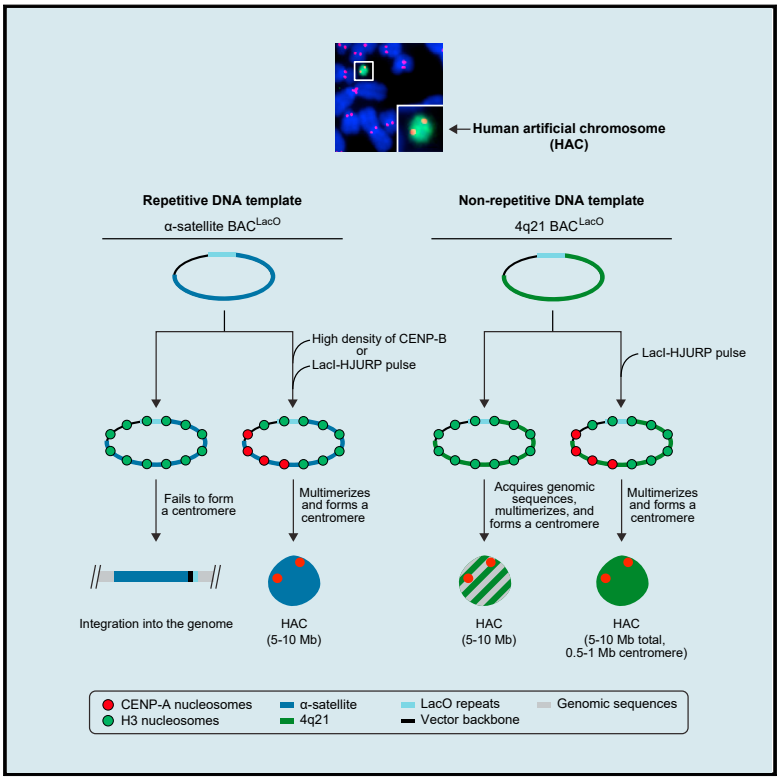
Article

Cell

Human Artificial Chromosomes that Bypass Centromeric DNA

Graphical Abstract



Highlights

- Development of human artificial chromosomes (HACs) where CENP-A chromatin is seeded
- Seeding CENP-A nucleosome assembly induces centromere formation
- Seeding centromeric chromatin bypasses sequence elements in repetitive centromere DNA
- Non-repetitive HAC templates ease initial construction and downstream genomic analyses

Authors

Glennis A. Logsdon, Craig W. Gambogi, Mikhail A. Liskovykh, ..., Karen H. Miga, Patrick Heun, Ben E. Black

Correspondence

blackbe@pennmedicine.upenn.edu

In Brief

Development of human artificial chromosomes that bypass centromeric DNA removes a key barrier limiting mammalian synthetic genome efforts.

Logsdon et al., 2019, Cell *178*, 624–639
July 25, 2019
https://doi.org/10.1016/j.cell.2019.06.006

CellPress

Cell

Article

Human Artificial Chromosomes that Bypass Centromeric DNA

Glennis A. Logsdon,[1,5] Craig W. Gambogi,[1] Mikhail A. Liskovykh,[2] Evelyne J. Barrey,[3] Vladimir Larionov,[2] Karen H. Miga,[4] Patrick Heun,[3] and Ben E. Black[1,6,*]

[1]Department of Biochemistry and Biophysics, Graduate Program in Biochemistry and Molecular Biophysics, and Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[2]Developmental Therapeutics Branch, National Cancer Institute, Bethesda, MD 20892, USA

[3]Wellcome Trust Centre for Cell Biology, Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK

[4]Center for Biomolecular Science and Engineering, University of California, Santa Cruz, Santa Cruz, CA 95064, USA

[5]Present address: Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA

[6]Lead Contact

*Correspondence: blackbe@pennmedicine.upenn.edu

https://doi.org/10.1016/j.cell.2019.06.006

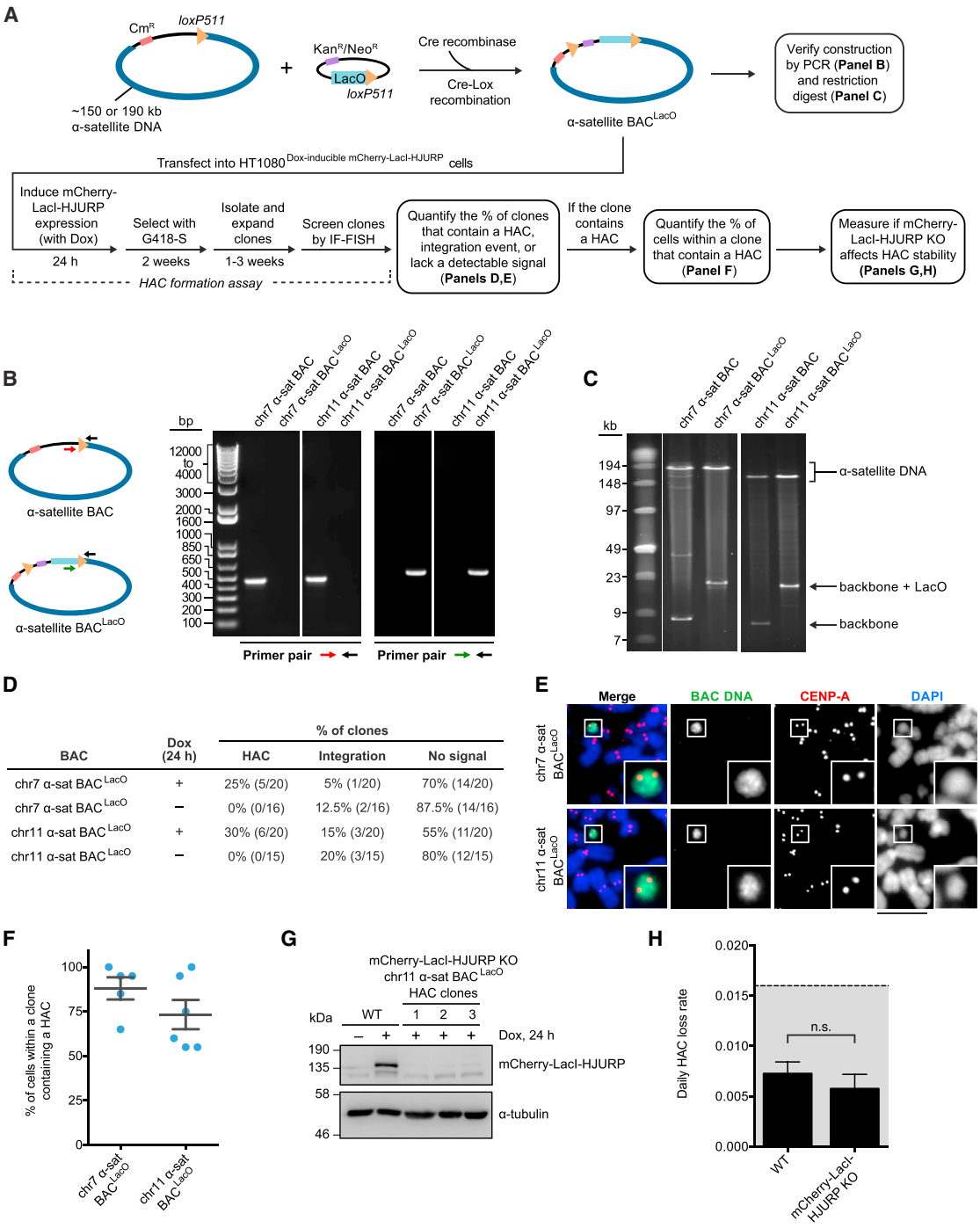
SUMMARY

Recent breakthroughs with synthetic budding yeast chromosomes expedite the creation of synthetic mammalian chromosomes and genomes. Mammals, unlike budding yeast, depend on the histone H3 variant, CENP-A, to epigenetically specify the location of the centromere—the locus essential for chromosome segregation. Prior human artificial chromosomes (HACs) required large arrays of centromeric α-satellite repeats harboring binding sites for the DNA sequence-specific binding protein, CENP-B. We report the development of a type of HAC that functions independently of these constraints. Formed by an initial CENP-A nucleosome seeding strategy, a construct lacking repetitive centromeric DNA formed several self-sufficient HACs that showed no uptake of genomic DNA. In contrast to traditional α-satellite HAC formation, the non-repetitive construct can form functional HACs without CENP-B or initial CENP-A nucleosome seeding, revealing distinct paths to centromere formation for different DNA sequence types. Our developments streamline the construction and characterization of HACs to facilitate mammalian synthetic genome efforts.

INTRODUCTION

Artificial chromosomes, either those built from isolated (Schueler et al., 2001) or synthetic (Basu et al., 2005; Ohzeki et al., 2002; Richardson et al., 2017) sequences, have the potential to transform synthetic biology and permit the development of numerous radical advancements in medicine (Boeke et al., 2016). The early stages of an ambitious project to generate an entire set of synthetic human chromosomes, termed the Human Genome Project-Write (Boeke et al., 2016), is building on recent success with synthetic budding yeast chromosomes (Annaluru et al., 2014; Richardson et al., 2017). Among many potential hurdles to translate success from yeast to mammals, the centromere likely represents the biggest challenge. Centromeres are the loci present once per natural chromosome that guide their segregation at cell division (McKinley and Cheeseman, 2016). While in budding yeast these loci are small (~125 bp) genetic elements, most other eukaryotes, including mammals, have an essential epigenetic contribution to their specification. This has provided an explanation to the originally paradoxical observation that the DNA typically found at human centromeres (α-satellite) is neither necessary nor sufficient for centromere identity and function (Eichler, 1999). For instance, centromere sequences can be silent (e.g., on one of the two megabase-sized regions of α-satellite on a so-called pseudodicentric chromosome) (Earnshaw and Migeon, 1985; Warburton et al., 1997) or completely bypassed when a new centromere (e.g., a neocentromere) (Depinet et al., 1997; Hasson et al., 2011; du Sart et al., 1997; Warburton et al., 1997) is formed. Instead, nucleosomes in which the histone variant, CENP-A, replaces canonical H3 epigenetically specify centromere location (Black and Cleveland, 2011; McKinley and Cheeseman, 2016).

Human artificial chromosomes (HACs) were first generated more than 20 years ago (Harrington et al., 1997), and through subsequent innovations, it became clear that the establishment of centromeric chromatin with CENP-A nucleosomes is what defines a functional HAC (Ebersole et al., 2000; Grimes et al., 2002; Ikeno et al., 1998; Mejía et al., 2002; Ohzeki et al., 2012; Okada et al., 2007). After the rare instance when a functional centromere is established, it is then faithfully propagated through the well-established epigenetic pathway that includes the dedicated centromere chromatin assembly protein, HJURP (Dunleavy et al., 2009; Foltz et al., 2009) (for a review, see McKinley and Cheeseman, 2016). A non-essential centromere protein, CENP-B—the only known sequence-specific DNA binding protein at mammalian centromeres, recognizing the 17-mer "CENP-B box" recognition element—plays an essential role in HAC formation (Ohzeki et al., 2002; Okada et al., 2007). This is presumably through its interactions with the CENP-A nucleosome and the key centromere protein, CENP-C (Fachinetti et al., 2013, 2015). Indeed, a classic study using the repetitive centromere DNA from the



BAC	Dox (24 h)	% of clones		
		HAC	Integration	No signal
chr7 α-sat BAC^{LacO}	+	25% (5/20)	5% (1/20)	70% (14/20)
chr7 α-sat BAC^{LacO}	–	0% (0/16)	12.5% (2/16)	87.5% (14/16)
chr11 α-sat BAC^{LacO}	+	30% (6/20)	15% (3/20)	55% (11/20)
chr11 α-sat BAC^{LacO}	–	0% (0/15)	20% (3/15)	80% (12/15)

Figure 1. HAC Formation Is Stimulated by Seeding a Round of CENP-A Nucleosome Assembly with HJURP

(A) Steps in building and testing HACs.

(B) PCR analysis of α-satellite BAC^{LacO} constructs.

(C) Restriction digest of BAC^{LacO} constructs to liberate individual parts.

(D) Results of the HAC formation assays.

(E) Representative images of chr7 and chr11 α-satellite BAC^{LacO} HACs. Insets are 2.5× magnifications. Bar, 10 μm.

(F) Quantification of the percentage of cells containing an α-satellite BAC^{LacO} HAC within each HAC-positive clone. The mean value (± SEM) is shown for each BAC^{LacO} construct.

(G) Immunoblots of the indicated cell lines.

(legend continued on next page)

X chromosome found that only regions of α-satellite with a high density of functional CENP-B boxes generated functional HACs (Schueler et al., 2001). These and other findings have led to two assumed universal rules for HAC formation: (1) a requirement for the specific forms of α-satellite with a high density of CENP-B boxes (Ohzeki et al., 2002; Schueler et al., 2001), and (2) the expression of CENP-B (Okada et al., 2007).

Bypassing these two rules (e.g., by forming a HAC on non-repetitive DNA constructs) would have several clear benefits. First, HAC construction would be greatly facilitated. Traditional HACs contain 50–200 kb of highly repetitive DNA (Ebersole et al., 2000; Grimes et al., 2002; Ikeno et al., 1998; Mejía et al., 2002; Ohzeki et al., 2012; Okada et al., 2007), which greatly complicates handling at all steps, from their initial construction to their clonal stability during bacterial propagation. Second, mapping the chromatin features of HACs using sequencing-based approaches would become possible. For instance, it is imperative to know where functional centromeres are located relative to other functional genetic elements that the HACs are engineered to carry. The highly repetitive sequences on traditional HACs unfortunately prohibit any useful genomic methodologies to define their composition and organization. Third, non-repetitive sequences would allow mammalian synthetic chromosomes to be generated by employing some of the fundamental principles used in recent yeast synthetic chromosome construction, where DNA repeats were removed to make their designed sequences compatible with recombination-based assembly (Richardson et al., 2017).

In considering a new generation of HAC design, alternative systems have emerged to form new centromeres through the artificial seeding of nascent CENP-A nucleosomes (Barnhart et al., 2011; Chen et al., 2014; Hori et al., 2013; Logsdon et al., 2015; Mendiburo et al., 2011; Ohzeki et al., 2012; Tachiwana et al., 2015) (reviewed in Barrey and Heun, 2017). One of these approaches, in fruit fly cells, built upon the earlier notion of epigenetic centromeric chromatin spreading (Maggert and Karpen, 2001). Initial CENP-A nucleosome assembly targeted locally at an array of Lac operator (LacO) sites eventually led to spreading of the centromere via natural centromeric chromatin assembly to the remainder of a small plasmid that did not contain any natural centromeric sequences (Mendiburo et al., 2011). While this plasmid does not align on the metaphase plate at cell division and does not yield very high stability through cell divisions (compared to HACs, for instance), it formed a functional mitotic kinetochore—the proteinaceous complex that forms at a mitotic centromere—and directed interactions with the microtubule-based spindle (Mendiburo et al., 2011). Taken together, these studies open the possibility that the requirements of α-satellite DNA and CENP-B for HAC formation, mentioned above, could be circumvented.

Here, we improve HAC technology with a collection of HACs that include repetitive centromeric sequences or non-repetitive genomic sequences, testing each type for their dependence on seeding CENP-A nucleosome assembly. We employ gene editing of centromere components to elucidate the molecular requirements for the establishment and propagation of different types of HAC DNA templates, and we utilize genomic approaches to gain a highly resolved understanding of HAC copy number as well as genetic and epigenetic composition.

RESULTS

Seeding HACs with CENP-A Nucleosomes

We first generated BAC constructs containing α-satellite sequences that are deemed nonfunctional in natural chromosomes due to a low density of CENP-B boxes (Hayden et al., 2013). A successful strategy to make these sequences functional to form a HAC is to first manipulate the constructs to increase the density of CENP-B boxes (Hayden et al., 2013). We devised an alternative strategy to avoid manipulation of the α-satellite sequences, themselves, by artificially driving an initial round of CENP-A chromatin assembly on an adjacent site on the construct. Our general strategy was to assemble constructs consisting of BACs harboring an array of LacO repeats immediately adjacent to human genomic DNA sequences (hereafter termed BAC^{LacO}) (Figure 1A). Then, to the LacO array, we targeted mCherry-LacI-HJURP, inducibly expressed from a genomically integrated transgene. This targeting would potentially initiate the assembly of CENP-A nucleosomes directly onto the BAC and facilitate the spreading of CENP-A nucleosomes to the neighboring sequences (Mendiburo et al., 2011). We engineered two BAC^{LacO} vectors containing α-satellite DNA coming from CENP-A-poor regions of the centromere on chromosomes (chr) 7 and 11 (Figure 1A). Our cloning strategy positioned the LacO repeats within 300 bp of the α-satellite sequence, keeping this distance small to potentially permit efficient spreading of centromeric chromatin. We isolated α-satellite BAC^{LacO} constructs that successfully recombined (Figure 1B) and retained both the repetitive α-satellite and LacO arrays (Figure 1C). Using established methodologies to isolate and identify HACs (Grimes et al., 2002; Okamoto et al., 2007), we found that a pulse of mCherry-LacI-HJURP expression was sufficient to stimulate HAC formation (Figures 1D and 1E). Because we obtained nearly identical results on two independent α-satellite sequences (Figures 1D–1F), we conclude that our strategy would stimulate HAC formation on broad classes of α-satellite higher order repeats. As expected, there was no HAC formation in the absence of the round of CENP-A chromatin assembly directed by the pulse of mCherry-LacI-HJURP (Figure 1D), indicating that the presence of the LacO array, itself, does not drive centromere formation on BAC^{LacO} constructs.

The HACs we formed were highly penetrant within a clonal cell population (Figure 1F), likely due to the pulse of mCherry-LacI-HJURP driving efficient and rapid centromere acquisition that can then be propagated independently of the initial HJURP-mediated seeding of CENP-A nucleosome assembly.

(H) Quantification of the daily HAC loss rate in WT or mCherry-LacI-HJURP KO cells after culturing without G418-S for 60 days. The mean daily loss rate (± SEM) is shown. n = 60 WT cells and 180 mCherry-LacI-HJURP KO cells, pooled from three independent experiments for each indicated cell type. The WT cells are from a chr11 α-satellite BAC^{LacO} clone, and the mCherry-LacI-HJURP KO cells are pooled from three derivative chr11 α-satellite BAC^{LacO} clones. n.s., not significant.

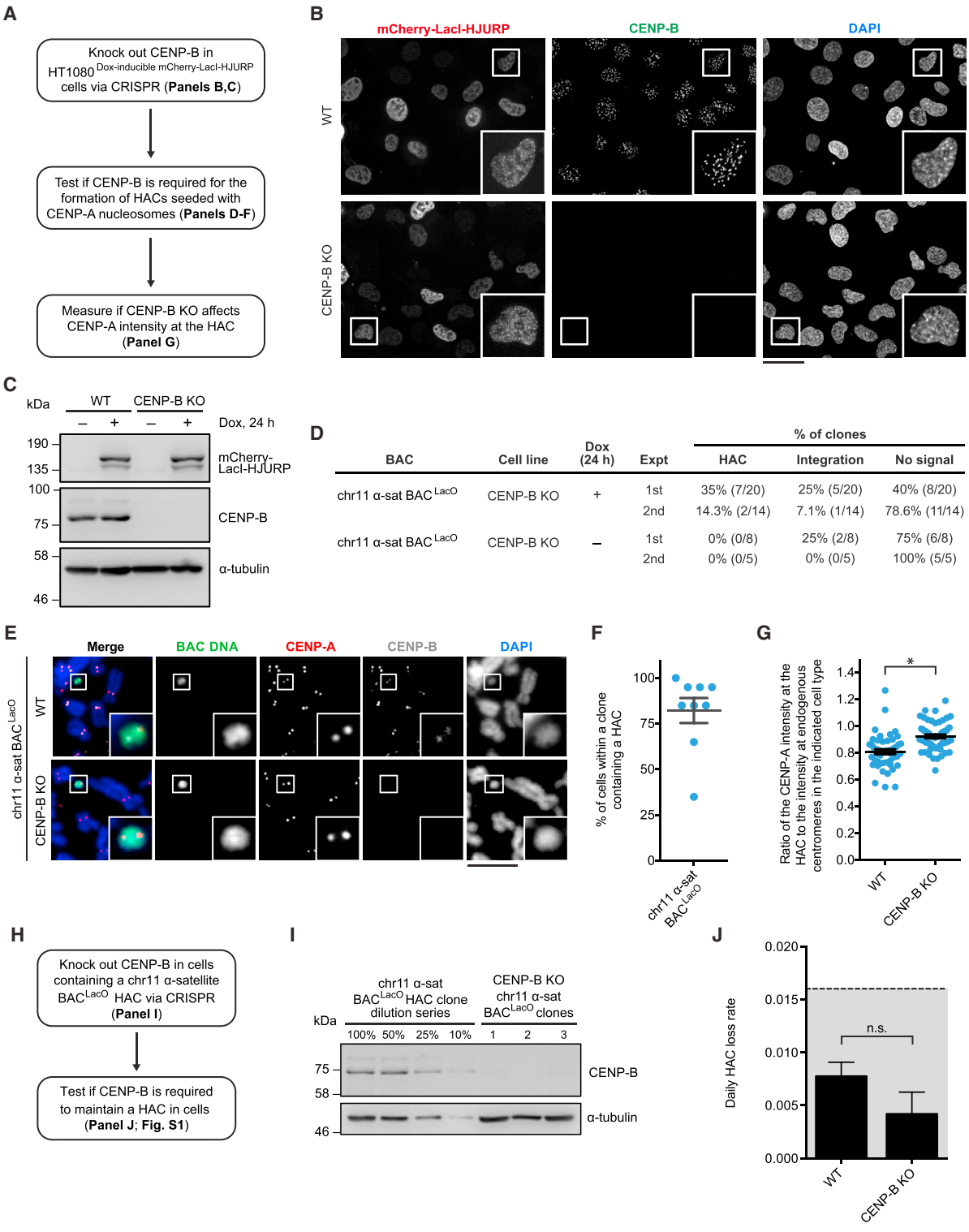


Figure 2. CENP-B Is Not Required for Formation or Maintenance of HACs Seeded with CENP-A Nucleosomes

(A) Steps to test whether or not CENP-B participates in the formation of HACs seeded with CENP-A nucleosomes.

(B) Representative images of the indicated cell lines following 24 h of dox treatment. Insets: 2.5× magnifications. Bar, 40 μm.

(C) Immunoblots of the indicated cell lines.

(D) Results of the HAC formation assays.

(E) Representative images of HACs formed in WT and CENP-B KO cells. Insets: 2.5× magnifications. Bar, 10 μm.

(legend continued on next page)

Alternatively, we considered that low, leaky expression of mCherry-LacI-HJURP continues to drive CENP-A nucleosome assembly on the HAC, thereby stabilizing the HAC in the cell. Therefore, we tested if genetically ablating mCherry-LacI-HJURP expression via CRISPR/Cas9-mediated gene editing affects HAC stability (Figure 1A). Choosing a cell line in which the chr11 α-satellite BAC^{LacO} HAC is present in ≥95% of cells, we derived three monoclonal cell lines in which mCherry-LacI-HJURP expression has been disrupted (Figure 1G). Using the standard approach for measuring HAC maintenance (Nakano et al., 2008; Ohzeki et al., 2012; Schueler et al., 2001), wherein all clones were cultured without antibiotic selection (G418-S) for 60 days, we found that the absence of mCherry-LacI-HJURP did not affect the daily HAC loss rate (Figure 1H). These daily HAC loss rates are similarly low as those reported for "conventional" HACs (Figure 1H, the range is shaded in gray (Ebersole et al., 2000; Ikeno et al., 1998)). Thus, the action of seeding CENP-A nucleosome assembly is limited to centromere establishment. After that, the centromere on the HAC is epigenetically maintained in the same manner as on natural chromosomes.

CENP-B-Independent HAC Formation and Maintenance

We next directly tested whether CENP-B expression—one of the universal requirements for conventional HAC formation (Okada et al., 2007)—could be bypassed by seeding CENP-A nucleosome assembly. To do so, we disrupted the CENP-B gene prior to performing a new set of HAC formation assays (Figures 2A–2C). We found that chr11 α-satellite BAC^{LacO} HACs form in the absence of CENP-B (Figure 2D,E). Because HAC formation on this construct is dependent on induction of mCherry-LacI-HJURP (Figure 2D,E), we conclude that seeding CENP-A nucleosomes onto the α-satellite DNA bypasses the requirement of CENP-B for centromere formation. Further, the absence of CENP-B did not affect the high number of cells containing a HAC (Figure 2F) (82% ± 7% in CENP-B knockout [KO] cells versus 73% ± 8% in wild type [WT] cells, shown in Figure 1F) or substantially alter the amount of CENP-A on the centromere of the HAC relative to those on natural chromosomes (Figure 2G) (note there is a small but measurable increase in the CENP-B KO cells). Thus, our experiments indicate that the absence of CENP-B has no detectable negative effect upon forming a HAC via seeding of CENP-A nucleosomes.

Because prior efforts with conventional HACs failed to form any functional centromeres in the absence of CENP-B (Okada et al., 2007), there are no data to indicate whether or not CENP-B is also important for HAC maintenance. To address this issue, we performed a HAC maintenance assay with a cell line containing a chr11 α-satellite BAC^{LacO} HAC and three monoclonal cell line derivatives of it in which we disrupted the CENP-B gene (Figures 2H and 2I). We found that the absence of CENP-B did not affect the daily HAC loss rate of the α-satellite HACs (Figure 2J). Further, CENP-A was retained at the HAC in the absence of CENP-B through our 60-day assay (Figure S1). Thus, we conclude that CENP-B is also dispensable for the maintenance of a HAC.

HACs that Lack α-Satellite DNA

The most prominent proposal for the role of α-satellite DNA in HAC formation is that a high density of CENP-B boxes facilitates early steps in centromere formation (Fujita et al., 2015; Ohzeki et al., 2002; Okada et al., 2007; Schueler et al., 2001). Because seeding CENP-A nucleosome assembly bypasses the requirement of CENP-B for centromere formation (Figure 2), we hypothesized that, likewise, the requirement for α-satellite DNA might be bypassed. To test this, we built and performed a small-scale HAC formation screen with a set of BACs containing an array of LacO repeats adjacent to non-α-satellite human genomic sequences (Figures S2A and S2B). We chose sequences for our initial screening based on proximity to known neocentromeres (Amor et al., 2004; Hasson et al., 2011, 2013), and we also included a clone several Mbp distal to a well-studied neocentromere (PD-NC4) (Table S1). One construct in the screen, 4q21 BAC^{LacO}, formed several HACs (Figures 3A–3C and S2C–S2E). In stark contrast to the α-satellite versions that we tested (Figures 1 and 2), we found that 4q21 BAC^{LacO} also reproducibly formed HACs in the absence of the induction of mCherry-LacI-HJURP expression (Figures 3A–3C). We considered that non-α-satellite sequences might be particularly sensitive to leaky expression of mCherry-LacI-HJURP in the absence of doxycycline. Thus, we generated a version of the 4q21 BAC that is identical to 4q21 BAC^{LacO} but lacks the LacO array (Figure S2F) and found that it also forms HACs (Figures 3A, 3D, and 3E). This eliminated the possibility of a dependence on any leaky mCherry-LacI-HJURP expression or on any other property imparted by the LacO array itself. Because the only sequences to-date to form a HAC in the absence of seeding CENP-A nucleosomes require CENP-B (Okada et al., 2007), we also considered the possibility that 4q21 BAC^{LacO} HACs somehow form via a CENP-B-dependent centromere formation pathway. To directly test this, we performed HAC formation assays with 4q21 BAC^{LacO} in our cell line where the CENP-B gene had been disrupted (Figures 2B and 2C) and found that HAC formation occurred in the absence of CENP-B (Figures 3A, 3F, and 3G). Thus, we conclude that the non-repetitive, non-centromeric 4q21 BAC^{LacO} construct forms a HAC in a CENP-B-independent manner.

(F) Quantification of the percentage of CENP-B KO cells containing a HAC within each clone. The mean value (± SEM) is shown.
(G) Quantification of CENP-A intensity at HACs formed in WT and CENP-B KO cells relative to the intensity at centromeres on endogenous chromosomes. The mean ratio (± SEM) is shown. n = 50 WT cells and 57 CENP-B KO cells, pooled from 3 independent clones for each indicated cell type. An asterisk indicates $p < 0.05$.
(H) Steps to test if CENP-B is important for maintenance of HACs that formed upon seeding of CENP-A nucleosomes.
(I) Immunoblots of the indicated cell lines.
(J) Quantification of the daily HAC loss rate in WT or mCherry-Lac-HJURP KO cells after culturing without G418-S for 60 days (shading as in Figure 1H). The mean daily loss rate (± SEM) is shown. n = 60 WT cells and 180 CENP-B KO cells, pooled from 3 independent experiments for each indicated cell type. The WT cells are from a chr11 α-satellite BAC^{LacO} clone, and the CENP-B KO cells are pooled from three derivative chr11 α-satellite BAC^{LacO} clones. n.s., not significant.
See also Figure S1.

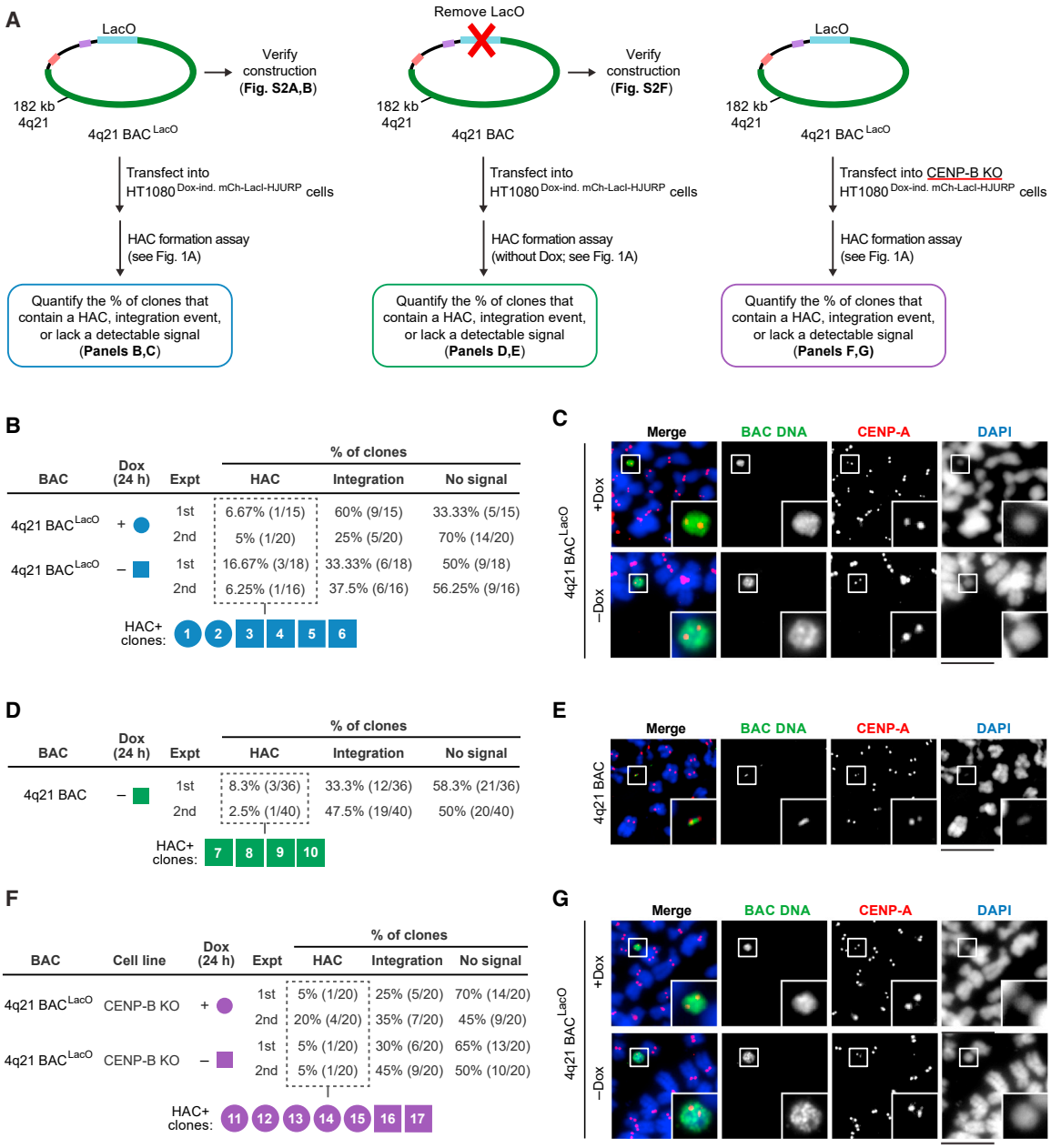


Figure 3. Formation of a HAC on a Template Lacking α-Satellite DNA Even without Seeding CENP-A Nucleosome Assembly or CENP-B

(A) Three tests of a non-α-satellite sequence for its ability to form a HAC.

(B) Results of the HAC formation assays with 4q21 BAC^{LacO} with and without seeding CENP-A nucleosome assembly.

(C) Representative images of the 4q21 BAC^{LacO} HACs formed with and without seeding CENP-A nucleosome assembly.

(D) Results of the HAC formation assays with the 4q21 BAC (i.e., a construct lacking any LacO repeats).

(E) Representative images of the 4q21 BAC HACs formed without any residual CENP-A nucleosome seeding by mCherry-LacI-HJURP.

(F) Results of the HAC formation assays with 4q21 BAC^{LacO} in CENP-B KO mCherry-LacI-HJURP HT1080 cells.

(G) Representative images of the 4q21 BAC^{LacO} HACs formed in the CENP-B KO cells. Insets: 2.5× magnifications. Bar, 10 μm (C, E, and G).

See also Figure S2 and Table S1.

Taken together, this series of HAC formation assays with non-α-satellite DNA constructs clearly indicate that centromere formation must be different from the CENP-B-dependent pathway used by traditional HACs (Ebersole et al., 2000; Grimes et al., 2002; Harrington et al., 1997; Ikeno et al., 1998; Mejía et al., 2002; Ohzeki et al., 2002, 2012; Okada et al., 2007; Schueler et al., 2001) or our new CENP-B-independent HACs that require seeding CENP-A nucleosome assembly (Figures 1 and 2).

Multiple Pathways for HAC Formation on a Non-repetitive DNA Template

We developed a tripartite strategy (Figure 4A) to investigate the pathway for centromere formation for each of the 17 clones isolated through our collection of 4q21-based HAC experiments (Figure 3) (clones 1–6 from 4q21 BAC^{LacO} in WT cells; clones 7–10 from 4q21 BAC in WT cells; clones 11–17 from 4q21 BAC^{LacO} in CENP-B KO cells).

First, using immunofluorescence to detect CENP-B protein and fluorescence *in situ* hybridization (FISH) to detect the HACs, we found that four of the ten clones that formed in the WT (CENP-B-positive) background had no detectable CENP-B protein (Figures 4B, 4C, and S3A) (clones 1, 3, 4, and 7). The other six of the ten clones had detectable CENP-B, with widely varying levels of acquired native centromere sequences likely housing some or all of the functional centromeric chromatin.

Second, using FISH to detect functional centromeric chromatin on HACs detected with the expression of HA epitope-tagged LacI, we found that seven of the ten remaining clones generated with 4q21 BAC^{LacO} had no detectable acquisition of functional centromeric chromatin (Figures 4D, 4E, and S3B) (clones 1, 4, 11–14, and 17; note that clone 7 was generated with a 4q21 BAC construct that lacks a binding site for the HA epitope-tagged LacI, so it could not be included in the second step of our analysis). Two other HACs appeared to form with the acquisition of high levels of functional centromeric chromatin (clones 15 and 16) and another HAC formed with the acquisition of only very little functional centromeric chromatin (clone 3) (Figures 4D, 4E, and S3B).

Third, eight out of the original seventeen 4q21-based HACs whose formation could not be attributed to the acquisition of functional centromeric chromatin in the first two steps of our analysis were subjected to CENP-A chromatin immunoprecipitation sequencing (ChIP-seq) (Figure 4F). By comparing the reads in each HAC-containing cell line to the parental cell line lacking a HAC, we assigned all of the reads coming from the HAC to either the 4q21-containing BAC sequences or the rest of the human reference genome (Table S2). As with prior analysis of human neocentromeres (Hasson et al., 2013), there is a massive increase in CENP-A ChIP-seq reads from the functional centromere on the HAC relative to what is observed in parental cells lacking a HAC. Thus, we assigned all 4q21 CENP-A ChIP reads to the HAC. Using this strategy, we found that four of the HACs (clones 1, 11, 12, and 14) have centromeres residing on DNA essentially entirely comprised of 4q21-containing BAC sequences, while the other four (clones 4, 7, 13, and 17) have acquired genomic sequences upon which at least a portion of the functional centromere (defined by the presence of CENP-A nucleosomes) resides (Figure 4F; Table S2). Both types of HACs (those with centromeres exclusively on the 4q21 sequence and those with acquired genomic sequences) multimerized with rearrangements at unique locations relative to one another but always within non-repetitive regions (i.e., outside of the LacO array) of 4q21 BAC^{LacO} (Figures S3C–S3E).

Our ChIP-seq studies revealed that the centromere on the four HACs essentially entirely comprised of 4q21 BAC^{LacO} vary widely in location (Figure 5A). CENP-A has maximal enrichment on different sequences within the construct, indicating that there is unlikely to be a small number of preferred sequences within the HAC that confer a propensity to establish functional centromeric chromatin. In two of the four clones (clones 11 and 14), the highest peaks of CENP-A enrichment are exclusively on the 4q21 genomic sequence, while on the other two (clones 1 and 12), the highest peaks also include the prokaryotic backbone of 4q21 BAC^{LacO} (Figure 5A). Using a genome-wide sequencing approach we recently applied to studies of centromere strength in mice (Iwata-Otsubo et al., 2017), we analyzed the total input mononucleosome populations isolated after micrococcal nuclease (MNase) digestion of chromatin (Figure S4) and found a substantial enrichment for sequences from chromosome 4q21 (Figure S5). Using established copy number variation analysis tools (Xie and Tammi, 2009), we found that each of the four HACs had substantially multimerized (multimerization varied from 41- to 55-fold, depending on the clone) (Figures 5A and S5; Table S3), consistent with the finding that these HACs exist as large entities in cells that are easily detectable by DAPI staining next to their natural counterparts (Figure 3C). This analysis also revealed that the sharp boundaries of CENP-A localization in some locations on the HACs (for instance in clone 11) (Figure 5A) are not due to amplifications of only specific regions of 4q21 BAC^{LacO} (Figure S4).

In contrast to the four HACs essentially entirely comprised of 4q21 BAC^{LacO}, a clone (clone 13) that had acquired genomic sequences (Figure 4F) showed discreet CENP-A enrichment in several locations within 4q21 (Figure 5B). Additionally, it showed enrichment on acquired sequences that map to α-satellite DNA that normally does not harbor functional centromeric sequences (Figures 4D–4F; Table S4). Another clone (clone 17) had no strong sites of CENP-A enrichment within 4q21, but rather had acquired α-satellite DNA that normally does not harbor functional centromeric sequences (Figures 4D and 4F; Table S4) as well as two other non-centromeric sequences within the genome: one from 3q13 and another from 9q22 (Figure 5B; Table S4). This clone has apparent genomic rearrangements (Figure S5C), and deeper sequencing of the bulk nucleosome reads from this clone revealed several rearrangements not present in the parental cell line that are proximal to the 9q22 region incorporated into the HAC (Figure S6A). It seems likely, therefore, that genomic integrity was compromised in the cell that received 4q21 BAC^{LacO} and originated this particular HAC (clone 17). Together, our findings indicate that, unlike constructs containing α-satellite DNA, non-repetitive constructs can form HACs either directly (Figures 4F and 5A) or by acquiring one or several genomic sequences (Figures 4F and 5B) upon which functional centromeric chromatin is assembled to confer HAC establishment.

For the HACs that have a centromere that we can account for entirely with the 4q21 BAC^{LacO} sequences (clones 1, 11, 12, and 14) by our conventional ChIP-seq analysis (Figure 5A), we found that, like other prior HACs formed with circular constructs (Ebersole et al., 2000; Grimes et al., 2001), none of them had acquired telomeric sequences (Figures S6B and S6C). We devised a strategy to further probe these four clones for potential junctions with α-satellite sequences (Figure 5C). Our strategy employs the recently completed reference models of all autosomal and allosomal human centromeres (Miga et al., 2014) and searches

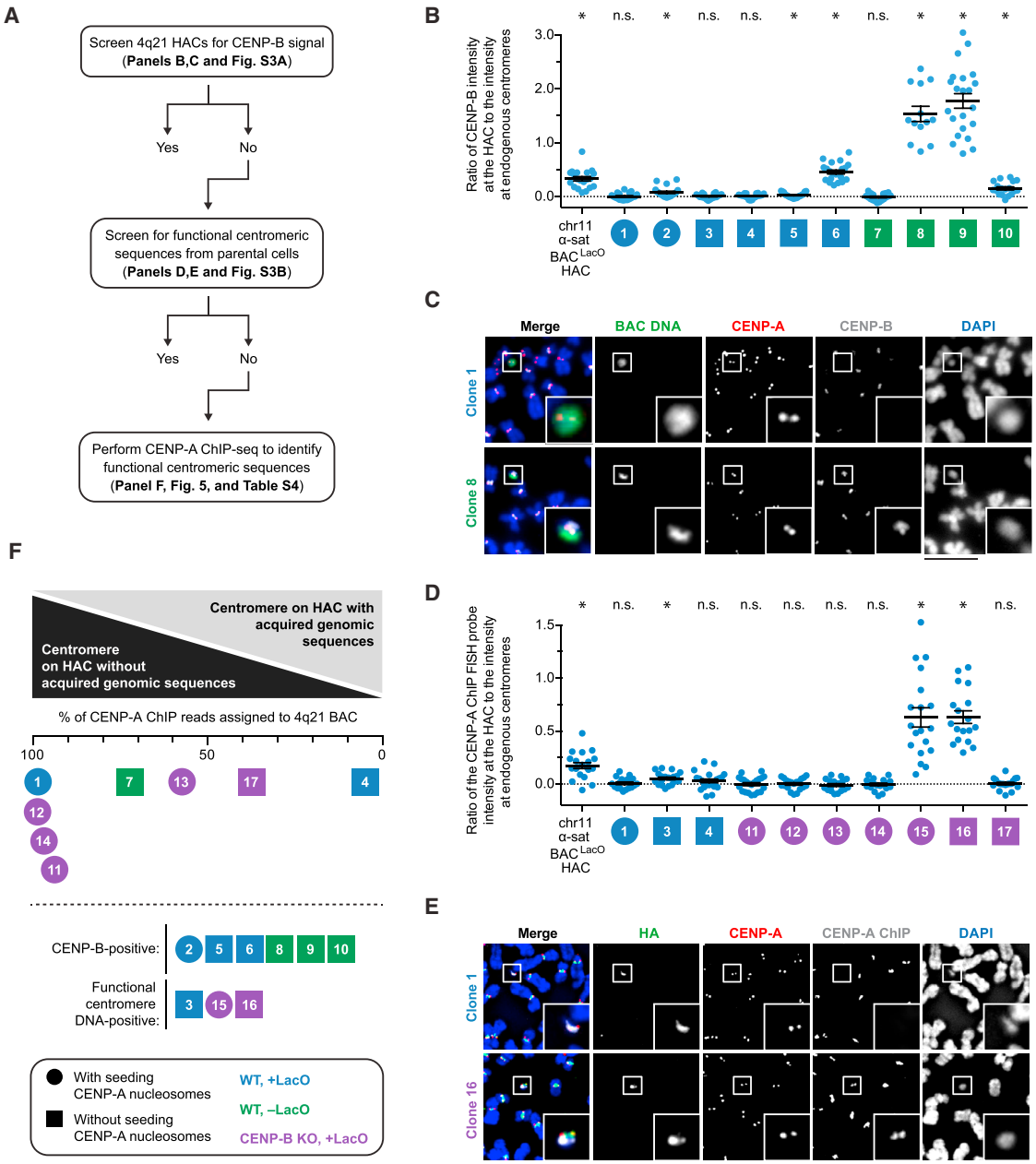


Figure 4. Seeding CENP-A Nucleosome Assembly Dictates the Pathway to Centromere Formation

(A) Steps to test whether the 4q21 HACs have acquired CENP-B protein or functional α-satellite DNA.

(B) Quantification of the intensity of CENP-B at chr11 α-satellite BAC^{LacO} and 4q21 HACs relative to the intensity at endogenous centromeres. Each data point represents a measurement taken at a single HAC. The mean ratio (± SEM) is shown. n = 20, 19, 20, 20, 20, 19, 21, 18, 13, 22, and 18 HACs for the clones shown, in order. p is < 0.0001, 0.8566, 0.0019, 0.6401, 0.2215, 0.0343, < 0.0001, 0.6269, < 0.0001, < 0.0001, < 0.0001 for the clones shown, in order, based on a one-sample t test with a hypothetical mean of 0. Clones with a p value < 0.05 are marked with an asterisk; clones with a p value ≥ 0.05 are marked as not significant (n.s.).

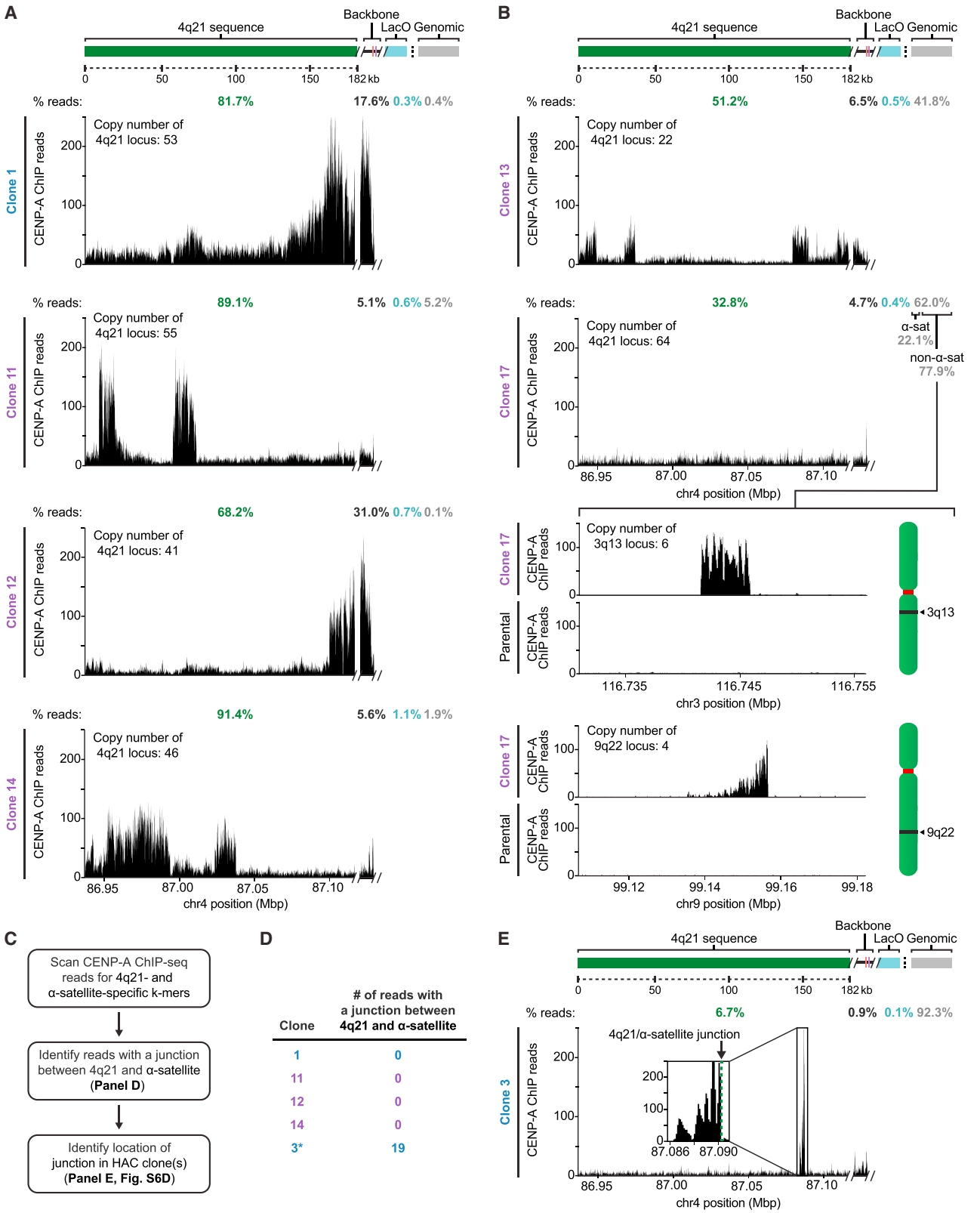
(C) Representative images of a 4q21 HAC that has acquired CENP-B-bound sequences (clone 8) and one that has not (clone 1).

(D) Quantification of the intensity of a CENP-A ChIP probe at chr11 α-satellite BAC^{LacO} and 4q21 HACs relative to the intensity at endogenous centromeres. Each data point represents a measurement taken at a single HAC. The mean ratio (± SEM) is shown. n = 20, 18, 20, 22, 22, 20, 18, 19, 19, 18, and 19 HACs for the clones shown, in order. p is < 0.0001, 0.5642, 0.0005, 0.1028, 0.9098, 0.9602, 0.4708, 0.7553, < 0.0001, < 0.0001, 0.7278 for the clones shown, in order, based on a one-sample t test with a hypothetical mean of 0. Clones with a p value 0.05 are marked with an asterisk; clones with a p value ≥ 0.05 are marked as not significant (n.s.).

(E) Representative images of a 4q21 HAC that has acquired CENP-A-associated sequences (clone 16) and one that has not (clone 1). The HACs are detected with HA-LacI, which binds the LacO repeats present in the HACs. Insets: 2.5× magnifications. Bar, 10 μm (C and E).

(F) Summary of the quantitative analysis of all 4q21 HAC clones.

See also Figure S3 and Table S2.



(legend on next page)

for sequences in our ChIP-seq dataset containing k-mers for both 4q21 and any α-satellite DNA. In addition to the four HACs (clones 1, 11, 12, and 14), we included clone 3 that we found contained a small, but detectable, FISH signal for functional centromeric DNA (Figures 4D and S3B). We found that none of the four HACs we interrogated contains a single detectable junction with α-satellite DNA (Figure 5D) within the pool of >6,000,000 reads analyzed per HAC. Clone 3, however, contained 19 such junction reads (Figure 5D). These 19 reads vary in length on one or both ends of the read, but contain precisely the same junction site between 4q21 and a sequence that is from within a monomer of α-satellite DNA (Figure S6D). Indeed, this junction site maps to one side of the single strong peak of CENP-A nucleosome enrichment on 4q21 (Figure 5E). Thus, using an approach that is capable of readily identifying the presence of a junction and defines its site at single-nucleotide resolution, we failed to find any evidence of such junctions in the four HACs we identified that formed a centromere without acquiring genomic sequences.

To investigate the stability and organization of HACs that have formed without acquiring genomic sequences, we focused our attention on detailed analysis of two clones, clones 1 and 11, that have clearly distinct CENP-A enrichment patterns from one another (Figure 5A). First, we found that the daily loss rate over 60 days in cell culture (Figures 6A and 6B) is similarly low for both HACs compared to the other HACs measured in this study (Figures 1H and 2J). Next, we sought to define the organization of each HAC with regard to their CENP-A occupancy in a way consistent with both their discrete paired sister centromere morphology (Figure 4C) and with the CENP-A ChIP-seq data mapped to the input 4q21 BAC^{LacO} sequence (Figure 5A). One possibility is that the sites of ChIP-seq enrichment represent low CENP-A occupancy per amplified copy of the 4q21 BAC^{LacO} that might coalesce in three dimensions on the HAC. Alternatively, only one or a small number of copies of 4q21 BAC^{LacO} house CENP-A, with the vast majority devoid of centromeric nucleosomes. To distinguish between these possibilities, we used stretched chromatin fibers (Blower et al., 2002; Iwata-Otsubo et al., 2017) to visualize HAC centromeres at high resolution (Figures 6C and 6D). We used HA-tagged LacI to identify the HAC and mark the portion of each copy of 4q21 BAC^{LacO} containing the LacO array (Figure 6D, visualized as a single green focus). For both clones 1 and 11, CENP-A largely occupies space on the fiber between these foci, consistent with our ChIP-seq mapping (Figure 5A), and only a small fraction of BACs (each copy of 4q21 BAC DNA is represented by a gap between LacO arrays) are occupied by CENP-A (Figures 6C and 6D). We also note that, for both clones, we observed several examples of fibers like the representative images shown where there is a major and minor site of CENP-A enrichment (closer together for clone 1 than for clone 11) (Figure 6D). Cumulatively, CENP-A nucleosomes occupy a region of high density similar to that on neighboring centromeres from endogenous chromosomes (Figure 6D; marked with an asterisk). Taking into consideration the 5.3–9.0 Mb size of the HACs (Table S3) and that our quantitation (Figure 6C) may somewhat overestimate the total fraction of the HAC occupied by CENP-A (see STAR Methods), we conclude that a typical copy of these HACs have regions of CENP-A that discontinuously span 0.5–1 Mb of neighboring copies of 4q21 BAC^{LacO}. Further, we conclude that the major CENP-A ChIP-seq peaks (Figure 5A) each represent the position of CENP-A enrichment on one or a small number of individual copies of 4q21 BAC^{LacO}.

Together, our mapping (Figures 5A and 5B), junction searching approaches (Figures 5C–5E), and chromosome stretching experiments (Figures 6C and 6D) highlight how HACs that bypass centromeric DNA allow for a comprehensive understanding of centromeric chromatin localization as well as HAC composition, organization, and copy number in a manner that is not imaginable with HACs built from centromeric DNA repeats.

DISCUSSION

Centromere formation on HACs has long been thought to require α-satellite DNA with a high density of CENP-B boxes (Ikeno et al., 1998; Ohzeki et al., 2002; Okada et al., 2007; Schueler et al., 2001), with proposals that a high density of local CENP-B on the naked DNA facilitates nearby nascent assembly of CENP-A nucleosomes (Okada et al., 2007) or stabilizes them upon formation (Fujita et al., 2015). Here, we report two other ways to establish a centromere during HAC formation. The first is a directed approach with local seeding of CENP-A nucleosomes on repetitive α-satellite DNA, taking advantage of a growing wealth of knowledge about the

Figure 5. Genomic Analysis of 4q21 HACs Reveals Precise Location of Newly Formed Centromeres

(A) CENP-A ChIP-seq analysis of the 4q21 HACs that formed without acquisition of CENP-A, functional α-satellite, or additional sequences from host chromosomes. CENP-A is localized along the 4q21 sequence and can be also be found on the backbone of the BAC. In all four clones, >90% of reads enriched with CENP-A align to the BAC sequence. The copy number of the 4q21 locus is shown for each HAC and includes the endogenous locus (that contributes 2.7 copies on average, as determined by IF-FISH).

(B) CENP-A ChIP-seq analysis of two 4q21 HACs that had acquired additional sequences from the host chromosomes during HAC formation. In clone 13, the HAC had acquired sequences predominantly from the chr10 centromere, while clone 17 had acquired sequences from the chr10 centromere and two non-centromeric loci, 3q13 and 9q22. The copy number of the 4q21, 3q13, and 9q22 loci are shown for each HAC and includes the endogenous locus (that contributes 2.7, 4.0, and 3.2 copies on average for the 4q21, 3q13, and 9q22 loci, respectively, as determined by IF-FISH).

(C) Steps to identify CENP-A ChIP-seq reads harboring a junction between 4q21 and α-satellite sequences in 4q21 HAC clones.

(D) Summary of the read junction analysis showing that none of the 4q21 HAC clones that had formed without acquisition of CENP-A, functional α-satellite, or additional sequences from host chromosomes contained reads with a junction between 4q21 and α-satellite. However, a clone that had acquired functional centromeric chromatin (denoted by an asterisk) contained 19 reads with such junctions.

(E) CENP-A ChIP-seq analysis of clone 3, which contains 19 reads spanning the junction between 4q21 and α-satellite (shown in Figure S6D). The location of the junction is indicated by a dashed green line.

See also Figures S4 and S5 and Tables S3, S4, and S5.

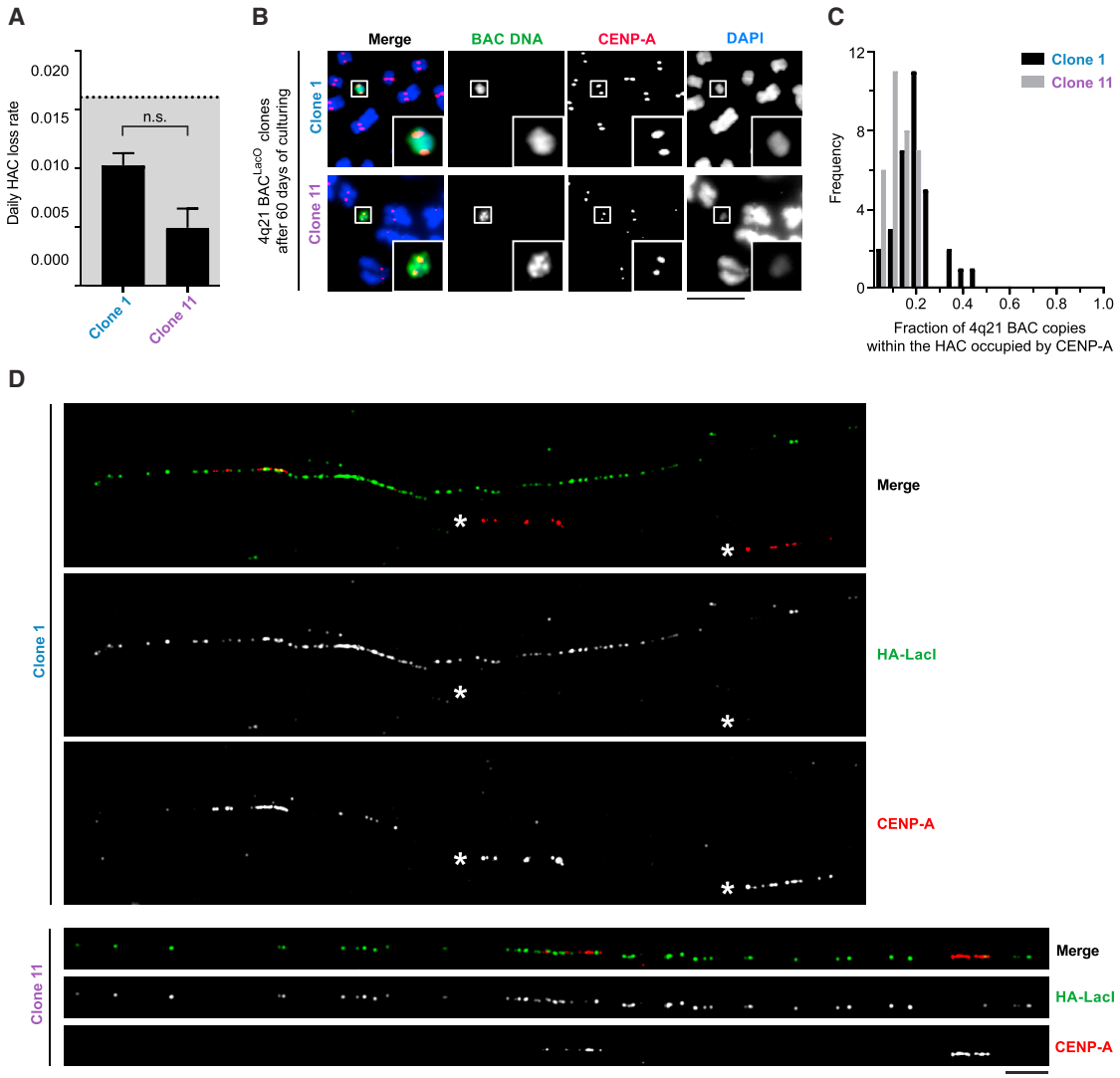


Figure 6. Stable 4q21 HACs that Have Not Acquired Genomic Sequences During Formation Harbor a Centromere with a High Local Density of CENP-A

(A) Quantitation of the daily HAC loss rate in clones 1 and 11 after culturing without G418-S for 60 days (shading as in Figure 1H). The mean daily loss rate (± SEM) is shown. n = 120 cells pooled from 3 independent experiments for each clone. n.s., not significant.

(B) Representative images of the HACs in clones 1 and 11 after 60 days of culturing in the absence of G418-S. Insets: 2.5× magnifications. Bar, 10 μm.

(C) Histogram of the fraction of 4q21 BAC copies that are occupied by CENP-A within clones 1 and 11. 4q21 BAC copies were visualized on physically stretched chromatin fibers with immunodetection of CENP-A and expression of HA-LacI in cells. The 4q21 BAC DNA is represented by gaps between resolvable foci of HA-LacI.

(D) Representative example of a stretched copy of the HAC in clones 1 and 11. CENP-A occupies discrete regions of neighboring copies of the 4q21 BAC^{LacO}, spanning a similar cumulative distance of stretched chromatin as in neighboring centromeres in the same field (denoted by asterisks). Bar, 5 μm.

CENP-A nucleosome assembly pathway (Barnhart et al., 2011; Dunleavy et al., 2009; Foltz et al., 2009; Logsdon et al., 2015; Mendiburo et al., 2011). More surprisingly, however, we found that HACs can form on non-repetitive sequences without a requirement for seeding CENP-A nucleosomes, CENP-B boxes, or the expression of CENP-B, itself. Our functional tests with HACs are especially important to inform the further development of recent proposals for DNA sequence-based contributions to centromere identity (Kasinathan and Henikoff, 2018) and strength (Iwata-Otsubo et al., 2017).

The HACs we report that do not acquire α-satellite sequences during centromere formation (Figures 4F and 5A) are able to epigenetically maintain centromere identity in a manner we propose is analogous to non-repetitive neocentromeres in the human population (Alonso et al., 2003, 2010; Hasson et al., 2011). Non-repetitive DNA that does not require seeding CENP-A nucleosome assembly might have been missed in previous HAC studies for several reasons. Sequences besides α-satellite were assumed to be implausible in human cells because of the reported CENP-B requirement (Ohzeki et al., 2002; Okada et al., 2007), the failure of α-satellite DNA with a low density of CENP-B

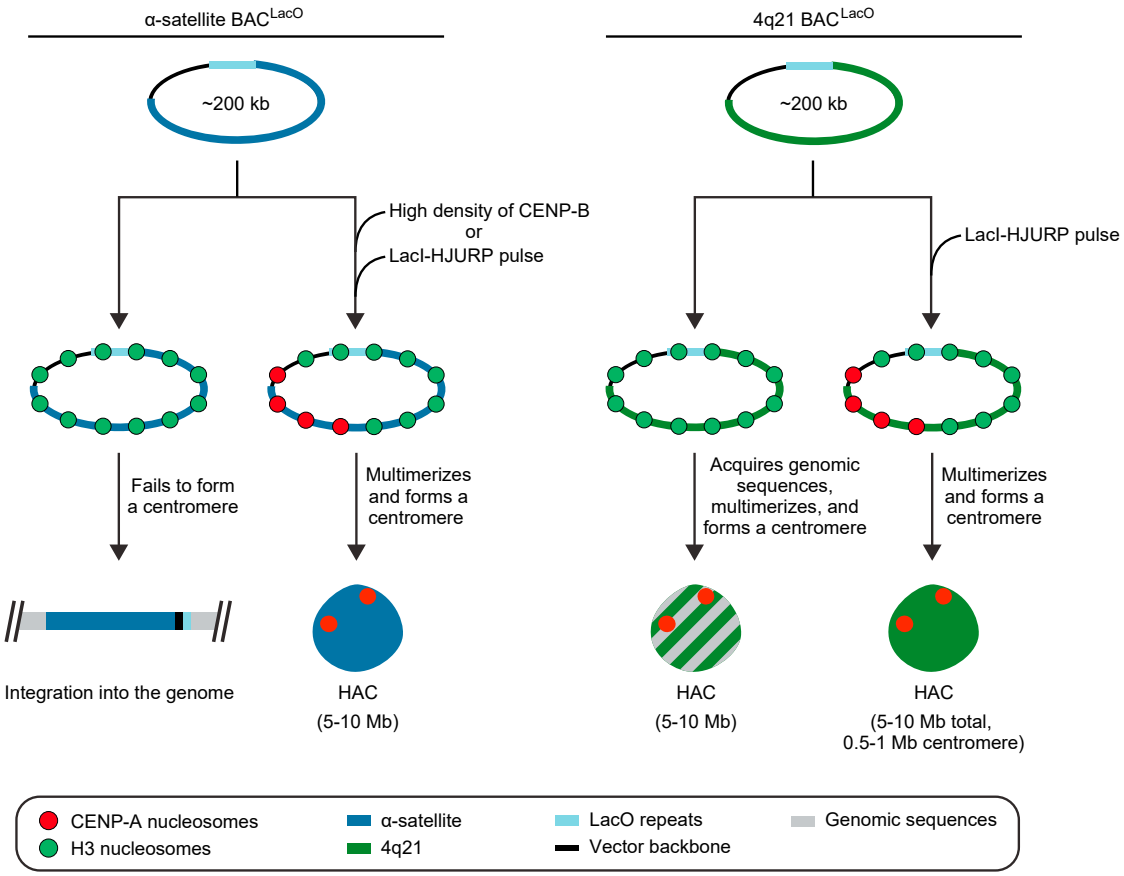


Figure 7. Pathways to HAC Formation

Cartoon drawing summarizing the findings in this study. In the absence of a high density of CENP-B binding or CENP-A nucleosome seeding, the α-satellite BAC^{LacO} vectors fail to form a centromere and are subsequently integrated into the genome. Alternatively, when α-satellite BAC^{LacO} vectors have a high density of CENP-B binding or are epigenetically seeded with CENP-A nucleosomes via LacI-HJURP, they assemble centromeric chromatin, multimerize, and form a functional HAC. When non-α-satellite vectors such as 4q21 BAC^{LacO} are introduced to human cells, they also can integrate into the genome (not depicted here), but strikingly can form HACs without CENP-B boxes or seeding of centromeric nucleosome assembly. In the absence of CENP-A nucleosome seeding via LacI-HJURP, 4q21 BAC^{LacO} vectors acquire host genomic sequences, which impart centromere competency to the vector and lead to the formation of a HAC. These sequences often consist of both α-satellite and non-α-satellite from host chromosomes. However, when CENP-A nucleosome assembly is directly seeded onto the 4q21 BAC^{LacO} vectors, centromeric chromatin is assembled onto the vector, it multimerizes, and a HAC is formed. The CENP-A nucleosomes reside on a small fraction of the 4q21 BAC^{LacO} vector sequences on the HAC, taking up the equivalent of 0.5–1.0 Mb of sequence, which is similar to the size observed at normal, endogenous human centromeres.

See also Figure S7.

boxes (Schueler et al., 2001), and the failure of two different non-repetitive genomic sequences (from chromosomes 10, Saffery et al., 2001; and X, Grimes et al., 2002; respectively). Indeed, HAC formation on 4q21 BAC^{LacO} is relatively rare, occurring in only one or a few clones isolated in a typical HAC experiment (Figure 3). Thus, previous conclusions likely precluded the exploration of non-repetitive sequences.

We propose a model wherein there are three types of human genomic sequences that can form a HAC (Figure S7). The first two are α-satellite DNA with either a high or low density of functional CENP-B boxes. We envision that both types of α-satellite DNA are similar in their inherent resistance to initial CENP-A nucleosome assembly and/or resistance to establishing a self-propagating centromere because they are susceptible to CENP-A nucleosome displacement by invading heterochromatin. The resistance can be overcome by either a high density of CENP-B binding (Figure 7) (Nakano et al., 2008; Schueler et al., 2001) and accelerated further by CENP-A overexpression (Pesenti et al., 2018) or local targeting of CENP-A nucleosome assembly with HJURP (Figure 7). Once the initial resistance is overcome, the natural epigenetic centromere propagation mechanism takes hold, wherein the high local concentration of existing CENP-A nucleosomes directs the formation of nascent CENP-A nucleosomes on nearby DNA (Black and Cleveland, 2011; McKinley and Cheeseman, 2016). On first blush, our proposal that the sequences found at all normal human centromeres would be inherently resistant to centromere formation seems paradoxical. One must remember, though, that centromere movement is very slow relative to the timescale of cell divisions. Epialleles of CENP-A location within a "sea" of highly repetitive

DNA (most of which is packaged by canonical nucleosomes) are reported to be heritable through the germline (Maloney et al., 2012). Taking our findings into account, centromeres appear to evolve to restrict the pace of movement of CENP-A-containing chromatin, providing a potential explanation to the paradox of why α-satellite DNA is found at all normal human centromeres even though it is not required for centromere identity and function (Eichler, 1999). If α-satellite repeats were inherently neutral or permissive to new centromere formation, then the high local density of CENP-A nucleosomes generated by its self-templated mechanism for propagation might be compromised by the rapid attraction to any of the other α-satellite repeats at a given centromere. Rather, once established, centromeres are built to be stable chromatin domains.

The third type of sequence that can form a HAC lacks repetitive DNA or any CENP-B boxes but is competent for centromere formation (Figure S7). When centromere formation occurs on 4q21 BAC^{LacO} via seeding CENP-A nucleosome assembly, the HACs frequently form without acquiring other DNA sequences (Figure 7). Without seeding of CENP-A nucleosomes, HACs still form, albeit through acquiring CENP-B-positive α-satellite, CENP-B-negative α-satellite, or mixtures of genomic sequences including regions where new centromeres form on non-repetitive sequences (Figure 7). This DNA acquisition pathway to centromere formation was never observed with our α-satellite-based HACs. Thus, the sequence of input naked DNA HAC templates strongly impacts and extends the possible routes to centromere formation and potential ultimate success in generating functional HACs.

In terms of the overall outlook for HACs, our findings reveal surprising flexibility in how one can form centromeres in mammalian cells, indicating that it is possible to surmount two major limitations of HACs (low formation efficiency as shown in Basu and Willard, 2005 and the strict requirement for including substantial amounts of highly repetitive α-satellite DNA as shown in Grimes et al., 2002; Ikeno et al., 1998, and Ohzeki et al., 2002). First, HAC formation on α-satellite constructs can be substantially increased by a single pulse of epigenetic seeding of centromeric chromatin assembly. Strategies to increase HAC formation efficiency are desperately needed in order to achieve their potential in synthetic biology, especially if they are to serve as the basis for entirely synthetic chromosomes (Boeke et al., 2016). Second, HAC formation can occur without α-satellite DNA, and, when coupled to initial seeding of CENP-A nucleosome, frequently does so without acquiring other genomic sequences. With our methodology to circumvent the prior absolute requirement for CENP-B boxes/CENP-B protein (Ohzeki et al., 2002; Okada et al., 2007) and repetitive sequences, one could now envision simplifying schemes to make completely synthetic HAC templates (Nakano et al., 2008).

Prior studies have established that a typical route for HAC formation includes either simple or complex rearrangements that culminate in reassembly of a continuous molecule >1 Mb in size (Kouprina et al., 2012). The HACs in this study that underwent our full genomic analysis are each in the 5–10 Mb range and appear to have been formed with similar rearrangements and reassembly (Figures S3C and S3D). For the 4q21 BAC^{LacO} HACs, what stands out is their ability to form without detectable acquisition of genomic DNA, but only when there is prior seeding of CENP-A nucleosomes by HJURP targeting. This targeting event likely extends from the moment of introduction of BAC^{LacO} constructs through several initial cell cycles. LacO-tethered HJURP likely acts to both support nascent CENP-A nucleosome assembly and ward off heterochromatin (as has been reported upon cellular introduction of ectopic α-satellite DNA; Ohzeki et al., 2012), providing the time it takes for a self-propagating epigenetic centromere to form. Later, after clonal HACs are isolated, CENP-A nucleosome seeding by HJURP targeting is no longer necessary for the centromere on the HAC. The HAC is now propagated indistinguishably from other centromeres. The fact that non-repetitive DNA can act as a self-propagating centromere is central to our current understanding of the epigenetic underpinnings of centromere specification (i.e., what are termed neocentromeres, *de novo* centromeres, and/or evolutionary new centromeres; Fu et al., 2013; Liu et al., 2015; Montefalcone et al., 1999; Schneider et al., 2016; Shang et al., 2010; Tolomeo et al., 2017; Topp et al., 2009; Wade et al., 2009; Wang et al., 2014). The strategy we report in this paper provides a way to make a new centromere in the context of an artificial chromosome that should be widely applicable among mammals and more broadly to many eukaryotic systems.

Combining the methodologies from synthetic chromosome efforts in yeast with new innovations in HACs presents an attractive avenue for synthetic biology efforts. The development of tools and automation to synthesize and analyze yeast chromosomes should now be extended to mammalian systems to accelerate HAC development. For instance, HAC studies are currently slowed by the requirement of clonal isolation of many cell lines (e.g., 453 cell lines were isolated for this study alone) and very low-throughput analysis tools (i.e., combined IF-FISH to identify HAC-containing lines), all of which would benefit from streamlined methodologies and emerging instrumentation. Our study reveals that one promising avenue will be to develop non-repetitive HAC vectors that will allow annotation of copy number and organization of the functional centromere on every isolated clone. Centromeric α-satellite DNA is the most abundant highly repetitive DNA in humans, constituting ~3% of our genome (Miga, 2017). Its repetitive nature has substantially slowed progress in HAC development because it is difficult to synthesize, a major challenge to clone and amplify without unwanted recombination, and refractory to characterization using genomic approaches. Our study reveals molecular requirements for centromere establishment and demonstrates that α-satellite DNA can be bypassed altogether, thereby greatly facilitating the construction of HACs and expanding the toolbox for centromere biology studies, gene therapy applications, and synthetic biology efforts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY

- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
- METHOD DETAILS
  - Plasmid construction
  - Cre-Lox recombination of BACs
  - HAC formation assays
  - HAC maintenance assays
  - IF-FISH on metaphase chromosome spreads with BAC-specific probes
  - HA-LacI and EGFP-LacI lentivirus production
  - HA-LacI lentiviral transduction and IF-FISH on metaphase chromosome spreads
  - EGFP-LacI lentiviral transduction and IF-FISH on metaphase chromosome spreads with a telomere-specific probe
  - Immunoblots
  - Southern blots
  - Native CENP-A ChIP
  - Next-generation sequencing and data processing
  - Custom reference genome
  - CNV analysis
  - Distribution of CENP-A ChIP reads within 4q21 BAC^{LacO} HACs
  - HAC read junction analysis
  - Deep-sequencing and chimeric read analysis
  - IF on chromatin fibers
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Quantification of CENP-A, CENP-B, and CENP-A ChIP FISH probe intensity
  - Quantification of HAC chromatin fibers
  - Statistical information
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2019.06.006.

ACKNOWLEDGMENTS

We thank B. Sullivan (Duke), R. O'Neill (UConn), R. Greenberg (UPenn), and T. Zhang (UPenn) for advice and reagents. We also thank E. Makeyev (NTU, Singapore), F. Zhang (MIT), D. Foltz (Northwestern), and D. Cleveland (UCSD) for reagents. We thank E. Strome (NKU) for discussions and M. Lampson (UPenn) for comments on the manuscript. This work was supported by NIH (GM130302 to B.E.B.). G.A.L. acknowledges support from the UPenn Cell and Molecular Biology Training Grant (GM007229). E.J.B. and P.H. were supported by a European Research Council Starting-Consolidator Grant (311674-BioSynCEN), and P.H. was further supported by a Wellcome Trust Senior Researcher Fellowship (103897/Z/14). The Wellcome Trust Centre for Cell Biology is supported by core funding from the Wellcome Trust [092076, 203149].

AUTHOR CONTRIBUTIONS

G.A.L. and B.E.B. conceived the project (with input from E.J.B. and P.H.). G.A.L., C.W.G., and M.A.L. performed experiments. G.A.L., C.W.G., M.A.L., V.L., K.H.M., and B.E.B. analyzed data. G.A.L., C.W.G., E.J.B., P.H., and B.E.B. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 19, 2018
Revised: April 7, 2019
Accepted: June 3, 2019
Published: July 25, 2019

REFERENCES

Alonso, A., Mahmood, R., Li, S., Cheung, F., Yoda, K., and Warburton, P.E. (2003). Genomic microarray analysis reveals distinct locations for the CENP-A binding domains in three human chromosome 13q32 neocentromeres. Hum. Mol. Genet. *12*, 2711–2721.

Alonso, A., Hasson, D., Cheung, F., and Warburton, P.E. (2010). A paucity of heterochromatin at functional human neocentromeres. Epigenetics Chromatin *3*, 6.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Amor, D.J., Bentley, K., Ryan, J., Perry, J., Wong, L., Slater, H., and Choo, K.H.A. (2004). Human centromere repositioning "in progress". Proc. Natl. Acad. Sci. USA *101*, 6542–6547.

Annaluru, N., Muller, H., Mitchell, L.A., Ramalingam, S., Stracquadanio, G., Richardson, S.M., Dymond, J.S., Kuang, Z., Scheifele, L.Z., Cooper, E.M., et al. (2014). Total synthesis of a functional designer eukaryotic chromosome. Science *344*, 55–58.

Barnhart, M.C., Kuich, P.H.J.L., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J. Cell Biol. *194*, 229–243.

Barrey, E.J., and Heun, P. (2017). Artificial chromosomes and strategies to initiate epigenetic centromere establishment. In Centromeres and Kinetochores, B.E. Black, ed. (Springer International Publishing), pp. 193–212.

Bassett, E.A., DeNizio, J., Barnhart-Dailey, M.C., Panchenko, T., Sekulic, N., Rogers, D.J., Foltz, D.R., and Black, B.E. (2012). HJURP uses distinct CENP-A surfaces to recognize and to stabilize CENP-A/histone H4 for centromere assembly. Dev. Cell *22*, 749–762.

Basu, J., and Willard, H.F. (2005). Artificial and engineered chromosomes: non-integrating vectors for gene therapy. Trends Mol. Med. *11*, 251–258.

Basu, J., Stromberg, G., Compitello, G., Willard, H.F., and Van Bokkelen, G. (2005). Rapid creation of BAC-based human artificial chromosome vectors by transposition with synthetic alpha-satellite arrays. Nucleic Acids Res. *33*, 587–596.

Bickmore, W. (1999). Chromosome structural analysis: a practical approach (Oxford University Press).

Black, B.E., and Cleveland, D.W. (2011). Epigenetic centromere propagation and the nature of CENP-A nucleosomes. Cell *144*, 471–479.

Blower, M.D., Sullivan, B.A., and Karpen, G.H. (2002). Conserved organization of centromeric chromatin in flies and humans. Dev. Cell *2*, 319–330.

Bodor, D.L., Rodríguez, M.G., Moreno, N., and Jansen, L.E.T. (2012). Analysis of protein turnover by quantitative SNAP-based pulse-chase imaging. Curr. Protoc. Cell Biol. *Chapter 8*, Unit8.8.

Boeke, J.D., Church, G., Hessel, A., Kelley, N.J., Arkin, A., Cai, Y., Carlson, R., Chakravarti, A., Cornish, V.W., Holt, L., et al. (2016). GENOME ENGINEERING. The Genome Project-Write. Science *353*, 126–127.

Carlson-Stevermer, J., Goedland, M., Steyer, B., Movaghar, A., Lou, M., Kohlenberg, L., Prestil, R., and Saha, K. (2016). High-content analysis of CRISPR-Cas9 gene-edited human embryonic stem cells. Stem Cell Reports *6*, 109–120.

Chen, C.-C., Dechassa, M.L., Bettini, E., Ledoux, M.B., Belisario, C., Heun, P., Luger, K., and Mellone, B.G. (2014). CAL1 is the Drosophila CENP-A assembly factor. J. Cell Biol. *204*, 313–329.

Depinet, T.W., Zackowski, J.L., Earnshaw, W.C., Kaffe, S., Sekhon, G.S., Stallard, R., Sullivan, B.A., Vance, G.H., Van Dyke, D.L., Willard, H.F., et al. (1997). Characterization of neo-centromeres in marker chromosomes lacking detectable alpha-satellite DNA. Hum. Mol. Genet. *6*, 1195–1204.

du Sart, D., Cancilla, M.R., Earle, E., Mao, J.I., Saffery, R., Tainton, K.M., Kalitsis, P., Martyn, J., Barry, A.E., and Choo, K.H.A. (1997). A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. Nat. Genet. *16*, 144–153.

Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell *137*, 485–497.

Earnshaw, W.C., and Migeon, B.R. (1985). Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. Chromosoma *92*, 290–296.

Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F., and Cleveland, D.W. (1987). Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. J. Cell Biol. *104*, 817–829.

Ebersole, T.A., Ross, A., Clark, E., McGill, N., Schindelhauer, D., Cooke, H., and Grimes, B. (2000). Mammalian artificial chromosome formation from circular alphoid input DNA does not require telomere repeats. Hum. Mol. Genet. *9*, 1623–1631.

Eichler, E.E. (1999). Repetitive conundrums of centromere structure and function. Hum. Mol. Genet. *8*, 151–155.

Fachinetti, D., Folco, H.D., Nechemia-Arbely, Y., Valente, L.P., Nguyen, K., Wong, A.J., Zhu, Q., Holland, A.J., Desai, A., Jansen, L.E.T., and Cleveland, D.W. (2013). A two-step mechanism for epigenetic specification of centromere identity and function. Nat. Cell Biol. *15*, 1056–1066.

Fachinetti, D., Han, J.S., McMahon, M.A., Ly, P., Abdullah, A., Wong, A.J., and Cleveland, D.W. (2015). DNA sequence-specific binding of CENP-B enhances the fidelity of human centromere function. Dev. Cell *33*, 314–327.

Fachinetti, D., Logsdon, G.A., Abdullah, A., Selzer, E.B., Cleveland, D.W., and Black, B.E. (2017). CENP-A modifications on Ser68 and Lys124 are dispensable for establishment, maintenance, and long-term function of human centromeres. Dev. Cell *40*, 104–113.

Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., 3rd, Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-specific assembly of CENP-A nucleosomes is mediated by HJURP. Cell *137*, 472–484.

Fu, S., Lv, Z., Gao, Z., Wu, H., Pang, J., Zhang, B., Dong, Q., Guo, X., Wang, X.-J., Birchler, J.A., and Han, F. (2013). De novo centromere formation on a chromosome fragment in maize. Proc. Natl. Acad. Sci. USA *110*, 6033–6036.

Fujita, R., Otake, K., Arimura, Y., Horikoshi, N., Miya, Y., Shiga, T., Osakabe, A., Tachiwana, H., Ohzeki, J., Larionov, V., et al. (2015). Stable complex formation of CENP-B with the CENP-A nucleosome. Nucleic Acids Res. *43*, 4909–4922.

Grimes, B.R., Schindelhauer, D., McGill, N.I., Ross, A., Ebersole, T.A., and Cooke, H.J. (2001). Stable gene expression from a mammalian artificial chromosome. EMBO Rep. *2*, 910–914.

Grimes, B.R., Rhoades, A.A., and Willard, H.F. (2002). Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation. Mol. Ther. *5*, 798–805.

Harrington, J.J., Van Bokkelen, G., Mays, R.W., Gustashaw, K., and Willard, H.F. (1997). Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. Nat. Genet. *15*, 345–355.

Hasson, D., Alonso, A., Cheung, F., Tepperberg, J.H., Papenhausen, P.R., Engelen, J.J.M., and Warburton, P.E. (2011). Formation of novel CENP-A domains on tandem repetitive DNA and across chromosome breakpoints on human chromosome 8q21 neocentromeres. Chromosoma *120*, 621–632.

Hasson, D., Panchenko, T., Salimian, K.J., Salman, M.U., Sekulic, N., Alonso, A., Warburton, P.E., and Black, B.E. (2013). The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat. Struct. Mol. Biol. *20*, 687–695.

Hayden, K.E., Strome, E.D., Merrett, S.L., Lee, H.-R., Rudd, M.K., and Willard, H.F. (2013). Sequences associated with centromere competency in the human genome. Mol. Cell. Biol. *33*, 763–772.

Holland, A.J., Fachinetti, D., Han, J.S., and Cleveland, D.W. (2012). Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. Proc. Natl. Acad. Sci. USA *109*, E3350–E3357.

Hori, T., Shang, W.-H., Takeuchi, K., and Fukagawa, T. (2013). The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. J. Cell Biol. *200*, 45–60.

Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H., and Masumoto, H. (1998). Construction of YAC-based mammalian artificial chromosomes. Nat. Biotechnol. *16*, 431–439.

Iwata-Otsubo, A., Dawicki-McKenna, J.M., Akera, T., Falk, S.J., Chmátal, L., Yang, K., Sullivan, B.A., Schultz, R.M., Lampson, M.A., and Black, B.E. (2017). Expanded satellite repeats amplify a discrete CENP-A nucleosome assembly site on chromosomes that drive in female meiosis. Curr. Biol. *27*, 2365–2373.e8.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.

Kasinathan, S., and Henikoff, S. (2018). Non-B-form DNA is enriched at centromeres. Mol. Biol. Evol. *35*, 949–962.

Khandelia, P., Yap, K., and Makeyev, E.V. (2011). Streamlined platform for short hairpin RNA interference and transgenesis in cultured mammalian cells. Proc. Natl. Acad. Sci. USA *108*, 12799–12804.

Kouprina, N., Samoshkin, A., Erliandri, I., Nakano, M., Lee, H.-S., Fu, H., Iida, Y., Aladjem, M., Oshimura, M., Masumoto, H., et al. (2012). Organization of synthetic alphoid DNA array in human artificial chromosome (HAC) with a conditional centromere. ACS Synth. Biol. *1*, 590–601.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., et al. (2007). The diploid genome sequence of an individual human. PLoS Biol. *5*, e254.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liu, Y., Su, H., Pang, J., Gao, Z., Wang, X.-J., Birchler, J.A., and Han, F. (2015). Sequential de novo centromere formation and inactivation on a chromosomal fragment in maize. Proc. Natl. Acad. Sci. USA *112*, E1263–E1271.

Logsdon, G.A., Barrey, E.J., Bassett, E.A., DeNizio, J.E., Guo, L.Y., Panchenko, T., Dawicki-McKenna, J.M., Heun, P., and Black, B.E. (2015). Both tails and the centromere targeting domain of CENP-A are required for centromere establishment. J. Cell Biol. *208*, 521–531.

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

Maloney, K.A., Sullivan, L.L., Matheny, J.E., Strome, E.D., Merrett, S.L., Ferris, A., and Sullivan, B.A. (2012). Functional epialleles at an endogenous human centromere. Proc. Natl. Acad. Sci. USA *109*, 13704–13709.

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. Nat. Rev. Mol. Cell Biol. *17*, 16–29.

Mejía, J.E., Alazami, A., Willmott, A., Marschall, P., Levy, E., Earnshaw, W.C., and Larin, Z. (2002). Efficiency of de novo centromere formation in human artificial chromosomes. Genomics *79*, 297–304.

Mendiburo, M.J., Padeken, J., Fülöp, S., Schepers, A., and Heun, P. (2011). Drosophila CENH3 is sufficient for centromere formation. Science *334*, 686–690.

Miga, K.H. (2017). The promises and challenges of genomic studies of human centromeres. In Centromeres and Kinetochores, B.E. Black, ed. (Springer International Publishing), pp. 285–304.

Miga, K.H., Newton, Y., Jain, M., Altemose, N., Willard, H.F., and Kent, W.J. (2014). Centromere reference models for human chromosomes X and Y satellite arrays. Genome Res. *24*, 697–707.

Montefalcone, G., Tempesta, S., Rocchi, M., and Archidiacono, N. (1999). Centromere repositioning. Genome Res. *9*, 1184–1188.

Morgenstern, J.P., and Land, H. (1990). Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. Nucleic Acids Res. *18*, 3587–3596.

Nakano, M., Cardinale, S., Noskov, V.N., Gassmann, R., Vagnarelli, P., Kandels-Lewis, S., Larionov, V., Earnshaw, W.C., and Masumoto, H. (2008). Inactivation of a human kinetochore by specific targeting of chromatin modifiers. Dev. Cell *14*, 507–522.

Ohzeki, J., Nakano, M., Okada, T., and Masumoto, H. (2002). CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J. Cell Biol. *159*, 765–775.

Ohzeki, J., Bergmann, J.H., Kouprina, N., Noskov, V.N., Nakano, M., Kimura, H., Earnshaw, W.C., Larionov, V., and Masumoto, H. (2012). Breaking the HAC Barrier: histone H3K9 acetyl/methyl balance regulates CENP-A assembly. EMBO J. *31*, 2391–2402.

Okada, T., Ohzeki, J., Nakano, M., Yoda, K., Brinkley, W.R., Larionov, V., and Masumoto, H. (2007). CENP-B controls centromere formation depending on the chromatin context. Cell *131*, 1287–1300.

Okamoto, Y., Nakano, M., Ohzeki, J., Larionov, V., and Masumoto, H. (2007). A minimal CENP-A core is required for nucleation and maintenance of a functional human centromere. EMBO J. *26*, 1279–1291.

Pesenti, E., Kouprina, N., Liskovykh, M., Aurich-Costa, J., Larionov, V., Masumoto, H., Earnshaw, W.C., and Molina, O. (2018). Generation of a synthetic human chromosome with two centromeric domains for advanced epigenetic engineering studies. ACS Synth. Biol. *7*, 1116–1130.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187-91.

Richardson, S.M., Mitchell, L.A., Stracquadanio, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D., Huang, C.L.V., Chandrasegaran, S., Cai, Y., et al. (2017). Design of a synthetic yeast genome. Science *355*, 1040–1044.

Saffery, R., Wong, L.H., Irvine, D.V., Bateman, M.A., Griffiths, B., Cutts, S.M., Cancilla, M.R., Cendron, A.C., Stafford, A.J., and Choo, K.H.A. (2001). Construction of neocentromere-based human minichromosomes by telomere-associated chromosomal truncation. Proc. Natl. Acad. Sci. USA *98*, 5705–5710.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Schneider, K.L., Xie, Z., Wolfgruber, T.K., and Presting, G.G. (2016). Inbreeding drives maize centromere evolution. Proc. Natl. Acad. Sci. USA *113*, E987–E996.

Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K., and Willard, H.F. (2001). Genomic and genetic definition of a functional human centromere. Science *294*, 109–115.

Shang, W.-H., Hori, T., Toyoda, A., Kato, J., Popendorf, K., Sakakibara, Y., Fujiyama, A., and Fukagawa, T. (2010). Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. Genome Res. *20*, 1219–1228.

Sullivan, B.A. (2010). Optical mapping of protein–DNA complexes on chromatin fibers. In Fluorescence in Situ Hybridization (FISH): Protocols and Applications, J.M. Bridger and E.V. Volpi, eds. (Humana Press), pp. 99–115.

Tachiwana, H., Müller, S., Blümer, J., Klare, K., Musacchio, A., and Almouzni, G. (2015). HJURP involvement in de novo CenH3(CENP-A) and CENP-C recruitment. Cell Rep. *11*, 22–32.

Tolomeo, D., Capozzi, O., Stanyon, R.R., Archidiacono, N., D'Addabbo, P., Catacchio, C.R., Purgato, S., Perini, G., Schempp, W., Huddleston, J., et al. (2017). Epigenetic origin of evolutionary novel centromeres. Sci. Rep. *7*, 41980.

Topp, C.N., Okagaki, R.J., Melo, J.R., Kynast, R.G., Phillips, R.L., and Dawe, R.K. (2009). Identification of a maize neocentromere in an oat-maize addition line. Cytogenet. Genome Res. *124*, 228–238.

Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T.L., Adelson, D.L., Bailey, E., Bellone, R.R., et al.; Broad Institute Genome Sequencing Platform; Broad Institute Whole Genome Assembly Team (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. Science *326*, 865–867.

Wang, K., Wu, Y., Zhang, W., Dawe, R.K., and Jiang, J. (2014). Maize centromeres expand and adopt a uniform size in the genetic background of oat. Genome Res. *24*, 107–116.

Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., Gimelli, G., Warburton, D., Tyler-Smith, C., Sullivan, K.F., et al. (1997). Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr. Biol. *7*, 901–904.

Xie, C., and Tammi, M.T. (2009). CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics *10*, 80.