bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

DNA replication initiator proteins facilitate CENPA loading on early replicating compact chromatin

Lakshmi Sreekumar[1], Priya Jaitly[1], Yao Chen[2], Bhagya C. Thimmappa[1], Amartya Sanyal[2], Leelavati Narlikar[3], Rahul Siddharthan[4], Kaustuv Sanyal[1]

[1]Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India; [2] School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, [3]Department of Chemical Engineering, CSIR-National Chemical Laboratory, Pune 411008, India; [4]The Institute of Mathematical Sciences/HBNI, Taramani, Chennai 600113, India

*corresponding author

Kaustuv Sanyal
Molecular Biology & Genetics Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
Jakkur, Bangalore - 560064
India
Email: sanyal@jncasr.ac.in
Telephone : +91-80-2208 2878
Fax : +91-80-2208 2766
Homepage: http://www.jncasr.ac.in/sanyal

Present address: Bhagya C. Thimmappa, Department of Biochemistry, Robert-Cedergren Centre for Bioinformatics and Genomics, University of Montreal, 2900 Edouard-Montpetit, Montreal, H3T1J4, QC, Canada

Keywords: Neocentromere, Hi-C, Orc4, Mcm2, centromere, gene silencing

Running title: Inheritance of early replicating centromeric chromatin

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

Abstract (250 words)

The process of centromere formation enables the cell to conserve established genetic and epigenetic information from the previous cell cycle and reuse it for future episodes of chromosome segregation. CENPA asserts the role of an epigenetic requirement in maintaining active centromeres. Active centromeres are subject to position effects which can cause its site of assembly to drift occasionally. Determinants of neocentromere formation, when a native centromere is inactivated, remain elusive. To dissect factors for centromere/neocentromere formation, here, we employed the budding yeast *Candida albicans*, whose centromeres have unique and different DNA sequences, and exhibit classical epigenetic regulation. We used CENPA-mediated reversible silencing of a marker gene, *URA3*, as an assay to select cells with ectopic centromeres. We defined pericentric boundaries for *C. albicans* centromeres by Hi-C analysis and these were located in early replicating domains. The pericentric boundaries primed with CENPA served as sites of neocentromere formation in isolates with ectopic centromeres, indicating that the number of non-centromeric CENPA molecules determines neocentromere location. To understand the importance of early replication timing of centromeres, we identified genome-wide binding sites of the Origin Recognition Complex subunit, Orc4. A fraction of these Orc4 enriched regions located within tDNA, cluster towards early replicating regions, and frequently interact among themselves than the late replicating regions, demonstrating the spatiotemporal distribution of these regions. Strikingly, Orc4 is highly enriched at centromeres of *C. albicans* and along with the helicase component Mcm2, stabilizes the kinetochore, suggesting a role of pre-replication complex proteins as epigenetic determinants of centromere identity.

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Introduction

The centromeric histone H3 variant CENPA is known to assemble on various types of DNA sequences although their position is predominantly confined to centromeres, chromosomal regions responsible for faithful chromosome segregation. CENPA is an adaptor molecule between rapidly diverged centromere (CEN) DNA sequences and the less diverged kinetochore machinery (Mellone and Allshire 2003, Sullivan, Maloney et al. 2016). Centromere assembly requires the establishment of centromeric "memory" by incorporation of CENPA into underlying chromatin, and its subsequent stabilization by binding of kinetochore proteins (McNulty and Sullivan 2017). Extensive studies in higher eukaryotes suggest that there are potential sites on a chromosome which are capable of harboring a functional centromere, however they are kept dormant by the more *cis*-acting dominant centromere (Amor and Choo 2002). This genetically well-defined locus is known to "drift" along the length of a chromosome making them malleable structures. Vertebrate cell lines upon prolonged culturing are subject to "centromere drift" (Hori, Kagawa et al. 2017) similar to the repeat-associated regional centromeres of *Schizosaccharomyces pombe* that exhibit stochastic repositioning of CENPA within CEN chromatin as a consequence of an oversized centromeric core (Yao, Liu et al. 2013). The existence of such plasticity in CEN chromatin indicates that centromeres in vertebrates to the unicellular fission yeast are specified and propagated by sequence independent mechanisms.

The episodic occurrence of centromere activity at non-centromere sequences, neocentromeres, strongly suggests the epigenetic nature of centromeres. First observed in humans to rescue acentric fragments (Voullaire, Slater et al. 1993), neocentromeres across species share common features as well as certain species-specific attributes. *S. pombe* forms sub-telomeric neocentromeres (Ishii, Ogiyama et al. 2008) whereas in humans they are more prevalent at sub-metacentric regions (Warburton 2004). In contrast, most neocentromeres have been detected at CEN proximal loci in flies (Maggert and Karpen 2001), and chicken (Shang, Hori et al. 2013). The assembly of ectopic CENPA as a "CENPA-rich zone" surrounding the endogenous CEN and proximity of neocentromere hotspots to native CEN in these organisms indicates that CENPA is peppered on CEN adjacent loci and can get rapidly incorporated to the centromere in case of CENPA eviction (Haase, Mishra et al. 2013,

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Fukagawa and Earnshaw 2014). Apart from location of CENPA chromatin, transcription plays an important role in specifying CEN identity and maintenance. Centromeres are known to be "difficult to transcribe" regions. However, pervasive level of transcription helps in CEN function as studied extensively in *S. pombe* centromeres (Choi, Stralfors et al. 2011). Silencing of pericentromeric heterochromatin by the RNAi-mediated pathway helps to create a microenvironment for CENPA loading at the *S. pombe* central core (Allshire and Ekwall 2015, Catania, Pidoux et al. 2015). Even in humans, the transcripts generated from the higher order repeats (HOR) of the alpha-satellite DNA interact with CENPA, rendering structural stability to CEN chromatin (McNulty, Sullivan et al. 2017). A recent study in *Drosophila* elucidates that a transcription-coupled remodeling is required for CENPA incorporation (Bobkov, Gilbert et al. 2018). This reemphasizes the role of regulated transcription to maintain centromere structure and function.

Non-repetitive centromeres provide an excellent model to study characterization of centromeric chromatin. In the ascomycetous budding yeast *Candida albicans*, the presence of unique and different CEN sequences on every chromosome (Sanyal, Baum et al. 2004) and the activation of neocentromeres at pre-determined hotspots proximal to the native CEN location (Thakur and Sanyal 2013) together provide evidence that the underlying DNA sequence is neither necessary nor sufficient for centromere formation (Sanyal, Baum et al. 2004, Baum, Sanyal et al. 2006, Thakur and Sanyal 2013). CENPA localization on a transgene under selective conditions is known to correspond to its transcriptional status. Similar to *S. pombe* (Allshire, Javerzat et al. 1994), reversible silencing of the expression of a marker gene, *URA3*, captured by 5'FOA counter-selection, has been observed upon its integration at the CENPA binding region of the centromere in *C. albicans* endowing it a transcriptionally flexible status (Thakur and Sanyal 2013).

For propagation of CEN chromatin, DNA replication ensures accurate assembly of centromeric nucleosomes in a cell cycle specific manner. Replication origins are marked by the physical association of the pre-replication complex (pre-RC) comprising of the hexameric origin recognition complex (Orc1-6), minichromosome maintenance complex (Mcm2-7) and accessory proteins (Leonard and Mechali 2013). These initiator complexes occupy discrete sites on a chromosome and

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

are temporally regulated to ensure complete genome duplication in the S phase. In the budding yeast *Saccharomyces cerevisiae*, approximately 400 ORC binding sites have been identified, but only a subset of them 'fire' at a given time (Wyrick, Aparicio et al. 2001, Nieduszynski, Knox et al. 2006). This implies that not all ORC binding sites act as functional DNA replication origins in each cell cycle. In most eukaryotes, replication origins are defined more flexibly as they rely very little on a DNA sequence requirement for origin specification (Parker, Botchan et al. 2017). Based on the presence of active firing and passive dormant origins, the genome can be classified into early, mid and late replicating regions (Yamazaki, Hayano et al. 2013). Centromeres and replication origins are often seen to be juxtaposed to each other from bacteria like *Bacillus subtilis* (Livny, Yamaichi et al. 2007) to yeast species like *Yarrowia lipolytica* (Vernis, Abbas et al. 1997). This physical proximity aids in centromere cohesion as well as ensures proper kinetochore assembly (Natsume, Muller et al. 2013). Additionally, CEN replication timing is pivotal in CENPA loading, where early replication of CENs ensures replication coupled loading of CENPA in *S. cerevisiae* (Pearson, Yeh et al. 2004). One of the mechanisms for the early replication of CENs in budding yeast is mediated by the timely recruitment of Dbf4-dependent kinase (DDK) at kinetochores with the help of the Ctf19 complex, which loads replication initiator proteins to pericentromeric replication origins (Natsume, Muller et al. 2013). Also, replication fork termination at the centromere promotes centromere DNA loop formation and this is required for kinetochore assembly (Cook, Bennett et al. 2018). Centromeres are the earliest to replicate in every S-phase of the *C. albicans* cell cycle (Koren, Tsai et al. 2010) by virtue of the early replicating origins flanking the centromere (Mitra, Gomez-Raja et al. 2014). Deletion of CEN proximal origins is also known to abrogate centromere function and debilitate kinetochore stability in this organism (Mitra, Gomez-Raja et al. 2014). Hence, there is an intimate crosstalk of replication origins, initiator proteins and kinetochore components in maintaining genome stability.

C. albicans centromeres are defined on the basis of a CENPA binding region which spans a 3-5 kb unique DNA sequence on every chromosome (Sanyal, Baum et al. 2004). There is no functional evidence for the existence of a pericentromeric boundary region to restrict CENPA spreading in *C. albicans*, as seen in case of fission yeast centromeres (Karpen and Allshire 1997, Allshire and Ekwall

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

2015, Allshire and Madhani 2018). Unlike *S. pombe*, the genome of *C. albicans* does not encode an HP1/Swi6-like protein, a methyl transferase like Clr4 required for H3K9me2 and components of a fully functional RNAi machinery (Freire-Beneitez, Price et al. 2016). In this study, we defined the pericentric boundaries of *C. albicans* by Hi-C analysis and a transgene silencing assay. A CENPA-primed region within this pericentric boundary is found to serve as the neocentromere hotspot. By identifying genome-wide binding sites of Orc4, we show that the pericentric regions lie in early replicating highly interacting compact CEN-adjacent regions. We observe a strong physical association of Orc4 to native and neocentromeres in *C. albicans*. The absence of Orc4 compromised kinetochore integrity, a phenotype that we also observed upon depletion of another pre-RC component, Mcm2. Thus, the genetic interaction between CENPA, Orc4 and Mcm2 revealed a previously unidentified role of pre-RC components in maintaining active centromeres in this pathogenic yeast.

Results

Core CENPA-rich regions in *C. albicans* are flanked by a ~25 kb long unusual pericentric heterochromatin

The centromere DNA spans a region of 3-5 kb in *C. albicans* (Sanyal and Carbon 2002, Sanyal, Baum et al. 2004) bound by the CENPA homolog, Cse4 (Sanyal and Carbon 2002). The presence of replication origins and neocentromere hotspots within 30 kb of centromere 7, *CEN7* indicates that CEN proximal regions are important hubs that regulate centromere activity (Thakur and Sanyal 2013, Mitra, Gomez-Raja et al. 2014). We analyzed the Hi-C data of *C. albicans* from a previous report (Burrack, Hutton et al. 2016). Our analysis revealed that all the centromeres interacted with adjacent "pericentric" regions at a higher probability than regions distal from the centromere (Fig 1A). Also, the clustered centromeres of *C. albicans* interact both in *cis* (with the pericentric regions) (Fig. 1A) and in *trans* (with other centromeres) (Supplemental fig. S1A) at a higher probability forming a compact chromatin environment than the average genome interaction found in bulk chromatin

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

(Supplemental fig. 1B). Upon examining the intra-chromosomal interactions, we observed a 25 kb region centring on *CEN7* that closely interacts with the CENPA bound CEN mid-core (Fig 1B). To gain further insights into the pericentric regions, we sought to examine the transcriptional status of CEN-adjacent regions. We inserted the 1.4 kb *URA3* gene at ten independent *CEN7*-proximal loci (Fig. 1C) (see Supplemental table S1 for location of insertions) in a strain that has two differentially marked arms of Chr7, J200 (Sanyal, Baum et al. 2004). We also performed integrations at a *CEN7*-distal locus and a *CEN5*-proximal locus (Supplemental table S1). We plated approximately 1 million cells of each *URA3* integrant type on CM+5'FOA and replica plated 100 colonies resistant to 5'FOA on CM-Uri to obtain the rate of *URA3* silencing (Fig 1D). We also monitored the frequency of chromosome loss in these strains by examining the simultaneous loss of two markers, *ARG4* and *URA3* or *HIS1* and *URA3* (Supplemental table S2). We observed a steep decline in the percentage of reversible silencing of *URA3* (the ratio of the number of 5'FOA resistant colonies that grew on CM-Uri and the total number of 5'FOA resistant colonies analysed) from the *CEN7* core to the periphery. *URA3* when inserted at *CEN7* core exhibited a significantly higher rate of silencing than the peripheral insertions (Fig. 1E, see Supplemental table S2). The clear trend of exponential decay in reversible silencing of *URA3*, correlated with contact probabilities made by the central core to the neighbouring regions, indicating that the clustered centromeres of *C. albicans* interact with pericentric regions to form a compact nuclear subdomain (up to 25 kb), the frequency of which is ablated at loci distal to the central core (Supplemental figs. S1C, 1D).

Transgene silencing at the pericentromeres is associated with a transient ectopic kinetochore

Transcriptional silencing of *URA3* at the *C. albicans* CEN core is known to facilitate CENPA binding (Thakur and Sanyal 2013). We wanted to examine the consequence of *URA3* silencing in these pericentromeric insertions. ChIP experiments on the 5'FOA resistant colonies revealed that *URA3* is significantly enriched with CENPA when cells were grown in CM+5'FOA than CM-Uri indicating that transcriptional repression of *URA3* at pericentromeres favors CENPA binding in all the *URA3* insertions which yielded the 5'FOA resistant colonies (Fig. 2 (top panel), Supplemental Figs. S2A, S2C). We did not observe this phenomenon in the far-*CEN7* integrant (Supplemental Fig. S2B). We

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

expressed Protein A-tagged Mtw1 (Supplemental fig. S3A), the Mis12 homolog in *C. albicans*, (Roy, Burrack et al. 2011) and detected its significant enrichment on *URA3* in LSK437 (4L) and LSK440 (4R) indicating that *URA3* can form an ectopic centromere (ecCEN) when minimally transcribed (Fig. 2, bottom panel). The overlapping CENPA and Mtw1 binding regions were limited to the repressed *URA3* locus and did not extend to regions beyond it (Supplemental Figs. S3B, S3C).

We further wanted to determine whether ecCEN can be stably propagated through mitosis by withdrawing the selection. We serially passaged the initial 5'FOA resistant colonies from LSK404 *(4L/4L::URA3)* and LSK425 *(4R/4R::URA3)* in non-selective media (YPDU) for up to 20 generations (Supplemental Fig. S4A) (see Supplemental methods). We observed a gradual decline in the relative enrichment of CENPA at *URA3* with every doubling and after ~20 mitoses, the CENPA level was comparable to a state when cells were forced to express *URA3* (in CM-Uri) (Supplemental Fig. S4B). Additionally, we observed that if at any stage of passaging in non-selective media (YPDU), these cells were regrown in presence of selection (CM+5'FOA), they could reassemble the CENPA associated ecCEN on *URA3* (Supplemental Fig. S4C). Thus, transcriptional repression of a transgene within the 25 kb compact pericentromeric region favors the formation of a transient ectopic kinetochore.

Pre-existing CENPA molecules can prime a chromosomal location to form neocentromeres

Neocentromeres provide a way to study *de novo* centromere formation since they recapitulate all molecular events for centromere assembly under natural conditions on a non-native locus (Amor and Choo 2002, Craig, Wong et al. 2003, Marshall, Chueh et al. 2008). In *C. albicans*, neocentromeres are shown to get activated at CEN-proximal loci irrespective of the length of the centromere DNA deleted (Thakur and Sanyal 2013). This prompted us to examine that in the event of a centromere deletion whether a cell would prefer to form a neocentromere on a pre-determined hotspot or on a CENPA-primed region located at the pericentric region. To address the same, we replaced the core 4.5 kb CENPA-rich *CEN7* region (Ca21Chr7 424475-428994) with the 1.2 kb *HIS1* sequence independently in two 5'FOA resistant strains, LSK443 (*4L/4L::URA3*) and LSK456 (*4R/4R::URA3*). We screened for colonies where *URA3* and *HIS1* were located on the same homolog (in *cis*) using Southern hybridization (Supplemental Figs. S5A,5B) (Supplemental table. S3) and obtained multiple

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

transformants. We performed the same deletion in the corresponding 5'FOA sensitive *URA3* integrants and examined whether a CENPA primed region could assemble a functional kinetochore. ChIP-qPCR analysis in the 5'FOA resistant strain LSK465 (*4R/4R::URA3 CEN7/CEN7::HIS1)* (Fig 3A, 3B) and LSK450 (*4L/4L::URA3 CEN7/CEN7::HIS1)* (Supplemental Figs. S7A, S7B) revealed that two independent kinetochore proteins, CENPA and Mtw1, assemble at *URA3* and neighboring regions, apart from the native centromere. We confirmed neocentromere formation on this altered chromosome by CENPA ChIP-sequencing (Supplemental fig. S6), which revealed two new hotspots at CENPA-primed regions, *URA3nCEN7-I* and *URA3nCEN7-II* (Fig 3C, Supplemental Fig. S6C) (see Supplemental table S4 for neocentromere coordinates). On the other hand, in the 5'FOA sensitive strains, neocentromeres formed at one of the pre-determined hotspots, *nCEN7-II* (Supplemental Fig 7D). This alludes to the fact that an initial targeting of CENPA to a primed locus within the 25 kb compact region can render centromeric properties to that site if CENPA can enable its nucleation and the subsequent assembly of a functional kinetochore independent of selection or any other target mechanisms.

Orc4 binds to discrete regions uniformly across the *C. albicans* genome to ensure efficient completion of DNA replication in S phase

Nuclear organization is important to study replication architecture. The mitotic propagation of epigenetic marks is ensured by timely replication of the genome. However, to decipher the same, the precise location and timing of replication origins is pivotal. We utilized the binding of an evolutionarily conserved replication initiator protein, Orc4 to map putative replication origins in *C. albicans.* Orc4 in *C. albicans* is a 564-aa long protein (https://doi.org/10.1101/430892) that contains the AAA+ domain which belongs to the AAA+ family of ATPases (Walker, Saraste et al. 1982) associated with a variety of cellular activities (Supplemental fig. 8A). We raised polyclonal antibodies against a peptide sequence from the N-terminus of the native Orc4 (aa 20-33) (Supplemental fig. 8B) of *C. albicans* (see Supplemental methods). Western blot with the whole cell extract of *C. albicans* SC5314 *(ORC4/ORC4*) yielded a strong specific band at the expected molecular weight of approximately 64 kDa when probed with purified anti-Orc4 antibodies (Supplemental fig. 8C).

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Indirect immuno-fluorescence microscopy using anti-Orc4 antibodies revealed that Orc4 was found to be strictly localized to the nucleus at all stages of the *C. albicans* cell cycle (Fig. 4A), a feature of the ORC proteins found to be conserved in *S. cerevisiae* as well (Dutta and Bell 1997).

Orc4 is an evolutionarily conserved essential subunit of the origin recognition complex (ORC) across eukaryotes (Chuang and Kelly 1999, Dai, Chuang et al. 2005). A conditional mutant of *orc4* constructed by deleting one allele and replacing the endogenous promoter of the remaining *ORC4* allele with the repressive *MET3* promoter (Care, Trevethick et al. 1999), showed growth impairment of *C. albicans* cells (Fig. 4B). Hence, Orc4 is essential for viability in *C. albicans* as well. We confirmed the depletion of Orc4 protein levels from the cellular pool by performing a western blot analysis in the Orc4 repressed versus expressed conditions (Supplemental fig. 8D). Subsequently, we used the purified anti-Orc4 antibodies as a tool to map its binding sites in the *C. albicans* genome. ChIP sequencing in asynchronously grown cells of *C. albicans* using anti-Orc4 antibodies yielded a total of 417 discrete Orc4 binding sites with 414 of these belonging to various genomic loci (Fig. 4C, Supplemental fig. S8E) while the remaining three mapped to mitochondrial DNA. We validated one region on each of the eight chromosomes by ChIP-qPCR (Supplemental fig. S8F). Strikingly, all centromeres were found to be highly enriched with Orc4 (Supplemental fig. S8E). The length of Orc4 binding regions across the genome ranged from 200 bp to ~3 kb. Approximately 61% of the Orc4 binding regions in our study were present in genic regions (252/414) in *C. albicans* deviating from the trend observed in *S. cerevisiae* where most of the chromosomal origins are located at intergenic regions (Xu, Aparicio et al. 2006).

Orc4-bound regions in *C. albicans* lack a common DNA sequence motif but are spatiotemporally positioned across the genome

Conserved DNA sequence features at replication origins are common in the *Saccharomyces* group (Nieduszynski, Knox et al. 2006). We used the *de novo* motif discovery tool DIVERSITY (Mitra, Biswas et al. 2018) on the *C. albicans* Orc4 binding regions. DIVERSITY allows for the fact that the profiled protein may have multiple modes of DNA binding. Here, DIVERSITY reports four binding modes (Fig. 5A left). The first mode, mode A is a strong motif GAnTCGAAC, present in 50 such

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

regions, 49 of which were found to be located within tRNA gene bodies. The other three modes were low complexity motifs, TGATGA (mode B), CAnCAnCAn (mode C) and AGnAG (mode D). Strikingly, each of the 417 binding regions were associated with one of these motifs. Mode C has been identified before (Tsai, Baller et al. 2014). The association to tRNA genes has been demonstrated previously in a subset of *S. cerevisiae* replication origins as well (Wyrick, Aparicio et al. 2001). Taken together, this suggests that ORCs in *C. albicans* do not rely on a specific sequence feature for binding DNA.

Replication origins are spatially distributed and temporally regulated to ensure timely duplication of the genome as well as to avoid re-initiation events. Depending on the time of activation and efficiency, replication origins are classified as early and late domain/factories. To categorize the replication timing of Orc4 binding sites, we utilized the fully processed replication timing profile of *C. albicans* available from a previous study (Koren, Tsai et al. 2010) and overlaid the DIVERSITY motifs onto the timing profile (Supplemental fig. S9). We observed a significant advanced replication timing of the tRNA associated motifs (mode A) (Fig. 5A middle). The other three modes (B, C, D) display no significant bias towards an early replication score. Moreover, we could correlate early replication timing with an increased enrichment of Orc4 in these regions (Fig. 5A right). Additionally, all the motifs were located towards the local maxima of the timing peaks (Supplemental fig. S9).

To locate these regions within the nuclear space, we mapped the interactions made by ORC binding regions with each other using the Hi-C data from a previous study in *C. albicans* (Burrack, Hutton et al. 2016). All the ORC binding regions were aligned with an increasing order of their replication timing (early to late) and subsequent interactions were mapped. Similar analysis was performed for the whole genome of *C. albicans.* We observe that the overall "only-ORC" interactions are higher than the whole-genome "all" interactions, suggesting that ORC binding regions interact more than the average (Supplemental figs. S10A, S10B). Early replicating regions (Fig. 5B) show a significantly higher interaction among themselves, in agreement with previous observations in *Candida glabrata* (Descorps-Declere, Saguez et al. 2015). Given that regions in this heatmap are ordered by timing and not genomic proximity, this suggests that regions with a similar timing in replication tend to associate

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

together. Hi-C analysis also revealed that mode A containing sites, that show an early replication timing, form stronger interactions among themselves than all the other modes (Fig. 5B). Hence, it is highly likely that a subset of ORC binding regions identified in our study are the chromosomal origins in *C. albicans* as they associate with categorically distinct domains separated in space and time of replication, facilitating origin function and usage.

The strong physical association of Orc4 at *C. albicans* centromeres stabilizes CENPA

Apart from the discrete genomic loci across all chromosome arms, the strong binding of Orc4 on all centromeres in *C. albicans* was particularly striking. This hints towards the possible role of replication initiator complexes in influencing centromere location and function. Upon comparison of the Orc4 enrichment with the CENPA occupancy in *C. albicans*, we observe that there is a significant overlap in the binding regions of both these proteins, indicating a strong physical association of ORCs at all centromeres (Fig. 6A, Supplemental Fig. S11A) (Supplemental table S5). We additionally examined for the presence of Orc4 in non-native centromeres. ChIP-qPCR analysis in the 5'FOA resistant strain LSK443 revealed that similar to CENPA binding, the conditional ecCEN at *URA3* is enriched with Orc4 (Supplemental fig. S11B). To validate the association of Orc4 at functional centromeres, we explored its binding to strains forming neocentromeres. Neocentromeres activated at *nCEN7-II* hotspot upon deletion of the 4.5 kb CENPA rich region on *CEN7* showed a significant Orc4 enrichment on the altered homolog (Supplemental fig. S11C). These observations strongly suggest that Orc4 is associated with all active centromeres in *C. albicans*. To examine its role in centromere function, we assayed for CENPA localization in an *orc4* conditional mutant, LSK331 (*orc4/MET3prORC4 CSE4/CSE4-GFP-CSE4)*. Orc4 depletion caused severe chromosome mis-segregation (Supplemental fig. S13A). ChIP-qPCR analysis revealed a significant reduction of chromatin associated CENPA upon Orc4 depletion (Fig. 6B), which was corroborated by degradation of CENPA protein levels (Supplemental fig. S13B). However, depletion of CENPA did not significantly alter the levels of Orc4 bound to the centromere (Fig. 6C), indicating that Orc4 strictly regulates CENPA localization at the centromere but not vice-versa. Hence, Orc4 has a direct role in stabilizing CENPA, thereby influencing centromere activity and kinetochore segregation.

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

The helicase subunit, Mcm2 influences CENPA stability and kinetochore segregation

Even though ORCs flag mark replication origins in the genome, their subsequent activity is governed by assembly of the Mcm2-7 helicase that primes the complex for replication initiation. There are distinct subunits and subunits of the pre-replicative complex (pre-RC) which perform roles outside replication initiation. *MCM2* is annotated as an uncharacterized ORF (Orf19.4354) in the *Candida Genome Database* (candidagenome.org). BLAST analysis using *S. cerevisiae* Mcm2 as the query sequence revealed that this Orf19.4354 translates to a 101.2 kDa protein that contains the conserved Walker A, Walker B and the R finger motif which together constitute the MCM box (Forsburg 2004) (Supplemental fig. 12). In order to determine the essentiality of this gene in *C. albicans*, we constructed a conditional mutant of *mcm2*, LSK311 (*mcm2/MET3prMCM2 CSE4-GFP-CSE4/CSE4*) by deleting one allele and replacing the endogenous promoter of the remaining *MCM2* allele with the repressive *MET3* promoter (Care, Trevethick et al. 1999). Mcm2 was found to be essential for viability (Fig 7B). We could detect severe kinetochore segregation defects in this mutant post 6 h of depletion of the protein (Supplemental fig. 13C). Depletion of Mcm2 led to a reduction in CENPA at the centromere (Fig. 7C, Supplemental fig. 13D). Hence, Mcm2 is possibly helping in loading the CENPA-H4 dimer at the *C. albicans* centromere, similar to what has been reported in human cells (Huang, Stromme et al. 2015). Taken together, we establish how an intricate crosstalk of DNA replication initiator proteins and early replication program at the pericentric regions help load, stabilize and propagate centromeric chromatin in absence of any obvious DNA sequence cues in *C. albicans.*

Discussion

Seeding of CENPA on DNA, the stability of centromeric chromatin during the cell cycle and its subsequent propagation involves a plethora of factors ranging from the primary DNA sequence and the chromatin context to crosstalk with DNA replication initiator and DNA damage repair proteins. Specific protein binding sites aid centromere formation in genetically defined point centromeres

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

(Lechner and Carbon 1991). Additional mechanisms must operate at multi-dimensional levels to spatiotemporally define centromere activity to a defined region in epigenetically regulated regional centromeres in most other organisms. Centromeric heterochromatin is distinct from arm heterochromatin, in terms of the degree of compaction and presence of topological adjusters like cohesin, condensin and topoisomerase II (Bloom 2014). The cruciform structure adopted by the centromeric chromatin in budding yeast facilitates cohesin maintenance on duplicated sister chromatids and orients the centromere towards the spindle pole (Stephens, Haase et al. 2011) (Lawrimore, Doshi et al. 2018). Centromere clustering gradually increases with cell cycle progression due to sister-chromatid cohesion during replication, and cohesin mediated spindle-dependent clustering during anaphase (Lazar-Stefanita, Scolari et al. 2017). In this study, we determine the extent and functional consequence of centromeric chromatin compaction in *C. albicans.* The fact that inter-centromeric interactions are much stronger than the average genomic interactions facilitates the formation of a CENPA cloud in a 3-dimensional milieu to enrich the local CENPA concentration in the clustered CENs of *C. albicans*. In the *S. cerevisiae* point centromeres, the presence of core and accessory CENPA molecules at the native centromeres and pericentromeres, respectively, helps in rapid incorporation of CENPA into the CEN chromatin during rogue loss events (Haase, Mishra et al. 2013), suggesting the dynamic nature of pericentromeric nucleosomes. Hence, the identification of a highly interacting 25 kb pericentric region in *C. albicans* enables us to dissect functional underpinnings of pericentromeres and spatial segregation of chromatin properties, in this case, created by the pericentric heterochromatin that acts as the reservoir of CENPA molecules.

The strong reversible silencing of the transgene at the *C. albicans* central core, that is a readout of its flexible transcriptional status (Thakur and Sanyal 2013) is reminiscent of the repeat-associated centromere organization in *S. pombe,* where the central core shows variegated levels of marker gene expression whereas the outer repeats shut down the transgene expression due to heterochromatinization (Allshire, Javerzat et al. 1994, Karpen and Allshire 1997). Even though *S. pombe* outer repeats do not bind CENPA, they are considered an important component of a functional centromere (Clarke, Amstutz et al. 1986). Similarly, the pericentric regions identified in our study,

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

probably possess pericentric properties in *C. albicans*. In *S. pombe*, CENPA can assemble on a non-centromeric region by competing out H3 (Castillo, Mellone et al. 2007), and the frequency of reversible silencing can be increased by overexpressing CENPA. In our study, the strong negative selection imposed on cells by 5'FOA, enables us to isolate rare individuals from a heterogenous population of cells that can transiently incorporate CENPA at an ectopic locus. However, even under selective growth conditions, only a few cells can tolerate the ecCEN because of the presence of the more dominating native centromere locus (*CEN7*), eventually weeding out cells with ecCEN from the population. Formation of an ecCEN outside the native CEN strongly suggests the existence of that non-centromeric CENPA molecules interspersed with the H3 nucleosome. Unlike *S. pombe*, CENPA overexpression in *C. albicans* does not lead to its extended occupancy beyond centromeric chromatin, it merely increases its occupancy at the native locus (Berman 2012). It is to be noted here that the ecCEN that was obtained by growth in selective media did not have an over-expression of CENPA and still harbored two intact copies of native *CEN7*. CENPA associated chromatin, has self-propagating properties and hence relies on an epigenetic memory (Black, Brock et al. 2007). Our observation that any CENPA-primed region within the identified pericentric boundaries (25 kb centring on the CEN core) can initiate neocentromere formation, emphasizes the importance of the number of CENPA molecules required to nucleate the kinetochore assembly. However, we have limited understanding regarding the determinants that act is favor of a particular locus on a chromosome to have a "centromere correctness".

Incorporation of CENPA into replicated chromosomes is uncoupled with DNA replication in most organisms. CEN proximal replication origins facilitate early replication of CEN chromatin in *C. albicans* (Koren, Tsai et al. 2010, Mitra, Gomez-Raja et al. 2014). Naturally, the identification of genome-wide replication origins in *C. albicans* will reveal useful insights into the replication architecture of this organism. Towards this objective, first, our analysis of the Orc4 binding regions revealed the lack of a DNA sequence requirement for most of the ORC binding sites in *C. albicans*. A previous genome-wide study on identification of ORC binding regions in *C. albicans* utilized antibodies against the *S. cerevisiae* ORC complex (Tsai, Baller et al. 2014). The said study reported

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

~390 ORC binding sites which exhibited a 25% overlap with the Orc4 binding regions identified in our study. Since we used antibodies against an endogenous protein (CaOrc4) to map its binding sites in *C. albicans,* we present an authentic depiction of Orc4 binding regions in the genome. We do find a strong association of a fraction of these regions within many tRNA genes. tRNA genes along with histone genes and centromeres are known to exhibit conserved replication timing (Muller and Nieduszynski 2017). Moreover, in *S. cerevisiae* there is a statistically significant bias for codirectional transcription and replication of tRNA genes (Muller and Nieduszynski 2017). tDNAs cluster near centromeres and recovers stalled forks (Thompson, Haeusler et al. 2003). Since centromeres are early replicating in fungal genomes, the presence of tDNAs in its vicinity might transduce an early replication program. The various Orc4 binding DNA motifs identified in our study hints towards differential usage and specification of origins facilitated by multiple modes of ORC binding in *C. albicans.* Secondly, in spite of the sequence heterogeneity, these Orc4 binding regions could be classified based on their replication timing, wherein the early replicating regions form closely associated units and interact sparsely with the late replicating ones. This is reminiscent of the genome-wide replication landscape of *C. glabrata* origins (Descorps-Declere, Saguez et al. 2015). The stochastic activation of early origins, makes up for the uneven distribution of origins in the genome. The 'replication wave' progresses with the sequential activation of pericentromeric origins to the chromosome arm origins. Initiation events convert early origin clusters to replication foci during S phase. As DNA replication progresses, more replisomes are formed, chromosomes are sufficiently mobilized which makes long range interactions more favourable with time. Hence, one can speculate the existence of topologically distinct domains that are separated in location and time as S-phase progresses.

C. albicans centromeres do not possess a firing origin (Mitra, Gomez-Raja et al. 2014). Replication forks originating from the centromere flanking origins stall at the centromere in a kinetochore dependent manner and facilitate new CENPA loading. Furthermore, CENPA loading is facilitated by the physical interaction of repair proteins like Rad51, Rad52 with CENPA, that are transiently localized to the kinetochore upon replication fork stalling (Mitra, Gomez-Raja et al. 2014). Hence,

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

there is an interplay of replication-repair machinery in maintaining centromere identity in *C. albicans*. We postulate that an ORC-cloud (Fig. 8) facilitates ORC abundance at all centromeres of *C. albicans* and can be attributed to the early replication of centromeres in every S-phase which in turn influences the loading of new CENPA. The anaphase specific loading of new CENPA in *C. albicans* has been demonstrated earlier by fluorescence spectroscopic measurements (Shivaraju, Unruh et al. 2012). However, specific chaperones and molecular pathways involved in the same are undeciphered. We posit ORC to be an essential component for CENPA loading by maintaining a heterochromatin environment at the centromeric locus. Members of the pre-RC have established roles in cell cycle dependent dynamics at the centromere. Mcm2 is a known chaperone that hands over the old histones from the replication fork to anti-silencing function Asf1 to recycle old histones and deposits them to the newly synthesised DNA (Hammond, Stromme et al. 2017). Mcm2 and Asf1 cochaperone an H3-H4 dimer through histone-binding mode (Richet, Liu et al. 2015). This is true for both canonical H3 as well as H3 variants like H3.3 and CENPA (Huang, Stromme et al. 2015). A recent study indicates the role played by Mcm2 in mouse embryonic stem cells to symmetrically partition modified histones to daughter cells using its histone-binding mode (Petryk, Dalby et al. 2018). In humans, the S phase retention of CENPA is mitigated by its simultaneous interaction with the specific chaperone HJURP and Mcm2 (Zasadzinska, Huang et al. 2018), which together transmit CENPA nucleosomes upon its disassembly ahead of the replication fork. In the light of the existing evidence in metazoan systems and the results obtained in our study, Mcm2 emerges as an evolutionarily conserved factor required for eviction of old CENPA molecules and loading of newly synthesized ones (Fig. 8). Although our experiments demonstrate a strong genetic interaction between these two proteins, the physical interaction of Orc4-CENPA and Mcm2-CENPA is still speculative due to technical difficulties. We hypothesize that during CEN chromatin replication at S phase, ORCs maintain the heterochromatin environment of CEN when “old” CENPA is evicted (Fig. 8). During anaphase, centromeric ORCs are briefly displaced, to facilitate loading of “new” CENPA with the help of a specific chaperone such as Scm3/HJURP and Mcm2 which stabilizes the kinetochore complex. In the next cell cycle, Mcm2 associates with the MCM complex to license replication origins during G_1. ORCs in *S. cerevisiae* have established roles in heterochromatinization and MTL silencing (Foss, McNally et al. 1993,

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Hickman, Froyd et al. 2011). The centromere silencing mechanisms in *C. albicans* is relatively unknown as this organism lacks a functional RNAi machinery and H3K9me2 marks. In this regard, we envision the ORC family of proteins as a possible silencing factor for centromeres in this organism.

Methods

All the strains and primers are listed in Supplemental Tables S6 and S7, respectively. Protocols and experimental procedures have been mentioned in Supplementary information.

Data access

The sequencing data used in the study have been submitted to NBCI under the SRA accession number PRJNA477284.

Acknowledgements

We thank Clevergene Biocorp for ChIP-seq experiments and analysis. We also thank Dr. Prakash for animal facility and B. Suma for confocal microscopy, JNCASR. We thank Dr. Koren and Prof. Berman for the raw data of the replication timing experiment. We also thank Prof. Rajan Dighe for helping us in raising polyclonal antibodies. This work was supported by Council of Scientific and Industrial Research (CSIR), Govt. of India (grant number 09/733(0178)/2012-EMR-I to LSK, and Tata Innovation Fellowship, Dept. of Biotechnology, Govt of India to KS. AS is supported by NTU's Nanyang Assistant Professorship grant and Singapore Ministry of Education Academic Research Fund Tier 1 grant (RG46/16), LN is supported by DBT grant BT/ PR16240/BID/7/575/2016 and RS thanks the PRISM-II project at IMSc, funded by DAE. KS also gratefully acknowledges intramural funding from Jawaharlal Nehru Center for Advanced Scientific Research, Bangalore.

References:

Allshire, R. C. and K. Ekwall (2015). "Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe." Cold Spring Harb Perspect Biol **7**(7): a018770.

Allshire, R. C., J. P. Javerzat, N. J. Redhead and G. Cranston (1994). "Position effect variegation at fission yeast centromeres." Cell **76**(1): 157-169.

Allshire, R. C. and H. D. Madhani (2018). "Ten principles of heterochromatin formation and function." Nat Rev Mol Cell Biol **19**(4): 229-244.

Amor, D. J. and K. H. Choo (2002). "Neocentromeres: role in human disease, evolution, and centromere study." Am J Hum Genet **71**(4): 695-714.

Baum, M., K. Sanyal, P. K. Mishra, N. Thaler and J. Carbon (2006). "Formation of functional centromeric chromatin is specified epigenetically in Candida albicans." Proc Natl Acad Sci U S A **103**(40): 14877-14882.

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Berman, J. (2012). "Candida albicans." Curr Biol **22**(16): R620-622.
Black, B. E., M. A. Brock, S. Bedard, V. L. Woods, Jr. and D. W. Cleveland (2007). "An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes." Proc Natl Acad Sci U S A **104**(12): 5008-5013.
Bloom, K. S. (2014). "Centromeric heterochromatin: the primordial segregation machine." Annu Rev Genet **48**: 457-484.
Bobkov, G. O. M., N. Gilbert and P. Heun (2018). "Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation." J Cell Biol.
Burrack, L. S., H. F. Hutton, K. J. Matter, S. A. Clancey, I. Liachko, A. E. Plemmons, A. Saha, E. A. Power, B. Turman, M. A. Thevandavakkam, F. Ay, M. J. Dunham and J. Berman (2016). "Neocentromeres Provide Chromosome Segregation Accuracy and Centromere Clustering to Multiple Loci along a Candida albicans Chromosome." PLoS Genet **12**(9): e1006317.
Care, R. S., J. Trevethick, K. M. Binley and P. E. Sudbery (1999). "The MET3 promoter: a new tool for Candida albicans molecular genetics." Mol Microbiol **34**(4): 792-798.
Castillo, A. G., B. G. Mellone, J. F. Partridge, W. Richardson, G. L. Hamilton, R. C. Allshire and A. L. Pidoux (2007). "Plasticity of fission yeast CENP-A chromatin driven by relative levels of histone H3 and H4." PLoS Genet **3**(7): e121.
Catania, S., A. L. Pidoux and R. C. Allshire (2015). "Sequence features and transcriptional stalling within centromere DNA promote establishment of CENP-A chromatin." PLoS Genet **11**(3): e1004986.
Choi, E. S., A. Stralfors, A. G. Castillo, M. Durand-Dubief, K. Ekwall and R. C. Allshire (2011). "Identification of noncoding transcripts from within CENP-A chromatin at fission yeast centromeres." J Biol Chem **286**(26): 23600-23607.
Chuang, R. Y. and T. J. Kelly (1999). "The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks." Proc Natl Acad Sci U S A **96**(6): 2656-2661.
Clarke, L., H. Amstutz, B. Fishel and J. Carbon (1986). "Analysis of centromeric DNA in the fission yeast Schizosaccharomyces pombe." Proc Natl Acad Sci U S A **83**(21): 8253-8257.
Cook, D. M., M. Bennett, B. Friedman, J. Lawrimore, E. Yeh and K. Bloom (2018). "Fork pausing allows centromere DNA loop formation and kinetochore assembly." Proc Natl Acad Sci U S A.
Craig, J. M., L. H. Wong, A. W. Lo, E. Earle and K. H. Choo (2003). "Centromeric chromatin pliability and memory at a human neocentromere." EMBO J **22**(10): 2495-2504.
Dai, J., R. Y. Chuang and T. J. Kelly (2005). "DNA replication origins in the Schizosaccharomyces pombe genome." Proc Natl Acad Sci U S A **102**(2): 337-342.
Descorps-Declere, S., C. Saguez, A. Cournac, M. Marbouty, T. Rolland, L. Ma, C. Bouchier, I. Moszer, B. Dujon, R. Koszul and G. F. Richard (2015). "Genome-wide replication landscape of Candida glabrata." BMC Biol **13**: 69.
Dutta, A. and S. P. Bell (1997). "Initiation of DNA replication in eukaryotic cells." Annu Rev Cell Dev Biol **13**: 293-332.
Forsburg, S. L. (2004). "Eukaryotic MCM proteins: beyond replication initiation." Microbiol Mol Biol Rev **68**(1): 109-131.
Foss, M., F. J. McNally, P. Laurenson and J. Rine (1993). "Origin recognition complex (ORC) in transcriptional silencing and DNA replication in S. cerevisiae." Science **262**(5141): 1838-1844.
Freire-Beneitez, V., R. J. Price and A. Buscaino (2016). "The Chromatin of Candida albicans Pericentromeres Bears Features of Both Euchromatin and Heterochromatin." Front Microbiol **7**: 759.
Fukagawa, T. and W. C. Earnshaw (2014). "The centromere: chromatin foundation for the kinetochore machinery." Dev Cell **30**(5): 496-508.
Haase, J., P. K. Mishra, A. Stephens, R. Haggerty, C. Quammen, R. M. Taylor, 2nd, E. Yeh, M. A. Basrai and K. Bloom (2013). "A 3D map of the yeast kinetochore reveals the presence of core and accessory centromere-specific histone." Curr Biol **23**(19): 1939-1944.
Hammond, C. M., C. B. Stromme, H. Huang, D. J. Patel and A. Groth (2017). "Histone chaperone networks shaping chromatin function." Nat Rev Mol Cell Biol **18**(3): 141-158.

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Hickman, M. A., C. A. Froyd and L. N. Rusche (2011). "Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage." Eukaryot Cell **10**(9): 1183-1192.

Hori, T., N. Kagawa, A. Toyoda, A. Fujiyama, S. Misu, N. Monma, F. Makino, K. Ikeo and T. Fukagawa (2017). "Constitutive centromere-associated network controls centromere drift in vertebrate cells." J Cell Biol **216**(1): 101-113.

Huang, H., C. B. Stromme, G. Saredi, M. Hodl, A. Strandsby, C. Gonzalez-Aguilera, S. Chen, A. Groth and D. J. Patel (2015). "A unique binding mode enables MCM2 to chaperone histones H3-H4 at replication forks." Nat Struct Mol Biol **22**(8): 618-626.

Ishii, K., Y. Ogiyama, Y. Chikashige, S. Soejima, F. Masuda, T. Kakuma, Y. Hiraoka and K. Takahashi (2008). "Heterochromatin integrity affects chromosome reorganization after centromere dysfunction." Science **321**(5892): 1088-1091.

Karpen, G. H. and R. C. Allshire (1997). "The case for epigenetic effects on centromere identity and function." Trends Genet **13**(12): 489-496.

Koren, A., H. J. Tsai, I. Tirosh, L. S. Burrack, N. Barkai and J. Berman (2010). "Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase." PLoS Genet **6**(8): e1001068.

Lawrimore, J., A. Doshi, B. Friedman, E. Yeh and K. Bloom (2018). "Geometric partitioning of cohesin and condensin is a consequence of chromatin loops." Mol Biol Cell **29**(22): 2737-2750.

Lazar-Stefanita, L., V. F. Scolari, G. Mercy, H. Muller, T. M. Guerin, A. Thierry, J. Mozziconacci and R. Koszul (2017). "Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle." EMBO J **36**(18): 2684-2697.

Lechner, J. and J. Carbon (1991). "A 240 kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere." Cell **64**(4): 717-725.

Leonard, A. C. and M. Mechali (2013). "DNA replication origins." Cold Spring Harb Perspect Biol **5**(10): a010116.

Livny, J., Y. Yamaichi and M. K. Waldor (2007). "Distribution of centromere-like parS sites in bacteria: insights from comparative genomics." J Bacteriol **189**(23): 8693-8703.

Maggert, K. A. and G. H. Karpen (2001). "The activation of a neocentromere in Drosophila requires proximity to an endogenous centromere." Genetics **158**(4): 1615-1628.

Marshall, O. J., A. C. Chueh, L. H. Wong and K. H. Choo (2008). "Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution." Am J Hum Genet **82**(2): 261-282.

McNulty, S. M. and B. A. Sullivan (2017). "Centromere Silencing Mechanisms." Prog Mol Subcell Biol **56**: 233-255.

McNulty, S. M., L. L. Sullivan and B. A. Sullivan (2017). "Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C." Dev Cell **42**(3): 226-240 e226.

Mellone, B. G. and R. C. Allshire (2003). "Stretching it: putting the CEN(P-A) in centromere." Curr Opin Genet Dev **13**(2): 191-198.

Mitra, S., A. Biswas and L. Narlikar (2018). "DIVERSITY in binding, regulation, and evolution revealed from high-throughput ChIP." PLoS Comput Biol **14**(4): e1006090.

Mitra, S., J. Gomez-Raja, G. Larriba, D. D. Dubey and K. Sanyal (2014). "Rad51-Rad52 mediated maintenance of centromeric chromatin in Candida albicans." PLoS Genet **10**(4): e1004344.

Muller, C. A. and C. A. Nieduszynski (2017). "DNA replication timing influences gene expression level." J Cell Biol **216**(7): 1907-1914.

Natsume, T., C. A. Muller, Y. Katou, R. Retkute, M. Gierlinski, H. Araki, J. J. Blow, K. Shirahige, C. A. Nieduszynski and T. U. Tanaka (2013). "Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment." Mol Cell **50**(5): 661-674.

Nieduszynski, C. A., Y. Knox and A. D. Donaldson (2006). "Genome-wide identification of replication origins in yeast by comparative genomics." Genes Dev **20**(14): 1874-1879.

Parker, M. W., M. R. Botchan and J. M. Berger (2017). "Mechanisms and regulation of DNA replication initiation in eukaryotes." Crit Rev Biochem Mol Biol **52**(2): 107-144.

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Pearson, C. G., E. Yeh, M. Gardner, D. Odde, E. D. Salmon and K. Bloom (2004). "Stable kinetochore-microtubule attachment constrains centromere positioning in metaphase." Curr Biol **14**(21): 1962-1967.

Petryk, N., M. Dalby, A. Wenger, C. B. Stromme, A. Strandsby, R. Andersson and A. Groth (2018). "MCM2 promotes symmetric inheritance of modified histones during DNA replication." Science.

Richet, N., D. Liu, P. Legrand, C. Velours, A. Corpet, A. Gaubert, M. Bakail, G. Moal-Raisin, R. Guerois, C. Compper, A. Besle, B. Guichard, G. Almouzni and F. Ochsenbein (2015). "Structural insight into how the human helicase subunit MCM2 may act as a histone chaperone together with ASF1 at the replication fork." Nucleic Acids Res **43**(3): 1905-1917.

Roy, B., L. S. Burrack, M. A. Lone, J. Berman and K. Sanyal (2011). "CaMtw1, a member of the evolutionarily conserved Mis12 kinetochore protein family, is required for efficient inner kinetochore assembly in the pathogenic yeast Candida albicans." Mol Microbiol **80**(1): 14-32.

Sanyal, K., M. Baum and J. Carbon (2004). "Centromeric DNA sequences in the pathogenic yeast Candida albicans are all different and unique." Proc Natl Acad Sci U S A **101**(31): 11374-11379.

Sanyal, K. and J. Carbon (2002). "The CENP-A homolog CaCse4p in the pathogenic yeast Candida albicans is a centromere protein essential for chromosome transmission." Proc Natl Acad Sci U S A **99**(20): 12969-12974.

Shang, W. H., T. Hori, N. M. Martins, A. Toyoda, S. Misu, N. Monma, I. Hiratani, K. Maeshima, K. Ikeo, A. Fujiyama, H. Kimura, W. C. Earnshaw and T. Fukagawa (2013). "Chromosome engineering allows the efficient isolation of vertebrate neocentromeres." Dev Cell **24**(6): 635-648.

Shivaraju, M., J. R. Unruh, B. D. Slaughter, M. Mattingly, J. Berman and J. L. Gerton (2012). "Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast." Cell **150**(2): 304-316.

Stephens, A. D., J. Haase, L. Vicci, R. M. Taylor, 2nd and K. Bloom (2011). "Cohesin, condensin, and the intramolecular centromere loop together generate the mitotic chromatin spring." J Cell Biol **193**(7): 1167-1180.

Sullivan, L. L., K. A. Maloney, A. J. Towers, S. G. Gregory and B. A. Sullivan (2016). "Human centromere repositioning within euchromatin after partial chromosome deletion." Chromosome Res **24**(4): 451-466.

Thakur, J. and K. Sanyal (2013). "Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function at native physical chromosomal loci in Candida albicans." Genome Res **23**(4): 638-652.

Thompson, M., R. A. Haeusler, P. D. Good and D. R. Engelke (2003). "Nucleolar clustering of dispersed tRNA genes." Science **302**(5649): 1399-1401.

Tsai, H. J., J. A. Baller, I. Liachko, A. Koren, L. S. Burrack, M. A. Hickman, M. A. Thevandavakkam, L. N. Rusche and J. Berman (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): e01703-01714.

Vernis, L., A. Abbas, M. Chasles, C. M. Gaillardin, C. Brun, J. A. Huberman and P. Fournier (1997). "An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast Yarrowia lipolytica." Mol Cell Biol **17**(4): 1995-2004.

Voullaire, L. E., H. R. Slater, V. Petrovic and K. H. Choo (1993). "A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere?" Am J Hum Genet **52**(6): 1153-1163.

Walker, J. E., M. Saraste, M. J. Runswick and N. J. Gay (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." EMBO J **1**(8): 945-951.

Warburton, P. E. (2004). "Chromosomal dynamics of human neocentromere formation." Chromosome Res **12**(6): 617-626.

Wyrick, J. J., J. G. Aparicio, T. Chen, J. D. Barnett, E. G. Jennings, R. A. Young, S. P. Bell and O. M. Aparicio (2001). "Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins." Science **294**(5550): 2357-2360.

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Xu, W., J. G. Aparicio, O. M. Aparicio and S. Tavare (2006). "Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae." BMC Genomics **7**: 276.

Yamazaki, S., M. Hayano and H. Masai (2013). "Replication timing regulation of eukaryotic replicons: Rif1 as a global regulator of replication timing." Trends Genet **29**(8): 449-460.

Yao, J., X. Liu, T. Sakuno, W. Li, Y. Xi, P. Aravamudhan, A. Joglekar, Y. Watanabe and X. He (2013). "Plasticity and epigenetic inheritance of centromere-specific histone H3 (CENP-A)-containing nucleosome positioning in the fission yeast." J Biol Chem **288**(26): 19184-19196.

Zasadzinska, E., J. Huang, A. O. Bailey, L. Y. Guo, N. S. Lee, S. Srivastava, K. A. Wong, B. T. French, B. E. Black and D. R. Foltz (2018). "Inheritance of CENP-A Nucleosomes during DNA Replication Requires HJURP." Dev Cell.

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Figures

Fig.1

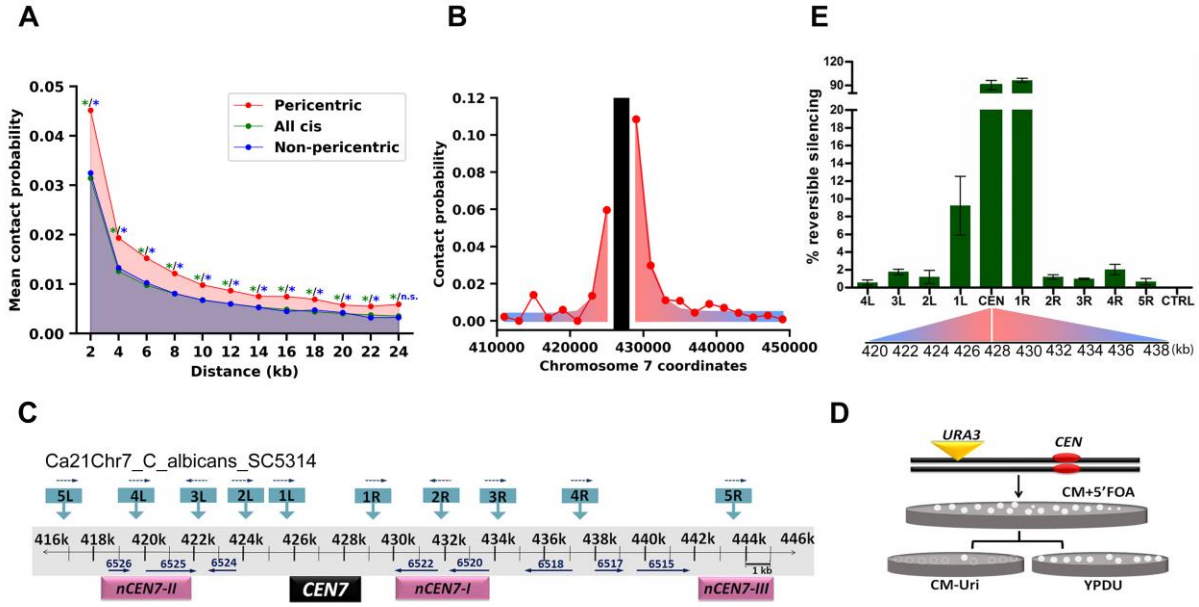

Fig 1. Centromeres in *C. albicans* are flanked by pericentromeric chromatin spanning ~25 kb centring the CENPA binding region. (A) The mean Hi-C contact probability (*y*-axis; bin size= 2 kb) separated at a given distance (*x*-axis) depicts significant increase in pericentric (centromeric region plus 10 kb upstream and downstream) interactions (red) over all *cis* interactions (green) and non-pericentric (a randomly selected region of same size as pericentric region and equidistant from the centromere of each chromosome) interactions (blue). (*: p-value<0.05; n.s.: not significant). **(B)** The 3C profile (bin size=2kb) anchored on centromeric bin (Chr7:426000-428000) showing contact probabilities (red dots) between the anchor bin (black) and its neighbouring bins on Chr7 indicates a strongly interacting 25 kb region. **(C)** A line diagram of ~30 kb region on Chr7 surrounding *CEN7* shows individual *URA3* insertion locations (blue arrows) and previously mapped neocentromere hotspots (*nCEN7-I*, *nCEN7-II*). Arrowheads and numbers indicate positions and identities of the ORFs respectively. **(D)** The assay strategy used to screen reversibly silenced colonies derived from the *URA3* integrants using 5'FOA. **(E)** A decline in the percentage of reversible silencing of *URA3* from mid-*CEN7* (CEN) to the pericentromeric integrants (4L,3L,2L…5R, CTRL) was observed with increasing distance from native *CEN*. The phase exponential decay curve is colour coded with the graph in (B). Coordinates for the respective insertions have been depicted below the black line.

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Fig 2

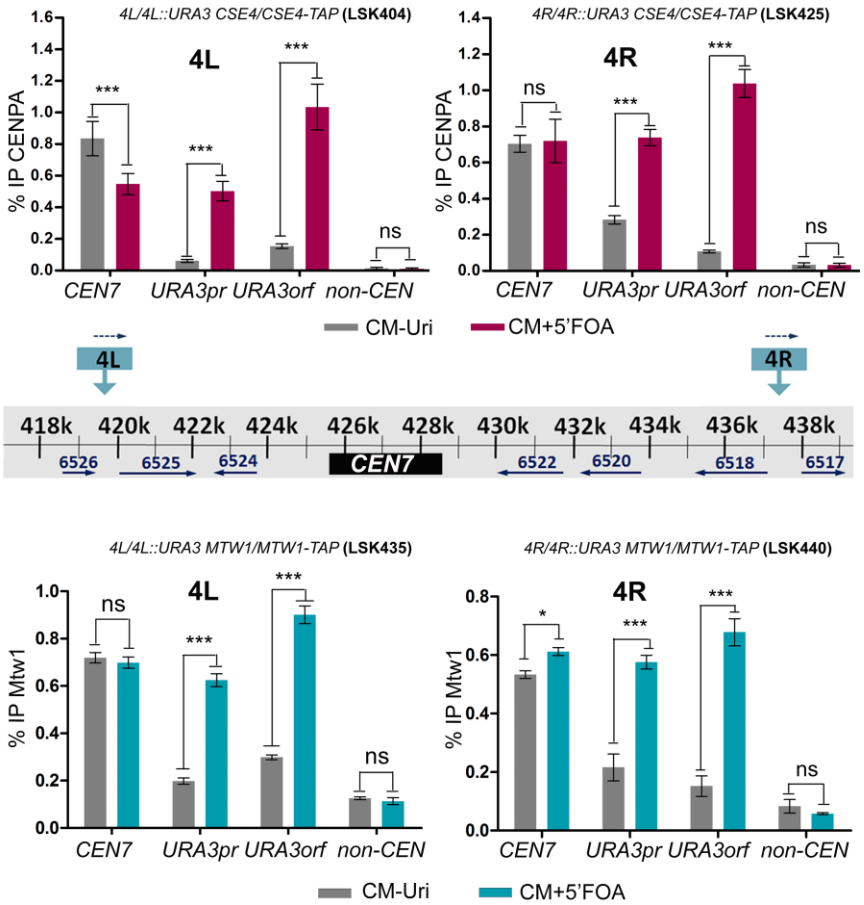

Fig 2. Transcriptional silencing of *URA3* at the pericentromeres favours formation of an ectopic kinetochore. A line diagram of the *URA3* insertions at locations 4L and 4R on Chr7 is shown (middle panel). Corresponding integrations are mentioned as graph titles. Anti-Protein A ChIP followed by qPCR analysis of the 5'FOA resistant colonies from the strains LSK404 (4L) and LSK425 (4R) was used to compare enrichment of CENPA on the indicated loci (*CEN7*, *URA3pr*, *URA3orf*, non-*CEN* region) in CM-Uri (grey) and CM+5'FOA (red) media (upper panel). Similar ChIP-qPCR assays (lower panels) performed on strains LSK435 (4L) and LSK440 (4R) showed significant enrichment of Mtw1 on the silent *URA3* locus (blue bar). Normalised CENPA and Mtw1 enrichment values indicate significant enrichment at *URA3* upon its transcriptional repression in 4L and 4R integrants (*** $p<0.001$, ** $p<0.01$, ns: $p>0.05$). (See Supplemental Figs S2 and S3 for ChIP-qPCR results of other *URA3* integrants)

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Fig. 3

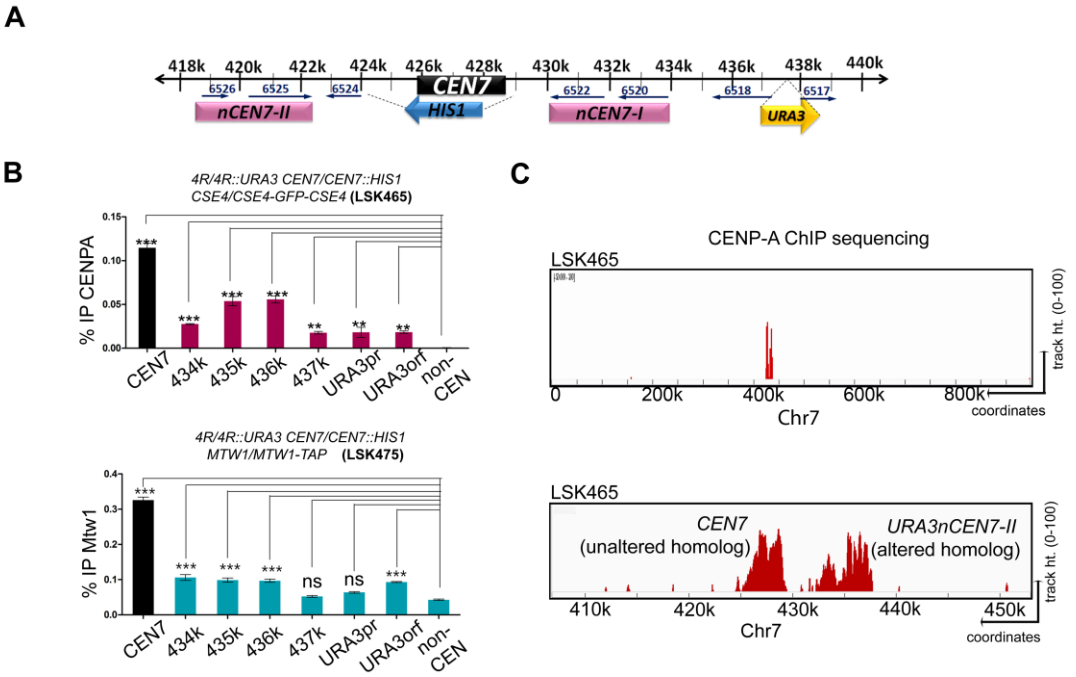

Fig 3. Pre-existing CENPA molecules can prime a chromosomal location for neocentromere formation. (A) In the diploid *C. albicans*, only one homolog of Chr7 where *CEN7* (CaChr7 424475-428994) has been replaced with *HIS1* in a *URA3* integrant strain (CaChr7 437729-437730) is shown. **(B)** Top panel indicates relative enrichment of CENPA at native *CEN7* from the unaltered homolog (black) and at neocentromere *URA3nCEN7-II* (red) in the 5'FOA resistant strain LSK465. Bottom panel indicates relative enrichment of Mtw1 at *CEN7* (black) and *URA3nCEN7-II* (blue) at the native centromere (427k) in the strain LSK465. Relative enrichment values of CENPA and Mtw1 indicate that the neocentromere formed on the altered homolog (*URA3nCEN7-II*) was mapped to a region surrounding the integration locus (CaChr7 435078-440387) error bars indicate SEM (***p<0.001, ** p<0.01, ns p>0.05). **(C)** CENPA ChIP-sequencing confirmed the presence of neocentromere in the strain LSK465, where the profile is a combination of two peaks (top), the one at *CEN7* is on the unaltered homolog the one at *URA3nCEN7-II)* is on the altered homolog. A 50 kb region (bottom) harbouring *CEN7* depicts the track height (as on IGV) on *y*-axis and coordinates on the *x*-axis.

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Fig 4.

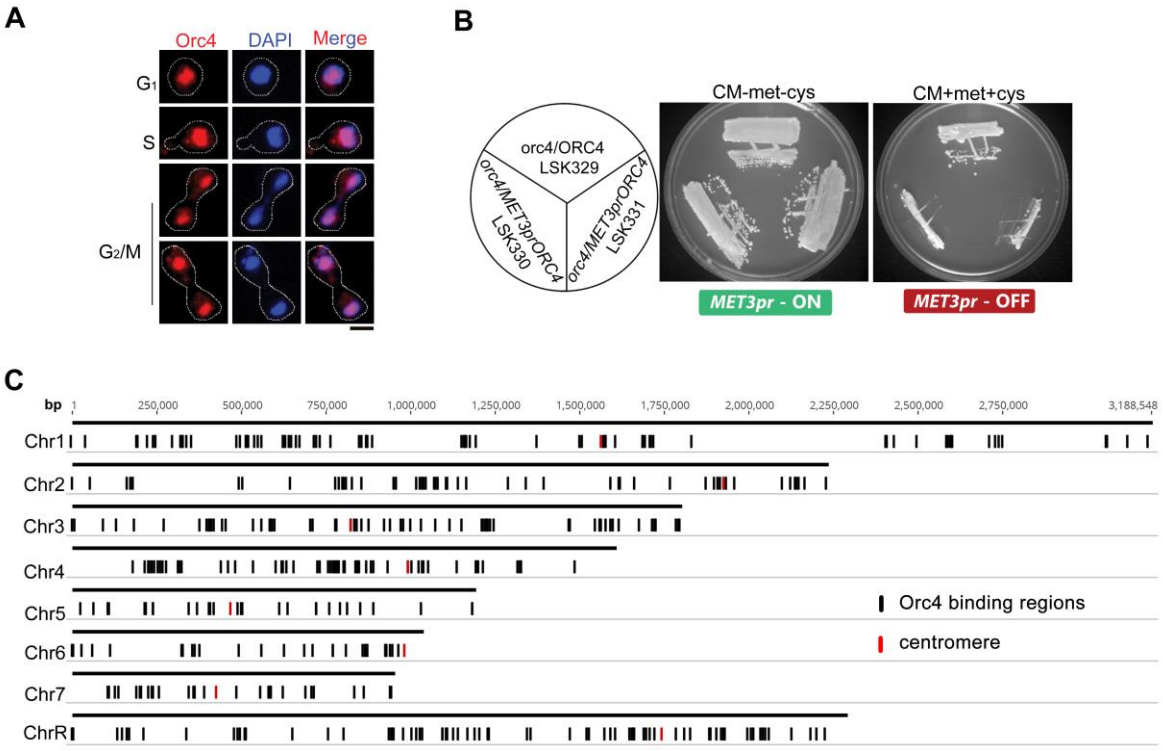

Fig 4. The essential pre-RC component Orc4 binds to discrete loci in the *C. albicans* genome. (A) Intracellular localization of Orc4 in *C. albicans* SC5314 cells stained with anti-Orc4 antibodies (red) and DAPI (blue) indicates Orc4 is localized to the nucleus throughout the cell cycle in *C. albicans.* Bar, 5 µm. (**B**) The promoter of *MET3*, is expressed in the absence of methionine and cysteine and repressed in the presence of both, was used for the controlled expression of *ORC4*. *C. albicans* cells with one deleted copy of *ORC4*, LSK329 (*ORC4*/*ORC4*::*FRT*), and two independent transformants of the conditional mutant of *orc4* (*ORC4*::*FRT*/*MET3prORC4*), LSK330 and LSK331 where the remaining wild-type copy was placed under the control of the *MET3* promoter were streaked on plates containing inducible (CM-met-cys) and repressible (CM+ 5 mM cys + 5 mM met) media photographed after 48 h of incubation at 30°C. (**C**) A whole genome view of Orc4 binding regions (black bars) on each of the eight *C. albicans* chromosomes including all eight centromeres (red bar).

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Fig 5.

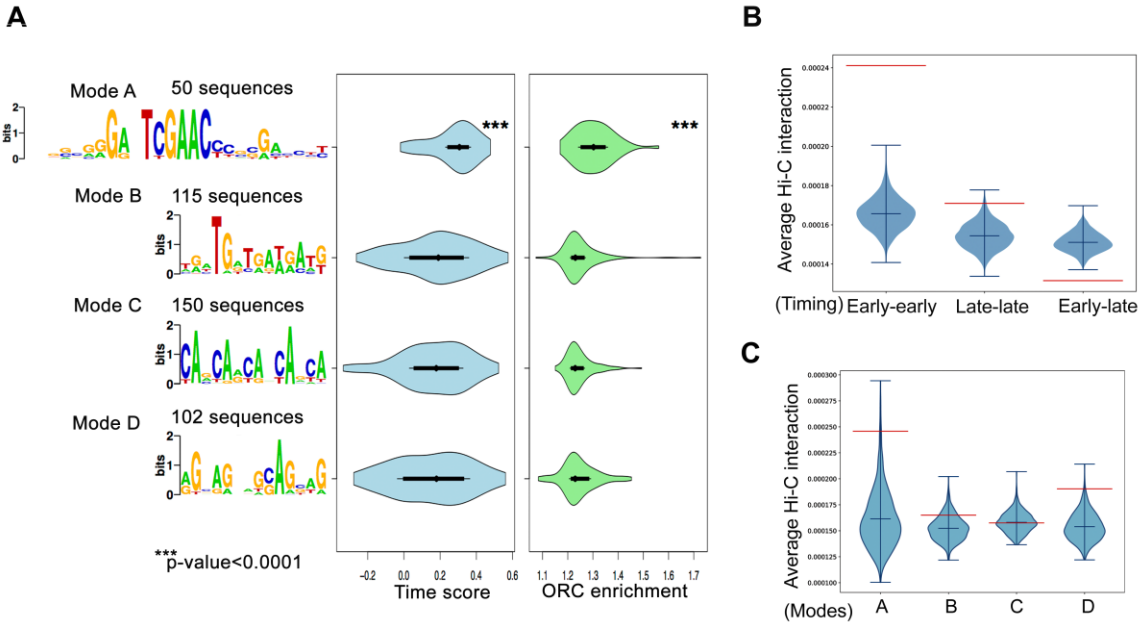

Fig 5. Spatiotemporal distribution of DNA sequence independent replication origins in *C. albicans* (A) The four different modes (motifs) identified by DIVERSITY (A, B, C, D) and their distribution across the 417 binding regions have been listed. Mode A corresponds to the tDNA motif. Violin plots depicting the replication timing profile obtained from (Koren, Tsai et al. 2010) for all four modes shows higher time scores for Mode A associated origins. Enrichment of Orc4 (based on the ChIP-sequencing results) shows a significant increase for Mode A associated origins. The other three modes depict enrichment values not significantly different from each other. (**B**) The average Hi-C interaction (red lines) of the early-early, late-late and early-late regions that have been classified based on their replication timing profile indicate higher interaction values among early replicating domains. Blue violins indicate mean interactions across 1000 randomizations. (**C**) Mode A containing regions show the highest average Hi-C interaction frequency (red lines) among themselves as compared to the three other modes. Blue violins indicate mean interactions across 1000 randomizations.

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Fig 6.

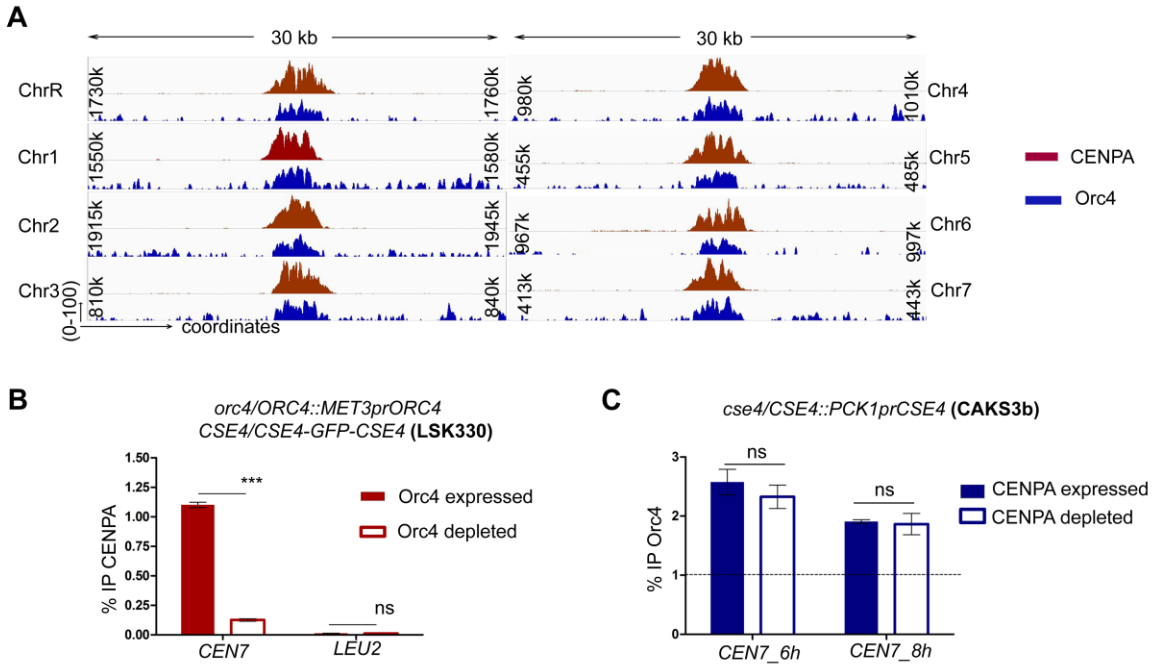

Fig 6. Centromeric localization of Orc4 stabilizes CENPA. (A) Orc4 ChIP sequencing revealed the highest binding of Orc4 at all eight *C. albicans* centromeres. The binding region of Orc4 (blue) showed a complete overlap with CENPA binding region (red). A 30 kb region harbouring each centromere (*x*-axis) was plotted against the subtracted ChIP sequencing reads (*y*-axis) for CENPA and Orc4. **(B)** ChIP-qPCR using anti-GFP antibodies (CENPA) revealed reduced CENPA enrichment at the centromere upon Orc4 depletion in the strain LSK330 either in CM-met-cys or CM+5mM met + 5mM cys for 24h. (**C**) ChIP-q PCR using anti-Orc4 antibodies revealed no significant reduction in centromeric Orc4 when CENPA was depleted for 6h and 8 h in YP with dextrose in the strain CAKS3b. Percent IP values are normalized to a non-centromeric control region, *LEU2*. Two-way ANOVA was used to determine statistical significance (***$p<0.001$, ns $p>0.05$).

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Fig 7.

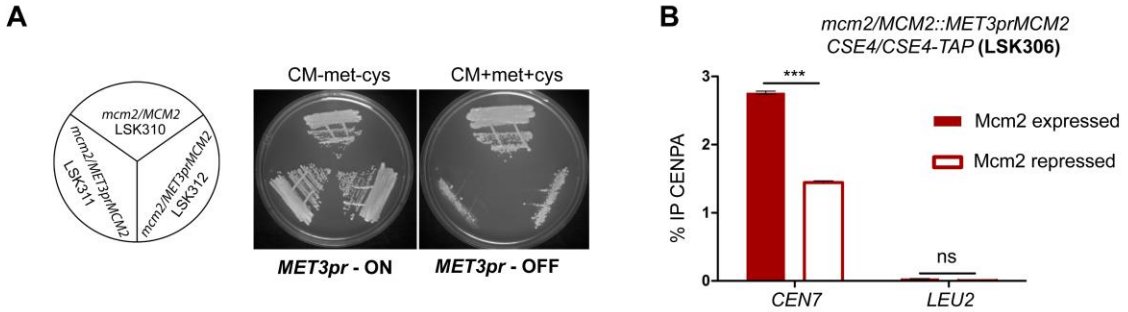

Fig 7. The helicase subunit Mcm2 is essential for viability and CENPA stability in *C. albicans*. **(A)** *C. albicans* cells where one deleted copy of *MCM2* has been deleted, LSK310 (*MCM2*/*MCM2*::*FRT*), and two independent transformants of the conditional mutant of *mcm2* (*MCM2*::*FRT*/*MET3prMCM2*), LSK311 and LSK312 where the remaining wild-type copy was placed under the control of the *MET3* promoter were streaked on plates containing inducible (CM-met-cys) and repressible (CM+5 mM cys + 5 mM met) media and photographed after 48 h of incubation at 30°C. (**B**) ChIP-qPCR using anti-protein A antibodies revealed significant reduction in CENPA at *CEN7* in the strain LSK306 when grown either in CM-met-cys or CM+5mM met+5mM cys for 6 h. ChIP was performed in biological triplicates. Two-way ANOVA was used to determine statistical significance (***$p<0.001$, ns $p>0.05$).

bioRxiv preprint doi: https://doi.org/10.1101/465880; this version posted November 8, 2018. 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 4.0 International license.

Fig 8.

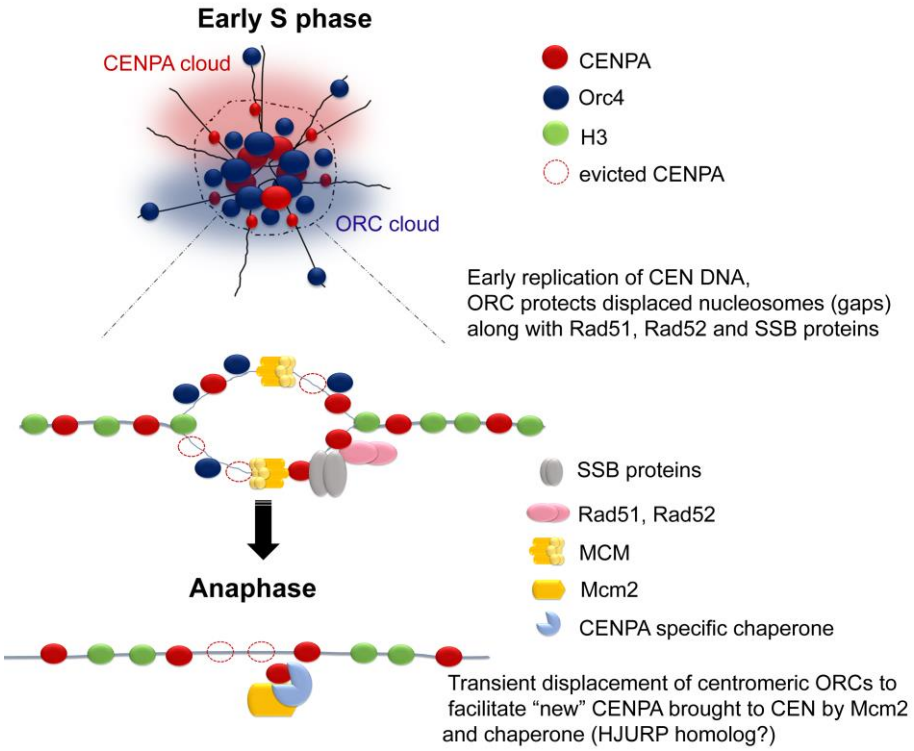

Fig 8. Propagation of CEN chromatin in *C. albicans*. CENPA rich centromeres are surrounded by the peripheral CENPA cloud in a compact chromatin environment. Along with the CENPA cloud, an ORC cloud exists at the centromeres to facilitate timely recruitment of initiator proteins for early CEN replication. The pericentromeres are interspersed with H3 and CENPA nucleosomes. In the early S phase, parental CENPA molecules are evicted from the central core and the gaps created are either occupied by a placeholder (unknown) or, the CEN DNA has to be protected by the physical proximity and association of proteins like Orc4, along with Rad51, Rad52 and single stranded DNA binding proteins. Orc4 maintains centromeric heterochromatin and facilitates early CEN DNA replication. The pre-RC soon replicates other parts of the genome till the end of S phase. During anaphase, Orc4, though otherwise constitutively associated with the CEN chromatin, undergoes a transient dissociation to facilitate new CENPA loading by a specific chaperone (unknown) that is co-chaperoned by Mcm2. Hence, the epigenetic marks are propagated to the subsequent cell cycles.