The octamer is the major form of CENP-A nucleosomes at human centromeres

Dan Hasson1,2,8, Tanya Panchenko3,4,8, Kevan J Salimian3,5,8, Mishah U Salman6, Nikolina Sekulic3, Alicia Alonso7, Peter E Warburton1,2 & Ben E Black3–5

The centromere is the chromosomal locus that ensures fidelity in genome transmission at cell division. Centromere protein A (CENP-A) is a histone H3 variant that specifies centromere location independently of DNA sequence. Conflicting evidence has emerged regarding the histone composition and stoichiometry of CENP-A nucleosomes. Here we show that the predominant form of the CENP-A particle at human centromeres is an octameric nucleosome. CENP-A nucleosomes are very highly phased on α-satellite 171-base-pair monomers at normal centromeres and also display strong positioning at neocentromeres. At either type of functional centromere, CENP-A nucleosomes exhibit similar DNA-wrapping behavior, as do octameric CENP-A nucleosomes reconstituted with recombiant components, having looser DNA termini than those on conventional nucleosomes containing canonical histone H3. Thus, the fundamental unit of the chromatin that epigenetically specifies centromere location in mammals is an octameric nucleosome with loose termini.

Faithful genome inheritance at cell division requires that each chromosome contain a single functional centromere1. The centromere is the site of assembly of the mitotic kinetochore—a massive complex of proteins that serves as the connection point to the microtubule-based spindle—and also serves as the site of final sister-chromatid cohesion1. Strong evidence suggests that CENP-A can provide the key epigenetic information to mark centromere location2–4 by distinguishing centromeres from the rest of the chromosome. Prime examples of the DNA sequence–independent nature of centromere inheritance are human neocentromeres that have been isolated out of the population, for which centromere function is uncoupled from the repetitive α-satellite DNA that typically houses CENP-A chromatin5–10. Fundamental unknowns remain regarding CENP-A nucleosomes, such as the histone composition and stoichiometry of the CENP-A particle and how much DNA it wraps.

There is now nearly a consensus on the point that purified recombinant CENP-A readily assembles into octameric nucleosomes in which two copies of CENP-A replace the two copies of canonical histone H3 (refs. 11–15). Reconstituted octameric CENP-A nucleosomes are known to have loose terminal-DNA contacts13,15,16. In addition to loose terminal-DNA wrapping, the CENP-A–targeting domain confers structural changes14,15, as well as conformational rigidity14,17, to the folded cores of reconstituted octameric nucleosomes. The relevance of all studies of recombinant nucleosomes to the context of native centromeric chromatin is unclear, however, because the field remains deeply divided over key issues on the nature of the protein–DNA particle into which CENP-A assembles in vivo18. Experiments involving isolation of CENP-A particles from various eukaryotic species have led to radically different models for the fundamental unit of centromeric chromatin including nonoctameric forms (for example, tetrasomes19, hemisomes20–23, hexasomes24 and so forth). Perhaps the two most intriguing and conflicting proposals for the major form of the CENP-A particles that specify centromere location in metazoans are for octameric nucleosomes and hemisomes. The two proposals suggest drastically different modes for how centromere-specifying chromatin particles are distinguished from bulk nucleosomes. Clear examples of when such molecular recognition is important at the centromere include the direct binding of CENP-A–-containing particles by the constitutive nonhistone centromere components CENP-N4,25 and CENP-C4,26.

To test the proposed models for the major form of the fundamental repeating unit of centromeric chromatin, we used native chromatin immunoprecipitation followed by sequencing (ChiP-seq) of CENP-A–containing particles from normal centromeres on α-satellite DNA and three naturally occurring neocentromeres.

RESULTS
CENP-A particles protect ~100–150 base pairs from nuclease

To investigate the nature of CENP-A–containing particles at functional human centromeres, we first considered the merits of a
micrococcal nuclease (MNase) digestion approach coupled to ChIP-seq. The MNase approach is attractive because it straightforwardly tests the specific predictions for how much DNA could be wrapped by octameric nucleosomes or hemisomes (Fig. 1a)\(^ {15,27}\). Since early nucleosome studies, MNase protection has been a standard for defining canonical nucleosomes\(^ {28,29}\). Crystallographic studies of canonical nucleosomes have defined how each histone dimer pair (H2A–H2B or H3–H4) has a single basic DNA-binding ridge that binds ~25–30 base pairs (bp) of DNA\(^ {27}\). The canonical histone octamer wraps ~100–120 bp of DNA in this way with approximately the final two turns (~20 bp) of terminal DNA stabilized by contacts with the αN helix of histone H3. Thus, in total, the canonical nucleosome core particle stably protects ~147 bp from MNase digestion. Tetrameric histone complexes of any sort have only enough DNA-wrapping surface to bind ~65 bp of DNA (Fig. 1b)\(^ {30–32}\). Before embarking on our ChIP-seq studies of CENP-A nucleosomes isolated from functional centromeres, we examined reconstituted CENP-A-containing complexes by using one of the same high-resolution, high-sensitivity detection methods with the native particles (described below). We found that recombinant CENP-A nucleosomes lack protection of crossed entry-exit DNA (that is, they do not protect a fragment corresponding to the ~165-bp peak protected by canonical nucleosomes containing conventional histone H3) and digest to three discrete peaks, one the size of the nucleosome core particle (~145 bp) and two of smaller size (~110 and ~130 bp; Fig. 1c and Supplementary Fig. 1). All of the fragment lengths protected by recombinant octameric CENP-A nucleosomes were substantially larger than any fragment from reconstituted (H3–H4) and (CENP-A–H4) tetrasomes, which protect ~65 bp (Fig. 1d). Thus, structural models and experiments with reconstituted particles encouraged us to pursue a similar MNase strategy with native CENP-A particles to distinguish between the radically different configurations that have been proposed to be the fundamental unit of functional centromeric chromatin.

Native ChIP of CENP-A–containing particles from human cultured cells strongly enriched centromere DNA and yielded MNase-protected fragments in three major size classes (~110 bp, ~130 bp and ~150 bp; Fig. 2a–c). Native ChIP of canonical nucleosomes containing conventional histone H3 yielded a distribution of a single size class of MNase-protected fragments that expectedly matched the input bulk nucleosomes (Fig. 2a and Supplementary Fig. 2a), which indicates that the smaller fragments observed for CENP-A–containing particles are not due to additional fragmentation during immunoprecipitation. If the smaller fragments (~110 bp) are derived from digestion of the nucleosome termini, as in our experiments with recombinant CENP-A octameric nucleosomes (Fig. 1c), then excessive digestion should remove the termini and leave a stable ~110-bp core fragment undigested. We tested this notion by repeating the isolation of CENP-A nucleosomes and using a low or high concentration of MNase. As expected, treatment with a high concentration of MNase yielded more heavily digested bulk chromatin with a higher mononucleosome/dinucleosome ratio and mononucleosomes that are trimmed down to core particles (Fig. 2d). Such treatment diminished the abundance of larger (~130–160 bp) and increased that of smaller (100–120 bp) MNase-protected DNA fragments from isolated CENP-A particles (Fig. 2e). These findings suggest that native CENP-A particles have a stable core with transiently unwrapping ends that are digested in a manner that is sensitive to the concentration of MNase.

Figure 1 Structure-based predictions for MNase protection and experimental outcomes with CENP-A–containing particles assembled with recombinant components. (a) Molecular models (cartoon representations, right) of the indicated proposed DNA–protein particles, demonstrating the expected length of DNA protected following MNase digestion. CENP-A and histones are color-coded. (b) Electrostatic surface potential maps depicting the predicted path of DNA wrapping a CENP-A–containing tetrasome or hemisome, where positively charged surfaces are colored in blue and negatively charged surfaces in red. A length of 64–66 bp completely covers the DNA-wrapping surface of either tetrameric configuration. (c,d) MNase digestion profiles of octameric CENP-A–or histone H3–containing mononucleosomes (c) or tetrasomes (d) reconstituted on a 200-bp template.
Transient unwrapping of nucleosome terminal DNA suggests that chemical protein-DNA cross-linking would lock the DNA to the CENP-A–containing histone octamer. Indeed, standard formaldehyde cross-linking, as is used in diverse chromatin studies\(^3,\!^3,\!^4\), yielded CENP-A–containing particles with a single distribution of MNase-protected fragments of ~150–170 bp, nearly identical to that of solubilized bulk nucleosomes (Supplementary Fig. 2b–e). These CENP-A particles were isolated out of nucleosome preparations that contained all detectable CENP-A protein (Supplementary Fig. 2b) and were specifically enriched for α-satellite DNA (Supplementary Fig. 2c) to a similar extent as were native preparations (Fig. 2c). Further, similar results were obtained for two independent cell types, one derived from healthy tissue (PD-NC4) and one derived from a tumor (HeLa; Supplementary Figs. 2d,e). Together, our findings suggest that we are monitoring the DNA-wrapping behavior of the major form of CENP-A nucleosomes and that, in doing so, our approach represents a highly sensitive means to probe centromere chromatin architecture.

**CENP-A–nucleosome positions on complex neocentromeric DNA**

We considered that the sub-145-bp MNase-protected fragments on natively prepared CENP-A particles could be caused either by the physical properties conferred by the incorporation of CENP-A into nucleosomes that make the terminal DNA susceptible to MNase digestion or by the properties imposed by the sequence or higher-order structure of the α-satellite DNA (where the monomer repeat unit is 171 bp\(^18,\!^35\)) upon which CENP-A is assembled at normal human centromeres. Neocentromeres provide a prime tool to investigate functional CENP-A nucleosomes in the absence of any effects imposed by α-satellite DNA. We used patient-derived cell lines with one neocentromere for ChIP-seq studies. Two of the neocentromeres map to single-copy, complex DNA sequences\(^6,\!^8\) (Fig. 3a,c), whereas the other is present on a repeat sequence where the ~12-kilobase (kb) monomer sequence is completely unrelated to α-satellites\(^10\) (Fig. 3b). We mapped paired-end CENP-A–nucleosome sequences and found strong enrichment at each of the neocentromeres we examined (Fig. 3a–c and Supplementary Table 1), which was in good agreement with earlier mapping efforts\(^6,\!^8,\!^{10}\). The vast majority of CENP-A nucleosomes at the neocentromere fell in the three CENP-A–nucleosome size classes (Fig. 2a and Supplementary Fig. 3c; three bins: 100–119 bp, 120–139 bp, 140–160 bp), and we found that all three size classes mapped to the same positions (Fig. 3d–i, Supplementary Figs. 3–5 and Supplementary Table 2).

The finding that the small (~110-bp), medium (~130-bp) and large (~150-bp) fragments localize to the same genomic positions (as opposed to distinct ones) is consistent with the notion that CENP-A nucleosomes have DNA termini that transiently unwrap and are thus prone to variable terminal nuclease digestion. Indeed, colocalization of all three size classes is evident for both quantitative global analysis of the entire neocentromere regions (Fig. 3d,f,h, Supplementary Figs. 3d, 4a and 5a and Supplementary Table 2) and for local analysis of CENP-A–nucleosome sites (Fig. 3e,i, Supplementary Figs. 3f–m, 4b–i and 5b–g and Supplementary Table 2). Initial removal of duplicate reads yielded similar results (Supplementary Fig. 3e,n), which indicates that the DNA-wrapping behavior that we observe is entirely attributable to the positioning of CENP-A–containing particles. Despite originating at diverse genomic locations on separate chromosomes, and despite the highly variable sizes and patterns of CENP-A–nucleosome enrichment (Fig. 3a–c), the wrapping behavior of individual CENP-A nucleosomes was markedly similar for all three of the neocentromeres we examined. In total, our analysis of neocentromeres suggests that the DNA-wrapping properties of CENP-A–containing particles are largely independent of sequence variation in complex DNA and can be attributed to the physical properties conferred by the presence of CENP-A.

**CENP-A nucleosomes on repetitive DNA of normal centromeres**

The highly repetitive nature of the DNA sequences found at normal centromeres raises the possibility that nucleosome positioning and DNA wrapping is more ordered on α-satellite DNA. Indeed, there are preferred MNase digestion sites of CENP-A–containing chromatin.

---

**Figure 2** Nuclease digestion of native CENP-A–containing particles resembles that of octamer nucleosomes with loose termini. (a) DNA length distributions of MNase-digested CENP-A native ChIP and bulk nucleosomes from the same preparation. (b) Fluorescence in situ hybridization using DNA from bulk nucleosomes or CENP-A native ChIP as probes. Bulk-nucleosome DNA labels the entire chromosome, whereas CENP-A probe labels solely centromeric regions, as shown. Scale bars, 10 μm. (c) Quantitative real-time PCR analysis comparing enrichment of CENP-A native ChIP DNA and bulk-nucleosome DNA. CENP-A ChIP sequences are enriched at α-satellite regions (α-satellite1, α-satellite2) but not at pericentric or promoter (ald) regions, as shown. Error bars, s.e.m. from three independent replicates. (d) Standard digestion (red) or overdigestion (blue, three-fold higher concentration of MNase used) of chromatin. (e) DNA length distributions of CENP-A native ChIP following standard digestion (red) or overdigestion (blue) of chromatin.

| Nucleosome length (bp) | Bulk nucleosomes | CENP-A native ChIP |
|------------------------|------------------|--------------------|
| 50                     | 50               | 50                 |
| 100                    | 100              | 100                |
| 150                    | 150              | 150                |
| 200                    | 200              | 200                |
| 250                    | 250              | 250                |
| 300                    | 300              | 300                |

Fold enrichment: 0, 5, 10, 15, 20.
within α-satellite monomers. Centromeres remain largely unannotated, and standard genomic-sequence filters discard these sequences. We developed a scheme that takes advantage of paired-end and long (100-bp) deep-sequencing reads (Supplementary Fig. 6a,b) and the fact that α-satellite monomers share >60% sequence identity to one another. Without both paired-end and long reads, it is impossible to identify the length or sequence of nucleosome-protected MNase fragments within such highly repetitive DNA. Our scheme was to align nucleosome sequences to a dimer α-satellite consensus sequence (Supplementary Fig. 6a,b). In doing so, we included all sequences that mapped within a single 140–160 bp monomer or spanned two monomers. For all three of the cell lines that we examined, we observed a biphasic behavior of CENP-A–nucleosome sequence alignments with a subset of sequences having an alignment value of 35–40% and with another of 260% (Fig. 4a and Supplementary Fig. 6c,g). The former subset represents the alignment value of random sequences (that is, sequences that do not originate from α-satellite DNA). The latter subset (260% identity shared with the α-satellite consensus) represents bona fide α-satellite sequences. CENP-A–nucleosome ChIP preparations were strongly enriched for α-satellite DNA, representing 35–52% of the sequences for each of the three cell lines used in this study (Fig. 4a and Supplementary Fig. 6c,g). Approximately 1.5% of bulk-nucleosome sequences were from α-satellite DNA and aligned with ≥60% identity, and this equated to 4–7 × 10^5 bulk-nucleosome sequences at centromeres (Fig. 4b and Supplementary Fig. 6d,h) for us to compare to their counterparts containing CENP-A.

The tripartite distribution of size classes of MNase digestion of CENP-A nucleosomes included 17–24% of the large bin (140–160 bp),

Figure 3 The three size classes of CENP-A nucleosomes localize to the same prominent positions on neocentromeres. (a–c) Bowtie-mapped paired-end CENP-A native ChIP-seq reads in three different human neocentromere-containing cell lines (PDNC4, MS4221 and IMS13q; ideograms, top) demonstrate the specificity of CENP-A native ChIP. The IMS13q neocentromere was formed on an aberrant chromosome with an inversion duplication. The MS4221 neocentromere contains repetitive DNA denoted by dotted lines with strikethrough. Mb, megabase. (d–i) Occupancy maps for the three different size classes of CENP-A nucleosomes along the length of the neocentromere for PDNC4 (d), MS4221 (f) and IMS13q (h) and within a subsection (2,500-bp window) in e, g and i, respectively.

Figure 4 CENP-A nucleosomes on the repetitive α-satellite DNA of normal centromeres have a tripartite distribution of nuclease-protected DNA fragments. (a,b) Alignment of CENP-A (a) or bulk-nucleosome (b) fragments to a dimer α-satellite consensus sequence. (c,d) Distribution of DNA lengths of all CENP-A (c) or bulk-nucleosome (d) fragments aligning to the α-satellite consensus sequence with ≥60% identity.
36–42% of the middle bin (120–139 bp) and 32–38% of the small bin (100–119 bp), with small variation observed between experiments performed in the three cell lines used in this study (Fig. 4c and Supplementary Fig. 6e,i). The tripartite distribution is in stark contrast to bulk nucleosomes on α-satellite DNA, for which MNase protection of 140–160 bp or slightly larger fragments predominated (Fig. 4d and Supplementary Fig. 6f)). Consistent with fully wrapped nucleosomes with or without crossed linker DNA at the entry–exit positions (Fig. 1a). Therefore, even when wrapped with nearly identical sequences—the closely related α-satellite DNA of normal centromeres—CENP-A nucleosomes exhibit distinctly shorter lengths of MNase protection than do their conventional counterparts with canonical histone H3.

Phasing of CENP-A nucleosomes on α-satellite DNA
To measure the degree of phasing of CENP-A nucleosomes on α-satellite DNA and to investigate the relationship between the three different size classes of DNA fragments protected from MNase digestion, we mapped our sequencing data back to the dimerized α-satellite sequence (Fig. 5). CENP-A nucleosomes were highly phased on α-satellite DNA, with the small (100–119 bp) and medium (120–139 bp) MNase-protected fragments showing the highest level of phasing (Fig. 5a and Supplementary Fig. 7a,d). The small- and medium-sized MNase-protected fragments share a 5′ digestion site ~15–20 bp 3′ of the position of the CENP-B box (a 17-bp binding site for the CENP-B protein\(^{37}\)), with the smallest fragments digested to be ~20 bp shorter than the medium-sized fragments at their 3′ ends (Fig. 5a and Supplementary Fig. 7a,d). Phasing of bulk nucleosomes on α-satellite DNA was less pronounced, but there was one clearly preferred site with MNase digestion near the 3′ end of the first CENP-B box and ~5–10 bp 5′ of the second CENP-B box (Fig. 5b and Supplementary Fig. 7b,e).

We predicted that CENP-A–containing and histone H3–containing octameric nucleosomes would have similar preferred sites on α-satellite DNA because the basic residues that contact nucleosomal DNA are largely conserved on the surface of the (CENP-A–H4)\(_2\) heterotetramer relative to (H3–H4)\(_2\) (Fig. 1)\(^{14,15}\). Upon plotting the midpoints of all nucleosome sequences that map to α-satellite...
DNA, we found that the most prominent position of the small (100–119 bp) MNase fragments from CENP-A nucleosomes was identical to the most prominent bulk-nucleosome position (Fig. 5c,d and Supplementary Fig. 7c,f; CENP-A ChIP compared to bulk nucleosomes). The midpoint of the middle-sized MNase fragments from CENP-A nucleosomes was shifted 10 bp 3′ of the midpoint of the small-sized CENP-A fragments and bulk nucleosomes (Fig. 5c,d and Supplementary Fig. 7c,f; CENP-A ChIP). Together, these data argue for a model for nucleosome positioning on α-satellite DNA wherein (i) canonical nucleosomes prefer a site between CENP-B boxes and maintain strong terminal DNA wrapping with their dyad axes positioned at or very near the midpoint peak we observed (Fig. 5d,f), (ii) the small-sized CENP-A fragments (Fig. 5c,e) represent MNase digestion of 15–20 bp from each end of a nucleosome with identical dyad-axis positioning, and (iii) the medium-sized CENP-A fragments (Fig. 5c,e) represent asymmetrically digested MNase products that have been cleaved by 15–20 bp at their 5′ end but not their 3′ end. Further, CENP-A nucleosomes at their most prominent position at centromeres did not strongly protect fragments >140 bp (Fig. 5). To the contrary, the >140 bp fragments protected by CENP-A nucleosomes were not well phased (Fig. 5a,c). Thus, in the context of their preferred biological context on the chromosome, CENP-A nucleosomes are strongly phased, and their propensity to unwrap DNA at their termini is accentuated, especially at the 5′ nucleosome entry-exit site (Fig. 5e; for CENP-A nucleosomes, the i′ site is almost always the site of cleavage, and the i site is very rarely used).

The role of CENP-B boxes in CENP-A–nucleosome phasing

Because our initial analysis (Fig. 5) suggested a strong relationship between the positioning of CENP-B boxes and the CENP-A nucleosome, we next investigated the extent to which CENP-A–nucleosome phasing is dependent upon functional CENP-B boxes. The mapping scheme we first used to examine the phasing of CENP-A nucleosomes on α-satellite DNA revealed that the most prominent CENP-A–nucleosome location in the genome yields MNase-digested fragments that exclude the location of the CENP-B box (Fig. 5a). Thus, such a mapping strategy based on consensus sequence does not allow us to directly assess the relationship of these nucleosome positions relative to functional CENP-B boxes that contain the key nucleotide sequence for recognition by the CENP-B protein37. Therefore, we further examined CENP-A–nucleosome positions in chromosome-specific higher-order repeat (HOR) α-satellite DNA sequences that have been identified for almost all human chromosomes, although many are poorly annotated in the human genome38. Most of these HORs contain a functional CENP-B box in some fraction of their monomers. Here we chose to examine the well-characterized 2-kb HOR from the X chromosome, which contains functional CENP-B boxes in 4 of its 12 monomers (Fig. 6a)38,39.

We compared this to the α-satellite HOR found on the Y chromosome, which does not contain any functional CENP-B boxes; in fact the Y chromosome is the only chromosome that does not show any binding of the CENP-B protein at its centromere40,41. Because we have contiguous end-to-end sequence reads for all of the CENP-A nucleosome–derived DNA sequences, we can effectively align them to these HORs. For instance, two of the neocentromere cell lines we used are derived from females and yield almost no CENP-A nucleosome–derived fragments that align with the Y-chromosome HOR (Table 1). One cell line, MS4221, is derived from a male and yields >500,000 CENP-A nucleosome–derived fragments that align with the Y-chromosome HOR (Table 1). Thus, our mapping strategy is extremely stringent and provides an attractive means to very faithfully and precisely assign CENP-A nucleosome–derived fragments to their location within annotated HORs.

Table 1  Number of CENP-A reads at HORs

| Cell line | Gender | X chromosome | Y chromosome |
|-----------|--------|--------------|--------------|
| PDNC4     | Female | 936,641      | 0            |
| MS4221    | Male   | 493,734      | 551,277      |
| IMS13q    | Female | 1,027,947    | 23           |
occurs (Fig. 6b). An additional contribution to CENP-A–nucleosome phasing by functional CENP-B boxes is suggested at the X-chromosome HOR where the monomer sequences that are more than one full monomer away from a functional CENP-B box appeared to contain a broader distribution of midpoints (Fig. 6a). At these locations, many CENP-A–nucleosome midpoint positions fell within the coordinates of the nonfunctional CENP-B boxes. Further, the difference between the prominent peaks of CENP-A positions and the valleys between them appeared to be more pronounced on the X than on the Y chromosome. Thus, aligning CENP-A–nucleosome positions on \( \alpha \)-satellite DNA suggests that a strong CENP-B-box–independent phasing component is encoded within the \( \alpha \)-satellite monomer and that additional ‘fine-tuning’ by CENP-B-box–dependent phasing may exist.

To examine the extent to which CENP-B-box–dependent phasing of CENP-A nucleosomes occurs, we mapped the X- and Y-chromosome HOR CENP-A–nucleosome sequences to the dimer \( \alpha \)-satellite consensus (Fig. 7). Notably, the phasing on the Y-chromosome HOR was specifically diminished relative to the X-chromosome HOR. (The X-chromosome HOR (Fig. 7a–c) showed very similar phasing to that of CENP-A–nucleosome sequences from all \( \alpha \)-satellite sequences (Fig. 5a,c and Supplementary Fig. 7a,c,d,f).) The CENP-B box–independent phasing on the Y-chromosome HOR (Fig. 7d,e) remained strong enough, however, to allow clear observation of the most prominent position(s) for each small, medium or large CENP-A nucleosome–derived fragments (Fig. 7e). These positions indicate that the central dyad of the preferred CENP-A–nucleosome position on the Y-chromosome HOR (Fig. 7f) is the same as deduced from our analysis on all \( \alpha \)-satellite sequences (Fig. 5e).

![Figure 7](image_url) CENP-A nucleosomes are less phased and gain symmetric MNase digestion on the Y-chromosome centromere that lacks functional CENP-B boxes. (a–c) Maps of X-chromosome HOR-aligned CENP-A sequences. (d–f) Maps of Y-chromosome HOR-aligned CENP-A sequences. Data and models are shown in the same manner as for the global analysis of CENP-A nucleosome–associated \( \alpha \)-satellite sequences (Fig. 5).
DISCUSSION
With regard to the fundamental unit of centromere-specifying chromatin, we report nuclease digestion experiments that demonstrate a strong similarity in the behaviors of octameric CENP-A–containing nucleosomes reconstituted with recombinant components and of the form present at functional human centromeres. We conclude that the predominant form of CENP-A–containing particles at functional centromeres is an octamer with loose terminal DNA, on the basis of several key findings: (i) the smallest CENP-A–containing particle protects ~110 bp from MNase digestion, which is ~30–50 bp longer than what could be accommodated by tetrameric models, (ii) three size classes of CENP-A particles all map to the same nucleosome positions on the complex DNA of neocentromeres, and (iii) CENP-A nucleosomes at normal centromeres share the same apparent dyad-axis positioning as do their conventional counterparts containing histone H3 on the 171-bp α-satellite DNA-repeat sequence.

Our findings do not exclude the possibility that a minor population of CENP-A–containing particles with special stoichiometry exists, nor do they exclude the possibility that other forms exist at particular steps during a cell cycle–coupled program of CENP-A–nucleosome maturation and propagation18. Mutation at the CENP-A–CENP-A interface abrogates CENP-A accumulation at centromeres19,20, which suggests that a particle with two copies of CENP-A is required at least transiently in this program. Atomic force microscopy (AFM) measurements indicate that CENP-A–containing particles isolated from phases outside of S phase are shorter than conventional nucleosomes but are of similar height at S phase22. These findings were interpreted as evidence for hemisomes as the predominant form through the majority of the cell cycle22. The use of AFM-based height measurements to differentiate between hemisomes and octameric nucleosomes from isolated CENP-A–containing particles may not be as straightforward as it originally seemed because reconstituted recombinant CENP-A–containing octameric nucleosomes are substantially shorter than their canonical counterparts containing conventional histone H3 (ref. 44). Further, and to this point, in addition to the neocentromere-containing cell lines derived from healthy tissue, our studies also include the same tumor-derived cell type used in the AFM study22, HeLa (Supplementary Fig. 2f). Under our culturing conditions, ~70% of the HeLa cell population was outside of S phase (Supplementary Fig. 2f). We observed DNA fragment lengths consistent with octameric CENP-A nucleosomes in HeLa (Supplementary Fig. 2e) with no evidence of the biphasic behavior predicted by a model where there are long periods of the cell cycle where CENP-A forms drastically different particles (for example, a hemisome and octameric-nucleosome switching model22). Therefore, because suboctameric forms are not highly populated in the genome, we conclude that such minor species would be present at very low levels or only very transiently during the cell cycle.

Our findings also uncovered notable coupling of the CENP-A nucleosome’s unwrapping of its terminal DNA with its strongly phased position within the 171-bp monomer unit of centromeric α-satellite DNA. We further conclude that CENP-B binding to the CENP-B box generates asymmetric unwrapping of CENP-A–nucleosome terminal DNA. CENP-A–containing or bulk nucleosomes are not positioned evenly between the sites of CENP-B boxes within α-satellite monomers. Rather, the site for the CENP-B box is immediately 5’ adjacent to the entry-exit site. Thus, this places the 3’ end of the CENP-B box very near to the nucleosome (Fig. 5e,f). CENP-B binding induces an ~60° bend in the DNA with the strongest kink induced 4 bp from the 3’ end of the CENP-B box45. We think it is very likely that this property of CENP-B contributes strongly to several chromatin features we observe on α-satellite DNA: (i) the general phasing observed for bulk nucleosomes, (ii) the enhanced phasing seen for CENP-A nucleosomes and (iii) the asymmetric unwrapping of nucleosome-terminal DNA that is exquisitely specific to CENP-A–containing nucleosomes that are bounded by CENP-B boxes. Regarding the last feature, it appears that CENP-A has evolved in a manner such that it is poised to have its nucleosomal termini unwrapped. We speculate that the physical relationship between CENP-A, CENP-B and α-satellite DNA is a product of coevolution. Whether at established centromere locations of highly repetitive DNA or at new centromere locations lacking repeats, however, CENP-A marks centromere location as part of an octameric nucleosome with loose termini.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. The sequencing data have been deposited in the GEO database under accession number GSE44724.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS
We thank O. Jabado (Icahn School of Medicine at Mount Sinai, New York, New York, USA) for help with Illumina sequencing, T. Patel (University of Pennsylvania, Philadelphia, Pennsylvania, USA (Upenn)) and B. Cole (Upenn) for advice on data analysis, D. Rogers (Upenn) and B. Gregory (Upenn) for advice, E. Bernstein (Icahn School of Medicine at Mount Sinai, New York, New York, USA) for mentoring and advising D.H., M. Lampson (Upenn) for comments on the manuscript, K. Luger (Colorado State University, Fort Collins, Colorado, USA), D. Cleveland (University of California, San Diego, La Jolla, California, USA), A. Straight (Stanford University, Stanford, California, USA) and D. Rhodes (Medical Research Council Laboratory of Molecular Biology, Cambridge, UK) for plasmids. A. Choo (Murdoch Children’s Research Institute, Victoria, Australia) for the cell line containing the PD-NC4 chromosome and R. Allshire (University of Edinburgh, Edinburgh, Scotland, UK) for sharing results before publication. This work was supported by US National Institutes of Health research grant GM082989 (B.E.B.), a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund (B.E.B.), a Rita Allen Foundation Scholar Award (B.E.B.), a predoctoral fellowship from the American Heart Association (K.J.S.) and a postdoctoral fellowship from the American Cancer Society (N.S.). T.P. acknowledges support from US National Institutes of Health grant GM08275 (Upenn Structural Biology Training Grant).

AUTHOR CONTRIBUTIONS
D.H. designed and performed experiments and analyzed data. K.J.S. and T.P. designed and performed experiments, developed new analytical tools, analyzed data and wrote the manuscript. M.U.S. developed new analytical tools. N.S. analyzed data and modeled nucleosomes. A.A. performed experiments and provided technical advice on ChIP experiments. P.E.W. directed the project, designed experiments and analyzed data. B.E.B. directed the project, designed experiments, analyzed data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Cleveland, D.W., Mao, Y. & Sullivan, K.F. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell 112, 407–421 (2003).
2. Mendiburu, M.J., Padeken, J., Fülpö, S., Schepers, A. & Heun, P. Drosophila CENH3 is sufficient for centromere formation. Science 334, 686–690 (2011).
3. Banhart, M.C. et al. HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J. Cell Biol. 194, 229–243 (2011).
4. 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).
5. Warburton, P.E. Chromosomal dynamics of human neocentromere formation. Chromosome Res. 12, 617–626 (2004).
6. Amor, D.J. et al. Human centromere repositioning ‘in progress’. Proc. Natl. Acad. Sci. USA 101, 6542–6547 (2004).
7. Bassett, E.A. et al. Epigenetic centromere specification directs aurora B accumulation but is insufficient to efficiently correct mitotic errors. J. Cell Biol. 190, 177–185 (2010).
8. 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).
9. Alonso, A. et al. Genomic microarray analysis reveals distinct locations for the CENP-A-binding domains in three human chromosome 13q32 neocentromeres. Hum. Mol. Genet. 12, 2711–2721 (2003).
10. Hassan, D. et al. Formation of novel CENP-A domains on tandem repetitive DNA and across chromosome breakpoints on human chromosome 8q21 neocentromeres. Chromosoma 120, 621–632 (2011).
11. Dechassa, M.L. et al. Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. Nat. Commun. 2, 313 (2011).
12. Conde e Silva, N. et al. Crystal structure of the human centromeric nucleosome containing CENP-A. Nature 476, 232–235 (2011).
13. Williams, J.S., Hayashi, T., Yanagida, M. & Russell, P. Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. Mol. Cell 33, 287–298 (2009).
14. Carrell, C.W., Silva, M.C.C., Godek, K.M., Jansen, L.E.T. & Straight, A.F. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. Nat. Cell Biol. 11, 896–902 (2009).
15. 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).
16. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389, 251–260 (1997).
17. Shaw, B.R., Herman, T.M., Kovacic, R.T., Beaudeau, G.S. & Van Holde, K.E. Analysis of subunit organization in chicken erythrocyte chromatin. Proc. Natl. Acad. Sci. USA 73, 505–509 (1976).
18. Axel, R. Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease. Biochemistry 14, 2921–2925 (1975).
19. Preuss, K.J. & Cleveland, D.W. Epigenetic centromere propagation and the nature of centromere-specific nucleosome from budding yeast. J. Biol. Chem. 286, 4021–4026 (2011).
20. Panchenko, T. et al. Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical turn.
21. Amor, D.J. et al. Human centromere repositioning ‘in progress’. Proc. Natl. Acad. Sci. USA 101, 6542–6547 (2004).
22. Bui, M. et al. Crystal structure of the CENP-A/H42 reveals physical features that mark centromeres. Nature 467, 347–351 (2010).
23. Tachiwana, H. et al. Structural determinants for generating centromeric chromatin. Nature 430, 578–582 (2004).
24. Williams, J.S., Hayashi, T., Yanagida, M. & Russell, P. Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. Mol. Cell 33, 287–298 (2009).
25. Dong, F. & Van Holde, K.E. Nucleosome positioning is determined by the (H3–H4) tetramer. Proc. Natl. Acad. Sci. USA 88, 10596–10600 (1991).
26. Read, C.M., Baldwin, J.P. & Crane-Robinson, C. Structure of subnucleosomal particles. tetramer (H3/H4)2,146 base pair DNA and hexameric (H3/H4)3(H2A/H2B):146 base pair DNA complexes. Biochemistry 24, 4435–4450 (1985).
27. Park, P.J. ChIP-seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. 10, 669–680 (2009).
28. Zhang, Z. & Pugh, B.F. High-resolution genome-wide mapping of the primary structure of chromatin. Cell 144, 175–186 (2011).
29. Willard, H.F. & Wayne, J.S. Hierarchical order in chromosome-specific human alpha satellite DNA. Trends Genet. 3, 192–198 (1987).
30. Vafa, O. & Sullivan, K.F. Chromatin containing CENP-A and alpha satellite DNA is a major component of the inner kinetochore plate. Curr. Biol. 7, 897–900 (1997).
31. Tanaka, Y. et al. Molecular cloning of cDNA for CENP-B, the major human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J. Biol. Chem. 109, 1963–1973 (1989).
32. Rudd, M.K. & Willard, H.F. Analysis of the centromeric regions of the human genome assembly. Trends Genet. 20, 529–534 (2004).
33. Wang, J.S. & Willard, H.F. Chromosome-specific alpha satellite DNA: nucleotide sequence analysis of the 2.0 kilobasepair repeat from the human X chromosome. Nucleic Acids Res. 13, 2731–2743 (1985).
34. Warburton, P.E. Epigenetic analysis of kinetochore assembly on variant human centromeres. Trends Genet. 17, 243–247 (2001).
35. Earnshaw, W.C. et al. Molecular cloning of CDNA for CENP-B, the major human centromere autoantigen. J. Cell Biol. 104, 817–829 (1987).
36. Zhang, W., Colmenares, S.U. & Karpen, G.H. Assembly of CENP-A/histone H4 for centromere assembly. J. Cell Biol. 181, 304–316 (2012).
37. Miell, M.A. et al. Crystal structure of the CENP-B protein-DNA complex: the DNA-
38. Rudd, M.K. & Willard, H.F. Analysis of the centromeric regions of the human genome assembly. Trends Genet. 20, 529–534 (2004).
39. Wang, J.S. & Willard, H.F. Chromosome-specific alpha satellite DNA: nucleotide sequence analysis of the 2.0 kilobasepair repeat from the human X chromosome. Nucleic Acids Res. 13, 2731–2743 (1985).
40. Warburton, P.E. Epigenetic analysis of kinetochore assembly on variant human centromeres. Trends Genet. 17, 243–247 (2001).
41. Earnshaw, W.C. et al. Molecular cloning of CDNA for CENP-B, the major human centromere autoantigen. J. Cell Biol. 104, 817–829 (1987).
42. Zhang, W., Colmenares, S.U. & Karpen, G.H. Assembly of CENP-A/histone H4 for centromere assembly. J. Cell Biol. 181, 304–316 (2012).
43. Miell, M.A. et al. Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. EMBO J. 20, 6612–6618 (2001).
ONLINE METHODS

Nucleosome reconstitution experiments. Tetrasomes, nucleosomes and nucleosomal arrays were reconstituted from purified components by using salt dialysis45. Briefly, human histones H3, H4, H2A and H2B were purified by monomer44 and mixed to form (H3–H4)2 tetramer and (H2A–H2B) dimer complexes44,47, whereas human (CENP-A–H4)2 was purified from a bicistronic vector as a tetramer17. The ‘601’ 1× 200-bp and ‘601’ 12× 200-bp DNA templates48,49 were both purified by anion-exchange chromatography. The indicated histone complexes were combined with the DNA in 2 M NaCl and dialyzed in steps: (i) TE (10 mM Tris, pH 7.8, 0.25 mM EDTA) supplemented with 1 M NaCl, followed by (ii) TE supplemented with 0.75 M NaCl and finally (iii) TE supplemented with 2.5 mM NaCl. Tetrasomes, nucleosomes or nucleosomal arrays were digested with 2 U/µg MNase (Roche) in the presence of 3 mM CaCl2 for 0.5 to 2 min. Each comparison shown between CENP-A– and histone H3–containing Chromatin immunoprecipitation. Agilent 2100 Bioanalyzer by using the DNA 1000 kit.

Sequencing libraries were generated and barcoded for multiplexing according to Illumina recommendations with a Qiagen MinElute column. For cross-linked ChIP, 2–5 × 107 cells were collected and resuspended in 2 ml of ice-cold buffer 1 (0.1 M NaCl, 50 mM MgCl2, 0.1 mM EGTA, 15 mM Tris, pH 7.5, 0.5 mM DTT, 0.1 mM PMSF and 1:1,000 protease inhibitor cocktail (Sigma)). Ice-cold buffer 1 supplemented with 0.1% IGEPAL (20 µl) was added, and samples were placed on ice for 10 min. The resulting 4 ml of nuclei was gently layered on top of 8 ml of ice-cold buffer III (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EGTA, 15 mM Tris, pH 7.5, 0.5 mM DTT, 0.1 mM PMSF and 1:1,000 protease inhibitor cocktail (Sigma)). Ice-cold buffer 1 supplemented with 0.1% IGEPAL (20 µl) was added, and samples were placed on ice for 10 min. The resulting 4 ml of nuclei was gently layered on top of 8 ml of ice-cold buffer III (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EGTA, 15 mM Tris, pH 7.5, 0.5 mM DTT, 0.1 mM PMSF and 1:1,000 protease inhibitor cocktail (Sigma)). The reaction was quenched with 5 mM EGTA on ice and centrifuged at 13,500 g for 10 min. The chromatin was resuspended in 10 mM EDTA, pH 8.0, 1 mM PMSF and 1:1,000 protease inhibitor cocktail (Sigma) to 400 ng/µl. MNase (Affymetrix) digestion reactions were carried out on chromatin (100 µg or more), using 0.9–2.8 U/µg chromatin in buffer A supplemented with 3 mM CaCl2 for 10 min at 37 °C. The reaction was quenched with 5 mM EGTA on ice and centrifuged at 13,500 g for 10 min. The chromatin was resuspended in 10 mM EDTA, pH 8.0, 1 mM PMSF and 1:1,000 protease inhibitor cocktail and rotated at 4 °C for 2–4 h. The mixture was adjusted to 500 mM NaCl, allowed to rotate for another 45 min and then centrifuged at 13,500 g for 10 min, yielding nucleosomes in the supernatant. Chromatin (100 µg or more) was diluted to 100 ng/µl with buffer B (20 mM Tris, pH 8.0, 5 mM EDTA, 500 mM NaCl and 0.2% Tween 20) and precleared with 60 µl 50% protein G–bead (GE Healthcare) slurry for 20 min at 4 °C. Precleared supernant (1–2 µg bulk nucleosomes) was saved for further processing. To the remaining supernatant, antibody was added and rotated overnight at 4 °C. Immuno-complexes were recovered by addition of 100 µl 50% protein G–bead slurry followed by rotation at 4 °C for 3 h. The beads were washed three times with buffer B and once with buffer B without Tween. For the input fraction, an equal volume of input recovery buffer (0.6 M NaCl, 20 mM EDTA, 20 mM Tris, pH 7.5, and 1% SDS) and 1 µl of RNase A (10 mg/ml) was added, followed by incubation for one hour at 37 °C. Proteinase K (100 µg/ml, Roche) was then added, and samples were incubated for another 3 h at 37 °C. For the ChIP fraction, 300 µl of ChIP recovery buffer (20 mM Tris, pH 7.5, 20 mM EDTA, 0.5% SDS and 500 µg/ml Proteinase K) was added and was incubated for 3–4 h at 56 °C. The resulting Proteinase K–treated samples were subjected to a phenol–chloroform extraction followed by purification with a Qiagen MinElute column. For cross-linked ChIP, 2–5 × 106 cells were processed with the SimpleChIP Enzymatic Chromatin IP Kit (Cell Signaling) by using the manufacturer’s recommendations. Unamplified bulk nucleosomes or ChIP DNA was analyzed by using an Agilent 2100 Bioanalyzer High Sensitivity Kit. The Bio-analyzer determines the quantity of DNA on the basis of fluorescence intensity. Antibodies used for ChIP were mouse α-CENP-A monoclonal (15 µg, ab13939, Abcam), rabbit α-H3K9me3 polyclonal (10 µg, ab8898, Abcam) and rabbit α-H3.3 polyclonal (17 µg, 09-838, Millipore).

Next-generation sequencing. Sequencing libraries were generated and barcoded for multiplexing according to Illumina recommendations with minor modifications. Briefly, 2–15 ng input or ChIP DNA was end-repaired and A-tailed. Illumina Truseq adaptors were ligated, libraries were size-selected to exclude nucleopolysemes, and the libraries were PCR-amplified by using Phusion polymerase. All steps in library preparation were carried out using New England Biolabs enzymes. Resulting libraries were submitted for 100-bp, paired-end Illumina sequencing on a HiSeq 2000 instrument.

ChIP-seq data processing. Paired-end ChIP-seq reads were aligned to the human genome build hg19 with Bowtie2 version 2.0.0 using paired-end mode. Reads were aligned by using a seed length of 50 bp, and only the single best alignment per read with up to two mismatches was reported in the SAM file. The aligned mate pairs were joined in MATLAB by using the ‘localalign’ function (to determine the overlapping region between the reads (requiring 295% overlap identity; Supplementary Fig. 3a). Duplicate-read removal was carried out by using the ‘rmdup’ command in SAMTools. To create nucleosome occupancy maps at neocentromeres, all joined reads were aligned to the neocentromere, and the number of reads that aligned with 100% identity were plotted for each particular base pair along the neocentromere coordinate (Supplementary Fig. 3b). For analysis of α-satellite DNA, all joined reads were aligned to the dimerized α-satellite consensus sequence, and those reads aligning with ≥ 60% identity were chosen for further analysis (Supplementary Fig. 6a,b).

Annotated α-satellite analysis. Paired-end ChIP-seq reads were aligned to the X-chromosome or Y-chromosome HOR with Bowtie2 version 2.0.0 by using paired-end mode. Reads were aligned by using a seed length of 50 bp, and only the single best alignment per read with zero mismatches was reported in the SAM file. The 2.0-kb X-chromosome HOR was previously described elsewhere39. The 5.8-kb Y-chromosome HOR was determined by performing dot plot analysis on the annotated portion of the centromere on chromosome Y in the human genome build hg19.

Statistical correlation analysis. Pairwise Pearson correlation coefficients between nucleosome occupancy maps of various size classes (and between randomly generated data sets) at the neocentromeres were determined by using MATLAB. P values were determined by using Student’s t test by transforming the correlations to a t statistic having n − 2 degrees of freedom.

Molecular modeling. Molecular models were generated by using PDB 1KX5 and 1ZBB for the H3-containing particles and 3AN2 for CENP-A–containing particles. Models of tetrasomes and hemisomes with crossed DNA were generated by using linker DNA from 1ZBB and minimized by using CNS50,51. The model of the CENP-A–nucleosome core particle was generated by using DNA from 1X55. The point in space of DNA crossing was determined as the sensus sequence, and those reads aligning with ≥ 2 degrees of freedom.

Fluorescence in situ hybridization and quantitative real-time PCR. ChIP DNA FISH probes were generated and used for metaphase FISH as previously described19. Quantitative real-time PCR (qPCR) was performed as previously described8.