HHS Public Access

Author manuscript

Nat Cell Biol. Author manuscript; available in PMC 2020 February 12.

Published in final edited form as:

Nat Cell Biol. 2019 June ; 21(6): 743–754. doi: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[1,2,*,#], **Karen H. Miga**[3,*], **Ofer Shoshani**[1], **Aaron Aslanian**[4], **Moira A. McMahon**[1], **Ah Young Lee**[1], **Daniele Fachinetti**[1,5], **John R. Yates III**[4], **Bing Ren**[1], **Don W. Cleveland**[1,#]

[1]Ludwig Institute for Cancer Research and Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, CA 92093 USA

[2]Present address: Department of Pharmacology & Chemical Biology, School of Medicine, University of Pittsburgh Cancer Institute (UPCI), University of Pittsburgh, Pittsburgh, PA 15213 USA

[3]Center for Biomolecular Science & Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA

[4]Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA

[5]Present address: Institut Curie, PSL Research University, CNRS, UMR 144, 26 rue d'Ulm, F-75005, Paris, France

Abstract

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 pre-replication loading. Thus, DNA replication acts as an error correction mechanism for maintaining centromere identity through its

[#]Correspondence should be addressed to DWC and/or YNA.

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 analyzed the sequencing data. M.A.M. and O.S. analyzed 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.Y.III performed mass spectrometry experiments and provided resources.

[*]These authors contributed equally

Conflict of Interests

The authors declare that they have no conflict of interest.

removal of non-centromeric CENP-A coupled with CCAN-mediated retention and precise reloading of centromeric CENP-A.

Introduction

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 a ~171 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 ~3% of α-satellite DNA repeats[7,8], and is flanked by pericentric heterochromatin containing H3K9me2/3[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 and 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[20,21] is active[22]. Temporal separation of centromeric DNA replication from new CENP-A chromatin assembly raises the important question of how is the centromeric epigenetic mark 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[18]. Moreover, endogenous CENP-A comprises only ~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 and mapping onto centromere reference models for the centromere of each human chromosome, we now identify 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 of the centromeres.

Results

CENP-A binding at 23 human centromere reference models

We produced HeLa cells expressing either i) CENP-A^{LAP}, a CENP-A variant carboxy-terminally fused to a localization [EYFP] and affinity [His] purification tag[27] at one endogenous CENP-A allele (Supplementary Fig. 1a, Fig. 1a) or ii) stably expressing an elevated level (4.5 times the level 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 TEV protease cleavage site (Supplementary Fig. 1b, Fig. 1a). Both CENP-A variants localize to centromeres (Supplementary Fig. 1c,d) and support long-term centromere identity and mediate high fidelity chromosome segregation in the absence of wild type CENP-A[7,8]. Importantly, while the HeLa cells we use have acquired an aneuploid genome, they are chromosomally stable, with high fidelity centromere function that has maintained the same karyotype over nearly two decades[28].

To immunopurify CENP-A-containing chromatin at G1 or G2, chromatin was isolated from synchronized cells (Fig. 1a; Supplementary Fig. 1e). Nuclease digestion was used to produce mono-nucleosomes 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, 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 TEV 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 ~3% of the genome[8,29]. While micro-capillary 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 centered 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; 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; Supplementary Fig. 2) using algorithm-based scripts [SICER and MACS[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 Figure S2 for the other 22). Sequences bound were largely unaffected by increasing CENP-A levels by 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 which 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 elevated 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[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. 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 and with the remaining 3 peaks only slightly diminished (Fig. 2; 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. 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 (Spplementary 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, 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 (Spplementary 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 chromatin immunoprecipitation following synchronization in G1 or mitosis (Supplementary Fig. 3b) in a second, nearly diploid human cell line (DLD-1) 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} [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 kb on chromosome 4[40] (87.278 to 87.578 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 >5-fold 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[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 with 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 corresponding 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 2-fold (compared to 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 and which are functionally related to transcriptional activity. Similarly, CENP-A expressed at endogenous levels was enriched 3-fold 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 Fig. 3d,f and Supplementary Fig. 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 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’ (HOT) regions defined by highly expressed regions of the genome (and that show binding of unrelated transcription factors without underlying sequence specificity[42]) were almost quantitatively removed in cells that had progressed through DNA replication (Supplementary Fig. 3h,i), demonstrating that enrichment of CENP-A at such highly expressed regions cannot be a consequence of non-specific binding to “hyper-ChIPable regions”[43].

Analysis of published CENP-A ChIP-seq datasets for HT1080 cells[44], a human epithelial fibrosarcoma cell line expressing FLAG-tagged CENP-A at ~3-fold 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-sequencing datasets from the HuRef human lymphoblastoid cell line[45] again 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 analyzed 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 replication machinery. CENP-A^{TAP} was affinity purified from mid S phase cells and CENP-A-bound DNAs were sequenced and mapped (Fig. 6a; Supplementary Fig. 5a). In parallel, newly synthesized DNA in synchronized cells was labeled by addition of bromodeoxyuridine (BrdU) for 1 hour at early (S0-S1), mid (S3-S4), and late S phase (S6-S7) (Fig. 6a; Supplementary Fig. 5a). Genomic DNA from each time point was sonicated (Supplementary Fig. 4b) 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 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; 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; 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 at mid-S phase (Fig. 6g,h). Similarly, ectopic CENP-A binding sites found in mid-S replicating regions remained 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 (Fig. 6d–f; Fig. 2; Supplementary Fig. 5h). 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 which 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 panel). A structural link that normally bridges multiple centromeric CENP-A nucleosomes and nucleates 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 mono-nucleosomal CENP-A chromatin affinity purified from late S/G2

(Table 1). Stable association with CENP-A was also seen for HJURP, multiple chromatin remodeling 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 and which is maintained even on mono-nucleosomes 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. 5hi, right panel). 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[57], is a binding partner in a CENP-A pre-nucleosomal 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 through late S/G2 (Fig. 7a). In striking contrast, the two other CAF1 subunits (CAF1 p150 and CAF1 p60) that are essential for *de novo* chromatin assembly *in vitro*[59], remained much more strongly associated with CENP-A nucleosomes in late S/G2 compared to 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 co-purified 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 to 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 possibililty 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 in order to test CENP-C's role specifically during centromeric DNA replication in late S phase (Fig. 7b), but without affecting deposition of new CENP-A that occurs in early G1.

Auxin addition 2 hours after release from thymidine block resulted in polyubiquitination and degradation of almost all CENP-C^{AE} within 15 minutes, as was evident by loss of fluorescence of the EYFP in CENP-C^{AE} (Fig. 7b–d and video S1). 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 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). At 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 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]) prior to 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 prior to 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 >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 identifies 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 re-loads CENP-A back onto the exact same centromeric DNA site (Figs. 2, 6f,g).

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, 6h–m), therefore precluding pre-mitotic 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 mono-nucleosomes at 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 to 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank A. Desai, P. Ly and C. Eissler for critical discussion, L.E.T. Jansen (Gulbenkian Institute, Portugal) for reagents, 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.

References

1. Wevrick R & Willard HF 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 U S A 86, 9394–8 (1989). [PubMed: 2594775]
2. Cleveland DW, Mao Y & Sullivan KF Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell 112, 407–21 (2003). [PubMed: 12600307]

3. Willard HF Chromosome-specific organization of human alpha satellite DNA. American journal of human genetics 37, 524–32 (1985). [PubMed: 2988334]
4. Manuelidis L & Wu JC Homology between human and simian repeated DNA. Nature 276, 92–4 (1978). [PubMed: 105293]
5. Earnshaw WC & Rothfield N Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma 91, 313–21 (1985). [PubMed: 2579778]
6. Palmer DK, O'Day K, Wener MH, Andrews BS & Margolis RLA 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J Cell Biol 104, 805–15 (1987). [PubMed: 3558482]
7. Bodor DL et al. The quantitative architecture of centromeric chromatin. eLife 3, e02137 (2014). [PubMed: 25027692]
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). [PubMed: 28235947]
9. Sullivan BA & Karpen GH Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nature structural & molecular biology 11, 1076–83 (2004).
10. Stimpson KM & Sullivan BA Epigenomics of centromere assembly and function. Current opinion in cell biology 22, 772–80 (2010). [PubMed: 20675111]
11. Marshall OJ, Chueh AC, Wong LH & Choo KH Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am J Hum Genet 82, 261–82 (2008). [PubMed: 18252209]
12. Fachinetti D et al. A two-step mechanism for epigenetic specification of centromere identity and function. Nature cell biology 15, 1056–66 (2013). [PubMed: 23873148]
13. Foltz DR et al. The human CENP-A centromeric nucleosome-associated complex. Nature cell biology 8, 458–69 (2006). [PubMed: 16622419]
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, 446–57 (2006). [PubMed: 16622420]
15. Hori T et al. CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. Cell 135, 1039–52 (2008). [PubMed: 19070575]
16. Hori T, Shang WH, Takeuchi K & Fukagawa T The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. The Journal of cell biology 200, 45–60 (2013). [PubMed: 23277427]
17. Padeganeh A et al. Octameric CENP-A nucleosomes are present at human centromeres throughout the cell cycle. Current biology: CB 23, 764–9 (2013). [PubMed: 23623556]
18. Jansen LE, Black BE, Foltz DR & Cleveland DW Propagation of centromeric chromatin requires exit from mitosis. The Journal of cell biology 176, 795–805 (2007). [PubMed: 17339380]
19. Nechemia-Arbely Y, Fachinetti D & Cleveland DW Replicating centromeric chromatin: spatial and temporal control of CENP-A assembly. Experimental cell research 318, 1353–60 (2012). [PubMed: 22561213]
20. Foltz DR et al. Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell 137, 472–84 (2009). [PubMed: 19410544]
21. Dunleavy EM et al. HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell 137, 485–97 (2009). [PubMed: 19410545]
22. Silva MC et al. Cdk activity couples epigenetic centromere inheritance to cell cycle progression. Developmental Cell 22, 52–63 (2012). [PubMed: 22169070]
23. Lacoste N et al. Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX. Molecular cell 53, 631–44 (2014). [PubMed: 24530302]
24. Van Hooser AA et al. Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. Journal of cell science 114, 3529–42 (2001). [PubMed: 11682612]
25. Shrestha RL et al. Mislocalization of centromeric histone H3 variant CENP-A contributes to chromosomal instability (CIN) in human cells. Oncotarget 8, 46781–46800 (2017). [PubMed: 28596481]

26. Filipescu D et al. Essential role for centromeric factors following p53 loss and oncogenic transformation. Genes Dev 31, 463–480 (2017). [PubMed: 28356341]
27. Mata JF, Lopes T, Gardner R & Jansen LE A rapid FACS-based strategy to isolate human gene knockin and knockout clones. PloS one 7, e32646 (2012). [PubMed: 22393430]
28. Landry JJ et al. The genomic and transcriptomic landscape of a HeLa cell line. G3 3, 1213–24 (2013). [PubMed: 23550136]
29. Hayden KE et al. Sequences associated with centromere competency in the human genome. Molecular and cellular biology 33, 763–72 (2013). [PubMed: 23230266]
30. Conde e Silva N et al. CENP-A-containing nucleosomes: easier disassembly versus exclusive centromeric localization. Journal of molecular biology 370, 555–73 (2007). [PubMed: 17524417]
31. Miga KH et al. Centromere reference models for human chromosomes X and Y satellite arrays. Genome research 24, 697–707 (2014). [PubMed: 24501022]
32. Schneider VA 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). [PubMed: 28396521]
33. Levy S et al. The diploid genome sequence of an individual human. PLoS biology 5, e254 (2007). [PubMed: 17803354]
34. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). Genome biology 9, R137 (2008). [PubMed: 18798982]
35. Zang C et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics 25, 1952–8 (2009). [PubMed: 19505939]
36. Maloney KA et al. Functional epialleles at an endogenous human centromere. Proc Natl Acad Sci U S A 109, 13704–9 (2012). [PubMed: 22847449]
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). [PubMed: 27918550]
38. Amor DJ & Choo KH Neocentromeres: role in human disease, evolution, and centromere study. American journal of human genetics 71, 695–714 (2002). [PubMed: 12196915]
39. Hasson D et al. The octamer is the major form of CENP-A nucleosomes at human centromeres. Nature structural & molecular biology 20, 687–95 (2013).
40. Amor DJ et al. Human centromere repositioning “in progress”. Proceedings of the National Academy of Sciences of the United States of America 101, 6542–7 (2004). [PubMed: 15084747]
41. Alonso A et al. Co-localization of CENP-C and CENP-H to discontinuous domains of CENP-A chromatin at human neocentromeres. Genome biology 8, R148 (2007). [PubMed: 17651496]
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). [PubMed: 26113264]
43. Teytelman L, Thurtle DM, Rine J & van Oudenaarden A Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. Proc Natl Acad Sci U S A 110, 18602–7 (2013). [PubMed: 24173036]
44. Thakur J & Henikoff S Unexpected conformational variations of the human centromeric chromatin complex. Genes Dev 32, 20–25 (2018). [PubMed: 29386331]
45. Henikoff JG, Thakur J, Kasinathan S & Henikoff S A unique chromatin complex occupies young alpha-satellite arrays of human centromeres. Sci Adv 1(2015).
46. Hansen RS et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A 107, 139–44 (2010). [PubMed: 19966280]
47. Bui M et al. Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. Cell 150, 317–26 (2012). [PubMed: 22817894]
48. Erliandri I et al. Replication of alpha-satellite DNA arrays in endogenous human centromeric regions and in human artificial chromosome. Nucleic Acids Res 42, 11502–16 (2014). [PubMed: 25228468]
49. Guse A, Carroll CW, Moree B, Fuller CJ & Straight AF In vitro centromere and kinetochore assembly on defined chromatin templates. Nature 477, 354–8 (2011). [PubMed: 21874020]
50. Carroll CW, Milks KJ & Straight AF Dual recognition of CENP-A nucleosomes is required for centromere assembly. J Cell Biol 189, 1143–55 (2010). [PubMed: 20566683]

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 e15 (2016). [PubMed: 27881301]
52. Rago F, Gascoigne KE & Cheeseman IM Distinct organization and regulation of the outer kinetochore KMN network downstream of CENP-C and CENP-T. Curr Biol 25, 671–7 (2015). [PubMed: 25660545]
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). [PubMed: 26124289]
54. Weir JR et al. Insights from biochemical reconstitution into the architecture of human kinetochores. Nature 537, 249–253 (2016). [PubMed: 27580032]
55. Hoffmann S et al. CENP-A Is Dispensable for Mitotic Centromere Function after Initial Centromere/Kinetochore Assembly. Cell Rep 17, 2394–2404 (2016). [PubMed: 27880912]
56. Smith S & Stillman B Stepwise assembly of chromatin during DNA replication in vitro. EMBO J 10, 971–80 (1991). [PubMed: 1849080]
57. Verreault A, Kaufman PD, Kobayashi R & Stillman B Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. Cell 87, 95–104 (1996). [PubMed: 8858152]
58. Hayashi T et al. Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. Cell 118, 715–29 (2004). [PubMed: 15369671]
59. Kaufman PD, 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, 1105–14 (1995). [PubMed: 7600578]
60. Huang H et al. A unique binding mode enables MCM2 to chaperone histones H3-H4 at replication forks. Nat Struct Mol Biol 22, 618–26 (2015). [PubMed: 26167883]
61. Zhang W et al. Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. Nature communications 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). [PubMed: 27062469]
63. Gascoigne KE et al. Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. Cell 145, 410–22 (2011). [PubMed: 21529714]
64. Zasadzinska E et al. Inheritance of CENP-A Nucleosomes during DNA Replication Requires HJURP. Dev Cell 47, 348–362 e7 (2018). [PubMed: 30293838]

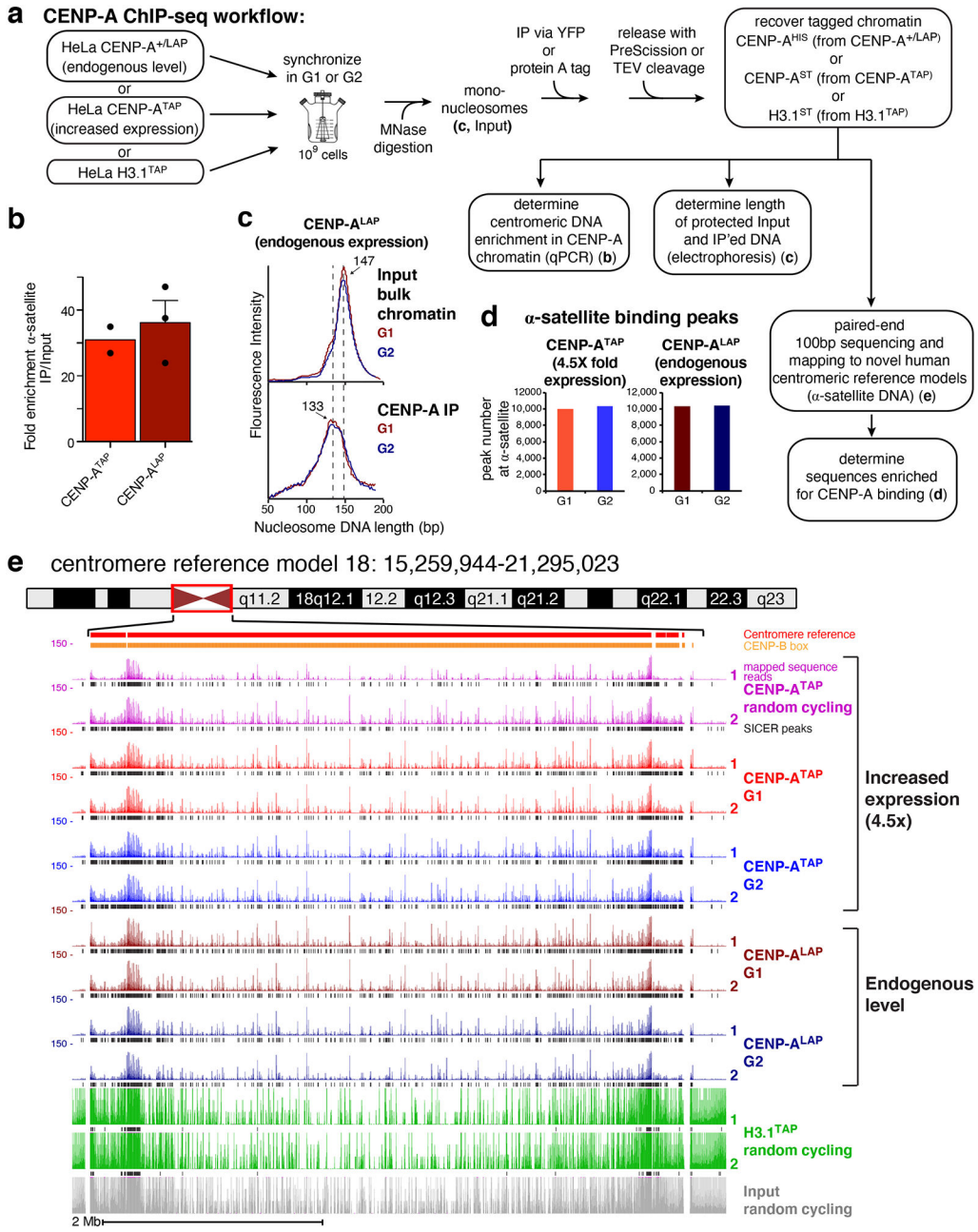

Figure 1. CENP-A ChIP-seq identifies CENP-A binding at reference centromeres of 23 human chromosomes.

(**a**) CENP-A ChIP-sequencing experimental design. (**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 mono-nucleosomes (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 at α-satellite DNA in CENP-A^{TAP} and CENP-A$^{+/LAP}$ cells at G1 and G2. The number represent peaks that are overlapping 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 multi-mapping and their exact linear order is not known. SICER peaks are shown in black below the raw read data. Two replicates are shown for each condition. Scale bar, 2Mb. Source data for b and d can be found in Supplementary Table 4.

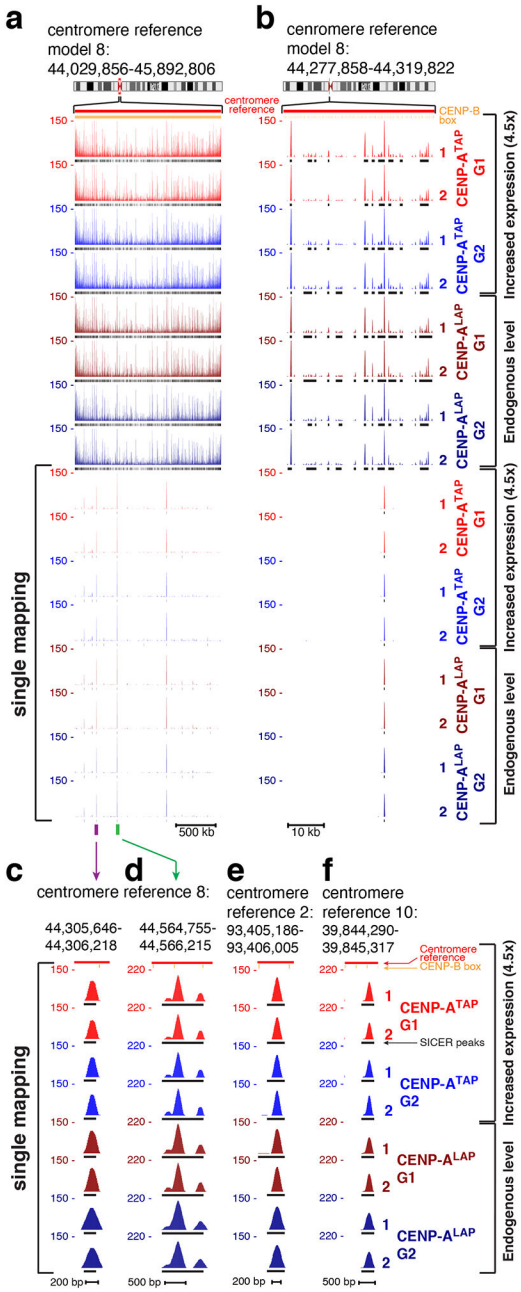

Figure 2. Retention of centromeric CENP-A through DNA replication.

(**a, b**) CENP-A ChIP-seq raw mapping data (colored) and SICER peaks (black lines, underneath) showing sequences bound by CENP-A (at both endogenous and increased expression levels) across centromere reference model of chromosome 8, before and after DNA replication. Upper part shows mapping of all reads (including reads that are multi-mapping) onto the repetitive α−satellite DNA. Lower part shows read mapping to sites that are single copy in the HuRef genome (single-mapping), after filtering out multi-mapping reads. Centromere reference location, red. CENP-B box location, orange. Scale bar, 500kb (**a**), 10kb (**b**). (**c**) High-resolution view of read mapping to a site that is single copy in centromere reference model of chromosome 8, marked by a purple bar in (**a**). Scale bar,

200bp. (**d**) High-resolution view of read mapping to a site that is single copy in centromere reference model of chromosome 8, marked by a green bar in (**a**). Scale bar, 500bp. (**e, f**) High-resolution view of read mapping to a site that is single copy in centromere reference model of chromosome 2 (**e**) and in the centromere reference model of chromosome 10 (**f**). Scale bar, 200bp (**e**) and 500bp (**f**). The experiment shown in (a-f) was repeated independently twice with similar results.

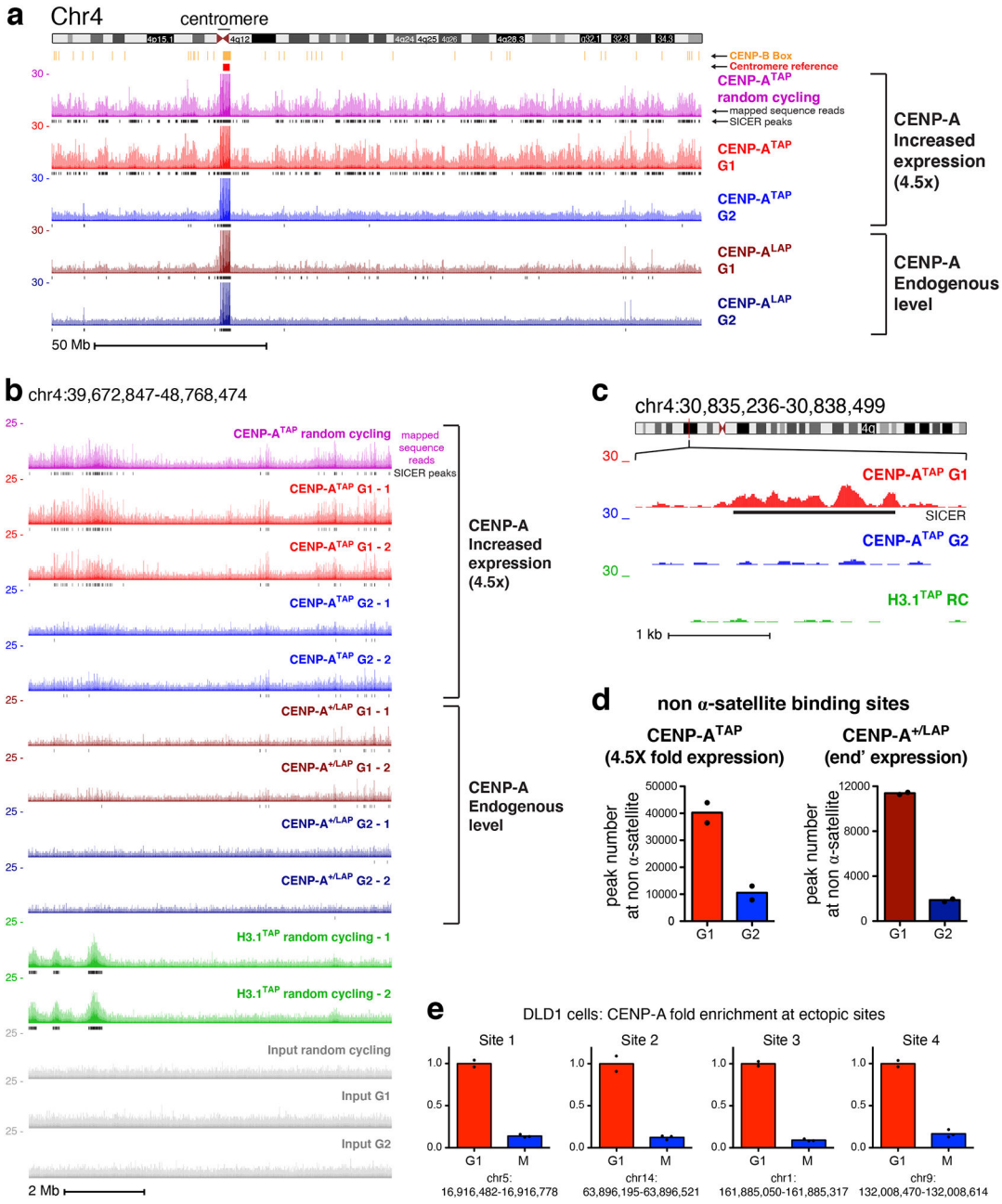

Figure 3. Sites of CENP-A assembly onto chromosome arms in early G1 are removed by G2.

(**a**) ChIP-sequencing raw mapping data (colored) 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. Centromere reference location, red. CENP-B box location, orange. Read counts were scale to 30 but reaches 150 at the centromere. Scale bar, 50Mb. (**b**) ChIP-sequencing data are shown 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}. Scale bar, 2Mb. (**c**) High resolution nucleosomal view of CENP-A^{TAP} mapping data at G1 and G2 at a non-centromeric site of chromosome 4. Scale bar, 1kb. 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 number 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} [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 qPCR 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. Resrults of two independent experiments for G1 and 3 for mitosis are shown. Source data for d and e can be found in Supplementary Table 4.

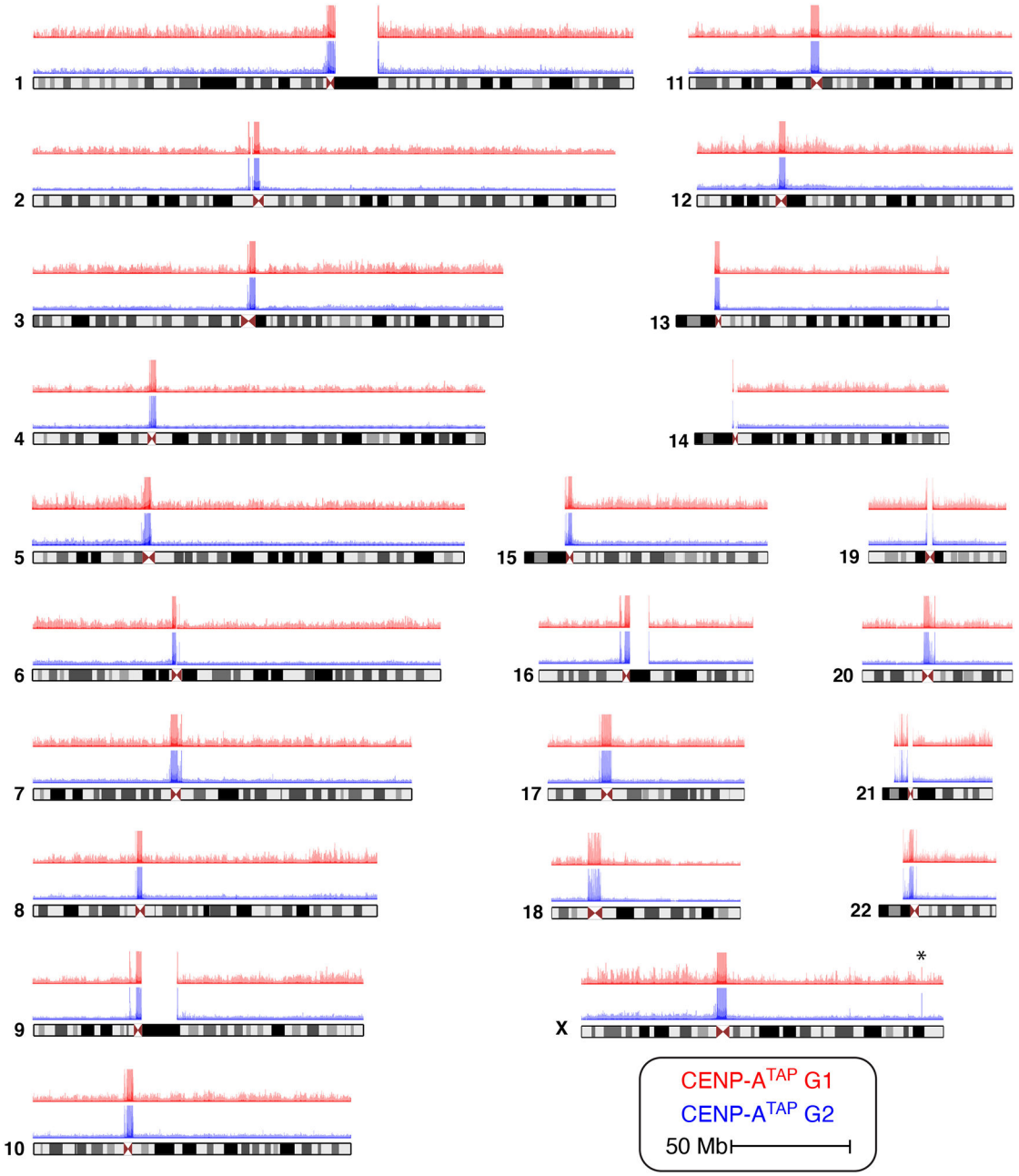

Figure 4. Ectopic CENP-A is removed following DNA replication from the arms of all 23 human chromosomes.

CENP-A^{TAP} ChIP-sequencing raw mapping data at G1 (red) and G2 (blue) 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.

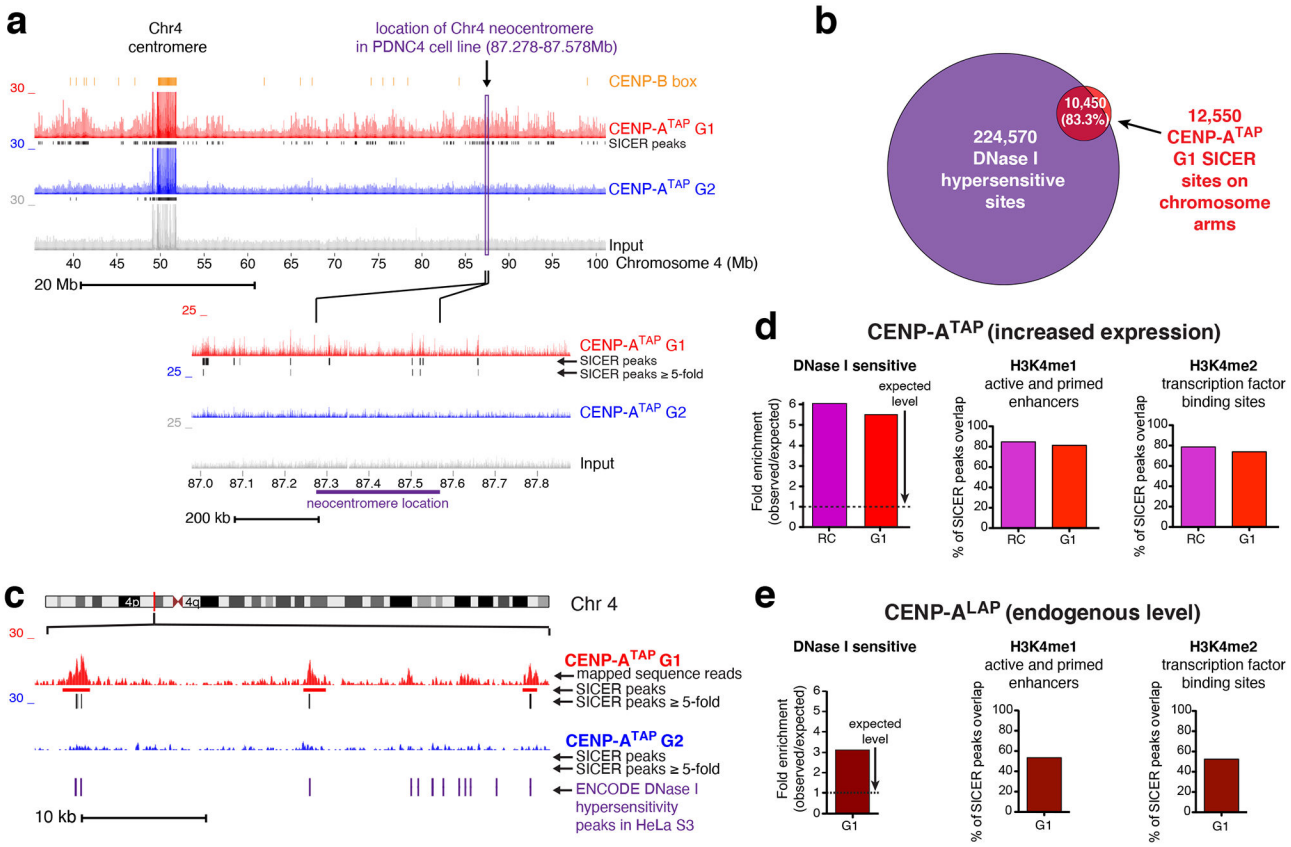

Figure 5. Neocentromeres are not positioned at sites of preferential CENP-A loading on chromosome arms.

(**a**) Read mapping data of CENP-A^{TAP} ChIP-sequencing 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**) Greater than 80% of CENP-A SICER peaks ≥ 5-fold in randomly cycling and G1 cells overlap with DNase I hypersensitive sites taken from ENCODE project. (**c**) Example from the chromosome 4 p-arm showing overlap of at least 100 bases between SICER peaks ≥ 5-fold and HeLa S3 DNase I hypersensitive sites taken from 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 supported between the two replicates were analyzed for their enrichment level at DNase I hypersensitive sites, with minimum overlap of 100 bases, compared to the level of enrichment at these sites by chance. SICER peaks ≥ 5-fold supported between the two replicates of CENP-A^{TAP} (d) or CENP-A^{LAP} (e) were analyzed for their overlap degree 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.

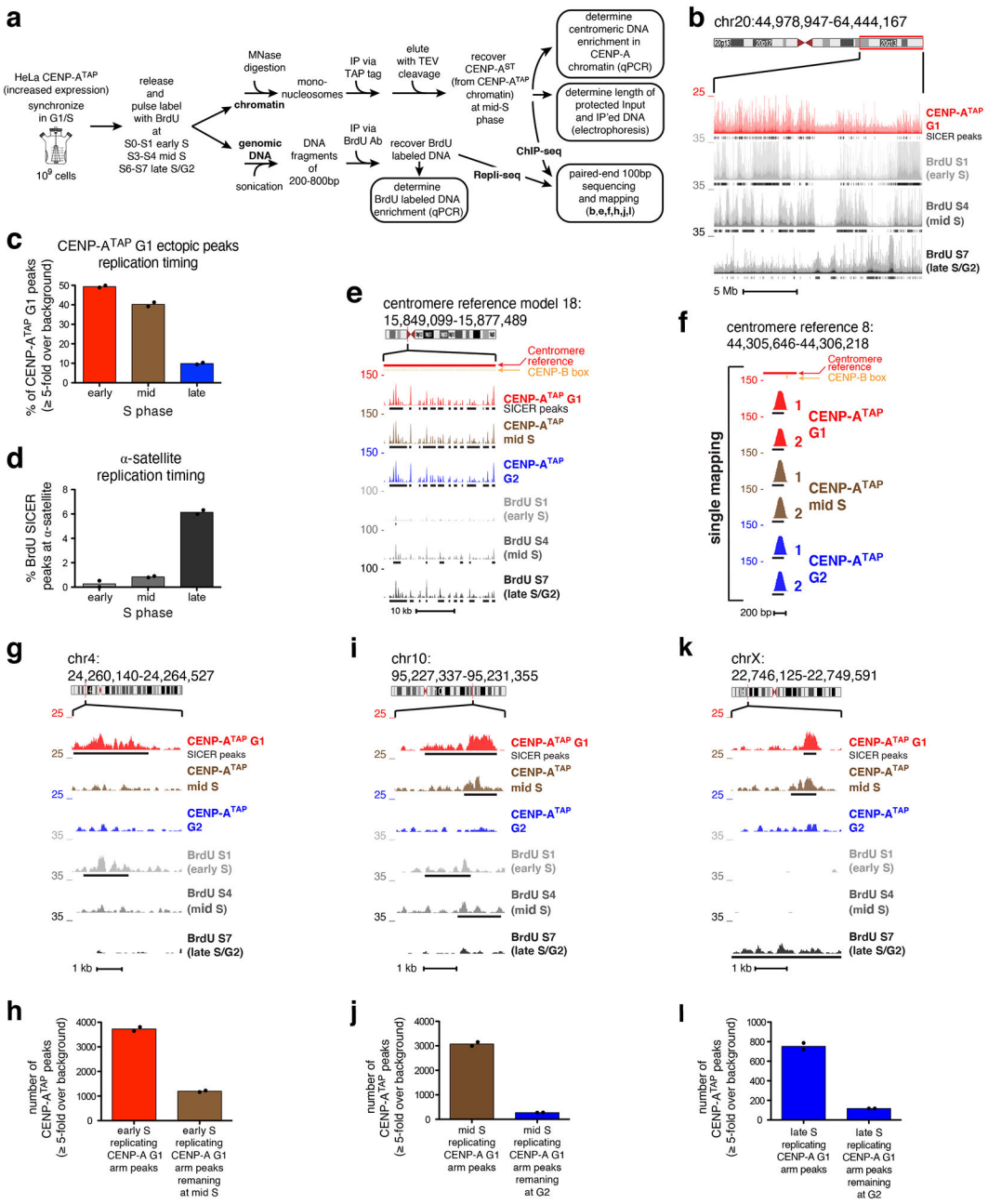

Figure 6. Ectopic CENP-A is removed contemporaneously with replication fork progression, while centromeric CENP-A is retained.

(**a**) Experimental design of CENP-A ChIP-seq combined with Repli-seq. (**b**) Raw mapping data of CENP-A^{TAP} ChIP-seq at G1 (red) and BrdU repli-seq (grey scale) at early S, mid S and late S/G2 at q-arm of chromosome 20. SICER peaks, black lines underneath raw mapping data. Scale bar, 5 Mb. (**c**) Percentage of ectopic G1 CENP-A^{TAP} ≥5-fold binding sites at early, mid, or late S replicating regions. (**d**) Percentage of BrdU SICER peaks at 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. SICER peaks, black lines underneath the raw mapping data. Centromere

reference location, red. CENP-B boxes, orange. Scale bar, 10kb. (**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 Figure 2c for CENP-A^{TAP} G1 and G2 is included for comparison. Scale bar, 200bp. (**g, i, k**) Raw mapping data (colored) and SICER peaks (black lines, underneath) of CENP-A^{TAP} ChIP-seq at G1, mid S phase and G2 and BrdU labeled repli-seq samples (grey scale) showing regions going through replication at 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. Scale bars, 1kb. (**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, and l can be found in Supplementary Table 4. Data shown in Fig 6b–l are from two biologically independent experiments.

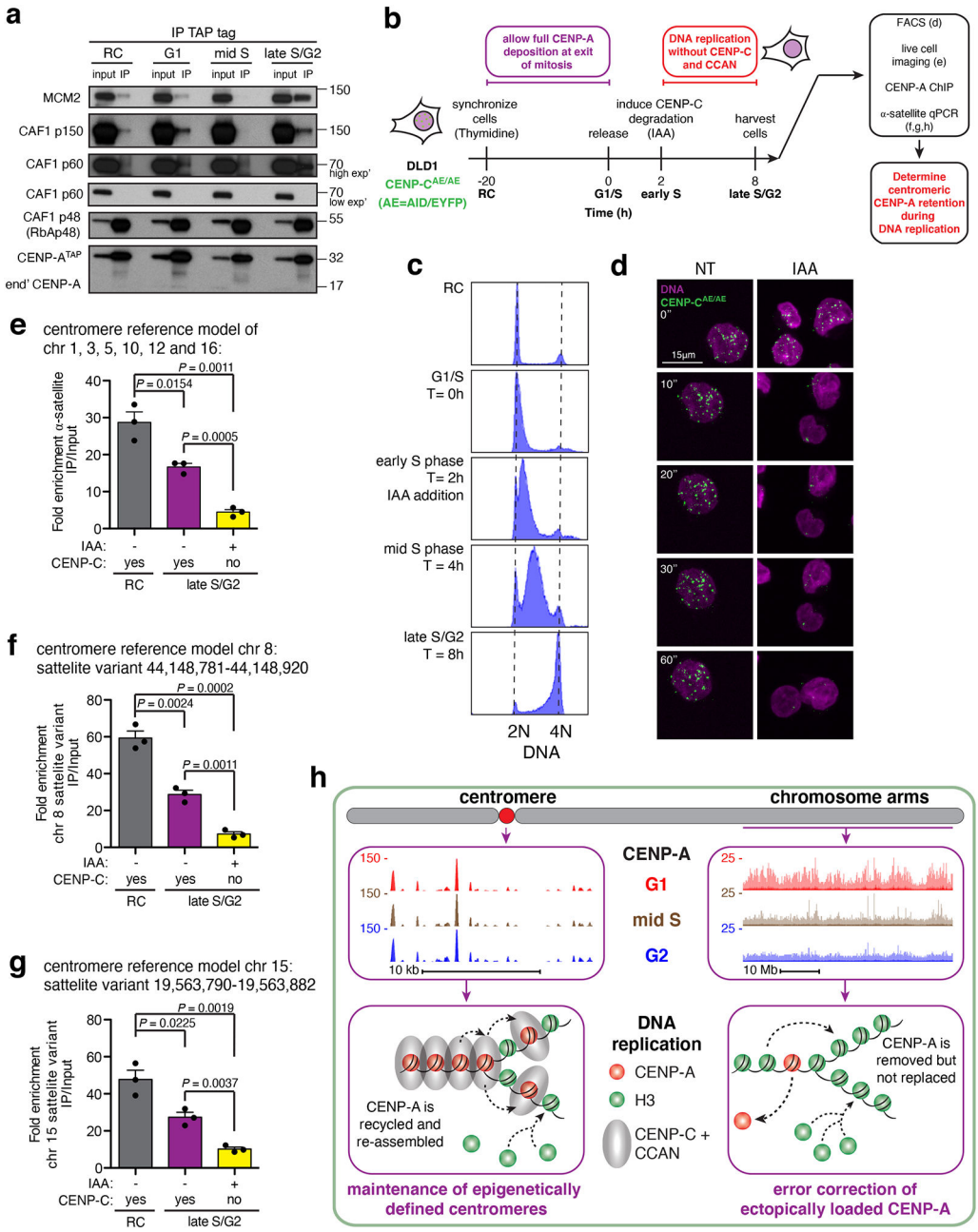

Figure 7. CENP-C and the CCAN complex are essential for the epigenetic maintenance of human centromeres during DNA replication.

(**a**) CENP-A^{TAP} was immunoprecipitated from micrococcal nuclease resistant chromatin isolated at 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**) 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 IAA at 0 minutes. DNA is labeled with sir-DNA, magenta. CENP-C$^{AE/AE}$, green. Scale bar, 15uM. The experiment was repeated independently three times with similar results. (**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.

Table 1.

Mass spectrometry of CENP-A individual nucleosomes reveal all the CCAN network components co-precipitated with CENP-A at late S/G2.

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 co-precipitated partners.

			Randomly cycling		Late S/G2	
			Spectra counts	Peptide counts	Spectra counts	Peptide counts
		CENP-A	107	22	70	21
		CENP-B	106	35	89	36
		CENP-V	11	7	6	5
CCAN	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