nature cell biology

ARTICLES

https://doi.org/10.1038/s41556-019-0331-4

DNA replication acts as an error correction mechanism to maintain centromere identity by restricting CENP-A to centromeres

Yael Nechemia-Arbely[1,4,7*], Karen H. Miga[2,7], Ofer Shoshani[1], Aaron Aslanian[3], Moira A. McMahon[1,5], Ah Young Lee[1], Daniele Fachinetti[1,6], John R. Yates III[3], Bing Ren[1] and Don W. Cleveland[1*]

Chromatin assembled with the histone H3 variant CENP-A is the heritable epigenetic determinant of human centromere identity. Using genome-wide mapping and reference models for 23 human centromeres, CENP-A binding sites are identified within the megabase-long, repetitive α-satellite DNAs at each centromere. CENP-A is shown in early G1 to be assembled into nucleosomes within each centromere and onto 11,390 transcriptionally active sites on the chromosome arms. DNA replication is demonstrated to remove ectopically loaded, non-centromeric CENP-A. In contrast, tethering of centromeric CENP-A to the sites of DNA replication through the constitutive centromere associated network (CCAN) is shown to enable precise reloading of centromere-bound CENP-A onto the same DNA sequences as in its initial prereplication loading. Thus, DNA replication acts as an error correction mechanism for maintaining centromere identity through its removal of non-centromeric CENP-A coupled with CCAN-mediated retention and precise reloading of centromeric CENP-A.

Correct chromosome segregation relies on a unique chromatin domain known as the centromere. Human centromeres are located on megabase-long[1] chromosomal regions and are comprised of tandemly repeated arrays of an approximately 171 base pair (bp) element, termed α-satellite DNA[2–4]. CENP-A is a histone H3 variant[5,6] that replaces histone H3 in chromatin assembled onto about 3% of α-satellite DNA repeats[7,8], and is flanked by pericentric heterochromatin containing H3K9me2/3 (ref. [9]). Nevertheless, α-satellite DNA sequences are neither sufficient nor essential for centromere identity[2,10], as demonstrated by several measures including identification of multiple examples of acquisition of a new centromere (referred to as a neocentromere) at a new location coupled with inactivation of the original centromere[11].

This has led to a consensus view that mammalian centromeres are defined by an epigenetic mark[2]. Use of gene replacement in human cells and fission yeast has identified the mark to be CENP-A-containing chromatin[12], which maintains and propagates centromere function indefinitely by recruiting CENP-C and the 16-subunit constitutive centromere associated network (CCAN)[13–16]. We[8] and others[17] have shown that the overwhelming majority of human CENP-A chromatin particles are octameric nucleosomes containing two molecules of CENP-A at all cell cycle points, with heterotypic CENP-A/histone H3-containing nucleosomes comprising at most 2% of CENP-A-containing chromatin[8].

During DNA replication, initially bound CENP-A is quantitatively redistributed to each daughter centromere[18], while incorporation of new CENP-A into chromatin occurs only after exit from mitosis[18,19], when its loading chaperone HJURP (ref. [20,21]) is active[22]. Temporal separation of centromeric DNA replication from new CENP-A chromatin assembly raises the important question of how the centromeric epigenetic mark is maintained across the cell cycle, when it would be expected to be dislodged by DNA replication and diluted at each centromere as no new CENP-A is assembled until the next G1 (ref. [18]). Moreover, endogenous CENP-A comprises only about 0.1% of the total histone H3 variants. Recognizing that a proportion of CENP-A is assembled at the centromeres with the remainder loaded onto sites on the chromosome arms[7,8,23], long-term maintenance of centromere identity and function requires limiting accumulation of non-centromeric CENP-A. Indeed, artificially increasing CENP-A expression in human cells increases ectopic deposition at non-centromeric sites, accompanied by chromosome segregation aberrations[23–26].

Using CENP-A chromatin immunoprecipitation (ChIP) and mapping onto centromere reference models for the centromere of each human chromosome, we now establish that DNA synthesis acts as an error correction mechanism to maintain epigenetically defined centromere identity by mediating precise reassembly of centromere-bound CENP-A chromatin, while removing ectopically loaded CENP-A found within transcriptionally active chromatin outside the centromeres.

Results

CENP-A binding at 23 human centromere reference models. We produced HeLa cells either (1) expressing CENP-A^{LAP}, a CENP-A variant carboxy-terminally fused to a localization (enhanced yellow fluorescent protein, EYFP) and affinity (His) purification tag[27] at one endogenous CENP-A allele (Supplementary Fig. 1a and Fig. 1a), or (2) stably expressing an elevated level (4.5 times the level

[1]Ludwig Institute for Cancer Research and Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, CA, USA. [2]Center for Biomolecular Science & Engineering, University of California Santa Cruz, Santa Cruz, CA, USA. [3]Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA, USA. [4]Present address: Department of Pharmacology & Chemical Biology, School of Medicine, University of Pittsburgh Cancer Institute (UPCI), University of Pittsburgh, Pittsburgh, PA, USA. [5]Present address: Ionis Pharmaceuticals, Carlsbad, CA, USA. [6]Present address: Institut Curie, PSL Research University, CNRS, UMR 144, Paris, France. [7]These authors contributed equally: Y. Nechemia-Arbely, K. H. Miga. *e-mail: arbelyy@upmc.edu; dcleveland@ucsd.edu

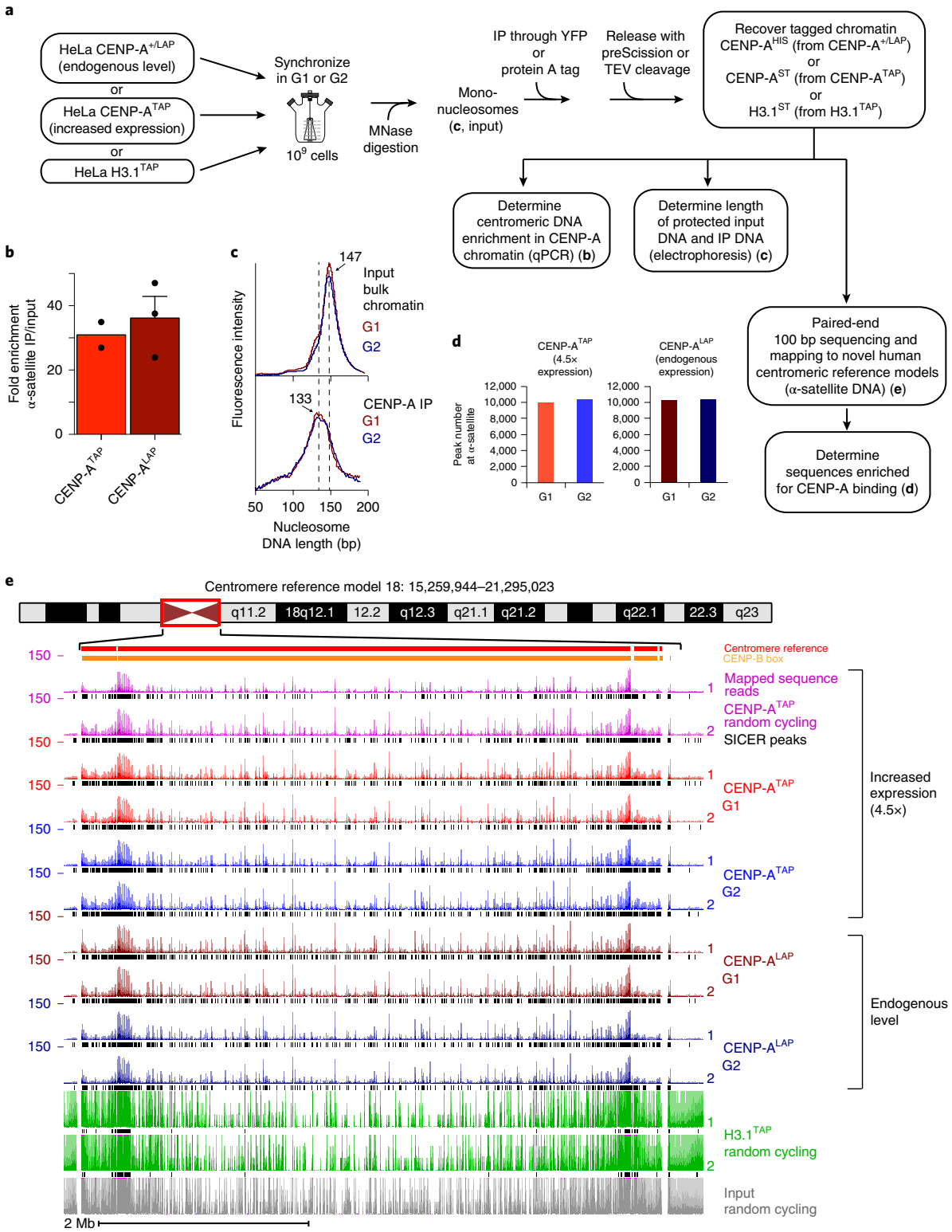

Fig. 1 | CENP-A ChIP-seq identifies CENP-A binding at reference centromeres of 23 human chromosomes. (**a**) CENP-A ChIP-seq experimental design. TEV, tobacco etch virus; qPCR, quantitative real-time PCR. (**b**) Quantitative real-time PCR for α-satellite DNA in chromosomes 1, 3, 5, 10, 12 and 16. $N=2$ for CENP-A^{TAP} and 3 for CENP-A^{LAP}, from biologically independent replicates. Error bars, s.e.m. (**c**) MNase digestion profile showing the nucleosomal DNA length distributions of bulk input mononucleosomes (upper panel) and purified CENP-A^{LAP} following native ChIP at G1 and G2. The experiment was repeated independently twice with similar results. (**d**) Number of CENP-A binding peaks for α-satellite DNA in CENP-A^{TAP} and CENP-A$^{+/LAP}$ cells at G1 and G2. The numbers represent peaks that overlap between the two biologically independent replicates. (**e**) CENP-A ChIP-seq shows CENP-A binding peaks at the centromere of chromosome 18 for CENP-A^{LAP} and CENP-A^{TAP} before and after DNA replication. CENP-A peaks across the reference model are a result of multimapping and their exact linear order is not known. SICER peaks are shown in black below the raw read data. *Y* axis shows read counts. Two replicates are shown for each condition. Source data for **b** and **d** can be found in Supplementary Table 4.

of CENP-A in parental cells) of CENP-A^{TAP}, a CENP-A fusion with carboxy-terminal tandem affinity purification (S protein and protein A) tags separated by a tobacco etch virus protease cleavage site (Supplementary Fig. 1b and Fig. 1a). Both CENP-A variants localize to centromeres (Supplementary Fig. 1c,d), support long-term centromere identity, and mediate high-fidelity chromosome segregation in the absence of wild-type CENP-A (refs. [7,8]). Importantly, while the HeLa cells we use have acquired an aneuploid genome, they are chromosomally stable, with a high-fidelity centromere function that has maintained the same karyotype over almost two decades[28].

To immunopurify CENP-A-containing chromatin at G1 or G2, chromatin was isolated from synchronized cells (Fig. 1a and Supplementary Fig. 1e). Nuclease digestion was used to produce mononucleosomes from chromatin isolated at G1 and G2, yielding the expected 147 bp of protected DNA length for nucleosomes assembled with histone H3 (Fig. 1a,c, upper panel). In parallel, chromatin was also isolated from randomly cycling cells stably expressing TAP-tagged histone H3.1 (H3.1^{TAP}—Supplementary Fig. 1b and Fig. 1a)[13]. CENP-A^{LAP}-, CENP-A^{TAP}- or H3.1^{TAP}-containing chromatin was then affinity purified and eluted under mild conditions with PreScission or tobacco etch virus protease cleavage (Fig. 1a).

α-satellite DNA sequences were enriched 30–35-fold in DNA isolated from CENP-A^{TAP} or CENP-A$^{+/LAP}$ cells (Fig. 1b), the expected enrichment since α-satellite DNA comprises about 3% of the genome[8,29]. While microcapillary electrophoresis of bulk input chromatin produced the expected 147 bp of protected DNA length for nucleosomes assembled with histone H3 (Fig. 1c, upper panel), isolated CENP-A^{LAP} chromatin expressed at endogenous CENP-A levels produced DNA lengths centred on 133 bp, before and after DNA replication (Fig. 1c, lower panel), as previously reported for octameric CENP-A-containing nucleosomes with DNA unwinding at entry and exit[8,30].

Affinity-purified CENP-A^{LAP}-, CENP-A^{TAP}- and H3.1^{TAP}-bound DNAs were sequenced and mapped (Fig. 1a,d and Supplementary Table 1) onto the centromere reference model of the human X chromosome[31] and centromere models (incorporated into the HuRef genome hg38) for each human autosome that include the observed variation in α-satellite higher-order repeat (HOR) array sequences[32,33]. Sequences bound by CENP-A were identified (Fig. 1d,e and Supplementary Fig. 2) using algorithm-based scripts (SICER and MACS, refs. [34,35]). CENP-A^{LAP} expressed at endogenous levels mapped with high reproducibility across the centromeric regions of all 23 reference centromeres (see Fig. 1e for chromosome 18 and Supplementary Fig. 2 for the other 22). Sequences bound were largely unaffected by increasing CENP-A levels 4.5-fold in CENP-A^{TAP} cells (Figs. 1e and 2a,b).

Centromeric α-satellite arrays varied widely in CENP-A binding, from 10.5-fold enrichment for array D3Z1 in cen3 to 213-fold for array GJ211930.1 in cen10, and with most enriched 20–40-fold relative to input DNAs (Supplementary Table 2). For the 6 (of 17) centromeres that contain more than one α-satellite array, only one array actively bound CENP-A (Supplementary Table 2). Multiple α-satellite arrays in 11 centromeres (Supplementary Table 2) showed enriched CENP-A binding in two or more arrays, consistent with CENP-A binding to a different array in each homologue, as previously shown for cen17 in two diploid human cell lines[36]. The increased levels of CENP-A in CENP-A^{TAP} cells did not increase the number of centromeric binding peaks (Fig. 1d,e), but did elevate CENP-A occupancy at some divergent monomeric α-satellite repeats (Supplementary Fig. 1f) or within HORs (Supplementary Fig. 1g), with both examples occurring in regions with few CENP-B boxes.

CENP-A nucleosomes are retained at centromeric loading sites after DNA replication. Despite the known redistribution of initially centromere-bound CENP-A onto each of the new daughter centromeres without addition of new CENP-A (ref. [18]), comparison of the sequences bound by CENP-A in G1 with those bound in G2 revealed that for all 23 centromeres, at both normal (CENP-A$^{+/LAP}$) and elevated (CENP-A^{TAP}) levels, CENP-A was bound to indistinguishable α-satellite sequences before and after DNA replication (Fig. 1e and Supplementary Fig. 2). Indeed, almost all (87%) of α-satellite binding peaks algorithmically identified for CENP-A^{TAP} during G1 remained at G2 (Supplementary Fig. 1h, top). A similarly high (89%) retention of CENP-A peaks found in G1 remained at G2 in CENP-A^{LAP} cells with CENP-A expressed at endogenous levels (Supplementary Fig. 1h, bottom). After filtering out multi-mapping reads, 96 single-copy, centromeric CENP-A binding sites were identified within the HORs of the 23 reference centromeres (Fig. 2). Remarkably, examination of these before and after DNA replication in CENP-A^{TAP} cells revealed quantitative retention of CENP-A in G2 in almost all (93 of 96) of these unique centromeric sites, with the remaining three peaks only slightly diminished (Fig. 2 and Supplementary Fig. 1i).

Ectopic CENP-A assembled onto chromosome arms in early G1 is removed by G2. In addition to the striking enrichment at centromeric α-satellites, genome-wide mapping of CENP-A-bound DNAs revealed preferential and highly reproducible incorporation into unique sequences, non-α-satellite sites on the arms of all 23 chromosomes (Figs. 3a,b and 4). At endogenous CENP-A levels, 11,390 ectopic sites were identified, 620 of which were loaded ≥5-fold over background (Fig. 3d; Supplementary Fig. 3a). Sites enriched for bound CENP-A were essentially identical in DNAs from randomly cycling cells or G1 cells (Fig. 3a,b). While a 4.5-fold increase in CENP-A levels in CENP-A^{TAP} cells did not increase the binding peaks (Fig. 1d) or the number of unique single-copy sites within centromeric HORs (Supplementary Fig. 1i, bottom), it drove an increased number of sites of CENP-A incorporation on the arms, producing 40,279 non-centromeric sites (Fig. 3d), 12,550 of which were loaded ≥5-fold over background (Supplementary Fig. 3a).

Remarkably, for all 23 human chromosomes and for CENP-A accumulated to endogenous (CENP-A^{LAP}) or increased (CENP-A^{TAP}) expression levels, passage from G1 to G2 almost eliminated enrichment of CENP-A binding to specific sites on the chromosome arms, while leaving α-satellite-bound sequences unaffected (Figs. 1d, 3 and 4). Loss by G2 of CENP-A binding at specific arm sites was highly reproducible (see experimental replicas in Fig. 3b). Scoring peak binding sites with thresholds of ≥5-, 10- or 100-fold of CENP-A binding over background, at least 90% of sites bound on chromosome arms in G1 in CENP-A^{TAP} cells were removed by early G2 (Supplementary Fig. 3a), and all of those still identified in G2 were substantially reduced in peak height.

Ectopic CENP-A removal after DNA replication was confirmed using CENP-A ChIP following synchronization in G1 or mitosis (Supplementary Fig. 3b) in a second, nearly diploid human cell line (DLD1) in which the two CENP-A alleles were modified to encode a degron-tagged, auxin-inducible degradable CENP-A^{AID} or a doxycycline-inducible CENP-A^{WT} (ref. [37]). Levels of CENP-A loaded at each of four ectopic sites before (G1) and after DNA replication revealed that almost all (85–90%) of ectopic CENP-A loaded in G1 was removed by mitotic entry (Fig. 3e).

Neocentromeres are not at sites of ectopic CENP-A loading. Recognizing that ectopic loading of CENP-A on chromosomes could be one component of neocentromere formation, we tested if the positions of known human neocentromeres[38] are sites of preferential ectopic CENP-A loading. Despite cytogenetic positioning of many reported neocentromeres[11], only two have been precisely mapped[39]. The first (PDNC4) spans 300 kilobases (kb) on chromosome 4 (ref. [40]) (87.278 to 87.578 megabases (Mb) in hg38). In CENP-A^{TAP} cells, only four sites of elevated CENP-A were present

in G1 within the genomic region of this neocentromere, only two of which had CENP-A loading more than fivefold over the background. Importantly, all sites were removed in G2-derived chromatin (Fig. 5a). An additional neocentromere (IMS13q) maps to a 100 kb region on chromosome 13 (ref. [41]) (97.047 to 97.147 Mb in hg38 assembly[39]). CENP-A binding to this region in CENP-A^{TAP} cells in G1 did not differ in density of peaks or peak heights from many similarly loaded sites scattered across the long arm of chromosome 13 (Supplementary Fig. 3c). Passage from G1 to G2 of CENP-A^{TAP} cells stripped almost all CENP-A bound that corresponded to the region of the IMS13q neocentromere. While we cannot exclude the possibility that other cell types have different epigenetic landscapes that affect the sites to which CENP-A binds at non-centromeric regions, our examination of these two best defined neocentromeres offers no support for neocentromere formation arising at the site of an inherent hotspot of ectopic CENP-A loading.

CENP-A is ectopically loaded at early G1 into open/active chromatin. The sites on the chromosome arms into which CENP-A was assembled in G1 in CENP-A^{TAP} cells were enriched twofold (compared with levels expected by chance) at promoters or enhancers of expressed genes, with a 2.5-fold enrichment at sites bound by the transcriptional repressor CTCF (Supplementary Fig. 3e), trends similar to previous reports for cells with increased CENP-A[23]. More than 80% of CENP-A^{TAP} binding sites on chromosome arms with peak heights ≥5-fold over background (Fig. 5b–d) overlapped with DNase I hypersensitive, accessible chromatin sites identified by the ENCODE project that are functionally related to transcriptional activity. Similarly, CENP-A expressed at endogenous levels was enriched threefold at DNase I hypersensitive sites (Fig. 5e) and promoters (Supplementary Fig. 3g).

Most (80%) of non-centromeric CENP-A^{TAP} binding peaks overlapped with H3k4me1 and H3k4me2 sites found in active and primed enhancers and at transcription factor binding sites identified by ENCODE, with a similar trend for CENP-A^{LAP} (Fig. 5d,e). Ectopic CENP-A^{TAP} or CENP-A^{LAP} (Supplementary Fig. 4a,b) peaks also showed a significant overlap with other marks of active transcription, including H2A.Z, H3K4me3, H3K27ac, H3K36me3 and H3K9ac. Conversely, both CENP-A^{TAP} and CENP-A^{LAP} were not enriched at H3k27me3 sites tightly associated with inactive gene promoters and facultatively repressed genes (Supplementary Figs. 3d,f and 4a,b). CENP-A binding peaks showed a mild (30–40%) overlap with histone modifications H4K20me1 and H3K79me2 (active transcription marks) or the H3K9me3 mark of transcription repression (Supplementary Fig. 4a,b). Overall, most (65% and 93%, respectively) of the ectopic CENP-A sites in cells with endogenous or elevated CENP-A were associated with any active transcription mark (Supplementary Fig. 4a,b), consistent with ectopic CENP-A preferentially bound to open, active chromatin.

Ectopic CENP-A in G1 in either CENP-A^{TAP} or CENP-A^{LAP} cells bound to 'high-occupancy target' regions defined by highly expressed regions of the genome (and that show binding of unrelated transcription factors without underlying sequence specificity[42])

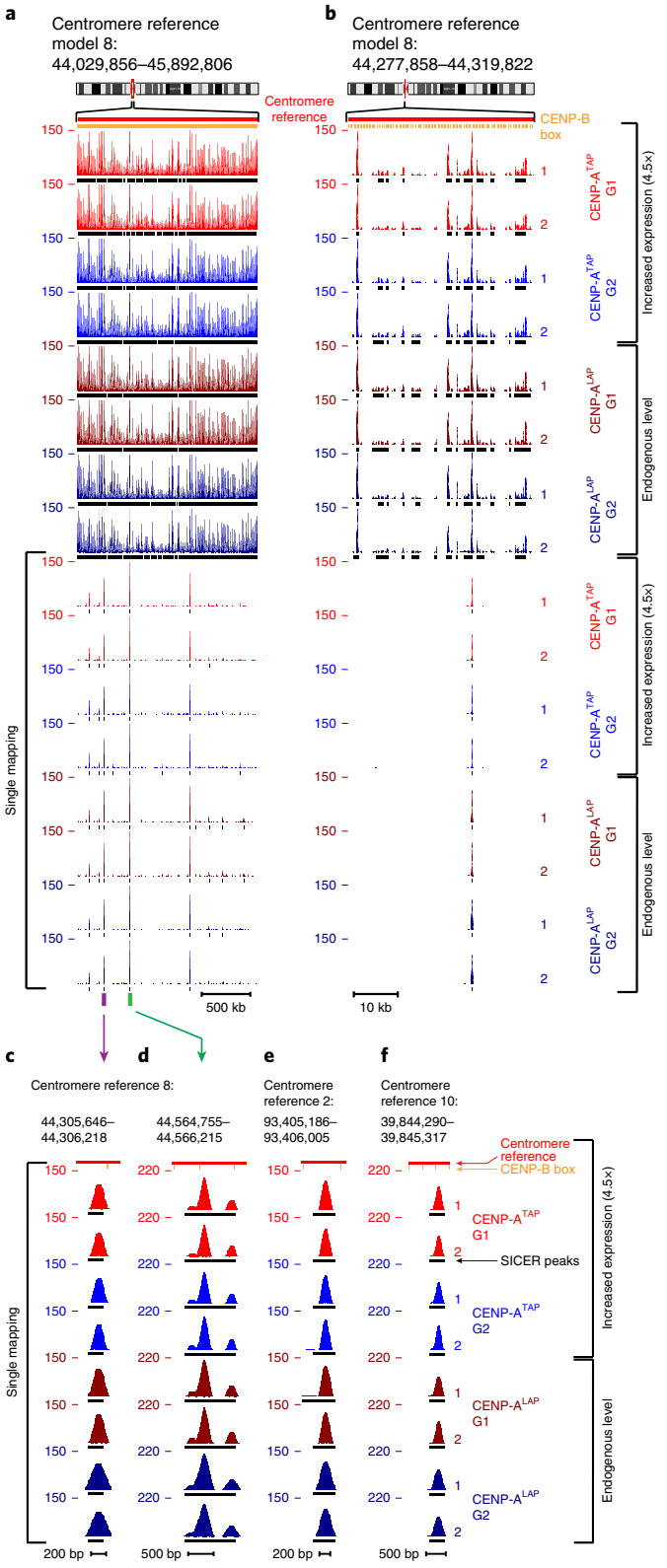

Fig. 2 | Retention of centromeric CENP-A through DNA replication. (**a**,**b**) CENP-A ChIP-seq raw mapping data (coloured) and SICER peaks (black lines, underneath) showing sequences bound by CENP-A (at both endogenous and increased expression levels) across the centromere reference model of chromosome 8, before and after DNA replication. Upper part, mapping of all reads (including reads that are multimapping) onto the repetitive α-satellite DNA. Lower part, read mapping to sites that are single copy in the HuRef genome (single mapping), after filtering out multimapping reads. Scale bar, 500 kb (**a**), 10 kb (**b**). (**c**) High-resolution view of read mapping to a site that is single copy in the centromere reference model of chromosome 8, marked by a purple bar in **a**. (**d**) High-resolution view of read mapping to a site that is single copy in the centromere reference model of chromosome 8, marked by a green bar in **a**. (**e**,**f**) High-resolution view of read mapping to a site that is single copy in the centromere reference model of chromosome 2 (**e**) and in the centromere reference model of chromosome 10 (**f**). The experiment shown in **a**–**f** was repeated independently twice with similar results.

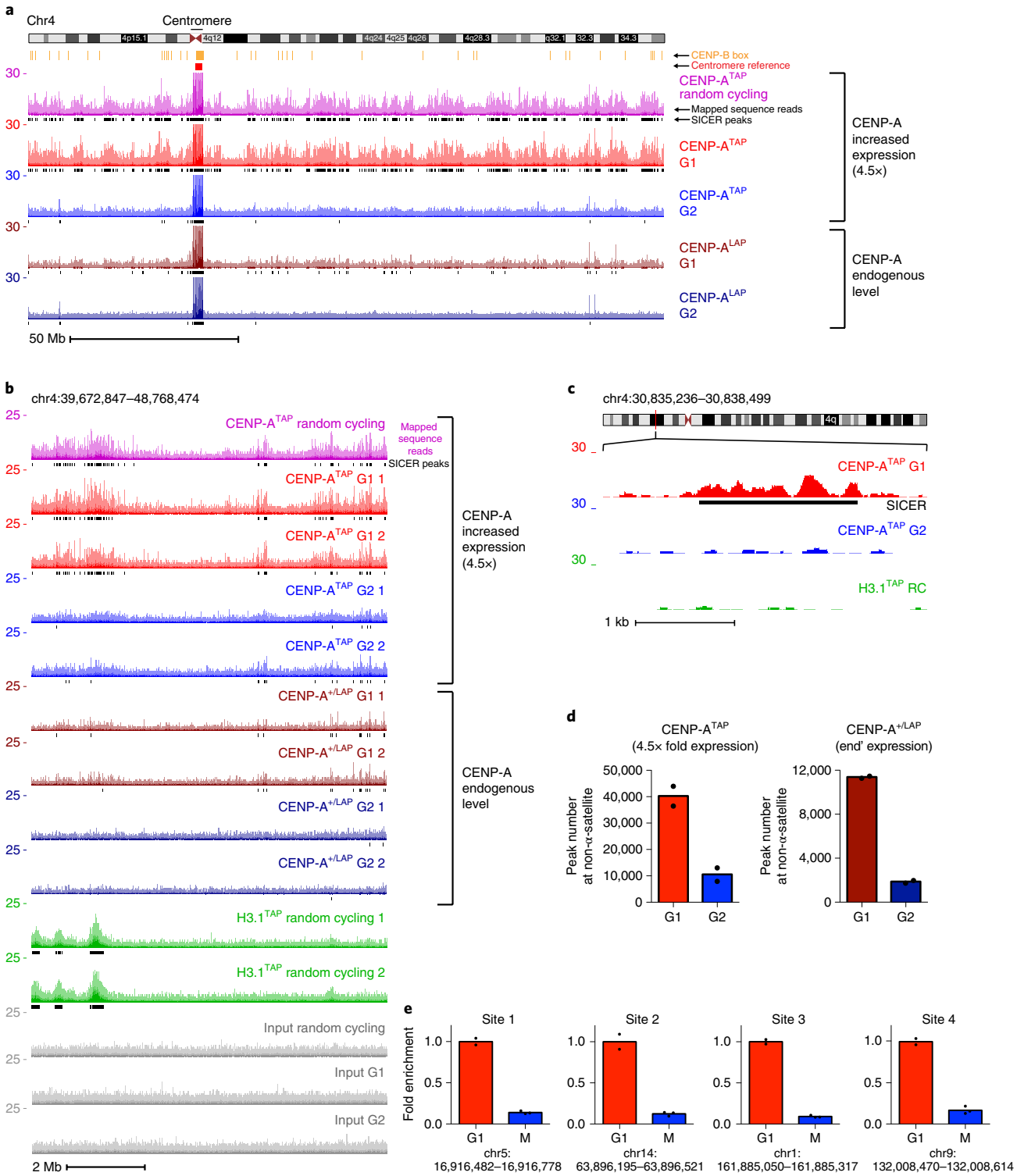

Fig. 3 | Sites of CENP-A assembly onto chromosome arms in early G1 are removed by G2. (**a**) ChIP-seq raw mapping data (coloured) and SICER peaks (black lines, underneath) showing sequences bound by CENP-A (at both endogenous and increased expression levels) across chromosome 4 before and after DNA replication. Read counts were scaled to 30 but reached 150 at the centromere. (**b**) ChIP-seq data for a region within the p-arm of chromosome 4, with two replicates for each time point, for CENP-A^{TAP} (increased CENP-A expression), CENP-A^{LAP} (endogenous level) and H3.1^{TAP}. (**c**) High-resolution nucleosomal view of CENP-A^{TAP} mapping data at G1 and G2 at a non-centromeric site of chromosome 4. The experiments in **a**–**c** were repeated independently twice with similar results. (**d**) Total number of non-α-satellite CENP-A binding sites for CENP-A^{TAP} and CENP-A^{LAP} at G1 and G2. The numbers represent the average of the two sequencing replicates per time point. (**e**) Quantitative real-time PCR following CENP-A ChIP from DLD1 cells with auxin-degradable CENP-A^{AID} and a doxycycline-inducible CENP-A^{WT} (ref. [37]) after synchronization in G1 or in mitosis (as shown in Supplementary Fig. 3b) for sites on the arms of chromosomes 1, 5, 9 and 14. Sites for quantitative real-time PCR were chosen based on identification of ectopic deposition of CENP-A at these locations in HeLa cells. Levels of CENP-A enrichment at mitosis were normalized to the level of enrichment at G1. Results of two independent experiments for G1 and three for mitosis are shown. Source data for **d** and **e** can be found in Supplementary Table 4.

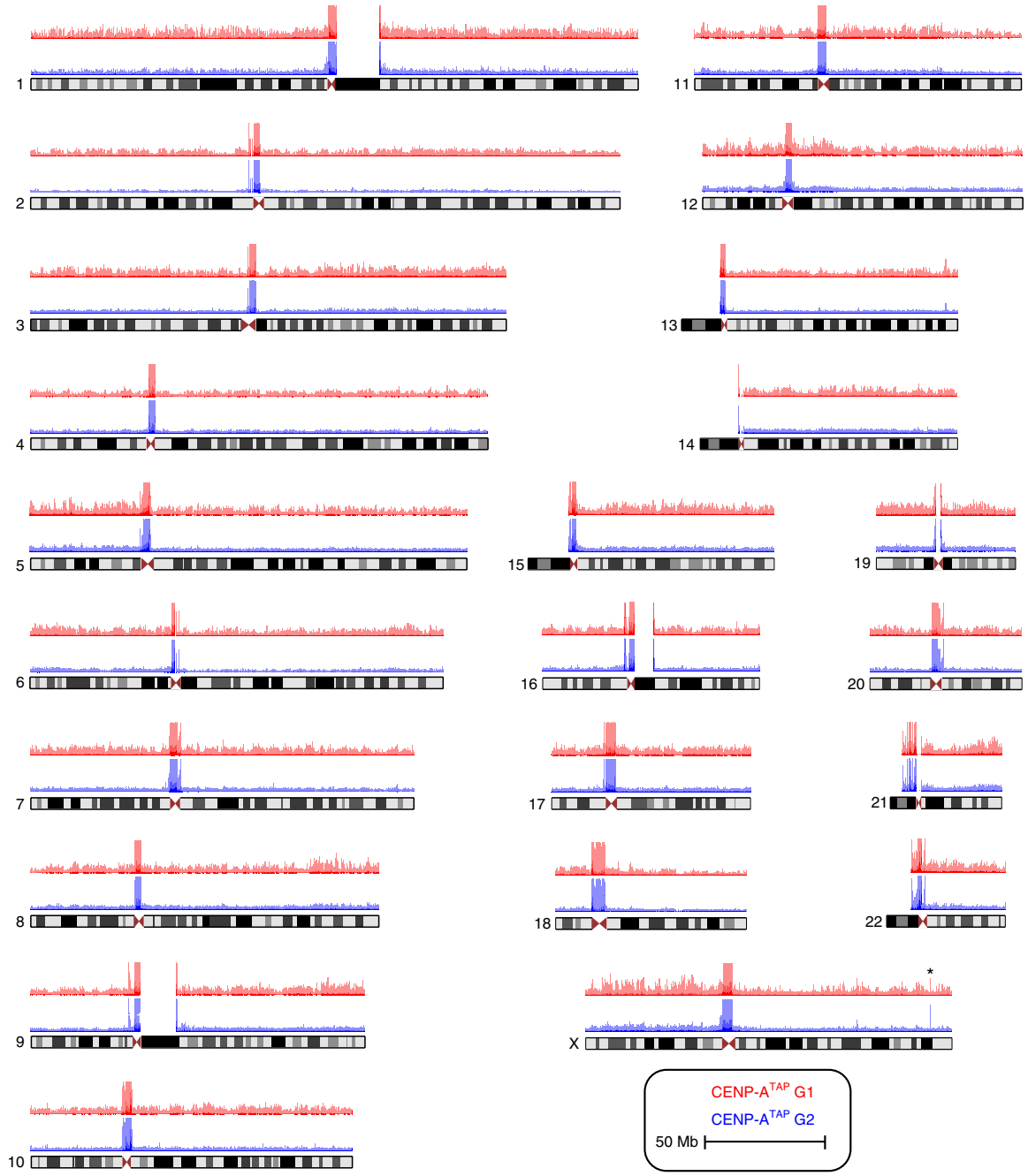

Fig. 4 | Ectopic CENP-A is removed following DNA replication from the arms of all 23 human chromosomes. CENP-A^{TAP} ChIP-seq raw mapping data at G1 and G2 for all human chromosomes. Chromosome X shows a spike of CENP-A enrichment not removed by G2 (marked by an asterisk). The experiment was repeated independently twice with similar results.

were almost quantitatively removed from cells that had progressed through DNA replication (Supplementary Fig. 3h,i), demonstrating that enrichment of CENP-A in such highly expressed regions cannot be a consequence of non-specific binding to 'hyper-ChIPable' regions[43].

Analysis of published CENP-A ChIP sequencing (ChIP-seq) datasets for HT1080 cells[44], a human epithelial fibrosarcoma cell line expressing Flag-tagged CENP-A at about three times the level of parental CENP-A, revealed similar trends: ectopic sites were enriched at DNase I hypersensitive sites and at transcription activation marks (H2A.Z and H3K4me1/2) but were not enriched at the transcription repression mark H3K27me3 (Supplementary Fig. 4c). Similarly, CENP-A ChIP-seq datasets from the HuRef human lymphoblastoid cell line[45] also revealed that the majority (51%) of ectopic CENP-A accumulated to its endogenous level was found at marks associated with transcription activation (including H2A.Z, H3K36me3 and H4K20me1). CENP-A was not enriched at the transcription repression mark H3K27me3 (Supplementary Fig. 4d), although in all three cell lines analysed ectopic CENP-A was enriched at H3K9me3 sites associated with heterochromatin of constitutively repressed genes (Supplementary Fig. 4a–d).

Ectopic, but not centromeric, CENP-A is removed by replication fork progression. We next tested whether removal by G2 of CENP-A assembled into nucleosomes at unique sites on the chromosome arms is mediated by the direct action of the DNA

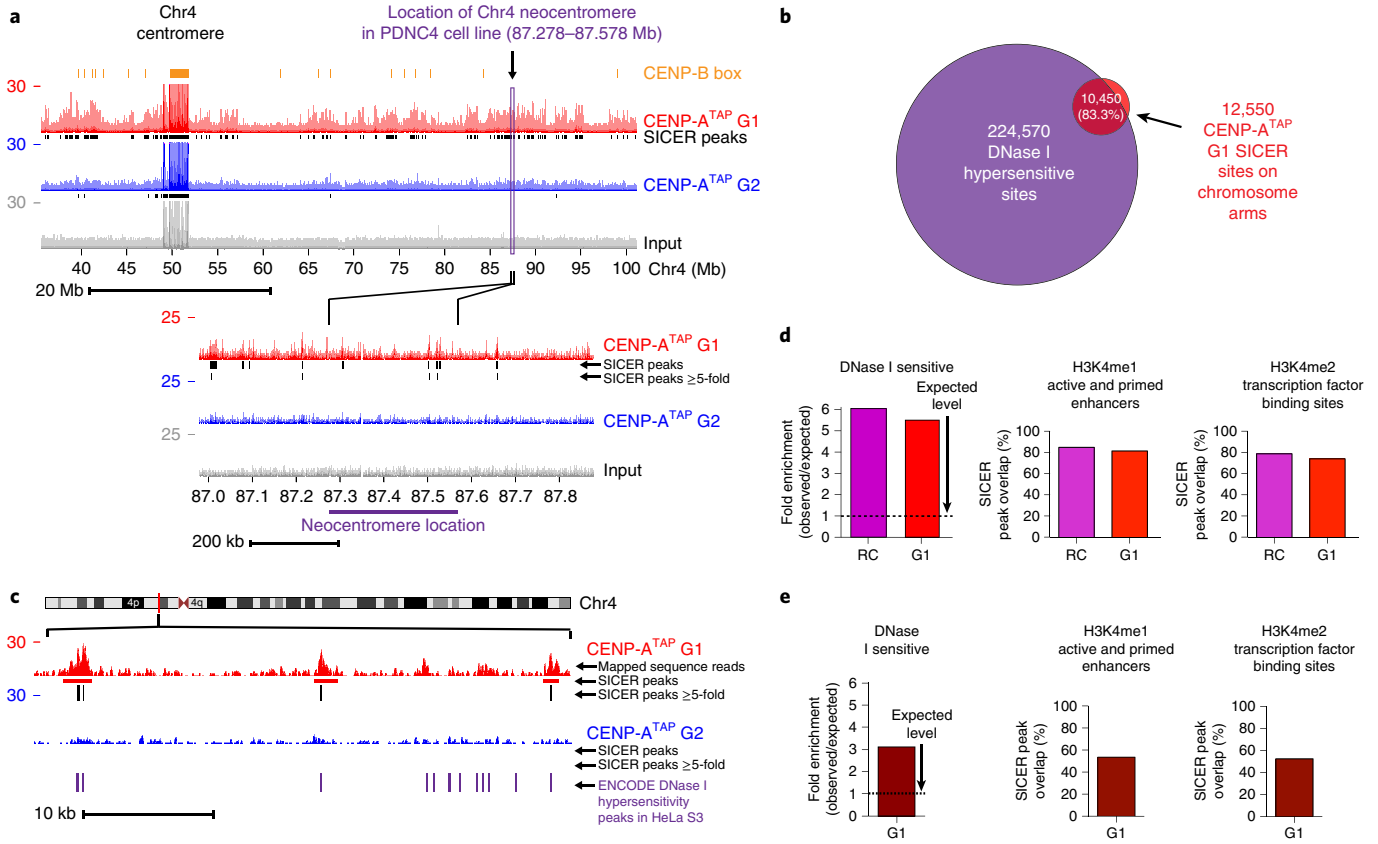

Fig. 5 | Neocentromeres are not positioned at sites of preferential CENP-A loading on chromosome arms. (**a**) Read mapping data of CENP-A^{TAP} ChIP-seq at G1 and G2, at the chromosomal location of a known patient-derived neocentromere[39] found in chromosome 4. The experiment was repeated independently twice with similar results. (**b**) More than 80% of CENP-A SICER peaks (≥5-fold over background) in randomly cycling and G1 cells overlap with DNase I hypersensitive sites taken from the ENCODE project. (**c**) Example from the chromosome 4 p-arm showing overlap of at least 100 bases between SICER peaks (≥5-fold over background) and HeLa S3 DNase I hypersensitive sites taken from the ENCODE project. The experiment was repeated independently twice with similar results. (**d**,**e**) CENP-A^{TAP} (**d**) or CENP-A^{LAP} (**e**) enrichment levels at DNase I hypersensitive sites, H3K4me1 and H3K4me2 sites. SICER peaks (≥5-fold over background) supported between the two replicates were analysed for their enrichment level at DNase I hypersensitive sites, with minimum overlap of 100 bases, compared with the level of enrichment at these sites by chance. SICER peaks (≥5-fold over background) supported between the two replicates of CENP-A^{TAP} (**d**) or CENP-A^{LAP} (**e**) were analysed for their degree of overlap with known sites of H3K4me1 and H3K4me2 binding taken from ENCODE HeLa S3 datasets. Source data for **d** and **e** can be found in Supplementary Table 4.

replication machinery. CENP-A^{TAP} was affinity purified from mid-S-phase cells and CENP-A-bound DNAs were sequenced and mapped (Fig. 6a and Supplementary Fig. 5a). In parallel, newly synthesized DNA in synchronized cells was labelled by addition of bromodeoxyuridine (BrdU) for 1 h at early (S0–S1), mid- (S3–S4) and late S phase (S6–S7) (Fig. 6a and Supplementary Fig. 5a). Genomic DNA from each time was sonicated (Supplementary Fig. 5b) and immunoprecipitated with a BrdU antibody (Fig. 6a). Eluted DNA was then sequenced and mapped to the genome (an approach known as Repli-seq[46]), yielding regions of early-, mid- and late-replicating chromatin (an example from a region of an arm of chromosome 20 is shown in Fig. 6b). Early replication timing was validated (Supplementary Fig. 5c) for two genes (MRGPRE and MMP15) previously reported to be early replicating (ENCODE Repli-seq[46]). Similarly, a gene (HBE1) and a centromeric region (Sat2) previously reported to be late replicating (ENCODE Repli-seq[46]) were confirmed to be replicated late (Supplementary Fig. 5d).

CENP-A chromatin immunoprecipitated from early- and mid-S-phase cells yielded levels of α-satellite DNA enrichment (Supplementary Fig. 5e) similar to those achieved at G1 phase (Fig. 1b). Furthermore, nucleosomal CENP-A chromatin produced by micrococcal nuclease digestion protected 133 bp of DNA at early and mid-S phase (Supplementary Fig. 5f), just as it did in G1 and G2 (Fig. 1c; see also ref. [8]), with no evidence for a structural change from hemisomes to nucleosomes and back to hemisomes during S phase as previously claimed[47]. Mapping of CENP-A binding sites on chromosome arms, combined with Repli-seq, revealed that almost all (91%) ectopic G1 CENP-A binding was found in early- or mid-S-replicating regions (Fig. 6b,c). While alphoid DNA sequences have been reported to replicate in mid-to-late[48] S phase, in our cells α-satellite-containing DNAs in all 23 centromeres were found almost exclusively to be late replicating (Fig. 6d).

Remarkably, throughout S phase, centromere-bound CENP-A found in G1 was completely retained across each reference centromere with the same sequence binding preferences (Fig. 6e and Supplementary Fig. 5g). Retention of CENP-A binding during DNA replication was also observed at the unique-sequence-binding sites within the HORs of each centromere: all 96 CENP-A^{TAP} G1 peaks at single-copy variants within α-satellite HORs remained bound by CENP-A (Fig. 6f and Supplementary Fig. 5h). In contrast, early-replicating ectopic CENP-A-binding sites were nearly quantitatively removed during or quickly after their replication and were no longer visible in mid-S phase (Fig. 6g,h). Similarly, ectopic CENP-A-binding sites found in mid-S-replicating regions remained

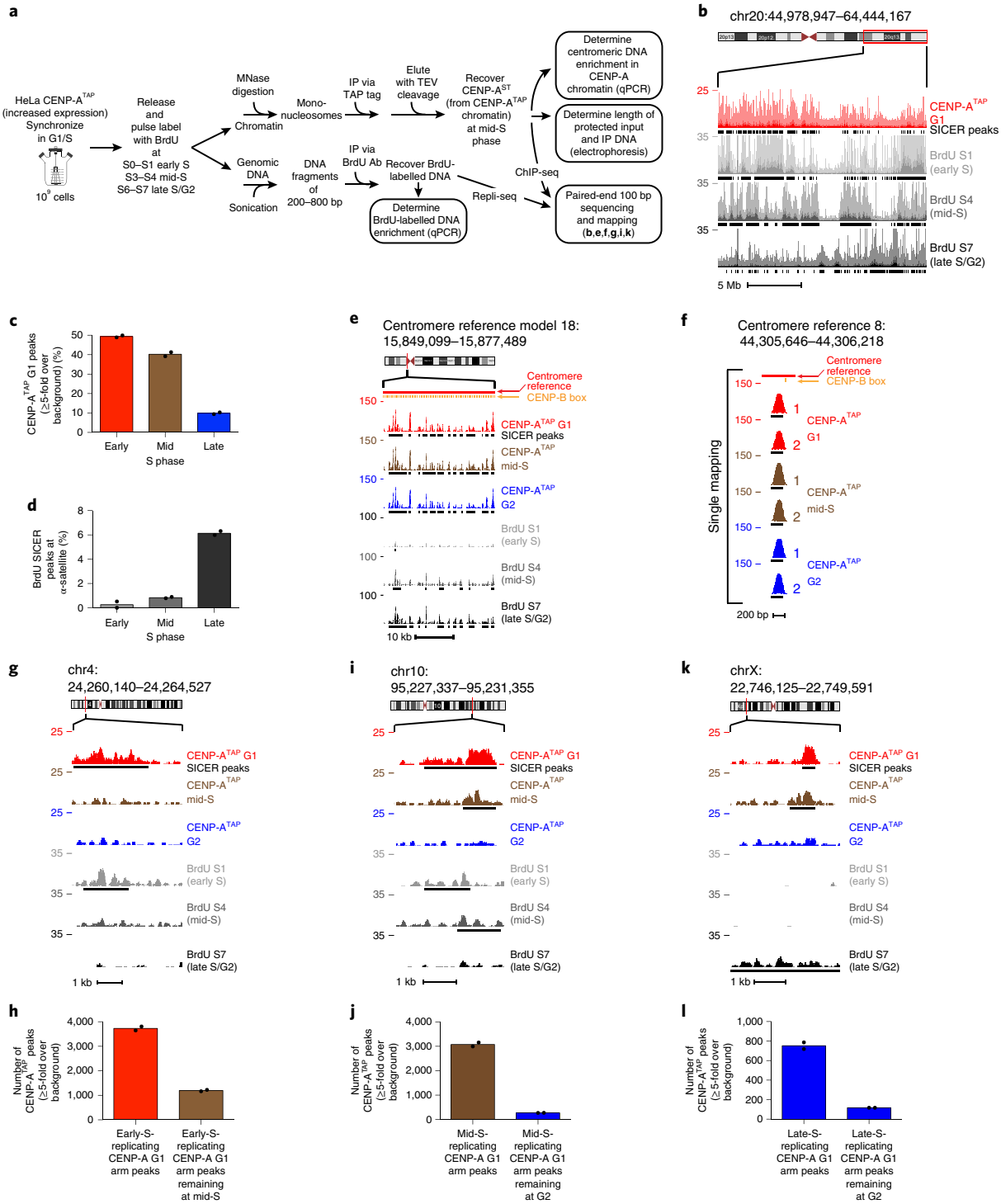

Fig. 6 | Ectopic CENP-A is removed contemporaneously with replication fork progression, while centromeric CENP-A is retained. (**a**) Experimental design of CENP-A ChIP-seq combined with Repli-seq. (**b**) Raw mapping data of CENP-A^{TAP} ChIP-seq at G1 and BrdU Repli-seq at early S, mid-S and late S/G2 at the q-arm of chromosome 20. (**c**) Percentage of ectopic G1 CENP-A^{TAP} ≥5-fold binding sites in early-, mid- or late-S-replicating regions. (**d**) Percentage of BrdU SICER peaks in α-satellite DNA found within early-, mid- or late-S-replicating regions. (**e**) CENP-A ChIP-seq raw mapping data at cen18 at G1, mid-S phase and G2, and BrdU Repli-seq at early S, mid-S and late S/G2. (**f**) High-resolution view of CENP-A mapping during DNA replication (mid-S) at a single-copy variant (marked by a purple bar in Fig. 2a) in the centromere reference model of chromosome 8. Data from Fig. 2c for CENP-A^{TAP} G1 and G2 are included for comparison. (**g**,**i**,**k**) Raw mapping data and SICER peaks of CENP-A^{TAP} ChIP-seq at G1, mid-S phase and G2, and BrdU-labelled Repli-seq samples showing regions going through replication in early S, mid-S and late S/G2 phase within the p-arm of chromosome 4 (**g**), the q-arm of chromosome 10 (**i**) and the p-arm of chromosome X. (**h**) Number of early replicating CENP-A^{TAP} G1 peaks (≥5-fold over background) retained at mid-S phase. (**j**) Number of mid-S-replicating CENP-A^{TAP} G1 peaks (≥5-fold over background) retained at G2. (**l**) Number of late-replicating CENP-A^{TAP} G1 peaks (≥5-fold over background) retained at G2. Source data for **c**,**d**,**h**,**j**,**l** can be found in Supplementary Table 4. Data shown in **b**–**l** are from two biologically independent experiments.

Table 1 | Mass spectrometry of CENP-A individual nucleosomes reveals all the CCAN network components coprecipitated with CENP-A at late S/G2

		Randomly cycling		Late S/G2	
		Spectrum counts	Peptide counts	Spectrum counts	Peptide counts
	CENP-A	107	22	70	21
	CENP-B	106	35	89	36
	CENP-V	11	7	6	5
1	CENP-C	542	94	324	94
2	CENP-L	68	19	37	16
3	CENP-N	96	31	61	31
4	CENP-H	49	17	28	11
5	CENP-I	181	32	90	32
6	CENP-K	206	32	93	28
7	CENP-M	22	7	14	6
8	CENP-O	21	8	14	9
9	CENP-P	61	13	64	15
10	CENP-Q	65	22	52	22
11	CENP-U	111	28	52	22
12	CENP-R	54	13	24	8
13	CENP-T	87	22	35	20
14	CENP-W	0	0	5	3
15	CENP-S	25	6	11	4
16	CENP-X	29	6	15	3

CENP-A^{TAP} was immunoprecipitated from the chromatin fraction of randomly cycling cells or late S/G2 synchronized cells followed by mass spectrometry to identify the coprecipitated partners.

at mid-S but were removed quickly after that and were absent by late S/G2 (Fig. 6i,j). For the 10% of ectopic CENP-A G1 peaks in late-S-replicating regions (Fig. 6c), almost all (85%) were removed by G2 (Fig. 6k,l), while late-replicating centromeric CENP-A peaks were retained, including the single-copy variants within the α-satellite HORs (Figs. 6d–f and 2 and Supplementary Fig. 5g,h). Thus, ectopic, but not centromeric, CENP-A-binding sites are removed as DNA replication progresses.

CENP-C/CCAN remain centromeric CENP-A associated during DNA replication. To comprehensively determine the components that associate with CENP-A chromatin during replication in late S, we used mass spectrometry following affinity purification of CENP-A nucleosomes (Supplementary Fig. 5i, left). A structural link that normally bridges multiple centromeric CENP-A nucleosomes and nucleates the full kinetochore assembly before mitotic entry is the 16-subunit constitutive centromere associated network (CCAN)[49–52]. This complex is anchored to CENP-A primarily through CENP-C[50,53,54] and sustained by CENP-B binding to CENP-B box sequences within α-satellite DNAs[55]. Remarkably, mass spectrometry identified that all 16 CCAN components[13,15] remained associated with mononucleosomal CENP-A chromatin that was affinity purified from late S/G2 (Table 1). Stable association with CENP-A was also seen for HJURP, multiple chromatin remodelling factors and nuclear chaperones, histones, centromere and kinetochore components, and other DNA replication proteins (Supplementary Fig. 5j–n). The continuing interaction during DNA replication of CCAN proteins with CENP-A, which is maintained even on mononucleosomes, provides strong experimental support that the CCAN complex tethers CCAN-bound centromeric CENP-A at or near the centromeric DNA replication forks, thereby enabling its efficient reincorporation after replication fork passage.

To test this further, the composition of CENP-A-containing nucleosomal complexes from G1 to late S/G2 was determined following affinity purification (via the TAP tag) of chromatin-bound CENP-A^{TAP} from a predominantly mononucleosome pool (Supplementary Fig. 5i, right). We initially focused on the chromatin assembly factor 1 (CAF1) complex, which is required for de novo chromatin assembly following DNA replication[56]. Its p48 subunit (also known as CAF1 subunit c, RbAp48 or RBBP4) binds histone H4 (ref. [57]), is a binding partner in a CENP-A prenucleosomal complex with HJURP and nucleophosmin (NPM1)[21] and maintains the deacetylated state of histones in the central core of centromeres after deposition[58]. Remarkably, CAF1 p48 co-immunopurified with CENP-A from G1 to late S/G2 (Fig. 7a). In striking contrast, the two other CAF1 subunits (CAF1 p150 and CAF1 p60), which are essential for de novo chromatin assembly in vitro[59], remained much more strongly associated with CENP-A nucleosomes in late S/G2 compared with mid-S (Fig. 7a). Additionally, MCM2, a core subunit of the DNA replicative helicase MCM2-7 complex that recycles old histones as the replication fork advances[60], was robustly copurified with CENP-A only in late-S-phase-derived chromatin, with no association detected in mid-S (Fig. 7a).

CENP-C is essential for the maintenance of centromeric CENP-A during DNA replication. The stable association only in late S phase (when all centromeric, but only a small minority of ectopically loaded CENP-A, is replicated) of CENP-A with MCM2 and the CAF1 subunits necessary for chromatin reassembly suggested that CENP-C and its CCAN complex tethered centromeric CENP-A near the replication forks and stabilized CENP-A binding to MCM2 and CAF1, thereby enabling CENP-A reassembly onto the daughter centromeres after DNA replication. We tested this possibility by rapidly depleting CENP-C just after S-phase entry in a human cell line (CENP-C$^{AE/AE}$) in which both CENP-C alleles were genome-engineered to produce CENP-C fused to both an auxin-inducible degron and EYFP[55]. Thymidine was used to synchronize these CENP-C$^{AE/AE}$ cells at the G1/S boundary (Fig. 7b,c). CENP-C degradation was induced just after S-phase entry to test CENP-C's role specifically during centromeric DNA replication in late S phase (Fig. 7b), but without affecting the deposition of new CENP-A that occurs in early G1.

Auxin addition 2 h after release from thymidine block resulted in polyubiquitination and degradation of almost all CENP-C^{AE} within 15 min, as was evident by the loss of fluorescence of the EYFP in CENP-C^{AE} (Fig. 7b–d and Supplementary Video 1). DNA replication was then allowed to continue without CENP-C and the CCAN complex it nucleates[53,55]. At the end of DNA replication, chromatin-bound CENP-A was immunoprecipitated and the enrichment of α-satellite-containing DNA was determined. In randomly cycling cells this resulted in a 30-fold enrichment of alphoid DNA (Fig. 7e). At the end of DNA replication and distribution of CENP-A to the two daughter centromeres in CENP-C-containing cells, alphoid DNA enrichment was reduced by half, as expected from doubling of centromeric DNA without addition of new CENP-A. However, degradation of CENP-C early in DNA replication led to loss by the end of S phase of most (73%) of CENP-A initially bound to α-satellite DNA (Fig. 7e).

CENP-C-dependent retention of centromeric CENP-A late in S phase was confirmed by examination of two specific α-satellite variants found within the HORs of the centromeres of chromosomes 8 (Fig. 7f) and 15 (Fig. 7g). Each of these satellite variants is represented only once in the human genome and each shows precise retention in G2 of CENP-A bound in G1 (Fig. 2). In both variants, α-satellite DNA was enriched 50–60-fold following CENP-A ChIP from randomly cycling cells, which was reduced to half as much

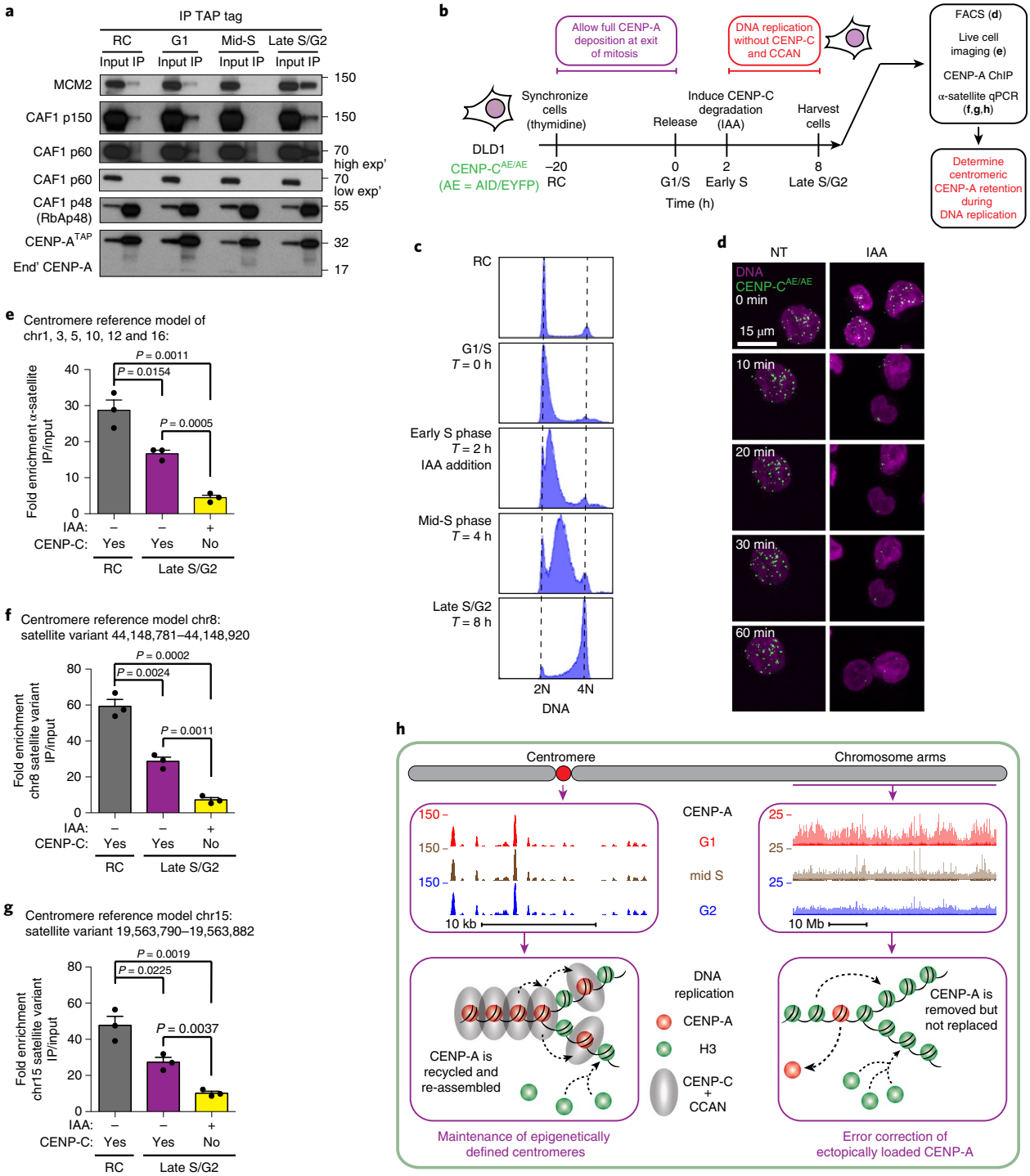

Fig. 7 | CENP-C and the CCAN complex are essential for the epigenetic maintenance of human centromeres during DNA replication. (**a**) CENP-A^{TAP} was immunoprecipitated from micrococcal-nuclease-resistant chromatin isolated in different cell cycle phases and the immunoprecipitates were examined by immunoblotting for CAF1 complex subunits and MCM2. The experiment was repeated independently twice with similar results. Unprocessed film scans of immunoblots can be found in Supplementary Fig. 6. (**b**) Experimental design to test the role of CENP-C and CCAN complex in retention of centromeric CENP-A during DNA replication. The experiment was repeated independently three times with similar results. (**c**) Fluorescence-activated cell sorting (FACS) analysis of DNA content showing the synchronization efficiency of CENP-C$^{AE/AE}$ DLD1 cells during the experiment shown in **b**. The experiment was repeated independently three times with similar results. (**d**) Live-cell imaging of CENP-C$^{AE/AE}$ DLD1 cells following addition of auxin (IAA) at 0 min. DNA is labelled with SiR-DNA. The experiment was repeated independently three times with similar results. NT, not treated. (**e**) Quantitative real-time PCR for α-satellite DNA in chromosomes 1, 3, 5, 10, 12 and 16. $N = 3$ from three biologically independent experiments. Error bars, s.e.m. P value determined using two-tailed t-test. (**f,g**) Quantitative real-time PCR for single-mapping α-satellite DNA variant in chromosome 8 (**f**) and chromosome 15 (**g**). $N = 3$ from three biologically independent experiments. Error bars, s.e.m. P value determined using two-tailed t-test. Source data for **e**–**g** can be found in Supplementary Table 4. (**h**) Model for maintaining centromeric CENP-A while removing it from non-centromeric sites on the chromosome arms during DNA replication to ensure maintenance of centromere identity across the cell cycle.

after DNA replication. Following CENP-C degradation in early S phase, however, CENP-A was not retained at either site during DNA replication (Fig. 7f,g). Taken together, these results demonstrate that depletion of CENP-C (and CCAN bound to CENP-A[53,55]) before centromere DNA replication results in loss of centromeric CENP-A by the end of S phase.

Discussion

Using reference models for 23 human centromeres, we have identified that during DNA replication CENP-A nucleosomes initially assembled onto centromeric α-satellite repeats are reassembled onto the same spectrum of α-satellite repeat sequences of each daughter centromere as are bound before DNA replication. Additionally, genome-wide mapping of sites of CENP-A assembly has identified that when CENP-A is expressed at endogenous levels the selectivity of the histone chaperone HJURP's loading in early G1 of new CENP-A at or near existing sites of centromeric CENP-A-containing chromatin is insufficient to prevent its loading onto more than 11,000 sites along the chromosome arms (Fig. 3d). We also show that the number of ectopic sites increases as CENP-A expression levels increase, as has been reported in multiple human cancers[39,61,62]. Sites of ectopic CENP-A are replicated in early and mid-S (Fig. 6c) and are nearly quantitatively removed as DNA replication progresses (Fig. 6g–l).

Taken together, our evidence demonstrates that DNA replication functions not only to duplicate centromeric DNA but also as an error correction mechanism to maintain epigenetically defined centromere position and identity by coupling centromeric CENP-A retention with its removal from assembly sites on the chromosome arms (Fig. 7h). Indeed, our data reveal that CENP-A loaded onto unique, single-copy sites within α-satellite DNAs of the 23 reference centromeres is precisely maintained at these sites during and after DNA replication, offering direct support that (at least for each of these single-copy sites) the replication machine reloads CENP-A onto the exact same centromeric DNA site (Figs. 2 and 6f).

DNA replication produces a very different situation for CENP-A initially assembled into nucleosomes on the chromosome arms. Sites of ectopically loaded CENP-A are nearly quantitatively stripped during DNA replication (Figs. 3, 4 and 6g–l), thereby precluding premitotic acquisition of CENP-A-dependent centromere function at non-centromeric sites and reinforcing centromere position and identity (Fig. 7h). Without such correction, ectopically loaded sites would be maintained cell cycle after cell cycle, potentially recruiting CENP-C and assembly of the CCAN complex[13–16]. Arm-associated CENP-A/CCAN would present a major problem for faithful assembly and function of a single centromere/kinetochore per chromosome, both by acquisition of partial centromere function and by competition with the authentic centromeres for the pool of available CCAN components. Indeed, high levels of CENP-A expression (1) lead to recruitment of detectable levels of 3 of 16 CCAN components (CENP-C, CENP-N and Mis18) assembled onto the arms[23,24,63], (2) are associated with ongoing chromosome segregation errors[25] and (3) have been reported in several cancers, where it has been proposed to be associated with increased invasiveness and poor prognosis[26,61,62].

As to the mechanism for retention during DNA replication of centromeric but not ectopically loaded CENP-A, our mass spectrometry analysis identifies a strong association of HJURP with CENP-A mononucleosomes in late S phase, comparable to the association identified in G1 (Supplementary Fig. 5l), supporting a probable role for HJURP in CENP-A retention, perhaps through interaction with MCM2-7 complex, as has previously been suggested[21]. This is consistent with evidence that HJURP can associate with MCM2 in a histone-independent manner[60], consistent with a possible co-chaperone relationship for CENP-A. Moreover, degradation of HJURP in early S reduces centromeric CENP-A retention through S phase[64].

Most importantly, our evidence demonstrates that the local reassembly of CENP-A within centromeric domains requires the continuing centromeric CENP-A association with CCAN complexes (Fig. 7), which act to tether disassembled CENP-A/H4 near the sites of centromere DNA replication. This local CENP-C/CCAN-dependent retention of CENP-A, coupled with the actions of the MCM2 replicative helicase, HJURP and CAF1, enables CENP-A's precise reassembly into chromatin within each daughter centromere, thereby maintaining epigenetically defined centromere identity.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41556-019-0331-4.

Received: 3 February 2018; Accepted: 18 April 2019;
Published online: 3 June 2019

References

1. Wevrick, R. & Willard, H. F. Long-range organization of tandem arrays of alpha satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability. *Proc. Natl Acad. Sci. USA* **86**, 9394–9398 (1989).
2. Cleveland, D. W., Mao, Y. & Sullivan, K. F. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112**, 407–421 (2003).
3. Willard, H. F. Chromosome-specific organization of human alpha satellite DNA. *Am. J. Hum. Genet.* **37**, 524–532 (1985).
4. Manuelidis, L. & Wu, J. C. Homology between human and simian repeated DNA. *Nature* **276**, 92–94 (1978).
5. Earnshaw, W. C. & Rothfield, N. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* **91**, 313–321 (1985).
6. Palmer, D. K., O'Day, K., Wener, M. H., Andrews, B. S. & Margolis, R. L. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J. Cell Biol.* **104**, 805–815 (1987).
7. Bodor, D. L. et al. The quantitative architecture of centromeric chromatin. *eLife* **3**, e02137 (2014).
8. Nechemia-Arbely, Y. et al. Human centromeric CENP-A chromatin is a homotypic, octameric nucleosome at all cell cycle points. *J. Cell Biol.* **216**, 607–621 (2017).
9. Sullivan, B. A. & Karpen, G. H. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat. Struct. Mol. Biol.* **11**, 1076–1083 (2004).
10. Stimpson, K. M. & Sullivan, B. A. Epigenomics of centromere assembly and function. *Curr. Opin. Cell Biol.* **22**, 772–780 (2010).
11. Marshall, O. J., Chueh, A. C., Wong, L. H. & Choo, K. H. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am. J. Hum. Genet.* **82**, 261–282 (2008).
12. Fachinetti, D. et al. A two-step mechanism for epigenetic specification of centromere identity and function. *Nat. Cell Biol.* **15**, 1056–1066 (2013).
13. Foltz, D. R. et al. The human CENP-A centromeric nucleosome-associated complex. *Nat. Cell Biol.* **8**, 458–469 (2006).
14. Okada, M. et al. The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat. Cell Biol.* **8**, 446–457 (2006).
15. Hori, T. et al. CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* **135**, 1039–1052 (2008).
16. Hori, T., Shang, W. H., Takeuchi, K. & Fukagawa, T. The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. *J. Cell Biol.* **200**, 45–60 (2013).
17. Padeganeh, A. et al. Octameric CENP-A nucleosomes are present at human centromeres throughout the cell cycle. *Curr. Biol.* **23**, 764–769 (2013).
18. Jansen, L. E., Black, B. E., Foltz, D. R. & Cleveland, D. W. Propagation of centromeric chromatin requires exit from mitosis. *J. Cell Biol.* **176**, 795–805 (2007).
19. Nechemia-Arbely, Y., Fachinetti, D. & Cleveland, D. W. Replicating centromeric chromatin: spatial and temporal control of CENP-A assembly. *Exp. Cell Res.* **318**, 1353–1360 (2012).
20. Foltz, D. R. et al. Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell* **137**, 472–484 (2009).

21. Dunleavy, E. M. et al. HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137**, 485–497 (2009).
22. Silva, M. C. et al. Cdk activity couples epigenetic centromere inheritance to cell cycle progression. *Dev. Cell* **22**, 52–63 (2012).
23. Lacoste, N. et al. Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX. *Mol. Cell* **53**, 631–644 (2014).
24. Van Hooser, A. A. et al. Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. *J. Cell Sci.* **114**, 3529–3542 (2001).
25. Shrestha, R. L. et al. Mislocalization of centromeric histone H3 variant CENP-A contributes to chromosomal instability (CIN) in human cells. *Oncotarget* **8**, 46781–46800 (2017).
26. Filipescu, D. et al. Essential role for centromeric factors following p53 loss and oncogenic transformation. *Genes Dev.* **31**, 463–480 (2017).
27. Mata, J. F., Lopes, T., Gardner, R. & Jansen, L. E. A rapid FACS-based strategy to isolate human gene knockin and knockout clones. *PloS ONE* **7**, e32646 (2012).
28. Landry, J. J. et al. The genomic and transcriptomic landscape of a HeLa cell line. *G3* **3**, 1213–1224 (2013).
29. Hayden, K. E. et al. Sequences associated with centromere competency in the human genome. *Mol. Cell. Biol.* **33**, 763–772 (2013).
30. Conde e Silva, N. et al. CENP-A-containing nucleosomes: easier disassembly versus exclusive centromeric localization. *J. Mol. Biol.* **370**, 555–573 (2007).
31. Miga, K. H. et al. Centromere reference models for human chromosomes X and Y satellite arrays. *Genome Res.* **24**, 697–707 (2014).
32. Schneider, V. A. et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).
33. Levy, S. et al. The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
34. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
35. Zang, C. et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958 (2009).
36. Maloney, K. A. et al. Functional epialleles at an endogenous human centromere. *Proc. Natl Acad. Sci. USA* **109**, 13704–13709 (2012).
37. Ly, P. et al. Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining. *Nat. Cell Biol.* **19**, 68–75 (2017).
38. Amor, D. J. & Choo, K. H. Neocentromeres: role in human disease, evolution, and centromere study. *Am. J. Hum. Genet.* **71**, 695–714 (2002).
39. Hasson, D. et al. The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat. Struct. Mol. Biol.* **20**, 687–695 (2013).
40. Amor, D. J. et al. Human centromere repositioning "in progress". *Proc. Natl Acad. Sci. USA* **101**, 6542–6547 (2004).
41. Alonso, A. et al. Co-localization of CENP-C and CENP-H to discontinuous domains of CENP-A chromatin at human neocentromeres. *Genome Biol.* **8**, R148 (2007).
42. Li, H. et al. Functional annotation of HOT regions in the human genome: implications for human disease and cancer. *Sci. Rep.* **5**, 11633 (2015).
43. Teytelman, L., Thurtle, D. M., Rine, J. & van Oudenaarden, A. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc. Natl Acad. Sci. USA* **110**, 18602–18607 (2013).
44. Thakur, J. & Henikoff, S. Unexpected conformational variations of the human centromeric chromatin complex. *Genes Dev.* **32**, 20–25 (2018).
45. Henikoff, J. G., Thakur, J., Kasinathan, S. & Henikoff, S. A unique chromatin complex occupies young alpha-satellite arrays of human centromeres. *Sci. Adv.* **1**, e1400234 (2015).
46. Hansen, R. S. et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
47. Bui, M. et al. Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. *Cell* **150**, 317–326 (2012).
48. Erliandri, I. et al. Replication of alpha-satellite DNA arrays in endogenous human centromeric regions and in human artificial chromosome. *Nucl. Acids Res.* **42**, 11502–11516 (2014).
49. Guse, A., Carroll, C. W., Moree, B., Fuller, C. J. & Straight, A. F. In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature* **477**, 354–358 (2011).
50. Carroll, C. W., Milks, K. J. & Straight, A. F. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J. Cell Biol.* **189**, 1143–1155 (2010).
51. Petrovic, A. et al. Structure of the MIS12 complex and molecular basis of its interaction with CENP-C at human kinetochores. *Cell* **167**, 1028–1040 (2016).
52. Rago, F., Gascoigne, K. E. & Cheeseman, I. M. Distinct organization and regulation of the outer kinetochore KMN network downstream of CENP-C and CENP-T. *Curr. Biol.* **25**, 671–677 (2015).
53. Klare, K. et al. CENP-C is a blueprint for constitutive centromere-associated network assembly within human kinetochores. *J. Cell Biol.* **210**, 11–22 (2015).
54. Weir, J. R. et al. Insights from biochemical reconstitution into the architecture of human kinetochores. *Nature* **537**, 249–253 (2016).
55. Hoffmann, S. et al. CENP-A is dispensable for mitotic centromere function after initial centromere/kinetochore assembly. *Cell Rep.* **17**, 2394–2404 (2016).
56. Smith, S. & Stillman, B. Stepwise assembly of chromatin during DNA replication in vitro. *EMBO J.* **10**, 971–980 (1991).
57. Verreault, A., Kaufman, P. D., Kobayashi, R. & Stillman, B. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**, 95–104 (1996).
58. Hayashi, T. et al. Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. *Cell* **118**, 715–729 (2004).
59. Kaufman, P. D., Kobayashi, R., Kessler, N. & Stillman, B. The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. *Cell* **81**, 1105–1114 (1995).
60. Huang, H. et al. A unique binding mode enables MCM2 to chaperone histones H3-H4 at replication forks. *Nat. Struct. Mol. Biol.* **22**, 618–626 (2015).
61. Zhang, W. et al. Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. *Nat. Commun.* **7**, 12619 (2016).
62. Sun, X. et al. Elevated expression of the centromere protein-A (CENP-A)-encoding gene as a prognostic and predictive biomarker in human cancers. *Int. J. Cancer* **139**, 899–907 (2016).
63. Gascoigne, K. E. et al. Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell* **145**, 410–422 (2011).
64. Zasadzinska, E. et al. Inheritance of CENP-A nucleosomes during DNA replication requires HJURP. *Dev. Cell* **47**, 348–362 e7 (2018).

Acknowledgements

We thank A. Desai, P. Ly and C. Eissler for critical discussion, L.E.T. Jansen (Gulbenkian Institute, Portugal) for reagents and D.-H. Kim for productive discussions and technical help. This work was supported by grants (R01 GM-074150 and R35 GM-122476) from the NIH to D.W.C., who receives salary support from Ludwig Cancer Research.

Author contributions

Y.N.-A. and D.W.C. conceived and designed experiments and wrote the manuscript. Y.N.-A. performed experiments. K.H.M. analysed the sequencing data. M.A.M. and O.S. analysed data and performed experiments. D.F. suggested experiments and provided key experimental input. A.Y.L. and B.R. prepared sequencing libraries and provided resources. A.A. and J.R.Y. performed mass spectrometry experiments and provided resources.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-019-0331-4.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to Y.N. or D.W.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Cell lines. Adherent HeLa cells stably expressing CENP-A^{TAP} or H3.1^{TAP} by retrovirus infection[13] or endogenously tagged CENP-A$^{+/LAP}$ by infection of a rAAV harbouring a LAP-targeting construct containing homology arms for CENP-A[27] were adapted to suspension growth by selecting surviving cells and were maintained in DMEM (Gibco) containing 10% fetal bovine serum (Omega Scientific), 100 U ml^{-1} penicillin, 100 U ml^{-1} streptomycin and 2 mM l-glutamine at 37 °C in a 5% CO_2 atmosphere with 21% oxygen. DLD1 cells with auxin-degradable CENP-A^{AID} and doxycycline-inducible CENP-A^{WT}[37] and DLD1 CENP-C$^{AE/AE}$ [55] were maintained in the same conditions. Cells were maintained and split every 4–5 d according to ATCC recommendations.

Cell synchronization. Cells were synchronized as previously described[8]. Briefly, suspension HeLa cells were treated with 2 mM thymidine in complete medium for 19 h, pelleted and washed twice in PBS, and released in complete medium containing 24 μM deoxycytidine for 9 h followed by addition of thymidine to a final concentration of 2 mM for 16 h, after which cells were released again into complete medium containing 24 μM deoxycytidine. For G2, cells were collected 7 h after release from the second thymidine block. For G1, thymidine was added for a third time, 7 h after the release, and cells were collected 11 h after this (a total of 18 h after the release from the second thymidine block). Adherent DLD1 cells with auxin-degradable CENP-A^{AID} and doxycycline-inducible CENP-A^{WT} (ref. [37]) were synchronized at G1 using 1.5 μM CDK4/6 inhibitor PD-0332991 (also known as Palbociclib) for 30 h or at mitosis using a single 2 mM thymidine block for 20 h and release into 0.1 mg ml^{-1} nocodazole. Adherent DLD1 CENP-C$^{AE/AE}$ (ref. [55]) were synchronized at G1/S with a single thymidine block, followed by washing twice in PBS, and releasing in complete medium containing 24 μM deoxycytidine. Two hours after release, CENP-C rapid degradation was induced by addition of IAA at a final concentration of 500 μM. Nocodazole (Sigma) was added at 0.1 mg ml^{-1} to prevent cells from going into the next cell cycle. Cells were collected 8 h after release.

Chromatin extraction and affinity purification. Chromatin was extracted from 1×10^9 nuclei of HeLa cells as previously described[8]. TAP- or LAP-tagged chromatin were purified in two steps. In the first step, native TAP-tagged chromatin was immunoprecipitated by incubating the bulk soluble mononucleosome pool with rabbit IgG (Sigma-Aldrich) coupled to Dynabeads M-270 epoxy (Thermo Fisher Scientific, 14301). Alternatively, CENP-A^{LAP} chromatin was immunoprecipitated using mouse anti-GFP antibody (clones 19C8 and 19F7, Monoclonal Antibody Core Facility at Memorial Sloan-Kettering Cancer Center, New York, USA) coupled to Dynabeads M-270 epoxy. Chromatin extracts were incubated with antibody-bound beads for 16 h at 4 °C. Bound complexes were washed once in buffer A (20 mM HEPES at pH 7.7, 20 mM KCl, 0.4 mM EDTA and 0.4 mM DTT), once in buffer A with 300 mM KCl and finally twice in buffer A with 300 mM KCl, 1 mM DTT and 0.1% Tween 20. In the second step, TAP–chromatin complexes were incubated for 16 h in final wash buffer with 50 μl recombinant tobacco etch virus protease, resulting in cleavage of the TAP tag and elution of the chromatin complexes from the beads. Alternatively, CENP-A^{LAP} chromatin was eluted from the beads by cleaving the LAP tag using PreScission protease (4 h, 4 °C). CENP-A-containing chromatin was immunoprecipitated from DLD1 cells with auxin-degradable CENP-A^{AID} and doxycycline-inducible CENP-A^{WT} (ref. [37]) using Abcam ab13939 CENP-A antibody coupled to Dynabeads M-270 epoxy.

DNA extraction. Following elution of the chromatin from the beads, Proteinase K (100 μg ml^{-1}) was added and samples were incubated for 2 h at 55 °C. DNA was purified from proteinase K-treated samples using a DNA purification kit following the manufacturer instructions (Promega, Madison, WI, USA) and was subsequently analysed either by running a 2% low-melting agarose (APEX) gel or with an Agilent 2100 Bioanalyzer using the DNA 1000 kit. The Bioanalyzer determines the quantity of DNA on the basis of fluorescence intensity.

Quantitative real-time PCR. Quantitative real-time PCR was performed using SYBR Green mix (Bio Rad) with a CFX384 Bio Rad Real Time System. Primer sequences used in this study were the following: MRGPRE (forward) 5′-CTGC GCGGATCTCATCTTCC-3′ and (reverse) 5′-GGCCCACGATGTAGCAGAA-3′; MMP15 (forward) 5′-GTGCTCGACGAAGAGACCAAG-3′ and (reverse) 5′-TTTCACTCGTACCCCGAACTG-3′; HBE1 (forward) 5′-ATGGTGCATTT TACTGCTGAGG-3′ and (reverse) 5′-GGGAGACGACAGGTTTCCAAA-3′; Sat2 (forward) 5′-TCGCATAGAATCGAATGGAA-3′ and (reverse) 5′-GCAT TCGAGTCCGTGGA-3′; α-satellite DNA (from chromosomes 1, 3, 5, 10, 12 and 16) (forward) 5′-CTAGACAGAAGAATTCTCAG-3′ and (reverse) 5′-CTG AAATCTCCACTTGC-3′ (ref. [41]); α-satellite DNA variant from chromosome 8 (forward) 5′-TGAATGCGAGAGAGAAGTAA-3′ and (reverse) 5′-TCAAAT ATATCCAAATATCCA-3′; α-satellite DNA variant from chromosome 15 (forward) 5′-GTTGCACATTCCGGTTCATACA-3′ and (reverse) 5′- TTTCAC CGTAGGCCTCAAAGGGCTCCAACT-3′; ectopic site 1 in chromosome 5 (forward) 5′- CCCTCC TGCCTGAAGATTTGAT-3′ and (reverse) 5′-AAAGCT TGGTGAGGGCAGTT-3′; ectopic site 2 in chromosome 14 (forward) 5′-GCT GTGTACTCCCGAACTCC-3′ and (reverse) 5′- GATCCTGTCCAG CTGCCAG; ectopic site 3 in chromosome 1 (forward) 5′- TCAGTTTGCACCATCCCCTG-3′ and (reverse) 5′-GCTCTGACTCATGCTCCTACTG-3′; ectopic site 4 in chromosome 9 (forward) 5′- AGTGCCCTCTGA ACGCTAAC-3′ and (reverse) 5′- ATTCCTCCCTGAGCTCCCAT-3′. Melting curve analysis was used to confirm primer specificity. To ensure linearity of the standard curve, reaction efficiencies over the appropriate dynamic range were calculated. Using the dCt method, we calculated fold enrichment of α-satellite DNA after immunopurification of CENP-A^{TAP} chromatin, compared with its level in the bulk input chromatin. For the Repli-seq experiment, we used the dCt method to calculate fold enrichment of replicated DNA after immunopurification of BrdU-labelled DNA compared with its level in the bulk input DNA. Reported values are the means of two independent biological replicates with technical duplicates that were averaged for each experiment. Error bars represent s.e.m. To determine CENP-A levels at ectopic sites at G1 and mitosis, we used the dCt method, to calculate fold enrichment of CENP-A-containing DNA compared with its level in the bulk input DNA. CENP-A levels at mitosis were normalized to levels at G1. Reported values are the means of two (G1) or three (mitosis) independent biological replicates with technical duplicates that were averaged for each experiment. Error bars represent s.e.m.

Immunoblotting. For immunoblot analysis, protein samples were separated by SDS–PAGE, transferred onto polyvinylidene difluoride membranes (Millipore) and then probed with the following antibodies: rabbit anti-CENP-A (Cell Signaling, 2186 s, 1:1,000), rabbit anti-CENP-B (Millipore, 07-735, 1:200), mouse anti-α-tubulin (Abcam, DM1A, 1:5,000), rabbit anti-CAF1 p150 (Santa Cruz, sc-10772, 1:500), rabbit anti-CAF1 p60 (Bethyl Laboratories, A301-085A, 1:1,000), rabbit anti-CAF1 p48 (Bethyl Laboratories, A301-206A, 1:1,000) and rabbit anti-MCM2 (Abcam, Ab4461, 1:1,000). Following incubation with horseradish peroxidase-labelled antibody (GE Healthcare, NA931V or NA934V), horseradish peroxidase was detected using enhanced chemiluminescence substrate (Thermo Scientific, 34080 or 34096).

Immunofluorescence and live-cell imaging. 1×10^6 suspension cells were centrifuged and resuspended with PBS. 10^5 cells were immobilized on glass slide by cytospin centrifugation for 3 min at 800 rpm. Cells were then fixed using ice-cold methanol at −20 °C for 10 min, followed by washing with cold PBS, and then incubated in Triton Block (0.2 M glycine, 2.5% fetal bovine serum, 0.1% Triton X-100, PBS) for 1 h. The following primary antibodies were used: mouse anti-GFP (Roche, 11814460001, 1:500), rabbit anti-CENP-B (Abcam 25734, 1:1,000) and human anti-centromere antibodies (ACA, Antibodies Inc, 15-234-0001, 1:500). The following secondary antibodies (Jackson Laboratories) were used for 45 min: donkey anti-human TR (1:300) and anti-mouse fluorescein isothiocyanate (FITC) (1:250). TAP fusion proteins were visualized by incubation with FITC-rabbit IgG (Jackson Laboratories, 1:200). Cells were then washed with 0.1% Triton X-100 in PBS, counterstained with 4,6-diamidino-2-phenylindole and mounted with mounting medium (Molecular Probes, P36934). Immunofluorescent images were acquired on a Deltavision Core system at ×60–100 magnification. 0.2 μm *Z*-stack deconvolved projections were generated using the softWoRx program. For live-cell imaging, cells were plated on high-optical-quality plastic slides (ibidi) and imaged using a CQ1 confocal quantitative image cytometer (Yokogawa) following addition of IAA. DNA was labelled with Sir-DNA (Cytoskeleton). IAA was added at the microscope stage.

Flow cytometry. Flow cytometry was used to determine the DNA content of the cells. 1×10^6 cells were collected, washed in PBS and fixed in 70% ethanol. Cells were then washed and DNA was stained by incubating cells for 30 min with 1% fetal bovine serum, 10 μg ml^{-1} propidium iodide and 0.25 mg ml^{-1} RNase A in PBS followed by FACS analysis for DNA content using a BD LSR II flow cytometer (BD Biosciences).

ChIP-seq library generation and sequencing. ChIP libraries were prepared following Illumina protocols with minor modifications. To reduce biases induced by PCR amplification of a repetitive region, libraries were prepared from 80–100 ng of input or ChIP DNA. The DNA was end-repaired and A-tailed and Illumina TruSeq adaptors were ligated. Libraries were run on a 2% agarose gel. Since the chromatin was digested to mononucleosomes, following adaptor ligation the library size was 250–280 bp. The libraries were size selected for 200–375 bp. The libraries were then PCR-amplified using only five or six PCR cycles since the starting DNA amount was high. Resulting libraries were sequenced using 100 bp, paired-end sequencing on a HiSeq 2000 instrument according to the manufacturer's instructions with some modifications (Illumina). Sequence reads are summarized in Supplementary Table 1.

Initial sequence processing and alignment. Illumina paired-end reads were merged to determine CENP-A- or H3-containing target fragments of varying length using PEAR software[65], with standard parameters ($P = 0.01$, minimum

overlap 10 bases, minimum assembly length 50 bp). Merged paired reads were mapped (BWA-MEM, standard parameters[66,67]) to the hg38 assembly (including alternative assemblies), which contains human α-satellite sequence models in each centromeric region (ref. [31]; BioProject PRJNA193213). The highly repeated sequences in the centromere reference models preclude distinguishing between centromeric and pericentromeric sequences, the order of repeats in the models is arbitrarily assigned, and portions of the centromeres of the acrocentric chromosomes 13, 14, 21 and 22, as well as portions of centromeres of chromosomes 1, 5 and 19, contain nearly identical arrays that cannot be distinguished. Alignments to repeats, or multiple regions in the reference genome with the same mapping score, were assigned randomly. That is, across an entire reference model a given read may have equal probability of mapping across most of the repeat copies; however, the final assignment is random and so the alignments are distributed across the array. Mapping was performed under the same protocol for all ChIP and input samples. Reads were determined to contain α-satellite if they overlapped sites (BEDTools: intersect[68]) in the genome previously annotated as α-satellite (UCSC Table Browser[69] was used to obtain a bed file of all sites annotated as ALR/α-satellite). Additionally, merged sequences were defined as containing α-satellite if they contained an exact match to at least two 18-base oligonucleotides specific to a previously published whole-genome sequencing read database of α-satellite, representing 2.6% of sequences from the HuRef genome[29,33]. Comparisons between the BWA mapping and 18-base oligonucleotide exact matching based strategies were highly concordant. Total α-satellite DNA content in hg38 assembly was estimated by using the UCSC RepeatMasker annotation[69]. To identify reads that aligned uniquely to low-frequency repeat variants, or base changes in a repeat unit that are only observed once within the hg38 reference model, we used mapping scores (MAPQ = 20, or the probability of correctly mapping a random read was 0.99). A summary of reads obtained is shown in Supplementary Table 1.

ChIP-seq peak calling. Enrichment peaks for ChIP experiments were determined using the SICER algorithm (v1.0.3)[35] using relevant input reads as background, with stringent parameters previously optimized for human CENP-A[23]: threshold for redundancy allowed for chip reads, 1; threshold for redundancy allowed for control reads, 1; window size, 200 bp; fragment size, 150 bp; shift in window length, 150; effective genome size as a fraction of the reference genome of hg38, 0.74; gap size, 400 bp; *e*-value for identification of candidate islands that exhibit clustering, 1,000; false discovery rate controlling significance, 0.00001. All sequencing samples were normalized to their matched input control, that is, each G1 sample was normalized to the G1 input and the G2 sequencing samples to the G2 input. In parallel, MACS peak calling was performed (macs14)[34], and wiggle tracks were created to represent the read depth of each dataset independently. Finally, we performed a final, rigorous evaluation of ectopic CENP-A peaks, or peaks predicted outside centromeric regions, using *k*-mer enrichment (previously described[29]). Each ectopic peak was reformatted into 50 bp sliding windows (in both orientations, with a slide of 1 bp). The normalized frequency of each 50-base oligonucleotide candidate window was evaluated in each ChIP-seq dataset relative to a normalized observed frequency in the corresponding background dataset. Scores were determined as the log-transformed normalized value of the ratio between ChIP-seq and background, and those with a score greater than or equal to 2 were included in our study as a high-confidence enrichment set.

Analysis of CENP-A peak overlap with functional annotation. Ectopic CENP-A peak calls, that is those that did not overlap with centromeric α-satellite DNA, for HeLa, HT1080b (ref. [44]) and HuRef (ref. [45]) cells, were evaluated for enrichment with functional annotation if they were supported between replicate ChIP-seq experiments and overlapped at least one enriched 50-base oligonucleotide with a log-transformed normalized ratio of ≥2, or with a minimum standard ratio of fivefold. Resulting high-confidence ectopic peak calls were intersected (BEDTools: intersect[68]) with select functional datasets in the genome (UCSC Table Browser[69]). Peaks within genes were determined as such if 90% (–*f* 0.9) of the SICER peak intersected with GRCh38 RefSeq genes (including introns and exons). Peaks were determined at promoters if they intersected 1,000 bp upstream and downstream of genes (with minimum overlap of 50% of the SICER peak). To evaluate the role of expression, gene annotation was catalogued further based on intersection with ENCODE HeLa expression data with RefSeq gene annotations (where 22,211 RefSeq genes (40.5% of total) demonstrated at least 10 average reads/gene; and highly expressed RefSeq genes (10,033, or 18.3% of total) are defined as ≥100 average reads/gene; UCSC Table Browser[69]). To investigate peak overlap with sites of CTCF enrichment, we intersected peaks with two ENCODE replicate datasets (GEO GSM749729 and GSM749739), with minimum overlap of 20 bp. To study the overlap with sites of open chromatin, peaks were intersected with DNase I hypersensitive ENCODE datasets (GEO GSM736564 and GSM736510), with minimum overlap of 100 bases. Enrichment at H3K27me3 sites was determined by intersecting CENP-A binding data with H3K27me3 ChIP-seq (GEO GSM945208), with minimum overlap of 100 bases. Results were evaluated relative to a simulated peak dataset to test if observed peak counts were higher than expected by chance. Simulations were repeated 100 times to provide basic summary statistics: average, s.d., maximum/minimum, relative enrichment value and empirical *P*-value. To study the features of CENP-A non-centromeric preferential sites, CENP-A peaks were intersected with Broad Peaks of the following publicly available ENCODE HeLa S3 datasets with minimum overlap of 100 bases: H3K4me1 (GEO GSM798322), H3K4me2 (GEO GSM733734), H3K4me3 (GEO GSM733682), H2A.z (GEO GSM1003483), H3K9me3 (GEO GSM1003480), H3K27ac (GEO GSM733684), H3K27me3 (GEO GSM733696), H3K36me3 (GEO GSM733711), H3K79me2 (GEO GSM733669), H3K9ac (GEO GSM733756) and H4K20me1 (GEO GSM733689). CENP-A SICER peaks were also intersected with ENCODE high-occupancy target regions in the human genome assembly[42] (GEO GSE54296).

Repli-seq experiments. BrdU-labelled DNA across S phase was prepared as previously described[46] with some modifications. Briefly, cells were synchronized using double thymidine block and release[8]. Following release from double thymidine block, cells were labelled with BrdU (Sigma, B5002) for 1 h by adding BrdU to the culture medium to a final concentration of 50 μM. BrdU was added immediately after release for labelling at early S (S0–S1), 3 h after release (S3) for mid-S (S3–S4) labelling or 6 h after release (S6) for labelling at late S (S6–S7). Following labelling with BrdU, genomic DNA was extracted, sonicated and heat denatured as previously described[46]. BrdU-labelled DNA was immunoprecipitated using an anti-BrdU antibody (BD Biosciences, 555627) coupled to magnetic Dyna M-270 epoxy beads (Thermo Fisher Scientific, 14301). Eluted single-stranded DNA was made into double-stranded DNA using random-prime extension (Thermo Fisher Scientific, Random Primers DNA labelling kit, 18187-013). Following cleanup of the double-stranded DNA (QIAgen QiaQuick PCR purification kit, 28104), the DNA was validated by performing quantitative real-time PCR using primers for MRGPRE, MMP15, HBE1 and Sat2 and comparison with the ENCODE HeLa S3 replication timing profile (GEO GSM923449). Libraries were then prepared as described above, and sequenced using the Illumina instrument according to the manufacturer's instructions, with the exception that, following adaptor ligation, Repli-seq libraries were size selected between 250 and 500 bp.

Mass spectrometry identification of proteins associating with CENP-A^{TAP} chromatin. CENP-A^{TAP} was immunoprecipitated from the chromatin fraction of randomly cycling cells or late-S-synchronized cells as described above. Following bead washes, beads were snap frozen in liquid nitrogen. Samples were prepared for mass spectrometry and run using LTQ Orbitrap Velos mass spectrometer (Thermo Scientific) as previously described[70]. Analysis was performed as previously described and is detailed in the Reporting Summary.

Statistics and reproducibility. For all experiments shown, *n* is indicated in the figure legends. Values represent the mean. Error bars, if shown, are s.e.m. (as indicated in the figure legends). For Fig. 6b,e,f,h,j,l, the experiment was repeated independently twice with similar results. For Fig. 6c,d,i,k,m, results of two biologically independent experiments are shown. Statistics source data for all graphical representations are available in Supplementary Table 4.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

ChIP-seq and Repli-seq datasets generated during the current study were deposited at GEO under primary accession number GSE111381. Figures 1, 2, 3, 4, 5 and 6 and Supplementary Figs. 1, 2, 3 and 5 have associated raw mapping data. Mass spectrometry data were deposited at the ProteomeXchange database under accession number PXD013385. Mass spectrometry data are found in Table 1 and Supplementary Fig. 5.

Previously published ChIP-seq data that were reanalysed in this study include HT1080b (ref. [44]) CENP-A^{FLAG} ChIP-seq (GSM2808187 and corresponding input dataset GSM2808190) and HuRef (ref. [45]) CENP-A native ChIP-seq (GSM1494428, GSM1494429, GSM1494430, GSM1494431, and corresponding input datasets GSM1494424, GSM1494425, GSM1494426, GSM1494427). Hela CENP-A^{TAP}, Hela CENP-A^{LAP}, HT1080b CENP-A^{FLAG} (ref. [44]) and HuRef (ref. [45]) endogenous CENP-A ChIP-seq datasets were intersected with the following publicly available ENCODE HeLa S3 datasets: DNase I (GSM736564 and GSM736510), H3K4me1 (GSM798322), H3K4me2 (GSM733734), H3K4me3 (GSM733682), H2A.z (GSM1003483), H3K9me3 (GSM1003480), H3K27ac (GSM733684), H3K27me3 (GSM945208), H3K36me3 (GSM733711), H3K79me2 (GSM733669), H3K9ac (GSM733756) and H4K20me1 (GSM733689). CENP-A^{TAP} and CENP-A^{LAP} SICER peaks were also intersected with ENCODE CTCF datasets (GSM749729 and GSM749739), and high-occupancy target region datasets (GSE54296).

Source data for Figs. 1, 3, 5, 6 and 7 and Supplementary Figs. 3, 4 and 5 have been provided as Supplementary Table 4. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

References

65. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620 (2014).
66. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
67. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
68. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
69. Karolchik, D. et al. The UCSC Table Browser data retrieval tool. *Nucl. Acids Res.* **32**, D493–D496 (2004).
70. Wang, Z., Wu, C., Aslanian, A., Yates, J. R.III. & Hunter, T. Defective RNA polymerase III is negatively regulated by the SUMO-Ubiquitin-Cdc48 pathway. *eLife* **7**, e35447 (2018).

nature research

Corresponding author(s): Prof. Don Cleveland
Dr. Yael Nechemia-Arbely

Last updated by author(s): Apr 8, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Microsoft Excel GraphPad Prism softWoRx (GE Healthcare) http://incelldownload.gehealthcare.com/bin/download_data/SoftWoRx/6.5.2/SoftWoRx.htm FACSDiva PEAR (Zhang et al., 2014) https://sco.h-its.org/exelixis/web/software/pear/doc.html Bwa-Mem (Li, 2013; Li and Durbin, 2010) https://github.com/lh3/bwa Xcalibur data system (Thermo Scientific)
Data analysis	Microsoft Excel (version 16.16.7) GraphPad Prism (version 5.0a) ModFit 3.0 Adobe Photoshop CS5.1 PEAR (Zhang et al., 2014) https://sco.h-its.org/exelixis/web/software/pear/doc.html Bwa-Mem (Li, 2013; Li and Durbin, 2010) https://github.com/lh3/bwa MACS (macs14) (Zhang et al., 2008) http://liulab.dfci.harvard.edu/MACS/ SICER (v1.0.3) (Zang et al., 2009) https://home.gwu.edu/~wpeng/Software.htm BEDTools: intersect (Quinlan and Hall, 2010) https://github.com/arq5x/bedtools softWoRx (GE Healthcare) http://incelldownload.gehealthcare.com/bin/download_data/SoftWoRx/6.5.2/SoftWoRx.htm RawXtract (version 1.9.9) (McDonald et al., 2004) http://fields.scripps.edu/yates/wp/?page_id=17 ProLuCID algorithm (Xu et al., 2015) http://fields.scripps.edu/yates/wp/?page_id=17 DTASelect (version 2.0) algorithm, http://fields.scripps.edu/yates/wp/?page_id=17

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The ChIP-seq and Repli-seq datasets generated during the current study were deposited at GEO under primary accession number GSE111381. Figs. 1-6, Figs. S1-S3 and Fig. S5 have associated raw mapping data.
The mass spectrometry datasets generated during the current study were deposited at ProteomeXchange database under accession PXD013385. Mass spectrometry data is found in Table 1 and Supplementary Fig. 5.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to determine sample size. Sample size was chosen based on previous experience and standards in the field.
Data exclusions	Data was excluded if FACS analysis of DNA content showed synchrony was not satisfactory. Representative IF cell images were chosen.
Replication	Synchronization experiments were repeated at lease three times with similar outcome. ChIP-sequencing, Repli-sequencing, Co-IP experiments were repeated two times each with similar outcome. ChIP-qPCR experiments were repeated 2-3 times in HeLa cells with similar outcomes, and 3 times in DLD1 cells.
Randomization	The experiments were not randomized as randomization was not relevant for this study.
Blinding	Investigators were not blinded during data collection and analysis because blinding was not possible due to the type of exp' (ChIP-sequencing and analysis)

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

All antibody information, including dilutions used for immunoblotting and immunofluorescence, is included in detail in Supplementary Table S3.
Primary antibodies used:
Antibody Species Assay and dilution Source Catalogue
Repli-seq: Mouse anti-BrdU, Becton-Dickinson Biosciences, CAT # 555627
ChIP-seq: CENP-ALAP Mouse anti-GFP, MSKCC antibody core facility clone 19C8 and clone 19F7
CENP-ATAP rabbit IgG, Sigma-Aldrich CAT # I5006
End' CENP-A: CENP-A, mouse, Abcam CAT # Ab13939

Immunoblot:
CENP-A, rabbit, 1:1,000 Cell Signaling Technology CAT # 2186
CAF1p150, rabbit 1:500 Santa Cruz CAT # sc-10772
CAF1p60, rabbit 1:1,000 Bethyl Laboratories CAT # A301-085A
CAF1p48, rabbit 1:1,000 Bethyl Laboratories CAT # A301-206A
MCM2, rabbit 1:1,000 Abcam CAT # Ab4461
Immunofluorescence:
CENP-B, rabbit 1:1,000 Abcam CAT # 25734
GFP, Mouse 1:500 Roche CAT # 11814460001
anti-centromere antibodies (ACA) human 1:500 Antibodies Inc CAT # 15-234-0001

Validation

Antibodies were validated by western blotting and/or immunofluorescence staining, as explained in details in Table S3.
BrdU Ab was used for Repli-seq previously (Hansen et al., 2010)
GFP monoclonal Ab was validated previously by us (Nechemia-Arbely et al., 2017) and others
rabbit IgG was validated previously by us (Foltz et al., 2006) and others
CENP-A (cell signaling) was validated in Fig. 6o, also previously used by us (Nechemia-Arbely 2017) and others
CENP-A (Abcam) was used for ChIP previously (Hasson et al., 2013)
CAF1p150 was validated in Fig. 7a; also see manufacturer's data sheet
CAF1p60 was validated in Fig. 7a; also see manufacturer's data sheet
CAF1p48 was validated in Fig. 7a; also see manufacturer's data sheet
MCM2 was validated in Fig. 7a; also see manufacturer's data sheet
CENP-B was validated in Fig. S1d and previously by us and others
GFP (Roche) was validated in Fig. S1c and previously by us (Nechemia-Arbely 2017) and others
anti-centromere antibodies (ACA) was validated in Fig. S1c and previously by us and others

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Human: HeLa CENP-ATAP (adherent) (Foltz et al., 2006)
Human: HeLa CENP-ATAP (suspension growth) (Nechemia-Arbely et al., 2017)
Human: HeLa CENP-A+/LAP (adherent) (Mata et al., 2012)
Human: HeLa CENP-A+/LAP (suspension growth) (Nechemia-Arbely et al., 2017)
Human: DLD1 CENP-C AID-EYFP/AID-EYFP (adherent) (Hoffmann et al., 2016)
Human: DLD1 cells with auxin degradable CENP-A/AID and a doxycycline-inducible CENP-A/WT (adherent) (Ly et al., 2017)

Authentication

SKY Karyotype analysis was used to authenticate HeLa cell lines.
DLD1 cells were not authenticated.

Mycoplasma contamination

Cells were regularly tested negative for mycoplasma.

Commonly misidentified lines
(See ICLAC register)

No cell lines used in this study were found in the database of commonly misidentified cell lines that is maintained by ICLAC and NCBI Biosample.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111381

Files in database submission

Input CENP-A-TAP RC bigWig
Input CENP-A TAP G1 bigWig
Input CENP-A TAP G2 bigWig
CENP-A-TAP RC-1 bigWig
CENP-A-TAP RC-2 bigWig
CENP-A-TAP G1-1 bigWig
CENP-A-TAP G1-2 bigWig
CENP-A-TAP midS-1 bigWig
CENP-A-TAP midS-2 bigWig
CENP-A-TAP G2-1 bigWig
CENP-A-TAP G2-2 bigWig
H3.1-TAP-1 bigWig
H3.1-TAP-2 bigWig
CENP-A-LAP G1-1 bigWig
CENP-A-LAP G1-2 bigWig
CENP-A-LAP G2-1 bigWig
CENP-A-LAP G2-2 bigWig
Input CENP-A-LAP G1 bigWig

Input CENP-A-LAP G2 bigWig
Repli-seq_BrdU lateS (S7)-2 bigWig
Repli-seq_BrdU lateS (S7)-1 bigWig
Repli-seq_BrdU midS (S4)-2 bigWig
Repli-seq_BrdU midS (S4)-1 bigWig
Repli-seq_BrdU earlyS (S1)-1 bigWig
Repli-seq_BrdU earlyS (S1)-2 bigWig
Repli-seq_BrdU Input lateS (S7) bigWig
Repli-seq_BrdU Input midS (S4) bigWig
Repli-seq_BrdU Input earlyS (S1) bigWig
hg38_CENPB_boxes bed

SICER_H3.1-TAP-1 bed
SICER_H3.1-TAP-2 bed
SICER_CENP-A-LAP G1-1 bed
SICER_CENP-A-LAP G1-2 bed
SICER_CENP-A-LAP G2-1 bed
SICER_CENP-A-LAP G2-2 bed
SICER_CENP-A-TAP RC-1 bed
SICER_CENP-A-TAP RC-2 bed
SICER_CENP-A-TAP G1-1 bed
SICER_CENP-A-TAP G1-2 bed
SICER_CENP-A-TAP midS-1 bed
SICER_CENP-A-TAP midS-2 bed
SICER_CENP-A-TAP G2-1 bed
SICER_CENP-A-TAP G2-2 bed
SICER_Repli-seq_BrdU earlyS (S1)-1 bed
SICER_Repli-seq_BrdU earlyS (S1)-2 bed
SICER_Repli-seq_BrdU midS (S4)-2 bed
SICER_Repli-seq_BrdU midS (S4) -1 bed
SICER_Repli-seq_BrdU lateS (S7)-2 bed
SICER_Repli-seq_BrdU lateS (S7)-1 bed

MACS_CENP-A-TAP G1-2 bed
MACS_CENP-A-TAP G2-1 bed
MACS_CENP-A-LAP G1-2 bed
MACS_CENP-A-LAP G2-2 bed
MACS_CENP-A-LAP G2-1 bed
MACS_CENP-A-LAP G1-1 bed
MACS_H3.1-TAP-1 bed
MACS_H3.1-TAP-2 bed
MACS_CENP-A-TAP G2-2 bed
MACS_CENP-A-TAP RC-2 bed
MACS_CENP-A-TAP G1-1 bed
MACS_CENP-A-TAP RC-1 bed

H3.1-TAP RC-2: logTratio ge 2 bigBed
H3.1-TAP RC-1: logTratio ge 2 bigBed
CENP-A-TAP G2-2: logTratio ge 2 bigBed
CENP-A-TAP G1-2: logTratio ge 2 bigBed
CENP-A-TAP RC-2: logTratio ge 2 bigBed
CENP-A-TAP G2-1: logTratio ge 2 bigBed
CENP-A-TAP G1-1: logTratio ge 2 bigBed
CENP-A-TAP RC-1: logTratio ge 2 bigBed
CENP-A-LAP G2-2: logTratio ge 2 bigBed
CENP-A-LAP G2-1: logTratio ge 2 bigBed
CENP-A-LAP G1-1: logTratio ge 2 bigBed
CENP-A-LAP G1-2: logTratio ge 2 bigBed

single copy variant_CENP-A-TAP midS-2 bigWig
single copy variant_CENP-A-TAP midS-1 bigWig
Single copy variants_Input CA-LAP G1 bigWig
Single copy variants_CENP-A-LAP G2-2 bigWig
Single copy variants_CENP-A-LAP G2-1 bigWig
Single copy variants_CENP-A-LAP G1-2 bigWig
Single copy variants_CENP-A-LAP G1-1 bigWig
single copy variant_H3.1-TAP RC-2 bigWig
single copy variant_H3.1-TAP RC-1 bigWig
single copy variant_CENP-A-TAP G2-2 bigWig
single copy variant_CENP-A-TAP G1-2 bigWig
single copy variant_CENP-A-TAP RC-2 bigWig
single copy variant_CENP-A-TAP G2-1 bigWig
single copy variant_CENP-A-TAP G1-1 bigWig
single copy variant_CENP-A-TAP RC-1 bigWig
SICER_Single copy variants_CENP-A-LAP G2-2 bed

SICER_Single copy variants_CENP-A-LAP G2-1 bed
SICER_Single copy variants_CENP-A-LAP G1-2 bed
SICER_Single copy variants_CENP-A-LAP G1-1 bed
SICER_single copy variant_H3.1-TAP RC-2 bed
SICER_single copy variant_H3.1-TAP RC-1 bed
SICER_single copy variant_CENP-A-TAP G1-1 bed
SICER_single copy variant_CENP-A-TAP G2-2 bed
SICER_single copy variant_CENP-A-TAP G1-2 bed
SICER_single copy variant_CENP-A-TAP RC-2 bed
SICER_single copy variant_CENP-A-TAP G2-1 bed
SICER_single copy variant_CENP-A-TAP RC-1 bed
SICER_single copy variant_CENP-A-TAP midS-1 bed
SICER_single copy variant_CENP-A-TAP midS-2 bed

Genome browser session
(e.g. UCSC)

https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr2%3A93392761-93439154&hgsid=716430491_n0fAAmCKVZ78JwopvxVDAv65PSss

Methodology

Replicates

For each cell line and cell cycle phase, two separate replicates of ChIP-seq or repli-seq were conducted in parallel.

Sequencing depth

All sequences were paired-end.
All sequencing information, including sequencing depth and length is provided in Table S1.

Antibodies

For ChIP-seq of CENP-ATAP we used magnetic beads coupled to rabbit IgG (Sigma-Aldrich, CAT # I5006).
For ChIP-seq of CENP-ALAP we used magnetic beads coupled to a mouse GFP antibody (MSKCC antibody core facility, clones 19C8 and 19F7).
For Repli-seq experiments we used magnetic beads coupled to a mouse BrdU antibody (Becton-Dickinson Biosciences, CAT # 555627).

Peak calling parameters

Paired-end reads were merged (using PEAR) and mapped (using Bwa-Mem, standard parameters). Enrichment peaks for ChIP experiments were determined using SICER algorithm (v1.0.3), using relevant input reads as background. Stringent parameters previously optimized for human CENP-A (Lacoste et al., 2014) were used: threshold for redundancy allowed for chip reads: 1, threshold for redundancy allowed for control reads: 1, window size: 200 bps, fragment size: 150 bps, shift in window length is 150, effective genome size as a fraction of the reference genome of hg38: 0.74, gap size: 400 bps, e-value for identification of candidate islands that exhibit clustering: 1000, and false discovery rate controlling significance: 0.00001. All sequencing samples were normalized to their matched input control, i.e. each G1 sample was normalized to the G1 input and the G2 sequencing samples to G2 input. In parallel, MACs peak calling was performed (macs14), and wiggle tracks were created to represent read depth of each dataset independently. Detailed description of read mapping and peak calling strategies is provided in the Methods section.

Data quality

A rigorous evaluation of ectopic CENP-A peaks (peaks predicted outside of centromeric regions) was performed using k-mer enrichment. Each ectopic peak was reformatted into 50-bp sliding windows (in both orientations, with slide of 1bp). The normalized frequency of each 50-mer candidate windows were evaluated in each ChIP-seq dataset relative to a normalized observed frequency in the corresponding background dataset. Scores were determined as the log transformed normalized value of the ratio between ChIP-seq and background, and those with a score greater than or equal to 2 were included in our study as a high-confident enrichment set. There were 41,767 ectopic CENP-ATAP peaks at G1 and 18,566 at G2, with FDR controlling significance: 0.00001. Out of the 41,767 G1 ectopic CENP-ATAP peaks, 12,550 peaks were >5-fold over background and only 1,260 peaks at G2. For CENP-ALAP, there were 11,390 ectopic CENP-ALAP peaks at G1, of which 620 were >5-fold over background. Detailed description of peak evaluation is provided in the Methods section.

Software

ChIP-seq and Repli-seq data were collected and analyzed using:
Merge paired-end reads using PEAR (Zhang et al., 2014) https://sco.h-its.org/exelixis/web/software/pear/doc.html

Map merged paired reads using Bwa-Mem (Li and Durbin, 2009; Li and Durbin, 2010; Li, 2013) https://github.com/lh3/bwa

Peak calling using MACS (macs14) (Zhang et al., 2008) http://liulab.dfci.harvard.edu/MACS/

Peak calling using SICER (v1.0.3) (Zang et al., 2009) https://home.gwu.edu/~wpeng/Software.htm

Intersection with alpha-satellite or with functional annotations using BEDTools: intersect (Quinlan and Hall, 2010) https://github.com/arq5x/bedtools

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	DLD1 and HeLa cells were synchronized using single or double thymidine block and release, respectively. Cells were harvested at different time points following release to obtain cells enriched at G1 (18 hours post release), early S phase (1 hour post release), mid S phase (4 hours post release), G2 (7 hours post release), or mitosis (10 hours post release). Cells were fixed with 70% ethanol, and then stained with propidium iodide to asses the DNA content of the cells. DNA content was analyzed using BD LSR II Flow Cytometer.
Instrument	BD LSR II Flow Cytometer (BD Biosciences) was used for data collection
Software	The data was collected by FACSDiva DNA content wasanalyzed by ModFit 3.0
Cell population abundance	At least 30,000 cells were acquired for each sample. 80-95% of the synchronized cells were at the desired cell cycle stage.
Gating strategy	FSC/SSC gate was used for gating the population of cells to exclude cell doublets and debris. Asynchronous growing cell population was first run to set up the axis. All test samples were then run in identical conditions.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.