vector backbone. A designed linker containing multiple restriction enzyme sites was inserted between inverted *C. albicans* telomere sequences on both sides of the kanamycin resistance gene (*KAN<sup>R</sup>*) marker to facilitate linearization of the plasmid and excision of this marker (Fig. 3A).

For the mini-ARS screen (Fig. 4A), three bona fide ORI fragments (1.2 to 3.2 kb) were PCR amplified from genomic DNA, extracted from the agarose gels, digested with AluI, and treated with alkaline phosphatase (CIP) to prevent fragment multimerization. In parallel, purified PCR products were treated with DNase I and then alkaline phosphatase, and 0.1- to 1-kb fragments were purified from the agarose gels and ligated into PvuII-digested pLN. Competent *E. coli* cells (Agilent XL2-Blue ultracompetent cells) were transformed with the ligated library DNA, and individual *E. coli* colonies carrying pLN-mini-ARS plasmids were randomly selected and transformed into *C. albicans*. Individual mini-ARs were tested for transformation efficiency, and the most efficient ones were analyzed by Sanger sequencing.

**Motif finding and analyses.** We used MEME SUITE (<http://meme.nbr.net/meme>) version 4.9.0 to find sequence motifs (46). The position-specific priors (PSP) were generated from a negative set of 1,000 random sequences, selected uniformly from whole genome sequence. The PSPs were used as a background model for motif discovery.

**Microarray data accession number.** ChIP-chip results from CaORC and probe information have been uploaded to the NCBI GEO database under accession no. GSE54923.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01703-14/-/DCSupplemental>.

Figure S1, TIF file, 0.1 MB.  
Figure S2, TIF file, 0.1 MB.  
Figure S3, TIF file, 0.1 MB.  
Figure S4, TIF file, 0.1 MB.  
Figure S5, TIF file, 0.1 MB.  
Table S1, XLSX file, 0.1 MB.  
Table S2, XLSX file, 0.1 MB.  
Table S3, XLSX file, 0.1 MB.  
Table S4, XLSX file, 0.1 MB.  
Table S5, XLSX file, 0.01 MB.

## ACKNOWLEDGMENTS

We thank Maryam Gerami-Nejad and Wenqiang Chang for assistance with plasmid constructions, and Maitreya Dunham, Chad Myers, Catherine Fox, Martin Kupiec, Duncan Clarke, Andrew Lane, Man Shun Fu, and other members in the Berman laboratory for helpful discussion and comments on the manuscript. We thank Stephen P. Bell for sharing anti-ORC and anti-MCM antibodies and Bonny Brewer for advice on DNA 2D gel analysis.

This work was supported by NIH/NIAID AI R01075096, AI R010624273, and ISF 340/13 to J.B., a 2011 Williston postdoctoral fellowship and grant PF-12-108-01-CCG from the American Cancer Society to L.S.B., NRSA postdoctoral fellowship F32GM096536-02 to M.A.H., NIH GM073991 to L.N.R., and F32 GM090561 to I.L.

## REFERENCES

- Newlon CS, Theis JF. 1993. The structure and function of yeast ARS elements. *Curr. Opin. Genet. Dev.* 3:752–758. [http://dx.doi.org/10.1016/S0959-437X\(05\)80094-2](http://dx.doi.org/10.1016/S0959-437X(05)80094-2).
- Bell SP, Stillman B. 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* 357:128–134. <http://dx.doi.org/10.1038/357128a0>.
- Aladjem MI. 2007. Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat. Rev. Genet.* 8:588–600. <http://dx.doi.org/10.1038/nrg2143>.
- Sasaki T, Gilbert DM. 2007. The many faces of the origin recognition complex. *Curr. Opin. Cell Biol.* 19:337–343. <http://dx.doi.org/10.1016/j.ccb.2007.04.007>.
- Shor E, Warren CL, Tietjen J, Hou Z, Müller U, Alborelli I, Gohard FH, Yemm AI, Borisov L, Broach JR, Weinreich M, Nieduszynski CA, Ansari AZ, Fox CA. 2009. The origin recognition complex interacts with a subset of metabolic genes tightly linked to origins of replication. *PLOS Genet.* 5:e1000755. <http://dx.doi.org/10.1371/journal.pgen.1000755>.
- Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM. 2001. Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* 294:2357–2360. <http://dx.doi.org/10.1126/science.1066101>.
- Clyne RK, Kelly TJ. 1995. Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. *EMBO J.* 14:6348–6357.
- Xu J, Yanagisawa Y, Tsankov AM, Hart C, Aoki K, Kommajosyula N, Steinmann KE, Bochicchio J, Russ C, Regev A, Rando OJ, Nusbaum C, Niki H, Milos P, Weng Z, Rhind N. 2012. Genome-wide identification and characterization of replication origins by deep sequencing. *Genome Biol.* 13:R27. <http://dx.doi.org/10.1186/gb-2012-13-4-r27>.
- Lee JK, Moon KY, Jiang Y, Hurwitz J. 2001. The *Schizosaccharomyces pombe* origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc. Natl. Acad. Sci. U. S. A.* 98:13589–13594. <http://dx.doi.org/10.1073/pnas.251530398>.
- Dai J, Chuang RY, Kelly TJ. 2005. DNA replication origins in the *Schizosaccharomyces pombe* genome. *Proc. Natl. Acad. Sci. U. S. A.* 102:337–342. <http://dx.doi.org/10.1073/pnas.0408811102>.
- Stinchcomb DT, Struhl K, Davis RW. 1979. Isolation and characterization of a yeast chromosomal replicator. *Nature* 282:39–43. <http://dx.doi.org/10.1038/282039a0>.
- Chan CS, Tye BK. 1980. Autonomously replicating sequences in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 77:6329–6333. <http://dx.doi.org/10.1073/pnas.77.11.6329>.
- Theis JF, Newlon CS. 1997. The ARS309 chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional ARS consensus sequence. *Proc. Natl. Acad. Sci. U. S. A.* 94:10786–10791. <http://dx.doi.org/10.1073/pnas.94.20.10786>.
- Nieduszynski CA, Knox Y, Donaldson AD. 2006. Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev.* 20:1874–1879. <http://dx.doi.org/10.1101/gad.385306>.
- Besnard E, Babled A, Lapasset L, Milhavel O, Parrinello H, Dantec C, Marin JM, Lemaître JM. 2012. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* 19:837–844. <http://dx.doi.org/10.1038/nsmb.2339>.
- Cayrou C, Coulombe P, Vigneron A, Stanojčić S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R, Méchali M. 2011. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* 21:1438–1449. <http://dx.doi.org/10.1101/gr.121830.111>.
- Karnani N, Taylor CM, Malhotra A, Dutta A. 2010. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol. Biol. Cell* 21:393–404. <http://dx.doi.org/10.1091/mbc.E09-08-0707>.
- MacAlpine HK, Gordán R, Powell SK, Hartemink AJ, MacAlpine DM. 2010. *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res.* 20:201–211. <http://dx.doi.org/10.1101/gr.097873.109>.
- Gilbert DM. 2012. Replication origins run (ultra) deep. *Nat. Struct. Mol. Biol.* 19:740–742. <http://dx.doi.org/10.1038/nsmb.2352>.
- Marahrens Y, Stillman B. 1992. A yeast chromosomal origin of DNA



- replication defined by multiple functional elements. *Science* 255:817–823. <http://dx.doi.org/10.1126/science.1536007>.
21. Lee DG, Bell SP. 1997. Architecture of the yeast origin recognition complex bound to origins of DNA replication. *Mol. Cell. Biol.* 17:7159–7168.
  22. Rao H, Stillman B. 1995. The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. *Proc. Natl. Acad. Sci. U. S. A.* 92:2224–2228. <http://dx.doi.org/10.1073/pnas.92.6.2224>.
  23. Liachko I, Bhaskar A, Lee C, Chung SCC, Tye B-K, Keich U. 2010. A comprehensive genome-wide map of autonomously replicating sequences in a naive genome. *PLoS Genet.* 6:e1000946. <http://dx.doi.org/10.1371/journal.pgen.1000946>.
  24. Di Rienzi SC, Lindstrom KC, Mann T, Noble WS, Raghuraman MK, Brewer BJ. 2012. Maintaining replication origins in the face of genomic change. *Genome Res.* 22:1940–1952. <http://dx.doi.org/10.1101/gr.138248.112>.
  25. Theis JF, Yang C, Schaefer CB, Newlon CS. 1999. DNA sequence and functional analysis of homologous ARS elements of *Saccharomyces cerevisiae* and *S. carlsbergensis*. *Genetics* 152:943–952.
  26. Liachko I, Tanaka E, Cox K, Chung SC, Yang L, Seher A, Hallas L, Cha E, Kang G, Pace H, Barrow J, Inada M, Tye BK, Keich U. 2011. Novel features of ARS selection in budding yeast *Lachancea kluyveri*. *BMC Genomics* 12:633. <http://dx.doi.org/10.1186/1471-2164-12-633>.
  27. Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM. 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev.* 24:748–753. <http://dx.doi.org/10.1101/gad.1913210>.
  28. Lipford JR, Bell SP. 2001. Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol. Cell* 7:21–30. [http://dx.doi.org/10.1016/S1097-2765\(01\)00151-4](http://dx.doi.org/10.1016/S1097-2765(01)00151-4).
  29. Berbenetz NM, Nislow C, Brown GW. 2010. Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. *PLoS Genet.* 6:e1001092. <http://dx.doi.org/10.1371/journal.pgen.1001092>.
  30. Eaton ML, Prinz JA, MacAlpine HK, Tretyakov G, Kharchenko PV, MacAlpine DM. 2011. Chromatin signatures of the *Drosophila* replication program. *Genome Res.* 21:164–174. <http://dx.doi.org/10.1101/gr.116038.110>.
  31. Koren A, Tsai HJ, Tirosh I, Burrack LS, Barkai N, Berman J. 2010. Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase. *PLoS Genet.* 6:e1001068. <http://dx.doi.org/10.1371/journal.pgen.1001068>.
  32. Beckerman J, Chibana H, Turner J, Magee PT. 2001. Single-copy IMH3 allele is sufficient to confer resistance to mycophenolic acid in *Candida albicans* and to mediate transformation of clinical *Candida* species. *Infect. Immun.* 69:108–114. <http://dx.doi.org/10.1128/IAI.69.1.108-114.2001>.
  33. Cannon RD, Jenkinson HF, Shepherd MG. 1990. Isolation and nucleotide sequence of an autonomously replicating sequence (ARS) element functional in *Candida albicans* and *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 221:210–218.
  34. Herreros E, García-Sáez MI, Nombela C, Sánchez M. 1992. A reorganized *Candida albicans* DNA sequence promoting homologous non-integrative genetic transformation. *Mol. Microbiol.* 6:3567–3574.
  35. Lombrana R, Almeida R, Revuelta I, Madeira S, Herranz G, Saiz N, Bastolla U, Gómez M. 2013. High-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian replication origins. *EMBO J.* 32:2631–2644. <http://dx.doi.org/10.1038/emboj.2013.195>.
  36. Siow CC, Nieduszynska SR, Müller CA, Nieduszynski CA. 2012. OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.* 40:D682–D686. <http://dx.doi.org/10.1093/nar/gkr769>.
  37. Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. 2010. The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol.* 8:e1000414. <http://dx.doi.org/10.1371/journal.pbio.1000414>.
  38. Brewer BJ. 1994. Intergenic DNA and the sequence requirements for replication initiation in eukaryotes. *Curr. Opin. Genet. Dev.* 4:196–202.
  39. Brewer BJ, Fangman WL. 1987. The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell* 51:463–471. [http://dx.doi.org/10.1016/0092-8674\(87\)90642-8](http://dx.doi.org/10.1016/0092-8674(87)90642-8).
  40. Côte P, Hogues H, Whiteway M. 2009. Transcriptional analysis of the *Candida albicans* cell cycle. *Mol. Biol. Cell* 20:3363–3373. <http://dx.doi.org/10.1091/mbc.E09-03-0210>.
  41. Kurtz MB, Cortelyou MW, Miller SM, Lai M, Kirsch DR. 1987. Development of autonomously replicating plasmids for *Candida albicans*. *Mol. Cell. Biol.* 7:209–217.
  42. Burke DT, Carle GF, Olson MV. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236:806–812. <http://dx.doi.org/10.1126/science.3033825>.
  43. Liachko I, Youngblood RA, Keich U, Dunham MJ. 2013. High-resolution mapping, characterization, and optimization of autonomously replicating sequences in yeast. *Genome Res.* 23:698–704. <http://dx.doi.org/10.1101/gr.144659.112>.
  44. Theis JF, Newlon CS. 2001. Two compound replication origins in *Saccharomyces cerevisiae* contain redundant origin recognition complex binding sites. *Mol. Cell. Biol.* 21:2790–2801. <http://dx.doi.org/10.1128/MCB.21.8.2790-2801.2001>.
  45. Bolon YT, Bielinsky AK. 2006. The spatial arrangement of ORC binding modules determines the functionality of replication origins in budding yeast. *Nucleic Acids Res.* 34:5069–5080. <http://dx.doi.org/10.1093/nar/gkl661>.
  46. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37:W202–W208. <http://dx.doi.org/10.1093/nar/gkp335>.
  47. Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* 4:e1000216. <http://dx.doi.org/10.1371/journal.pcbi.1000216>.
  48. Pohl TJ, Brewer BJ, Raghuraman MK. 2012. Functional centromeres determine the activation time of pericentric origins of DNA replication in *Saccharomyces cerevisiae*. *PLoS Genet.* 8:e1002677. <http://dx.doi.org/10.1371/journal.pgen.1002677>.
  49. Liachko I, Youngblood RA, Tsui K, Bubb KL, Queitsch C, Raghuraman MK, Nislow C, Brewer BJ, Dunham MJ. 2014. GC-rich DNA elements enable replication origin activity in the methylotrophic yeast *Pichia pastoris*. *PLoS Genet.* 10:e1004169. <http://dx.doi.org/10.1371/journal.pgen.1004169>.
  50. Mitra S, Gómez-Raja J, Larriba G, Dubey DD, Sanyal K. 2014. Rad51-Rad52 mediated maintenance of centromeric chromatin in *Candida albicans*. *PLoS Genet.* 10:e1004344. <http://dx.doi.org/10.1371/journal.pgen.1004344>.
  51. Wilson RB, Davis D, Mitchell AP. 1999. Rapid hypothesis testing with *Candida albicans* through gene disruption with short homology regions. *J. Bacteriol.* 181:1868–1874.
  52. Shen J, Guo W, Köhler JR. 2005. CaNAT1, a heterologous dominant selectable marker for transformation of *Candida albicans* and other pathogenic *Candida* species. *Infect. Immun.* 73:1239–1242. <http://dx.doi.org/10.1128/IAI.73.2.1239-1242.2005>.
  53. Ezhkova E, Tansey WP. 2006. Chromatin immunoprecipitation to study protein-DNA interactions in budding yeast. *Methods Mol. Biol.* 313:225–244. <http://dx.doi.org/10.1385/1-59259-958-3:225>.
  54. Ketel C, Wang HS, McClellan M, Bouchonville K, Selmecki A, Lahav T, Gerami-Nejad M, Berman J. 2009. Neocentromeres form efficiently at multiple possible loci in *Candida albicans*. *PLoS Genet.* 5:e1000400. <http://dx.doi.org/10.1371/journal.pgen.1000400>.
  55. Cleveland WS, Grosse E. 1991. Local regression models, p 309225–376. *In* Chambers JM, Hastie TJ (ed), *Statistical models in S*. Chapman and Hall, CRC Press, Boca Raton, FL.
  56. Friedman J, Hastie T, Tibshirani R. 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33:1–22.