

# Analysis of Complex DNA Rearrangements during Early Stages of HAC Formation

Elisa Pesenti,\* Mikhail Liskovyykh, Koei Okazaki, Alessio Mallozzi, Caitlin Reid, Maria Alba Abad, A. Arockia Jeyaprakash, Natalay Kouprina, Vladimir Larionov, Hiroshi Masumoto, and William C. Earnshaw\*



Cite This: *ACS Synth. Biol.* 2020, 9, 3267–3287



Read Online

ACCESS |



Metrics & More



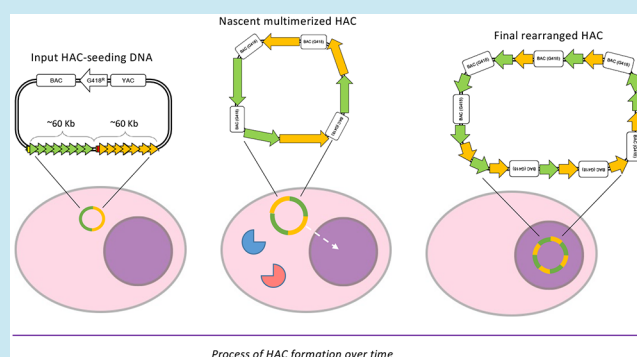
Article Recommendations



Supporting Information

**ABSTRACT:** Human artificial chromosomes (HACs) are important tools for epigenetic engineering, for measuring chromosome instability (CIN), and for possible gene therapy. However, their use in the latter is potentially limited because the input HAC-seeding DNA can undergo an unpredictable series of rearrangements during HAC formation. As a result, after transfection and HAC formation, each cell clone contains a HAC with a unique structure that cannot be precisely predicted from the structure of the HAC-seeding DNA. Although it has been reported that these rearrangements can happen, the timing and mechanism of their formation has yet to be described. Here we synthesized a HAC-seeding DNA with two distinct structural domains and introduced it into HT1080 cells. We characterized a number of HAC-containing clones and subclones to track DNA rearrangements during HAC establishment. We demonstrated that rearrangements can occur early during HAC formation. Subsequently, the established HAC genomic organization is stably maintained across many cell generations. Thus, early stages in HAC formation appear to at least occasionally involve a process of DNA shredding and shuffling that resembles chromothripsis, an important hallmark of many cancer types. Understanding these events during HAC formation has critical implications for future efforts aimed at synthesizing and exploiting synthetic human chromosomes.

**KEYWORDS:** centromere, human artificial chromosome, epigenetic engineering, mitosis, kinetochore, CENP-A



Human artificial chromosomes (HACs) are nonessential mini-chromosomes that replicate and segregate correctly in human cells.<sup>1</sup> HACs made using synthetic centromeric DNA have been extensively characterized and improved over the past decade.<sup>2–5</sup> They now represent an important tool to study epigenetic regulation of centromere structure and function,<sup>2,6–10</sup> to study full-length gene functions in mutant animal or human cells,<sup>5,11–16</sup> to measure chromosome instability (CIN), and to identify new targets for cancer therapy.<sup>17–19</sup> Also, synthetic HACs do not interfere with embryogenesis in mice, making them a promising tool for future gene therapeutic studies.<sup>15</sup> Synthetic HACs were originally designed using a “bottom up” approach to contain only predefined DNA arrays.<sup>1,20,21</sup> Their design allows the HAC centromeres to be easily modified and inactivated/removed by targeting with chimeric proteins specifically directed to the synthetic DNA.<sup>2,4,11,14</sup> However, they also present a number of challenges that must be overcome to enable them to be exploited fully.

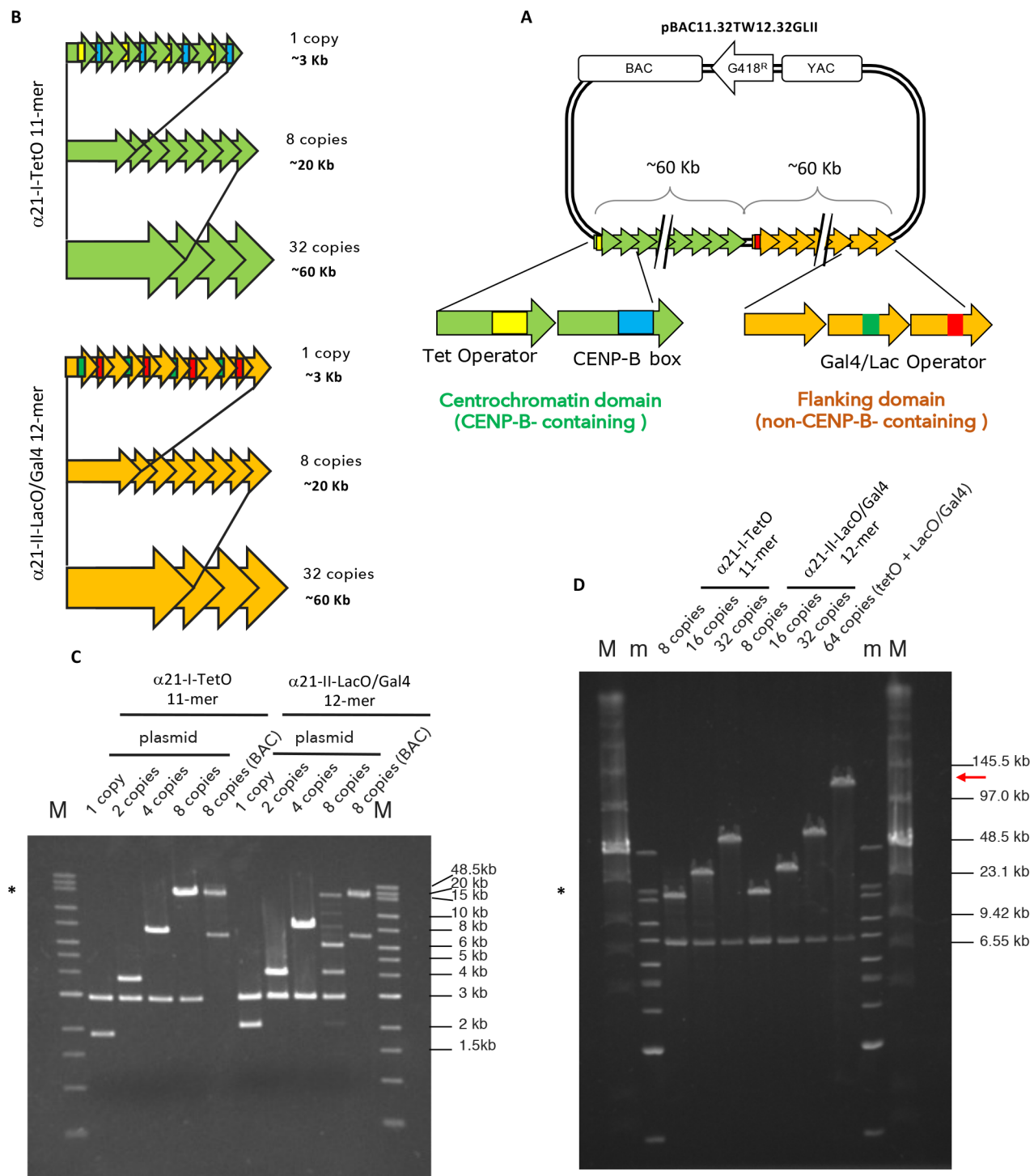
HACs form in a complex and as-yet incompletely characterized process after transfection of seeding DNA into human cells. During this relatively lengthy period (typically ~3

weeks), the HAC-seeding DNA undergoes spontaneous multimerization and in at least some cases may pass through a stage when it is transiently inserted into the arm of an endogenous chromosome.<sup>3,22</sup> The multimerization presumably allows it to attain a threshold size required for stable chromosome segregation.<sup>23</sup> Only the alphoid<sup>tetO</sup> HAC has previously been characterized in molecular detail. That analysis found complex rearrangements in the organization of its seeding DNA during this multimerization process, including inversions and deletions.<sup>3</sup> These rearrangements are unpredictable and uncontrollable as they occur during the clonal expansion before HAC-bearing cell lines are established and identified. This is the time during which the HAC-seeding DNA is forming a functional centromere, an absolute

Received: June 19, 2020

Published: December 8, 2020

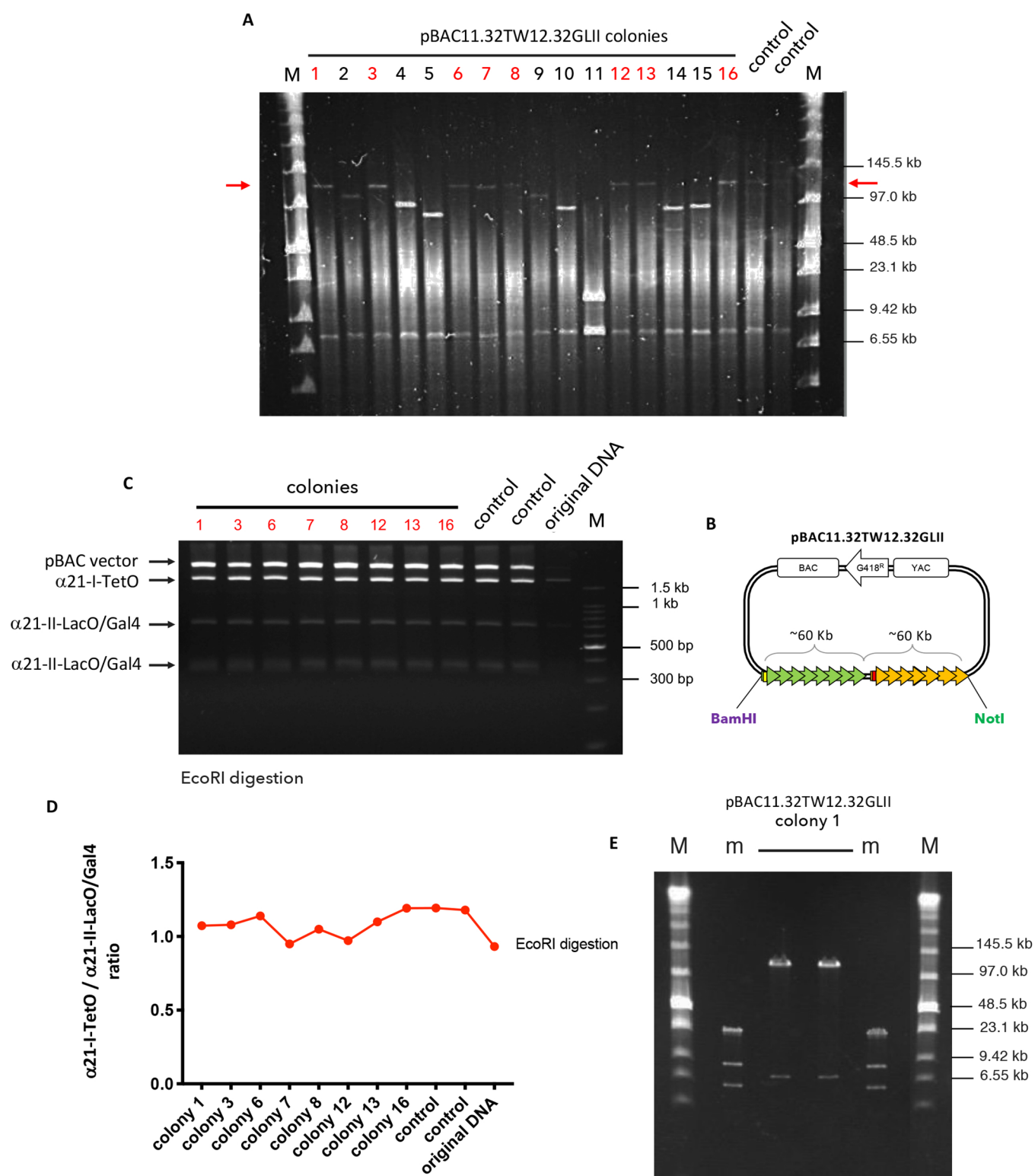




**Figure 1.** Generation of synthetic  $\alpha 21$ -I<sup>TetO</sup> and  $\alpha 21$ -II<sup>LacO/Gal4</sup> arrays. (A) Scheme of the pBAC11.32TW12.32GLII containing BAC and YAC cassettes, G418 resistance cassette, and synthetic DNA:  $\alpha 21$ -I<sup>TetO</sup> formed by high ordered repeats (HOR) monomers (green arrows) containing CENP-B boxes (blue) alternating with monomers containing TetO (yellow);  $\alpha 21$ -II<sup>LacO/Gal4</sup> formed by high ordered repeats (HOR) monomers (yellow arrows) containing Gal4 binding sequence (green) alternating with LacO (red). (B) Schematic of the assembly of the  $\alpha 21$ -I<sup>TetO</sup> and  $\alpha 21$ -II<sup>LacO/Gal4</sup> arrays. (C,D) PFGE analysis of the nascent  $\alpha 21$ -I<sup>TetO</sup> and  $\alpha 21$ -II<sup>LacO/Gal4</sup> arrays, cut with BamHI/NotI after each cycles of tandem ligation array amplification as described in Figure S2A (C) and Figure S2B (D). Expected sizes:  $\alpha 21$ -I<sup>TetO</sup> 11-mer 1 copy (1.9 kb), 8 copies (15.2 kb), 32 copies (60.8 kb);  $\alpha 21$ -II<sup>LacO/Gal4</sup> 12-mer 1 copy (2 kb), 8 copies (16 kb), 32 copies (64 kb). Plasmid vector is 2.9 kb, BAC vector is 7.1 kb. The asterisk (\*) indicates the fragments that have been cloned into BAC vector (8 copies, 16 kb); red arrow in D indicates the size of the final pBAC11.32TW12.32GLII (~120 kb) (m and M, markers).

requirement for the DNA to be stably maintained in cells. As a result, each HAC-bearing cell clone obtained after the selection process contains a HAC that could potentially have a DNA

organization different from its sister clones. Although it is known that rearrangements can happen,<sup>3</sup> how and when they occur remains unknown. Importantly, it is not known if these



**Figure 2.** Formation of input pBAC11.32TW12.32GLII DNA. (A) CHEF analysis of 16 bacterial DNA after transformation with pBAC11.32TW12.32GLII and NotI and BamHI digestion: red arrows indicate the size of the final vector (~120 kb); colonies labeled in red contain the inset of the desired length. DNA used for transfection as a control (in duplicate) (M marker). (B) Scheme of the pBAC11.32TW12.32GLII input DNA showing restriction sites for NotI and BamHI used to release the synthetic DNA. (C) PFGE analysis of selected bacterial colonies (in red) digested with EcoRI: each fragment originates from a different array (label on the left). DNA used for transfection as a control (in duplicate); original DNA as uncut sample (M marker). (D)  $\alpha 21$ -I-TetO and  $\alpha 21$ -II-LacO/Gal4 DNA ratio calculated with ImageJ on the intensity of the bands shown in C for each bacterial colony. Control and original DNA as in C. (E) CHEF analysis of bacterial colony #1 DNA (in duplicate) digested with NotI and BamHI to release the synthetic DNA (m and M, markers).

structural rearrangements are an obligate part of the selection process that occurs during centromere formation.

To investigate further how and when rearrangements happen, here we developed a novel alphoid<sup>2domain</sup> HAC, based on the structure of the alphoid<sup>tetO</sup> HAC. That first synthetic HAC was constructed from centromeric DNA with a dimeric structure.<sup>2</sup> One monomer of the alphoid<sup>tetO</sup> HAC was derived from the centromere of chromosome 17 type I  $\alpha$ -satellite DNA containing a CENP-B box. The other monomer was wholly synthetic alphoid DNA and carried a Tet operator (TetO) in place of the CENP-B box. The CENP-B box is found on all human chromosomes (except the Y) and is a 17 bp sequence recognized by the protein CENP-B.<sup>24,25</sup> This protein's function is still under investigation, but CENP-B binding is required for stable deposition of the centromeric histone H3-variant CENP-A when HAC-seeding DNA is introduced into cells.<sup>26–29</sup> The presence of the TetO sequence on the synthetic HAC allows the centromeric DNA to be targetable with chimeric Tet repressor (TetR)-fusion proteins that can manipulate the chromatin environment of the centromere and therefore modify the behavior of the HAC centromere.

Here, we have performed the first systematic study of rearrangements that occur during HAC formation and determined how they alter the epigenetic landscape in the HAC centromere and how they impact HAC segregation in mitosis. The seeding DNA of the new alphoid<sup>2domain</sup> HAC resembles that used to construct the alphoid<sup>hybrid</sup> HAC described earlier<sup>4</sup> but was much larger, as we hypothesized that this might minimize the need for rearrangements and amplification during HAC formation. The alphoid<sup>2domain</sup> HAC contains a CENP-B-containing centrochromatin array and a non-CENP-B-containing domain. The presence of two different domains allows simultaneous targeting of centromeric and flanking regions with different (TetR, LacI, and Gal4) fusion proteins and also makes it possible to track rearrangements within and between the arrays. Using this new alphoid<sup>2domain</sup> HAC, we demonstrate that dramatic DNA rearrangements can occur early during HAC formation and that once formed, they are stably maintained across many cell generations. Thus, a time-limited disruptive event of DNA shredding and shuffling, possibly involving a process resembling chromothripsis,<sup>30–32</sup> can occur early during centromere establishment in human cells.

## RESULTS

**Generation of Synthetic  $\alpha 21$ -I<sup>TetO</sup> and  $\alpha 21$ -II<sup>LacO/Gal4</sup> Arrays Using Tandem Ligation Array Amplification.** The alphoid<sup>2domain</sup> HAC is formed by two arrays, similarly to a previously constructed alphoid<sup>hybrid</sup> HAC,<sup>4</sup> but using a  $\sim 2.5\times$  larger ( $\sim 120$  kb) HAC-seeding construct. The CENP-B-containing centrochromatin array was designed to resemble the previously published alphoid<sup>tetO</sup> HAC,<sup>2</sup> but in this case, using 11-mer (1886 bp) high order repeats (HORs) of alphoid type I DNA from the centromere in human chromosome 21. Each monomer of this synthetic HOR contains either a 17 bp CENP-B box, essential for CENP-A deposition,<sup>29,33</sup> or a 39 bp tetracycline operator (TetO) targetable sequence, which is the binding site for *E. coli* tetracycline repressor (TetR). This dimer is the basic unit for the so-called  $\alpha 21$ -I<sup>TetO</sup> (TetO) array, which consists of alternating CENP-B-containing and TetO (non-CENP-B)-containing monomers (Figures 1A, S1A).

The other, non-CENP-B-containing, array is comprised of repeated segments of  $\alpha$ -satellite type II DNA, lacking CENP-B boxes. In endogenous chromosomes these sequences form the pericentromeric heterochromatin flanking the centromere. To allow targeting of this non CENP-B-containing array with different fusion proteins, LacO and Gal4-targetable sequences were embedded in the array, as previously described.<sup>4</sup> This allows its targeting by chimeric fusions to either *E. coli* lactose repressor (LacI) and/or the yeast Gal4 protein. We refer to this non-CENP-B-containing array as the  $\alpha 21$ -II<sup>LacO/Gal4</sup> (LacOGal4) array (Figure 1A and Figure S1A).

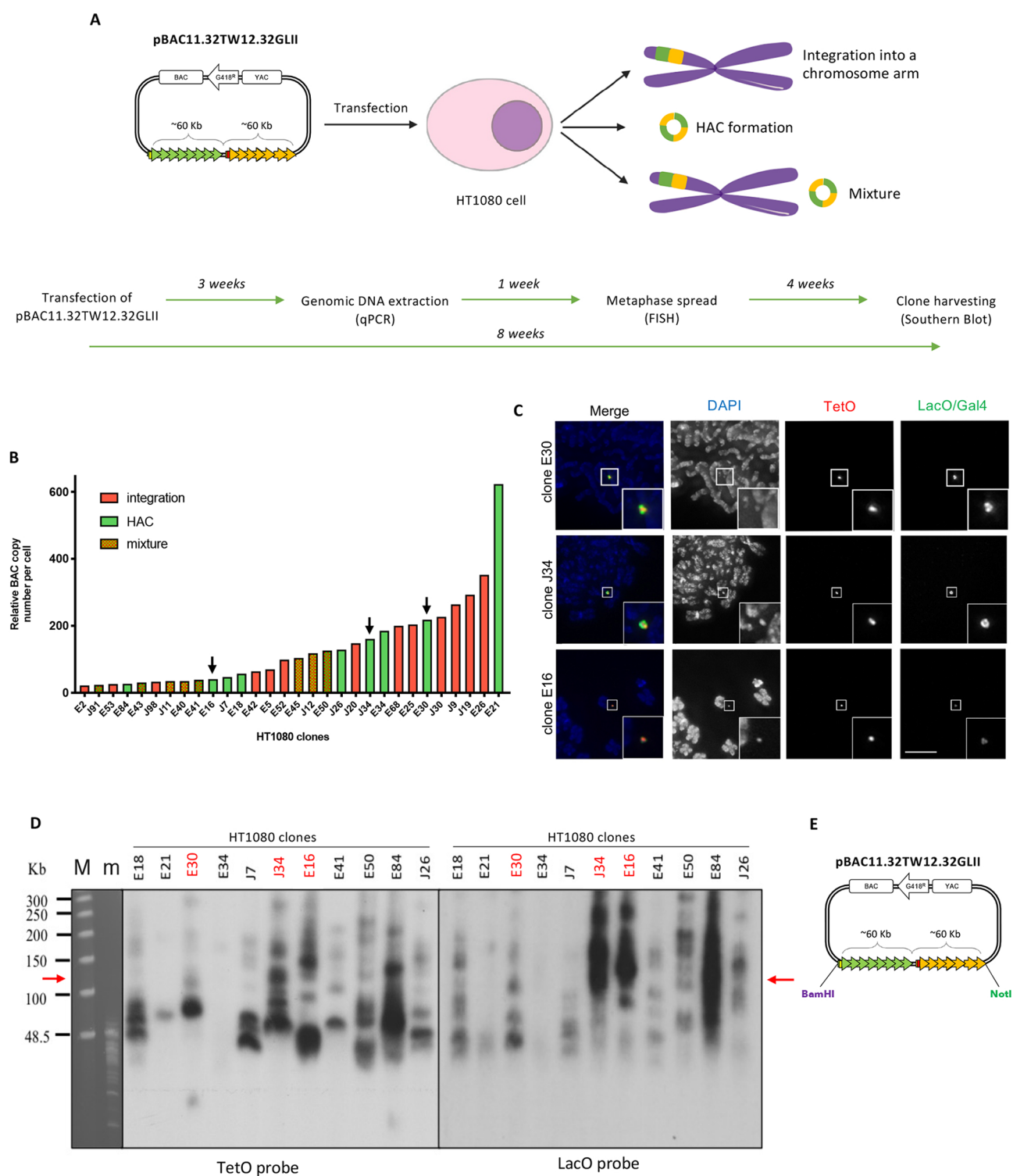
Our initial cloning efforts yielded a  $\alpha 21$ -I<sup>TetO</sup> 11-mer (1886 bp) and a  $\alpha 21$ -II<sup>LacO/Gal4</sup> 12-mer (2068 bp) in a plasmid backbone (Figure S1B,C). Each basic unit of this 11-mer or 12-mer was then elongated by tandem-ligation-amplification until fragments containing 8 copies were obtained (Figure 1B and Figure S2A). In this tandem-ligation-amplification, cycles of restriction enzyme digestion were performed and followed by ligation as shown in Figure S2A. Upon each cycle of ligation, the restriction site joining the two units was lost, so the next digestion occurred without cutting the nascent elongating array. In this case, cycles of SpeI/ScaI and NheI/ScaI digestions were performed (Figure S2A). After each round of restriction digestion and ligation, the nascent DNA was cut with BamHI and NotI, in order to separate the inset from the 2.9 kb vector, and subsequently analyzed by agarose gel electrophoresis (Figure 1C). Ultimately, the highest molecular weight band (16.6 kb for 8 copies, marked with \*) was excised and cloned into a BAC vector capable of more stably maintaining longer repetitive sequences. The structure of the BAC vector is shown in Figure S1D.

Starting from a BAC clone carrying 8 copies of the 11-mer and 12-mer, the tandem-ligation-amplification process was repeated until the insets reached 32 copies ( $\sim 60$  kb). To do that, cycles of SpeI/KasI and NheI/KasI digestions were performed (Figure S2B). As before, the nascent array was cut with BamHI and NotI after each reaction of restriction digestion and ligation to separate the inset from the 7.1 kb BAC vector and analyzed by agarose gel electrophoresis (Figure 1D).

Ultimately, the two complete  $\alpha 21$ -I<sup>TetO</sup> and  $\alpha 21$ -II<sup>LacO/Gal4</sup> arrays ( $\sim 60$  kb each) were joined together by tandem ligation into a pBAC vector containing a G148 resistance gene (Figure 1A,D and Figure S2B). As a result, we obtained the HAC-seeding construct pBAC11.32TW12.32GLII, carrying 32 copies of the  $\alpha 21$ -I<sup>TetO</sup> 11-mer and 32 copies of the  $\alpha 21$ -II<sup>LacO/Gal4</sup> 12-mer, with a total length of  $\sim 120$  kb (Figure 1D, red arrow). This input DNA was then amplified in bacteria prior to transfection into human HT1080 fibrosarcoma cells for HAC formation.

**Isolation of Input pBAC11.32TW12.32GLII DNA with Equal Amounts of  $\alpha 21$ -I<sup>TetO</sup> and  $\alpha 21$ -II<sup>LacO/Gal4</sup> Repeats.** To amplify the HAC-seeding construct, pBAC11.32TW12.32GLII DNA was electroporated into *E. coli* DH10B and the size of the array was determined by CHEF (contour-clamped homogeneous electric field) gel electrophoresis. In all, 16 BACs isolated from different bacterial clones were obtained from large scale bacterial cultures and digested with NotI and BamHI to release the HAC-seeding array from the BAC backbone (Figure 2A, B). Gel electrophoresis revealed that 8 out of 16 colonies (labeled in red in Figure 2A) maintained the original  $\sim 120$  kb length of the synthetic BAC DNA (Figure 2A, red arrow).





**Figure 3.** Screening of HT1080 colonies after transfection with pBAC11.32TW12.32GLII. (A) Scheme showing the possible fates of the pBAC11.32TW12.32GLII HAC seeding DNA after transfection in HT1080: in yellow and green (as integration or HAC) is represented the synthetic DNA. Timeline of the experiments performed from transfection into HT1080 cells. (B) BAC copy number (y axis) analyzed by qPCR in each HT1080 clone (x axis): only HT1080 clones containing >20 BAC copies are represented in the graph. HT1080 clones are represented in green (HAC), red (integration), or mixture (both) according to the results of the FISH screening, as shown in C. Black arrows indicate the clones shown in C and analyzed further. (C) Representative pictures of oligo-FISH staining of HT1080 clones: slides have been hybridized with DNA probes (TetO-dig/rhodamine  $\alpha$ -dig antibody, Gal4-biotin and LacO-biotin/Fitc-streptavidin). DAPI stains DNA. Scale bar = 10  $\mu$ m. (D) Southern blot of selected HT1080 clonal DNA (as labeled on top of the panel) digested with BamHI and separated by CHEF; the transferred membrane was hybridized with radioactively labeled TetO (left) or LacO (right) specific probes. Red arrows indicate the expected size of the band without rearrangements. Clones labeled in red have been screened further (M and m, markers). (E) Cartoon of the pBAC11.32TW12.32GLII input DNA showing restriction sites for NotI and BamHI.

To establish a HAC with an equal amount of CENP-B-containing and non-CENP-B-containing chromatin, it was important to confirm that the HAC-seeding pBAC11.32TW12.32GLII DNA carries an equal number of  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays. To do so, we digested the input DNA with EcoRI, a restriction enzyme that cuts both the arrays and the BAC vector (Figure S1B,C,D; only single-cut restriction enzymes are shown). We predicted *in silico* how the EcoRI restriction pattern should look (Figure S1E) and whether each band originates from the  $\alpha 21\text{-I}^{\text{TetO}}$  or the  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  array. The  $\alpha 21\text{-I}^{\text{TetO}}$  array cut with EcoRI should produce a fragment of 1880 bp, while the  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  array should produce fragments of 677, 370, 342, 340, and 339 bp. The vector yields a band of 7499 bp (Figure S1E).

The 8 colonies containing  $\sim 120$  kb BAC DNA were digested with EcoRI and the corresponding digests run on an agarose gel (Figure 2C). The results in Figure 2C match the prediction in Figure S1E. Analyzing the intensity of the corresponding bands on the agarose gel in Figure 2C using ImageJ, we scored the ratio between  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays. This confirmed that the BAC DNA contains equal amounts of  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  repeats (Figure 2D). Thus, BAC DNA from clone number 1 (number 1 of Figure 2A and 2C; Figure 2E, sample in duplicate) was chosen as our HAC-seeding input DNA for HAC formation in human cells.

**Screening of HT1080 Colonies Following Transfection with HAC-Seeding DNA.** HAC formation occurs following transfection of the HAC-seeding DNA into a suitable cell line, and colonies originating from single cells grow under selection. During the process, the input DNA is incorporated into the cell nucleus where it can undergo different fates: it can be integrated into a chromosome arm; it can form an autonomous HAC; the cell population can contain a mixture of both (Figure 3A); or, less frequently, the cells can acquire drug resistance but lose the remainder of the input DNA (not shown).<sup>34</sup> In order to form a HAC, the input DNA must multimerize to reach a threshold size for a stable chromosome.<sup>23</sup> This step occurs naturally after transfection and it is uncontrollable, leading to different levels of amplification of the input DNA within the cell. As a result, after transfection each single colony contains either a HAC, an integration, or a mixture of the two, with a different degree of amplification of the HAC-seeding DNA.<sup>34</sup>

As previously published,<sup>2,4</sup> we chose to transfect pBAC11.32TW12.32GLII into HT1080 cells. This fibrosarcoma cell line has a chromatin state permissive for HAC formation due to having a relatively low level of H3K9me3 as a result of decreased expression of Suv39h1 methyltransferase.<sup>10</sup> In cells with higher Suv39h1 expression, CENP-A assembles on HAC-seeding DNA, but is subsequently displaced by invading H3K9me3-containing heterochromatin.<sup>10</sup>

pBAC11.32TW12.32GLII from clone 1 (Figure 2E) was transfected into HT1080 cells and single cell clones were grown for 3 weeks in media containing Geneticin. We collected genomic DNA from 124 resistant colonies and measured the BAC copy number by qPCR to obtain an approximate measurement of the degree of amplification of HAC-seeding DNA. Primers specific for the  $\alpha 21\text{-I}^{\text{TetO}}$  HAC were designed and a different HAC with a known BAC copy number was used as standard.

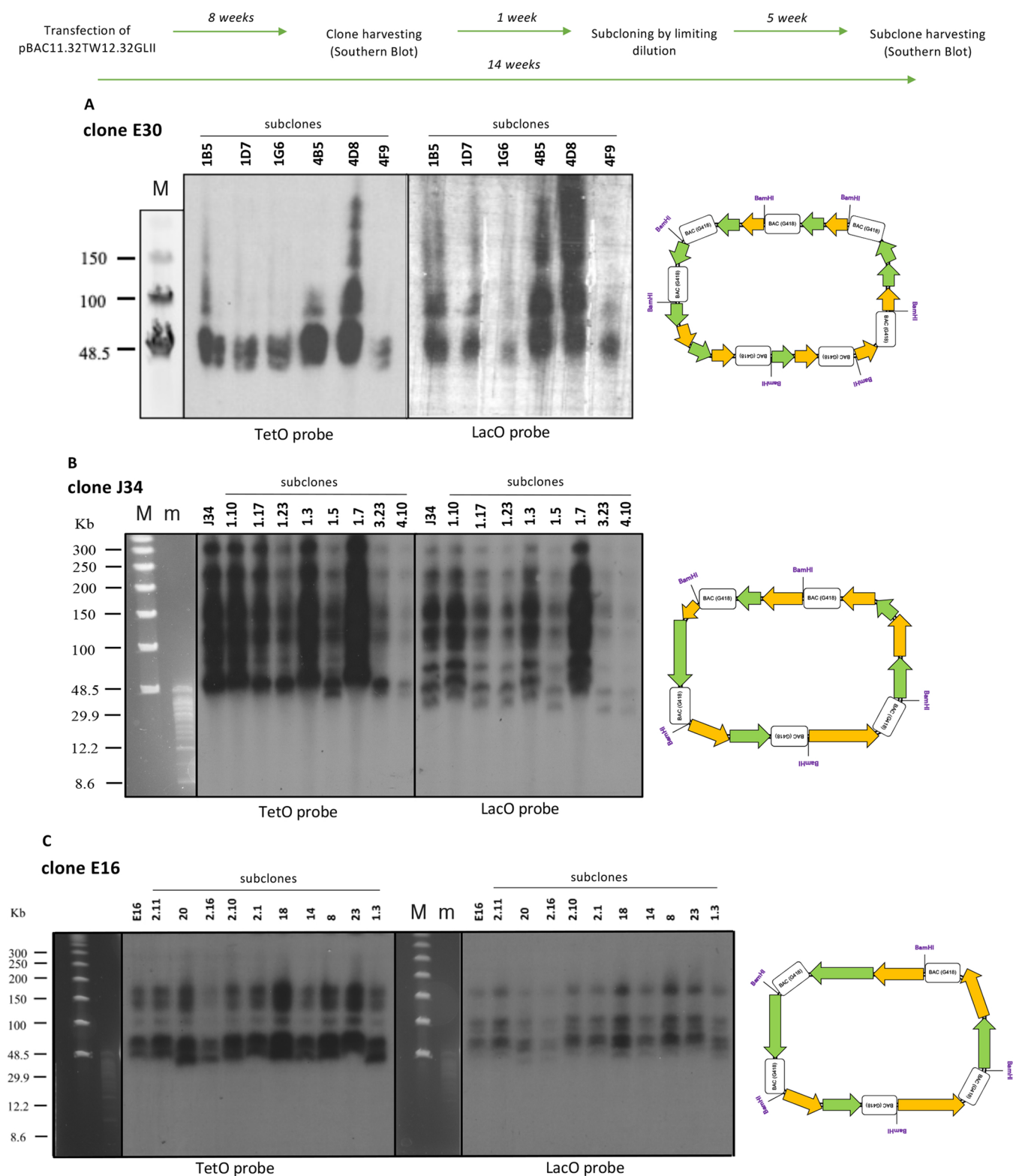
Thirty HT1080 colonies containing detectable amounts of HAC-seeding BAC sequences (copy numbers  $>20$ ), were then

screened by fluorescent *in situ* hybridization (FISH) for the presence of HACs, integrations, or mixtures of both (Figure 3B). FISH to detect pBAC11.32TW12.32GLII was performed 4 weeks after transfection (timeline in Figure 3) using TetO and LacO-specific oligos labeled with fluorochromes (see Methods for details). Figure 3B presents data of the qPCR analysis combined with the results of the FISH screening. Representative images from the FISH screening of selected HT1080 clones are shown in Figure 3C. HACs can be visualized as discrete spots by DAPI staining (Figure 3C). Interestingly, the size of each HAC estimated by FISH correlated the results of qPCR, with the larger HACs corresponding to higher BAC copy numbers and vice versa (Figure 3B black arrows, C). For simplicity, we discarded HAC-containing clones with more than 1 HAC.

In all HACs, the  $\alpha 21\text{-I}^{\text{TetO}}$  array seems to localize to the center of the HAC, where it is surrounded by the  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays (Figure 3C). This organization corresponds to that seen for previously published HACs,<sup>4</sup> and presumably reflects the HAC structure, with a CENP-B-containing centromere surrounded by pericentromeric heterochromatin.

**pBAC11.32TW12.32GLII Forms HACs More Efficiently than Previous HAC-Seeding Constructs.** In order to determine the fate of the HAC-seeding DNA, a minimum of 25 metaphases were screened by FISH for each clone. In the screening shown in Figure 3B, 30% (9/30) of HT1080 colonies contained only HACs, 43% (13/30) contained integrations, and 26.6% (8/30) contained a mixture of both. This frequency of HAC-containing colonies for the  $\alpha 21\text{-I}^{\text{TetO}}$  HAC is  $\sim 3$  times higher than in previous studies.<sup>2,4,33</sup> The relationship between the size of the HAC-seeding DNA and the efficiency of HAC formation is complex and may be influenced by the actual HAC-seeding sequences employed. Synthetic sequences seem to yield a lower efficiency of HAC formation than natural  $\alpha$ -HAC sequences. Examination of the literature reveals that the efficiency of HAC formation varies greatly depending on the type of  $\alpha$ -HAC centromeric DNA used for transfection (e.g., 32% HAC-forming colonies using DNA from chromosome 17 centromere, and only 4.3% of positive colonies using DNA from Y chromosome).<sup>35</sup> Ebersole and colleagues obtained a 10% efficiency of HAC colony formation using a  $\sim 120$  kb 5mer-based synthetic array.<sup>33</sup> In contrast the transfection with a  $\sim 120$  kb  $\alpha 21\text{-I}^{\text{TetO}}$  HAC-seeding DNA array yielded an efficiency of HAC formation of 4.3%.<sup>2</sup> In contrast, we observed an 11.7% efficiency of  $\alpha 21\text{-I}^{\text{TetO}}$  HAC formation using the much smaller  $\sim 60$  kb synthetic DNA.<sup>4</sup> In the present study, the efficiency of HAC formation by pBAC11.32TW12.32GLII (30%) was comparable to the frequency of HAC formation reported in our previous study when cells were stably cotransfected with the HAC seeding DNA plus CENP-A directed to the synthetic centromere.<sup>4</sup> It is possible that a combination of the structure and the size of the  $\alpha 21\text{-I}^{\text{TetO}}$  HAC DNA may increase the efficiency of CENP-A deposition on the centromeric DNA, although other factors cannot be ruled out.

**Analysis of Rearrangements of the pBAC11.32TW12.32GLII Arrays in HAC-Containing HT1080 Colonies.** We wished to determine whether the size amplification that occurred during early stages of  $\alpha 21\text{-I}^{\text{TetO}}$  HAC formation was also accompanied by rearrangements of the HAC-seeding DNA arrays. To perform a structural analysis of the  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays,



**Figure 4.** Analysis of rearrangements of the pBAC11.32TW12.32GLII arrays. Timeline of the experiments performed from transfection into HT1080 clones and subclones. (A,B,C) Southern blot of subclones from clone E30 (A), J34 (B) and E16 (C): DNA was digested with BamHI and separated by CHEF. The transferred membrane was hybridized with radioactively labeled TetO (left) or LacO (right) specific probes. Cartoons on the right represent the outcome of rearrangements in the corresponding clone: green arrows represent  $\alpha 21$ -T<sup>TetO</sup> array, yellow arrows represent  $\alpha 21$ -LacO/Gal4 array, boxes represent the BAC vector (M and m, markers).

we performed Southern blot analysis using TetO and LacO specific probes (Figure 3D). Cell clones were grown for 8 weeks, or approximately 50 doublings of the HT1080 cells,

prior to Southern blot analysis (timeline in Figure 3). Genomic DNA from 9 HAC-containing cell lines (and two HT1080 clones containing integrations as control) was digested with



BamHI, which has a unique site only on the vector backbone (Figure 3E). The DNA fragments were separated by CHEF gel electrophoresis and the membranes hybridized with TetO and LacO-specific probes.

If only simple multimerization of pBAC11.32TW12.32GLII occurred during HAC formation, the Southern blots should display a single ~120 kb band, corresponding to the size of digested input DNA. Importantly, since the restriction enzyme cuts only at the edge of the  $\alpha$ -satellite arrays, this should be the case regardless of whether the Southern blot analysis uses the TetO or LacO probe. Surprisingly, none of the analyzed clones showed this single ~120 kb band (Figure 3D, red arrow). Instead, each clone has a different number of DNA fragments of different sizes, and these also vary for each clone for the two probes. Many bands are smaller than the 120 kb input band, but some are considerably larger. Thus, the arrays of the HAC-seeding pBAC11.32TW12.32GLII DNA underwent a complex series of rearrangements during HAC formation, as described for the  $\text{alphoid}^{\text{TetO}}$  HAC.<sup>3</sup>

On the basis of the results displayed in Figure 3, we decided to further characterize three clones (E30, J34 and E16) that showed different levels of amplification by qPCR (black arrows in Figure 3B) and different numbers and sizes of rearrangements by Southern blotting (labeled in red in Figure 3D).

**pBAC11.32TW12.32GLII Undergoes Multiple Rearrangements during Early Stages of HAC Formation.** The rearrangement of HAC-seeding DNA was previously described for the single-domain  $\text{alphoid}^{\text{TetO}}$  HAC.<sup>3</sup> It was proposed that the HAC-seeding DNA structure may continue to change and evolve for weeks or possibly months after HAC transfection. This raises the possibility that the populations analyzed in Figure 3D might consist of mixtures of  $\text{alphoid}^{2\text{domain}}$  HACs with different structures. To test this hypothesis, the three clones E30, J34 and E16 were further subcloned to obtain homogeneous cell populations (timeline in Figure 4; 9 weeks or approximately 55 population doublings after transfection). Initially clone E21 was also subcloned but, unfortunately, we could not grow any subclone with stable HAC segregation, so E21 was excluded from the subsequent analysis.

$\text{Alphoid}^{2\text{domain}}$  HAC subclones were isolated by limiting dilution and screened by FISH for the presence and the number of the HACs in each clonal cell line (12 subclones for E30, 35 subclones for J34 and 25 subclones for E16; detailed timeline in Figure S3). Subclones with highly mis-segregating HACs were discarded and subclones with a higher percentage of single HACs per cell were studied further (6 subclones for E30, 8 subclones for J34 and 10 subclones for E16; Figure 4). As an example of the screening, Figure S3A shows representative images from FISH screening of 6 subclones from clone E30. Figure S3B shows the number of metaphases containing 0, 1, or 2 HACs in subclones from clone E30, with percentages indicating the cells with 1 HAC.

To study rearrangements in subsequent cell generations, genomic DNA from the selected subclones originating from E30, J34 and E16 was digested with BamHI and separated by CHEF gel electrophoresis (Figure 4). Surprisingly, Southern blot analysis revealed that all subclones were almost identical in the number and sizes of rearrangements. Furthermore, they all recapitulated the pattern of rearrangements seen in the original clone (Figure 4A–C). As for the original clones, hybridization with the TetO-specific probe consistently yielded

a different hybridization pattern from that seen with the LacO-specific probe on the same sample.

To strengthen our hypothesis that the HAC-seeding DNA undergoes a series of multiple rearrangements, we decided to investigate the structure of the  $\text{alphoid}^{2\text{domain}}$  HAC using FISH on DNA fibers (Figure S4). We hybridized stretched DNA from clone E30 (subclone 1B5), clone J34 (subclone 1.10) and clone E16 (subclone 23) with TetO and LacOGal4 specific probes. Images reveal that there is not a regular alternation of TetO and LacOGal4 spots on DNA fibers, as we would expect if the alternating TetO/LacO structure of the HAC-seeding DNA had been maintained. Instead, TetO and LacOGal4 spots on DNA fibers show various size and patterns, confirming that rearrangements occur during HAC formation (Figure S4).

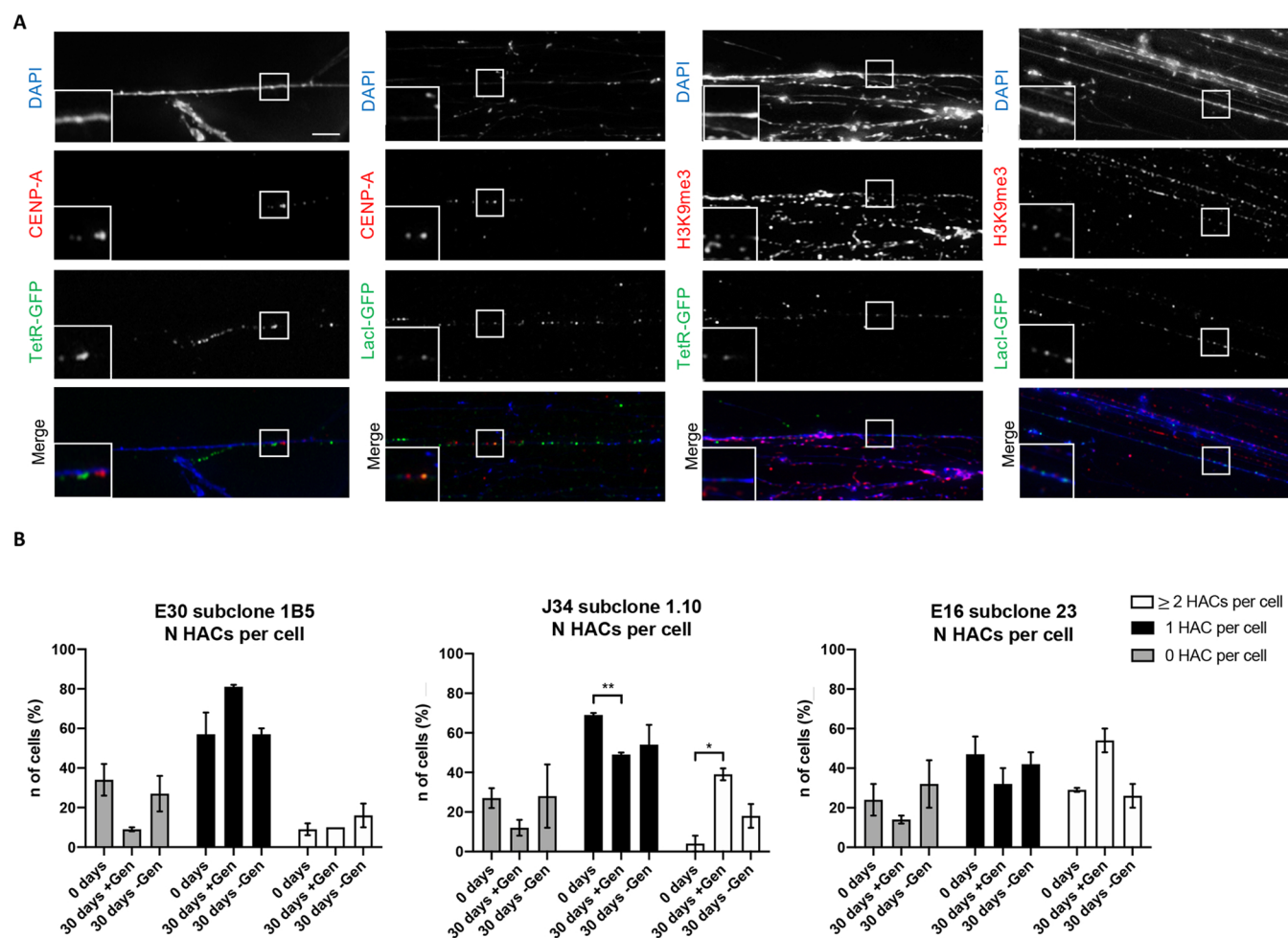
Taken together, these data show that the  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays of HAC-seeding DNA pBAC11.32TW12.32GLII independently undergo unique fragmentation, recombination and amplification events during the first 8 weeks of the  $\text{alphoid}^{2\text{domain}}$  HAC formation. Subsequently, these rearrangements appear to be maintained stably through cell generations, up to 14 weeks. This agrees with the observation that the HAC structure seems to be stable through multiple cycles of MMCT (microcell mediated chromosome transfer) in different cell lines.<sup>3,36</sup>

**Different HAC-Containing Clones Show Different Degrees of Rearrangements.** Not all the  $\text{alphoid}^{2\text{domain}}$  HAC clones analyzed underwent the same degrees of rearrangements. For example, subclones from clone E30 (Figure 4A) display a predominant band around 50 kb in all subclones, and only 2 of the 6 subclones clearly show fragments around 80–100 kb with both probes (subclones 4B5 and 4D8). In contrast, the LacO probe shows 3–4 bands with various signal strengths. This could reflect dimerization/multimerization of the 50 kb fragment observed in the blot. Thus, compared with the other clones analyzed, E30 seems to be less scrambled, with all fragments showing the same size (cartoon in Figure 4A). It therefore appears that during E30 HAC formation pBAC11.32TW12.32GLII underwent an early event in which both the  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays were shortened to roughly 40% of their initial lengths (easiest to imagine if a single deletion of the 120 kb construct occurred spanning the junction between the two arrays), but then were amplified while avoiding further rearrangements.

In marked contrast, clones J34 and E16 displayed a much larger number and variety of rearrangements, with fragments ranging from ~50 kb up to ~300 kb (Figure 4B, C and relative diagrams on the right). One possible explanation for this structure is that early during formation of those two HACs and following some initial amplification of the arrays, the nascent HAC-seeding DNA experienced multiple chromosome breaks and shuffling followed by religation, leading to fragments of different sizes.

Interestingly, the smaller array size in clone E30 (Figure 4A) correlates with its larger number of BAC copies (presumably resulting from a larger number of amplification cycles) quantitated by qPCR (Figure 3B). In contrast, J34 and E16 with more complex rearrangements, including those producing much larger fragments (Figure 4B,C), have lower BAC copy numbers (Figure 3B). These observations suggest that the rearrangements may have occurred very early, prior to completion of the multimerization that allowed the HAC to pass the minimum size threshold required to form a stable centromere/kinetochore.<sup>23</sup>





**Figure 5.** Visualization of  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays on chromatin fibers and mitotic stability of the HAC. (A) Representative pictures of IF staining on fibers of subclone J34 1.10: slides have been incubated with TetR-eYFP and LacI-eYFP expressed in *E. coli* and stained with  $\alpha$ -CENP-A mouse or  $\alpha$ -H3K9me3 mouse/TRITC  $\alpha$ -mouse antibody. DAPI stains the DNA. Scale bar = 5  $\mu\text{m}$ . (B) Number of metaphases (%) containing 0, 1, or  $\geq 2$  (2 or 3) HACs for subclones E30 1B5, J34 1.10 and E16 23, after spreading metaphases and hybridizing with TetO and LacO/Gal4 specific probes; total of 2 biological repeats, 50 metaphases for each condition were analyzed.  $\pm$  Gen indicates treatment for 30 days from day 0 with (+) or without (-) Geneticin. Error bars denote SEM. Statistical test: unpaired *t* test (\**P* < 0.05, \*\**P* < 0.01).

These data show that, as previously suggested,<sup>3</sup> during aliphoid<sup>2domain</sup> HAC formation, the predicted regular structure of the HAC-seeding DNA is disrupted by complex rearrangements whose mechanism remains unknown.

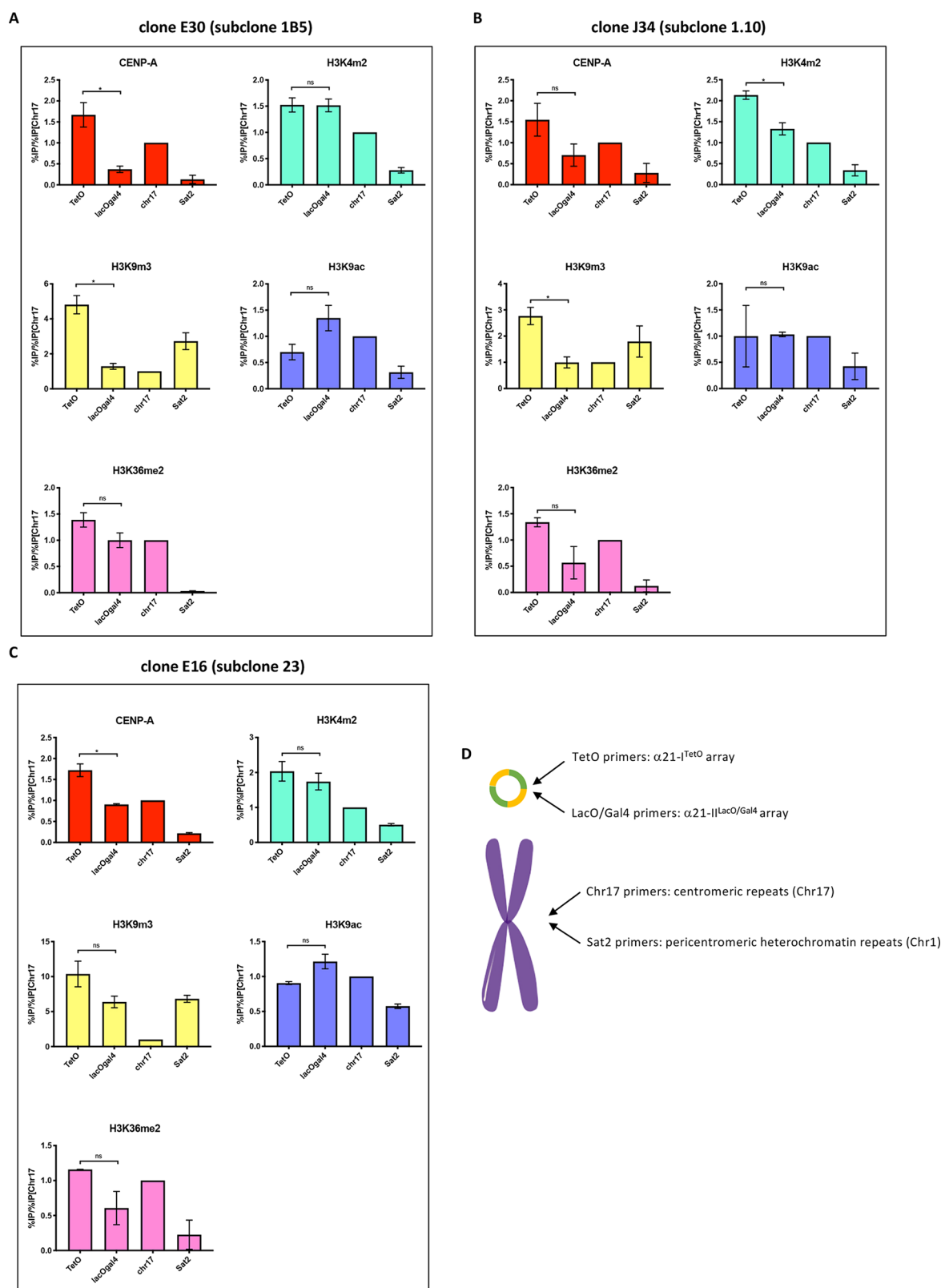
**Visualization of  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  Arrays on Chromatin Fibers.** We performed indirect immunofluorescence (IF) staining on stretched DNA fibers to confirm the presence of  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays, and to also determine the distribution of CENP-A and H3K9me3 on those fibers.

For each set of subclones, one was selected for further experiments, based on the percentage of cells bearing a single HAC (E30 subclone 1B5, J34 subclone 1.10 and E16 subclone 23; as example for clone E30, see Figure S3B). DNA fibers were prepared from the selected subclones and incubated with purified TetR-eYFP or LacI-eYFP fusion proteins expressed in *E. coli* to visualize the corresponding array (TetR-eYFP and LacI-eYFP expressed *in vivo*, both dissociate from the chromatin during fiber preparation) (Figure 5A and Figure S6). Staining of both arrays simultaneously was not possible, since the purified proteins were both tagged with GFP. Attempts to specifically stain fibers with mCherry-TetR

isolated from *E. coli* were not successful. Fibers were also stained using CENP-A or H3K9me3-specific antibodies (Figure 5A).

IF staining on fibers revealed that the  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays are both present along these stretched DNA fibers. CENP-A and H3K9me3 are adjacent to both arrays, with no apparent preference for one or the other array (representative images for J34 subclone 1.10 are shown in Figure 5A). The presence of CENP-A and H3K9me3 in close proximity to both arrays can be explained if the rearrangements during aliphoid<sup>2domain</sup> HAC formation lead to a “scrambled” structure of the  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays.

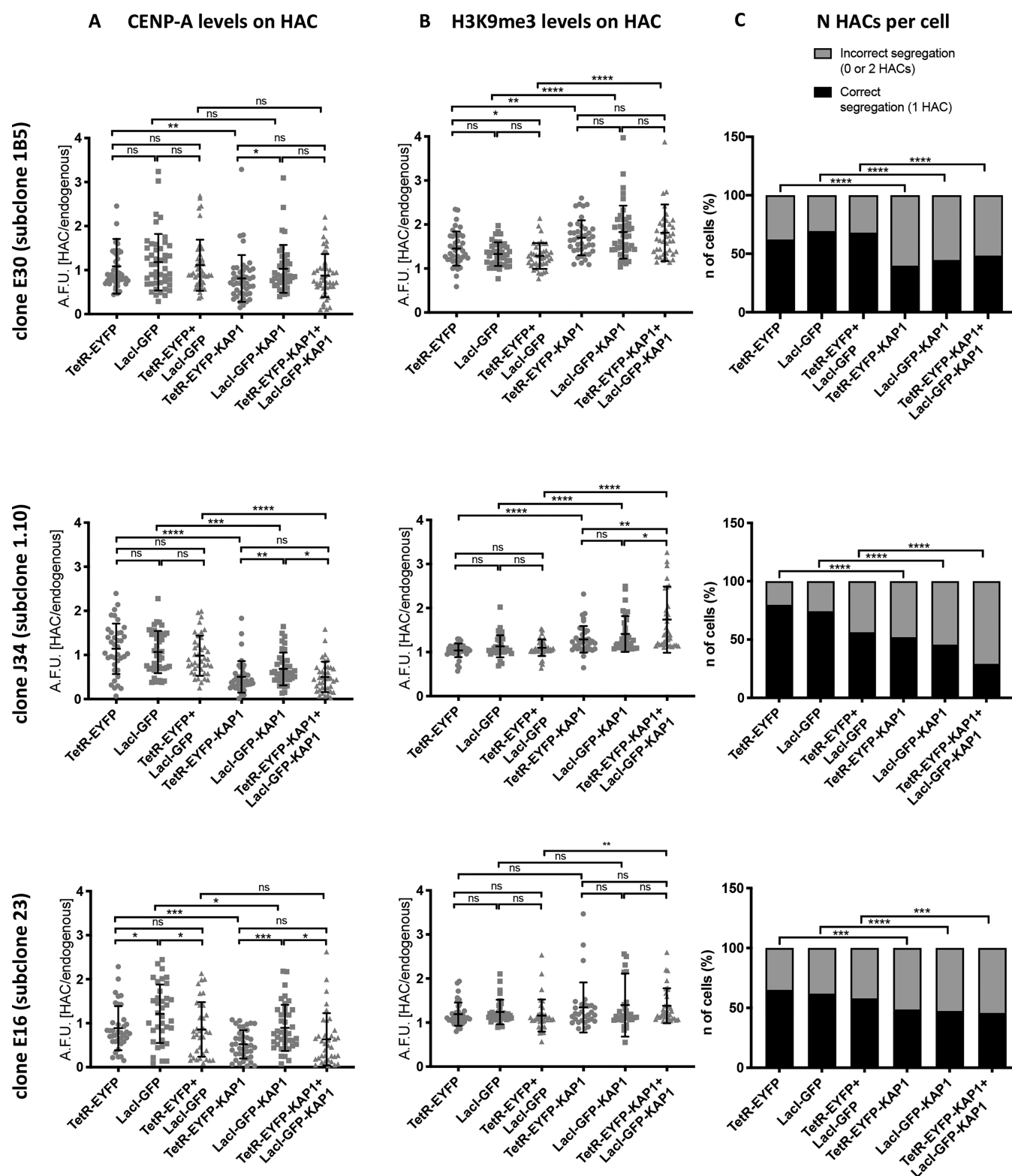
**Geneticin Selection Enriches the Number of HACs in the Cell Population.** To characterize how the mitotic stability of the aliphoid<sup>2domain</sup> HAC is affected by its structure, we performed a stability assay, counting the number of cells with different number of HACs over a period of 30 days with (+) and without (-) Geneticin (~25 cell divisions). Metaphase chromosome spreads from E30 subclone 1B5, J34 subclone 1.10 and E16 subclone 23 were analyzed by FISH and imaged at each time point using labeled oligos specific for



**Figure 6.** CENP-A accumulates preferentially on the  $\alpha 21\text{-I}^{\text{TetO}}$  array. (A,B,C) ChIP-qPCR analysis of CENP-A and indicated histone marks modifications in HT1080 subclones E30 1B5 (A), J34 1.10 (B), and E16 23 (C). The  $\alpha 21\text{-I}^{\text{TetO}}$  array (TetO),  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  array (LacOgal4), satellite D17Z1 (chr17), and degenerate satellite II (sat2) repeats were assessed. Values have been normalized against satellite D17Z1 (chr17). Total of 2 biological repeats,  $n \approx 5 \times 10^6$  cells each. Error bars denote SEM. Statistical test: Mann–Whitney test ( $*P < 0.05$ ). (D) Cartoon showing the localization of primers used in qPCR on the HAC and on endogenous corresponding chromosomes.

$\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$ , to count cells containing 0, 1, or  $\geq 2$  HACs.

At time point zero, each subclone had a characteristic number of cells with 0, 1, or  $\geq 2$  HACs (Figure 5B).



**Figure 7.**  $\alpha 21\text{-I}^{\text{TetO}}$  and  $\alpha 21\text{-II}^{\text{LacO/Gal4}}$  arrays are not functionally independent. (A,B) Quantification of HAC-associated CENP-A staining (A) and H3K9me3 staining (B) in individual cells of each indicated HT1080 subclone 48 h after transfection with the indicated fusion proteins; values plotted as A.F.U. (arbitrary forming units). Solid bars indicate the medians, and error bars represent the SD;  $n =$  two independent experiments for each staining;  $\sim 30$  cells analyzed in each experiment. Asterisks indicate a significant difference ( $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $****P < 0.0001$ ; Mann–Whitney test). (C) Number of interphase nuclei (%) showing correct segregation (1 HAC) or mis-segregation (0 or 2 HACs) of each indicated HT1080 subclone 48 h after transfection with the indicated fusion proteins; the presence of the HAC was detected by GFP signal.  $n = 500$  cells analyzed per condition ( $****P < 0.001$ ; Fisher's exact test).

Interestingly, 7 days after thawing and while growing in the presence of Geneticin (corresponding to time point zero), the

number of cells containing no HAC was higher than expected (20–40% of the cell population, gray bars) for all three

subclones. It is possible that after the stress of freezing/thawing, cells were not yet fully responding to the selection.

Importantly, all of the alphoid<sup>2domain</sup> HACs are extremely stable after 30 days growth in the absence of Geneticin, as shown by the number of cells bearing 1 HAC (black bars; daily loss rate of the HAC ( $R$ ) =  $4.76 (\pm 0.6) \times 10^{-3}$  for E30 subclone 1B5;  $8.69 (\pm 6.67) \times 10^{-3}$  for J34 subclone 1.10; and  $3.47 (\pm 1.66) \times 10^{-3}$  for E16 subclone 23; see [Methods](#)). These measured loss rates were comparable to those measured for previous HACs.<sup>2,4</sup> Thus, alphoid<sup>2domain</sup> HACs with dramatically different rearrangements of their seeding DNA are very similar in their ability to replicate and segregate correctly during the great majority of cell divisions.

Surprisingly, when cells grow for 30 days (+) Geneticin, they seem to acquire a selective advantage for increasing the HAC copy number, as shown by the number of cells bearing  $\geq 2$  HACs (white bars). The accumulation of HAC was particularly evident in J34 subclone 1.10 and E16 subclone 23, while E30 subclone 1B5 did not exhibit this increase ([Figure 5B](#)). Notably, the enrichment of cells with 1 or  $\geq 2$  HACs after 30 days (+) Geneticin was coupled for all subclones with a reduction in the number of cells with 0 HACs.

The enrichment in cells with  $\geq 2$  HACs could be explained if heterochromatin spreading silences the Geneticin resistance gene. In this case, Geneticin would select for cells in the population with an increased HAC copy number. Despite the small sample size, it is interesting to note that the alphoid<sup>2domain</sup> HACs with the more rearranged arrays (J34 1.10 and E16 23) were those where the copy number increased under selection, possibly indicating that the chromatin state is less stable. In contrast, the clone with the least rearranged structure (E30 1B5) showed the highest chromatin stability. Taken together these data suggest that the HAC DNA structure may have an impact on HAC chromatin stability over time.

**CENP-A Accumulates Preferentially on the  $\alpha 21$ -I<sup>TetO</sup> Array with CENP-B Boxes.** The IF staining on fibers in [Figure 5](#) shows CENP-A and H3K9me3 apparently localized on both the  $\alpha 21$ -I<sup>TetO</sup> and  $\alpha 21$ -II<sup>LacO/Gal4</sup> arrays. To better characterize the chromatin state of the two arrays on the HAC-seeding DNA, we performed chromatin immunoprecipitation (ChIP) for CENP-A and several indicative histone modifications using a set of well-characterized monoclonal antibodies,<sup>37</sup> followed by quantitative PCR (ChIP-qPCR) on genomic DNA from E30 subclone 1B5, J34 subclone 1.10 and E16 subclone 23 (scheme of the primers used for qPCR is presented in [Figure 6D](#)).

ChIP data were highly reproducible for the three subclones of the alphoid<sup>2domain</sup> HAC ([Figure 6A,B,C](#)). Thus, the overall chromatin organization was maintained despite differences in the level of rearrangements. CENP-A accumulated on the  $\alpha 21$ -I<sup>TetO</sup> array  $\sim 2$ – $3$  times more than on the  $\alpha 21$ -II<sup>LacO/Gal4</sup> array in all the three subclones. This is an average of  $\sim 1.5$  times more than on the endogenous centromere of chromosome 17, used as a control ([Figure 6A,B,C](#)). This contrasts with a previous study in which the alphoid<sup>hybrid</sup> HAC was apparently unable to maintain CENP-A only on the centromeric array.<sup>4</sup> It is possible that CENP-A deposition on the  $\alpha 21$ -I<sup>TetO</sup> array may be favored by the larger size of the input pBAC11.32TW12.32-GLII HAC-seeding DNA. CENP-A deposition also correlated with higher levels of H3K4me2 and H3K36me2, as expected for centrochromatin.<sup>7,9</sup> The  $\alpha 21$ -I<sup>TetO</sup> array also contained a relatively high level of H3K9me3 and a low level of H3K9ac

([Figure 6A,B,C](#)), revealing differences from the alphoid<sup>tetO</sup> HAC, which contained a single HAC-seeding array.<sup>2,7</sup>

Unexpectedly, the  $\alpha 21$ -II<sup>LacO/Gal4</sup> array had levels of H3K9ac and H3K4me2 (markers for actively transcribed chromatin)  $\sim$  twice those of the satellite II DNA used as a control. Consistent with this observation, levels of H3K9me3 on the  $\alpha 21$ -II<sup>LacO/Gal4</sup> array were  $\sim 2$ – $4$  times lower than on the  $\alpha 21$ -I<sup>TetO</sup> array ([Figure 6A,B,C](#)), revealing a generally open conformation of the chromatin. This was surprising, as we had initially expected this array, which lacks CENP-B boxes, to form heterochromatin. Our data suggest that a regular array of alphoid type II DNA lacking CENP-B boxes is not sufficient to establish pericentric heterochromatin. However, given the intermixing of sequences on the HAC-seeding DNA, we cannot exclude the possibility that the presence of strong heterochromatin might have been counter-selected due to its potentially harmful effects on expression of the Geneticin-resistance gene or to the juxtaposition with large numbers of CENP-B boxes.

Taken together these data suggest that  $\alpha 21$ -I<sup>TetO</sup> array recruits CENP-A and establishes a functional centromere in the alphoid<sup>2domain</sup> HAC, despite sustaining high levels of H3K9me3.

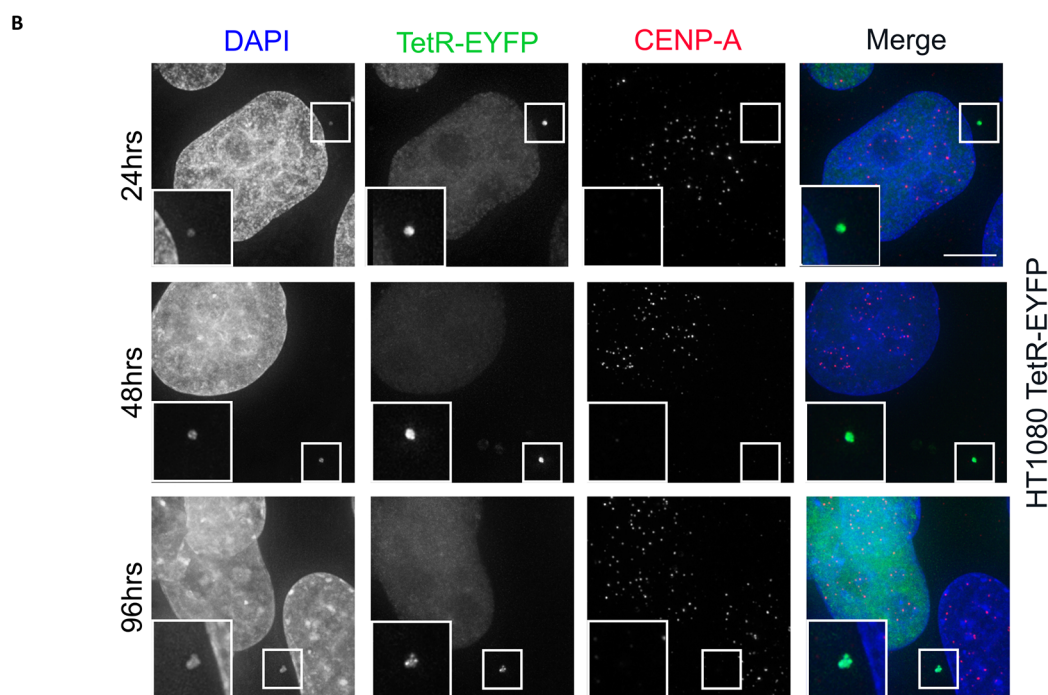
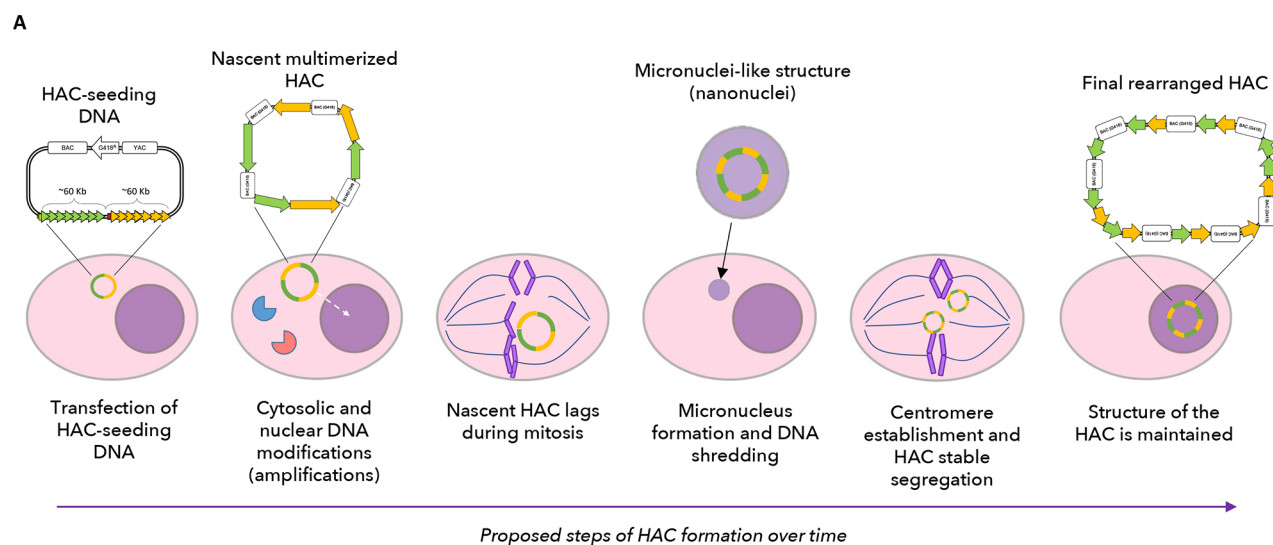
#### $\alpha 21$ -I<sup>TetO</sup> and $\alpha 21$ -II<sup>LacO/Gal4</sup> Arrays Do Not Form Functionally Independent Chromatin Domains.

To determine whether the molecular structure of the HAC impacts the function of the  $\alpha 21$ -I<sup>TetO</sup> and  $\alpha 21$ -II<sup>LacO/Gal4</sup> arrays, we asked whether the two arrays are functionally distinct. To do this, we transiently expressed KAP1 as a chimeric fusion to either TetR-eYFP or LacI-GFP. KAP1 is a scaffolding protein that recruits the CoREST complex, promoting a silent chromatin state and increasing the level of H3K9me3.<sup>38</sup> Previous studies revealed that KAP1 recruitment into the centromere causes a loss of CENP-A and inactivates the kinetochore.<sup>39</sup> Thus, if the two arrays on the alphoid<sup>2domain</sup> HAC are functionally independent, KAP1 recruitment should have an effect of the HAC centromere only when targeted to the  $\alpha 21$ -I<sup>TetO</sup> array.

We performed quantitative fluorescent analysis to measure the level of CENP-A and H3K9me3 on the alphoid<sup>2domain</sup> HAC after targeting KAP1-eYFP fusions to the two arrays both separately and simultaneously for 48 h, using the eYFP to localize the HAC arrays in interphase cells ([Figure S5A,B](#)). Targeting KAP1 to the  $\alpha 21$ -I<sup>TetO</sup> array led to a significant ( $\sim 2$  fold) decrease in CENP-A levels on the HAC for all three subclones analyzed ([Figure 7A](#)). This is similar to what was reported for the alphoid<sup>tetO</sup> HAC.<sup>6</sup> The decrease in CENP-A was accompanied by an increase in H3K9me3 levels when targeting KAP1 to the  $\alpha 21$ -I<sup>TetO</sup> array ([Figure 7B](#)). Different subclones showed different levels of H3K9me3 enrichment, possibly due to intrinsic variation in the H3K9me3 basal levels in each subclone ([Figure 6A,B,C](#)).

Targeting KAP1 to the  $\alpha 21$ -II<sup>LacO/Gal4</sup> array also resulted in a decrease in HAC-associated CENP-A, although the effect was milder than observed with tethering to the  $\alpha 21$ -I<sup>TetO</sup> array ([Figure 7A](#)). Thus, even though most CENP-A was associated with the  $\alpha 21$ -I<sup>TetO</sup> region, targeting proteins to the  $\alpha 21$ -II<sup>LacO/Gal4</sup> region still affected CENP-A levels. This confirms the proximity of the arrays and is consistent with the pattern of histone modifications observed in [Figure 5A](#). The increase in H3K9me3 levels seen after tethering KAP1 to  $\alpha 21$ -II<sup>LacO/Gal4</sup> array appeared to be more significant than tethering KAP1 to  $\alpha 21$ -I<sup>TetO</sup>. This could be explained by the initial lower level of





**Figure 8.** Early stages of  $\alpha$ loid<sup>2domain</sup> HAC formation. (A) Scheme representing how HAC-seeding DNA rearranges to form the  $\alpha$ loid<sup>2domain</sup> HAC. HAC-seeding DNA formed by  $\alpha$ 21-I<sup>TetO</sup> (green) and  $\alpha$ 21-II<sup>LacO/Gal4</sup> (yellow) arrays is transfected in HT1080 cells; under the effect of proteins of the DNA-sensing pathways (elements in red and blue), the nascent HAC gets initially rearranges and it increases in size due to “slippage” during replication (dashed arrow indicates nascent HAC entering the nucleus). During mitosis, the nascent HAC lags and it gets therefore incorporated into a micronucleus, resulting in additional massive rearrangements which lead to the final structure of the  $\alpha$ loid<sup>2domain</sup> HAC. (B) Indirect immunofluorescence of HT1080 cells constitutively expressing TetR-EYFP at indicated time points after transfection with HAC-seeding DNA. After fixation, cells were stained for CENP-A; DAPI stains the nuclei. Enlargements show nanonuclei containing HACs (green bright spots). Scale bar = 10  $\mu$ m.

H3K9me3 on the  $\alpha$ 21-II<sup>LacO/Gal4</sup> array (Figure 6A,B,C): there might be more unmodified H3K9 that can be converted to H3K9me3 upon the effect of KAP1.

Targeting KAP1 to both arrays simultaneously did not completely suppress kinetochore function as revealed by CENP-A levels, which are partly maintained. In the double tethering, neither CENP-A nor H3K9me3 levels differed greatly from the single tethering, rejecting the hypothesis that the arrays are independent and they would cooperate to

establish a state of “super-repression” when both targeted with KAP1 (Figure 7A,B).

In parallel with measuring the effects of KAP1 tethering on CENP-A and H3K9me3 levels, we also scored the effects of this tethering on centromere function (*e.g.*, HAC segregation in mitosis). Despite differences in levels of correctly or mis-segregating HACs in the initial cell populations, targeting KAP1 to one or both arrays always led to a significant increase in the number of mis-segregating HACs. Interestingly, segregation was significantly impaired even when overall

CENP-A levels were not greatly reduced by KAP1 (Figure 7C). This is probably because the percentage of mis-segregating cells is determined by scoring individual cells in which the level of CENP-A falls below a critical threshold, and it is not determined by the average CENP-A level in the cell population, as already described.<sup>40</sup>

Together, these observations lead to the conclusion that CENP-A, H3K4me2 and H3K36me2, which are all necessary for kinetochore maintenance and function, are enriched on the  $\alpha 21$ -I<sup>TetO</sup> array in the alphoid<sup>2domain</sup> HAC. Nevertheless, the close proximity and scrambled structure of the two arrays allows the chromatin modifier KAP1 to act simultaneously on both arrays.

## DISCUSSION

We have generated several alphoid<sup>2domain</sup> HACs by transfecting HT1080 cells with pBAC11.32TW12.32GLII, a HAC-seeding DNA of ~120 kb. This HAC-seeding DNA contains two distinct  $\alpha$ -satellite DNA arrays: one rich in binding sites for TetR and CENP-B and one lacking CENP-B boxes but having binding sites for LacI and Gal4. We had expected that the former might form centrochromatin and the latter heterochromatin, but experimental results revealed another outcome.

The new alphoid<sup>2domain</sup> HAC shows two important differences from previous generations of synthetic HACs (alphoid<sup>tetO</sup> HAC and alphoid<sup>hybrid</sup> HAC).<sup>2,4</sup> First, the efficiency of alphoid<sup>2domain</sup> HAC formation in HT1080 was higher than that typically seen with other HACs.<sup>2,4</sup> Indeed, it was comparable to results obtained when cotransfecting the HAC-seeding DNA plus CENP-A specifically targeted to the synthetic centromere.<sup>4</sup> It therefore appears that this longer HAC-seeding DNA may be more efficient at promoting stable CENP-A deposition. Second, ChIP-qPCR analysis revealed that CENP-A accumulated preferentially on the CENP-B-containing array of the alphoid<sup>2domain</sup> HAC. This was not observed with the previous alphoid<sup>hybrid</sup> HAC, which was formed from a smaller HAC-seeding DNA.<sup>4</sup>

Surprisingly, H3K9me3 was also recruited to the CENP-B-containing array on the alphoid<sup>2domain</sup> HAC. Previous results have revealed that CENP-B can have a dual role in recruiting centrochromatin or heterochromatin markers depending on the context.<sup>10,27,41</sup> We speculate that the alphoid<sup>2domain</sup> HAC shows 3 types of chromatin. Some of the CENP-B-rich arrays form classical centrochromatin<sup>42</sup> containing CENP-A, H3K4me2 and H3K36me2, but others form H3K9me3-rich heterochromatin, which previous studies have shown to be incompatible with centrochromatin. Thus, CENP-A-containing arrays are likely interspersed with H3K9me3-bearing arrays. Surprisingly, the non-CENP-B array did not form the predicted heterochromatin, but instead appeared to form relatively "open" euchromatin. It is possible that heterochromatin failed to form as a result of selective pressure to avoid silencing the Geneticin resistance gene. Following DNA rearrangements, TetO, LacOGal4 regions and the Geneticin resistance gene can end up being near one another, potentially selecting against rearrangements in which heterochromatin forms and spreads inactivating Geneticin resistance gene. Another possibility is that the rearrangements bring relatively high levels of CENP-B boxes close to the alphoid type II DNA, and this somehow interferes with the ability of the latter to nucleate heterochromatin.<sup>41,43</sup>

Studies of the alphoid<sup>tetO</sup> HAC revealed that HAC-seeding DNA can undergo dramatic reorganization during HAC

establishment.<sup>3</sup> However, the timing and the causes of this phenomenon were unknown. Our Southern blot analysis of various clones of alphoid<sup>2domain</sup> HAC-bearing HT1080 cells reveals that the HAC-seeding DNA in each clone has undergone a unique pattern of rearrangements, both in the size and in the number of fragments observed after restriction digestion and probing for the arrays present in pBAC11.32TW12.32GLII. This highly cell-specific pattern is acquired by each cell in the first 8 weeks following transfection with the HAC-seeding DNA, apparently before completion of the multimerization/amplification that allows the transfected DNA to surpass the size threshold required for stable segregation in mitosis.<sup>23</sup> The specific pattern of rearrangements is stably inherited by HAC-containing subclones, as shown by Southern blot analysis performed 14 weeks after HAC seeding DNA transfection and in agreement with previous reports of HAC stability during MMCT (microcell mediated cell transfer).<sup>3,36</sup> These observations indicate that early during the process of centromere formation, the HAC-seeding DNA encounters a limited series of events that lead to deletions, additions and shuffling of its arrays, but that subsequent to centromere formation the HAC genome is stabilized. Importantly, the larger size of the seeding DNA did not prevent the final structure of the HAC from being amplified or rearranged, as hypothesized prior to this study.

We propose three discrete steps at which modifications on the HAC-seeding DNA possibly occur: in the cytosol, shortly after entry of the HAC-seeding DNA (first step), in the nucleus during replication (second step), and as a consequence of micronucleus formation (third step) (Figure 8A). Not every HAC-seeding DNA will necessarily undergo all three steps, but we suggest that they can all cooperate to form the rearranged mature HAC.

It has been reported extensively in the literature that exogenous DNA is naturally altered upon transfection into cells. Transfected exogenous DNA can undergo mutations, deletions, formation of concatemers, or be eventually lost prior to entering the nucleus.<sup>44–49</sup> Most of these modifications happen early after transfection, before the transfected DNA replicates.<sup>50</sup> This is the result of DNA-sensing pathways:<sup>51</sup> for example, the DNA sensor cGAS binds cytosolic DNA and produces the second messenger cGAMP, which binds STING and leads to activation of an inflammatory response.<sup>52–54</sup> Indeed, it was recently reported that cGAS has an affinity for  $\alpha$ -satellite DNA.<sup>55</sup> Transfected DNA can also undergo double strand breaks (DSBs) triggering a DNA damage repair response<sup>46</sup> that can involve non-homologous end joining (NHEJ) or homologous recombination (HR).

We suggest that transfected HAC-seeding DNA activates these cytosolic responses and that this results in an initial round of DNA alterations (first step in our proposed model). This rearranged DNA then enters the nucleus where it undergoes a process of amplification either due to recombination or "slippage" during replication due to its repetitive sequence, leading to the formation of concatemers (second step). If, during these early stages the HAC-seeding DNA fails to reach a length sufficient to establish a functional centromere or accurate regulation of sister chromatid cohesion, then at the subsequent mitosis the nascent HAC would fail to segregate properly, likely ending up as a lagging chromosome in anaphase. Such lagging chromosomes typically lead to chromosome bridges and micronucleus formation.<sup>32,56,57</sup> Both outcomes have been associated with chromothripsis, a

disruptive event of shredding and shuffling of the DNA, that is associated with cancer development.<sup>30–32,58</sup> During chromothripsis, a region of the genome is cut in tens to hundreds of pieces by an as-yet unknown agent (although rupture of the nuclear envelope in micronuclei has been reported to lead to cGAS accumulation<sup>59</sup>), and the fragments are rejoined randomly by NHEJ, generating a patchwork of DNA fragments.<sup>30</sup>

If the alphoid<sup>2domain</sup> HAC-seeding DNA undergoes chromothripsis, large numbers of rearrangements could potentially occur in a very short period of time (third step). If the rearranged HAC subsequently attains the minimum size for stable mitotic segregation, this could explain the origin of sister HAC-containing cell lines, each with a unique set of rearrangements. In an effort to investigate early stages of HAC formation, we decided to follow by microscopy the HAC-seeding DNA shortly after transfection. Images obtained at 24, 48, and 96 h after transfection in two different cell lines (HT1080 and HT1080 constitutively expressing TetR-EYFP<sup>41</sup>) show that the forming alphoid<sup>2domain</sup> HAC can be found in small micronucleus-like structures termed nanonuclei that were previously observed when the centromere of the alphoid<sup>tetO</sup> HAC was inactivated by altering its epigenetic status<sup>2</sup> (Figure 8B; more images in Figure S7). Those nanonuclei were postulated to be micronuclei containing a single HAC formed when the HAC failed to segregate properly in mitosis. Nanonuclei are negative for CENP-A staining; in the prior study because the centromere had been inactivated and here because the centromere has not yet been established (Figure 8B; more images in Figure S7). The present results are consistent with a previous report that CENP-A accumulates on nascent HACs from 4 days after transfection.<sup>29</sup> This experiment is consistent with our proposed model.

This model suggests that the fate of the HAC seeding-DNA may depend on the phase in the cell cycle when transfection happens, as this could result in a longer or shorter exposure of the DNA molecule to cytosol. Furthermore, NHEJ is more active during the G1 phase of the cell cycle, when sister chromatids are not yet formed.<sup>60</sup> The question of the relationship between HAC DNA reorganization and the cell cycle timing of DNA transfection remains an important one for future study.

The results described here have important implications for ongoing efforts to build synthetic human chromosomes by de novo synthesis,<sup>61</sup> as has been done with great success for budding yeast.<sup>62–64</sup> Unlike budding yeast, which has a point centromere,<sup>65</sup> metazoans have regional centromeres that require establishment of a proper epigenetic environment for their function and stability.<sup>66–71</sup> Our data suggest that this process of centromere formation is frequently associated with DNA rearrangements. It would be extremely unfortunate if human chromosomes synthesized at great cost and effort were to become scrambled in an uncontrollable fashion upon their introduction to human host cells during the process of centromere establishment.

Importantly, once centromere function is established, the associated DNA arrays appear to be much more stable. In fact, although some minor mutations were detected in some clones when the ~90 kb *BRCA1* gene was inserted into an established HAC vector (possibly the result of step 1), there was no detected chromothripsis.<sup>12</sup> Therefore, alternative strategies for building synthetic chromosomes—e.g., assembly of synthetic human chromosomes by building upon available HACs with a

multi-integrase site adjacent to the TetO-array—may avoid these complex DNA rearrangements.<sup>72</sup>

It will be important in future studies to use the alphoid<sup>2domain</sup> HAC system to establish suitable conditions for conservation of the organization of chromosome-sized DNA molecules introduced into human cells during centromere establishment. Routes that can be explored include the transfection of cell lines in specific stages of the cell cycle when the DNA may be less prone to undergo rearrangements (e.g., mitosis); the cotransfection CENP-A or with TetR-linked coactivators of kinetochore establishment (e.g., CENP-A chaperones) or the inactivation of cGAS and other molecules of the cytoplasmic DNA-sensing pathways. Future studies with HACs will allow us to determine whether transfected HAC-seeding DNA does undergo chromothripsis, and if so, how to minimize this. The HAC system will also be useful for studies to optimize procedures to increase the efficiency of centromere activation and establishment of properly regulated cohesion on exogenous DNA. Only when these technical issues have been resolved it will be possible to form predetermined artificial and synthetic chromosomes in human cells.

## METHODS

**Construction of BAC Carrying  $\alpha 21$ -I<sup>TetO</sup> and  $\alpha 21$ -II<sup>LacO/Gal4</sup> Arrays by Tandem Ligation Amplification.** The BAC clone pBAC11.32TW12.32GLII, carrying 32 copies of the  $\alpha 21$ -II<sup>LacO/Gal4</sup> 12-mer following 32 copies of the  $\alpha 21$ -I<sup>TetO</sup> 11-mer was constructed in three steps describing below.

1. Construction of  $\alpha 21$ -II<sup>LacO/Gal4</sup> alphoid 12-mer and insertion into pBluescript vector.  $\alpha 21$ -II<sup>LacO/Gal4</sup> alphoid 12-mer has been designed based on alphoid type II DNA of chromosome #21, and it has been synthesized by GENEART. The SpeI and NheI sites are located respectively at left and right ends of the  $\alpha 21$ -II<sup>LacO/Gal4</sup> 12-mer to be inserted into pBluescript vector. The vector and the 12-mer were joined using the homologous recombination-based method (GENEART Seamless Cloning and Assembly Kit, ThermoFisher Scientific). The resulting plasmid carries one copy of  $\alpha 21$ -II<sup>LacO/Gal4</sup> 12-mer accompanying unique NheI and unique SpeI site at the ends.
2. Extension of the  $\alpha 21$ -II<sup>LacO/Gal4</sup> 12-mer insert in the plasmid vector by repeating the tandem ligation. To extend the length of the alphoid insert, the tandem ligation was repeated until the plasmid harbored 8 copies of the 12-mer using SpeI, NheI, and ScaI restriction enzymes (NEB).<sup>29,73</sup> Therefore, the band of the highest molecular weight (16.6 kb for 8 copies) was excised after PFGE and cloned into the BAC vector. The  $\alpha 21$ -I<sup>TetO</sup> 11-mer was designed based on the sequence of type I alphoid 11-mer of chromosome #21 centromere.<sup>21</sup>
3. Extension of the  $\alpha 21$ -II<sup>LacO/Gal4</sup> 12-mer insert in the BAC vector. Starting from the BAC clone carrying 8 copies of the  $\alpha 21$ -II<sup>LacO/Gal4</sup> 12-mer, the tandem ligation was repeated until the 12-mer insert reached 32 copies using SpeI, NheI, and KasI restriction enzymes (NEB). Finally, 32 copies of  $\alpha 21$ -I<sup>TetO</sup> 11-mer has been cut out from the BAC vector and ligated into the same vector of  $\alpha 21$ -II<sup>LacO/Gal4</sup> 12-mer to obtain the final product, pBAC11.32TW12.32GLII.

After each cloning step, the forming arrays were digested with *Bam*HI and *Not*I restriction enzymes (NEB) and



analyzed on 1% agarose gel electrophoresis using 100 bp DNA Ladder, Quick-Load 1 kb Extend DNA ladder or Low Range PFG Marker (NEB).

**Quantitative PCR (qPCR) to Detect BAC Copy Number.** Cells from HAC-containing clones have been harvested and genomic DNA has been collected using Maxwell DNA purification kit (Promega). qPCR analysis have been performed using SYBR Green Master Mix (Roche) and the following primers: N11F5: 5'-GGGATCACTAGCAAT-AAAAGGTAGAC-3' and N11R6: 5'-TCCTTCTGCTCGTTTTTATGGC-3' for the BAC synthetic DNA; 11-10R: 5'-AGGGAATGTCTTCCCATAAAAAC-3' and mCbox-4: 5'-GTCTACCTTTTATTTGAATTCCCG-3' for the alphoid chr21 array as control. As a standard, DNA from a previously characterized HAC-containing cell line (H21) with a known number of BAC copies ( $n = 125$ ) has been diluted with serial dilutions and amplified with the same primers.

**Cell Culture, Transfection, HAC Formation, and Subcloning.** Human HT1080 and HT1080 constitutively expressing TetR-EYFP<sup>41</sup> cells were cultured in DMEM supplemented with 10% FBS (Labtech) plus 100 U/mL penicillin G and 100  $\mu\text{g}/\text{mL}$  of streptomycin sulfate (Invitrogen). Cells were grown at 37 °C in 5%CO<sub>2</sub> in a humidified atmosphere. Transfection of pBAC11.32TW12.32-GLII DNA was performed using Viafect (Promega) following the manufacturer's instructions. For transfections of cells growing in 6-well plates, transfection complexes containing 10  $\mu\text{L}$  of Viafect reagent and 1  $\mu\text{g}$  of plasmid DNA were prepared in 200  $\mu\text{L}$  of OptiMEM (Invitrogen). After 5 min of incubation at room temperature, 200  $\mu\text{L}$  of transfection complexes was added dropwise in 2 mL of media. After 6 h, the media was changed to the wells and transfected cells were selected adding 400  $\mu\text{g}/\text{mL}$  of Geneticin (Thermo Fisher) and grown for 2–3 weeks until separate resistant colonies were present. Resistant colonies were isolated manually and moved into 24-well plates. Isolated clones were expanded in the presence of 400  $\mu\text{g}/\text{mL}$  of Geneticin. For targeting experiments with TerR-KAP1 and LacI-KAP1, cells have been transfected using Xtremegene-9 (Roche) according to manufacturer's instructions. For transfections in 12-well plates, transfection complexes containing 3  $\mu\text{L}$  of Xtremegene-9 reagent and 500 ng of plasmid DNA were prepared in 100  $\mu\text{L}$  of OptiMEM (Invitrogen). After 20 min of incubation at room temperature, 100  $\mu\text{L}$  of transfection complexes was added dropwise in 1 mL of media. For cotransfection, 500 ng of each plasmid has been transfected in the same reaction.

**Southern Blot Hybridization Analysis.** Southern blot hybridization was performed as described previously<sup>36</sup> with minor changes. Genomic DNA was prepared in agarose plugs (Low Melt Agarose, Biorad),  $0.5 \times 10^6$  cells per plug; plugs have been treated with Proteinase K (CHEF genomic DNA plug kit, Biorad) and restriction-digested by BamHI (NEB) overnight in the buffer recommended by the manufacturer. The digested DNA was CHEF (CHEF Mapper, Bio-Rad) separated (autoprogram, 5–250 kb range, 16 h transfer), transferred to membrane (Amersham Hybond-N+), and blot-hybridized with an 82 bp probe for TetO and a 74 bp probe for LacO and Gal4 containing P-32 CTP. Radioactive labeled DNA sequences for the probes were synthesized by PCR using the primers and synthetic DNA fragment as a template (tetO\_south\_21\_M: 5'-TTTGTGGAAGTGGACATTTA-CTAGCAGCAGAGCTCTCCCTATCAGTGATAGAGAC-TAGCCATAAAAATAGACAGAAGCATT-3', tetO\_

south\_21\_F1: 5'-TTTGTGGAAGTGGACATTTTC-3', tetO\_south\_21\_R1: 5'-AATGCTTCTGTCTATTTTTTA-3'; lacO\_south3\_M: 5'-TGTGGAAGTGGACATTTCCGACCACATGTGGAATTGTGAGCGGATAACAATTTGTG-GCCATAAAAATAGACAGA-3', lacO\_south3\_F1: 5'-TGTGGAAGT GGACATTTCCGA-3', lacO\_south3\_R1: 5'-TCTGTCTATTTTTATGGGCC-3'; gal4\_south1\_M: 5'-AATGGACATTTCCGACGGAGGACAGTCTCCGTCGACGGAGGACAGTCTCCGCATAAAAATCTA-3', gal4\_south1\_F1: 5'-AATGGACATTTCCGACG-3', gal4\_south1\_R1: 5'-TAGATTTTATGCGGAG-3').

The membrane was incubated for 2 h at 65 °C for prehybridization in Church's buffer (0.5 M Na-phosphate buffer containing 7% SDS and 100  $\mu\text{g}/\text{mL}$  of unlabeled salmon sperm carrier DNA). The labeled probe was heat denatured in a boiling water for 5 min, cooled, added to the hybridization Church's buffer, and allowed to hybridize for 48 h at 65 °C. Blots were washed once in 2 $\times$  SSC (300 mM NaCl, 30 mM sodium citrate, pH 7.0)/0.05% SDS for 20 min at 30 °C, once in 2 $\times$  SSC/0.05% SDS for 10 min at 65 °C and then three times in 2 $\times$  SSC/0.05% SDS for 5 min at 65 °C. Blots were exposed to X-ray film 2–48 h at –80 °C.

**Expression and Purification of Recombinant TetR/LacI-eYFP.** TetR and LacI were cloned as C-terminally His-tagged proteins in a pET23a vector and proteins were purified following a previously described procedure.<sup>9</sup> Briefly, the vectors were transformed in *E. coli* BL21 Gold cells and colonies grown at 37 °C until OD<sub>600</sub>1 in Super Broth containing ampicillin. The cultures were then induced with 0.35 mM IPTG overnight at 18 °C and cell pellets were lysed in a buffer containing 20 mM Tris HCl pH 7.5, 500 mM NaCl, 35 mM imidazole and 2 mM 2-mercaptoethanol. Proteins were affinity-purified using a Ni-NTA column (GE Healthcare), washed with high salt buffer (20 mM Tris HCl pH 7.5, 1000 mM NaCl, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 2 mM ATP, 35 mM imidazole, and 2 mM 2-mercaptoethanol) and eluted with 20 mM Tris HCl pH 7.5, 150 mM NaCl, 400 mM imidazole, and 2 mM 2-mercaptoethanol. The pure eluted fractions were pooled and dialyzed overnight against storage buffer (20 mM Tris HCl pH 7.5, 150 mM NaCl, 5% glycerol, and 2 mM 2-mercaptoethanol). Sample quality was analyzed by 15% SDS-PAGE stained with Coomassie Blue. The final protein concentrations that were used for the IF staining on fibers were 1.2 and 1.7 mg/mL for TetR-eYFP and LacI-eYFP, respectively.

**Fluorescent *in Situ* Hybridization (FISH) and DNA Fibers Preparation.** Metaphase chromosomes from HT1080 were obtained following a standard protocol: 3 h before harvesting, cells were treated with Colcemid (Invitrogen) at a final concentration of 0.1  $\mu\text{g}/\text{mL}$ . Collected cells were resuspended in warm hypotonic solution (75 mM KCl) for 20 min at 37 °C and fixed in methanol:acetic acid (3:1). Slides were kept at –20 °C until they were processed for FISH. To obtain stretched chromatin fibers,  $2 \times 10^6$  cells were centrifuged, and the pellets were washed in 1 $\times$  PBS. Ten  $\mu\text{L}$  drops have been placed on slides and let dry. Once the slides were mounted on the Shandon Sequenza cover plates (Thermo Scientific), DNA fibers were released applying a lysis solution (700 mM NaOH in ethanol) and fixed in methanol. Slides have been kept in PBS at 4 °C until stained.

For oligo-FISH staining, oligonucleotides recognizing the TetO sequence (5'-ACTAGCAGCAGAGCTCTCCCTATCAGTGATAGAGACTAG-3') labeled with Digoxigenin,



and oligonucleotides recognizing both LacO (5'-CATGT-GGAATTGTGAGCGGATAACAATTTGTGG-3') and Gal4 (5'-TCGACGGAGGACAGTCCTCCG-3') sequences labeled with Biotin were purchased (Sigma). Oligonucleotides were mixed at 100 ng/ $\mu$ L and resuspended in hybridization buffer (50% formamide, 10% dextran sulfate, 2 $\times$  SSC (300 mM NaCl, 30 mM sodium citrate, pH 7.0)) and 50  $\mu$ g/mL of salmon sperm DNA (Sigma). FISH was carried out following standard procedures. Slides were denatured in 70% formamide/2 $\times$  SSC at 68  $^{\circ}$ C for 45 s (70  $^{\circ}$ C for 1 min for fibers) and hybridized in a humid chamber at 37  $^{\circ}$ C for 2 h. Slides were then washed in 20%formamide/2 $\times$  SSC for 5 min and in 2 $\times$  SSC/0.1% Tween-20 for 5 min at 37  $^{\circ}$ C. Oligonucleotide probes were detected with rhodamine-conjugated antidigoxigenin (Roche) and fluorescein-conjugated streptavidin (Vector Laboratories) incubated for 30 min at 37  $^{\circ}$ C. Slides were mounted with Vectashield (Vector Laboratories) containing 4',6-diamidino-2-phenylindole (DAPI) for chromosome counterstaining. IF staining on fibers, slides have been blocked with 1% BSA/1 $\times$  PBS/0.1% Triton X-100 and then incubated with TetR-GFP or LacI-GFP proteins purified from *E. coli* in blocking buffer overnight at 4  $^{\circ}$ C, together with  $\alpha$ -CENP-A and  $\alpha$ -H3K9me3 antibodies as described in immunofluorescence (IF) protocol. Slides have been washed, incubated with secondary antibodies and sealed as immunofluorescence (IF) staining standard protocol. For the IF staining on fibers, we aimed to count 20–30 fibers for each staining.

**Analysis of the HAC Stability.** HAC-containing subclones have been thawed and maintained in culture with 400  $\mu$ g/mL of Geneticin (Thermo Fisher) for 7 days. At day 0, metaphase chromosomes have been spread on slides and labeled for FISH as described. At day 0, cell cultures have been split into two batches: one batch has been kept in culture with 400  $\mu$ g/mL of Geneticin (Thermo Fisher) for 30 days, while the other has been kept in culture with simple DMEM/10% FBS/1% PenStrepto. At day 30, metaphase chromosomes from each batch have been spread on slides and labeled for FISH as described. Metaphases at day 0 and day 30 have been scored for the presence of 0, 1, 2, or >2 HACs. The daily loss rate of the HAC (R) was calculated using the formula  $N_n = N_0 \times (1 - R)^n$ , where  $N_0$  is the number of metaphase chromosome spreads showing a HAC in the cells cultured under selection and  $N_n$  is the number of HAC-containing metaphase chromosome spreads after  $n$  days of culture in the absence of selection.

**Indirect Immunofluorescence (IF) Staining and Microscopy Analysis.** Indirect immunofluorescence (IF) staining of cells fixed in 3.7% formaldehyde/1 $\times$  PBS was performed at 37  $^{\circ}$ C for 10 min following standard procedures. The following antibodies were used: mouse anti-CENP-A (clone A1, 1:500<sup>37</sup>) and rabbit anti-H3K9me3 (abcam 8898; 1:300). Microscope images were acquired on a DeltaVision Core system (Applied Precision) using an Olympus IX-71 inverted microscope stand with an Olympus UPlanSApo 100 $\times$  oil immersion objective (numerical aperture (NA) 1.4) and an LED light source. Camera (Photometrics Cool Snap HQ), shutter, and stage were controlled through SoftWorx (Applied Precision). Z-series were collected with a spacing of 0.2  $\mu$ m, and image stacks were subsequently deconvolved in SoftWorx. CENP-A and H3K9me3 signal quantification at the HAC was performed using ImageJ. HAC was visualized in the green channel thanks to the tethering of the corresponding

fluorescent protein. For CENP-A signal quantification, a custom-made macro in ImageJ (National Institutes of Health, Bethesda, MD) was used.<sup>74</sup>

**Chromatin Immunoprecipitation and Quantitative PCR (ChIP-qPCR).** Exponentially growing cells were harvested with TrypLE Express (Gibco) and resuspended in D-PBS (Gibco) up to a concentration of 1  $\times$  10<sup>6</sup> cells/mL. They were cross-linked with 1% formaldehyde solution (Sigma) for 5 min at room temperature, followed by quenching with 2.5 M glycine for 5 min at room temperature. Cells at a concentration of 5  $\times$  10<sup>6</sup> cells/mL were lysed in lysis buffer (10 mM Tris pH 8.0; 10 mM NaCl; 0.5% NP-40) containing protease inhibitors (1  $\mu$ g/mL of CLAP; 0.5  $\mu$ g/mL of Aprotinin; 1 mM PMSF) for 10 min on ice. Cells pellets have been snap-frozen in liquid N and stored in -80  $^{\circ}$ C until processed. Nuclei were resuspended in lysis buffer with protease inhibitors (50 mM Tris pH 8.0; 2 mM EDTA; 0.2% SDS; 134 mM NaCl; 0.88% Triton X-100; 0.088% Na-deoxycholate) and incubated with 400 U/mL MNase for 30 min at 20  $^{\circ}$ C. Chromatin was sheared by sonication in a Bioruptor sonicator (Diagenode) for 9 cycles (30s ON/30s OFF) at high setting and 4  $^{\circ}$ C in dilution buffer 1 enriched with. The collected supernatants after sonication were diluted with 300  $\mu$ L of dilution buffer 1, 500  $\mu$ L of dilution buffer 2 (50 mM Tris pH 8.0; 167 mM NaCl; 1.1% Triton X-100; 0.11% Na-deoxycholate), and 500  $\mu$ L of RIPA buffer containing 150  $\mu$ L of NaCl (RIPA-150) and protease inhibitors. Antimouse IgG Dynabeads (Invitrogen) were conjugated with the relevant antibodies for 6 h with RIPA-150/0.5% BSA at 4  $^{\circ}$ C and washed twice with RIPA-150/0.5% BSA. 500  $\mu$ L of collected sheared chromatin was incubated with the beads at 4  $^{\circ}$ C overnight. Beads were afterward washed twice with RIPA-150 and RIPA buffer containing 500 mM NaCl (RIPA-500) and a final wash with TE pH 8.0. Antibody/chromatin complexes were de-cross-linked with 10% Chelex-100 resin (BioRad) in water at 93  $^{\circ}$ C and incubated with RNase A and Proteinase K. DNA was subsequently recovered using magnetic rack. DNA from the chromatin immunoprecipitates (ChIPed DNA) and input DNA was subject to real-time PCR analysis using a SYBR Green Mastermix (Roche) on a LightCycler480 system (Roche). For each primer pair, a standard curve was prepared from the input material and included on every plate to calculate the percent of precipitated DNA relative to the input material. ChIPed DNA was subjected to RT-PCR using the following oligonucleotides: tetO-Fw (5'-CCACTCCCTA-TCAGTGATAGAGAA-3'), tetO-Rv (5'-TCGACTTCTG-TTTAGTTCGTGCG-3') for the  $\alpha$ 21-I-tetO domain of the hybrid HAC, lacOgal4-Fw (5'-TATGGTGTGACGGAGGACA-3'), and lacOgal4-Rv (5'-CCGCTCACAATTCACATGTG-3') for the  $\alpha$ 21-II-lacOgal4 domain of the hybrid HAC, chr17-Fw (5'-TTGTGGTTTGTGGTGAAAA-3') and chr17-Rv (5'-CTCAAAGCGCTCCAAATCTC-3') for the alphoid chr17 array, sat2-Fw (5'-TCGCATAGAATCGAA-TGGAA-3') and sat2-Rv (5'-GCATTCGAGTCCGTGGA-3') for the pericentromeric alphoid chr1.

**IF during Early Stages of Alphoid<sup>2</sup>domain HAC Formation.** HT1080 and HT1080 constitutively expressing TetR-EYFP,<sup>41</sup> both growing on coverslips, were transfected with pBAC11.32TW12.32GLI using Viafect (Promega) as already described. Transfected cells were fixed at the stated time points following IF procedures and stained using mouse anti-CENP-A (clone A1, 1:500<sup>37</sup>). Microscope images were acquired on a DeltaVision Core system (Applied Precision).

For HT1080, cells were transfected with LacI-GFP plasmid to localize the HAC DNA 18–24 h prior fixation.

## ■ ASSOCIATED CONTENT


### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acssynbio.0c00326>.

Supplementary Figures S1–S7 (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

Elisa Pesenti – Wellcome Trust Centre for Cell Biology, Edinburgh EH9 3BF, United Kingdom;  [orcid.org/0000-0002-1958-5411](https://orcid.org/0000-0002-1958-5411); Email: [elisa.pesenti@ed.ac.uk](mailto:elisa.pesenti@ed.ac.uk)

William C. Earnshaw – Wellcome Trust Centre for Cell Biology, Edinburgh EH9 3BF, United Kingdom; Email: [bill.earnshaw@ed.ac.uk](mailto:bill.earnshaw@ed.ac.uk)

### Authors

Mikhail Liskovych – National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States


Koei Okazaki – Kazusa DNA Research Institute, Kisarazu 292-0818, Japan

Alessio Mallozzi – Wellcome Trust Centre for Cell Biology, Edinburgh EH9 3BF, United Kingdom

Caitlin Reid – Wellcome Trust Centre for Cell Biology, Edinburgh EH9 3BF, United Kingdom

Maria Alba Abad – Wellcome Trust Centre for Cell Biology, Edinburgh EH9 3BF, United Kingdom

A. Arockia Jeyaprakash – Wellcome Trust Centre for Cell Biology, Edinburgh EH9 3BF, United Kingdom

Natalay Kouprina – National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;  [orcid.org/0000-0001-8690-5625](https://orcid.org/0000-0001-8690-5625)

Vladimir Larionov – National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States

Hiroshi Masumoto – Kazusa DNA Research Institute, Kisarazu 292-0818, Japan

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acssynbio.0c00326>

### Author Contributions

E.P. and W.C.E. designed the experiments; E.P. performed the experiments, with the help of A.M. and C.R.; M.L. performed Southern blot experiments; K.O. constructed the HAC seeding DNA; M.A.A. expressed bacteria purified proteins; M.L., H.M., V.L., N.K., and W.C.E. contributed with critical discussions and new experimental tools; E.P. analyzed the data; E.P. and W.C.E. wrote the paper.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was funded by the UK Research Councils' Synthetic Biology for Growth program [core grant BB/M018040/1 to the BBSRC/EPSRC/MRC Synthetic Biological Research Centre] (E.P.). Additional funding of the Earnshaw lab is provided by Wellcome, of which W.C.E. is a Principal Research Fellow [grant number 073915; imaging was performed in Centre Optical Instrumentation Laboratory (COIL), which is

supported by a Core Grant (203149)] to the Wellcome Centre for Cell Biology at the University of Edinburgh. Additional experiments were supported by MEXT KAKENHI [grant numbers 16H04747, 16H01414, and 18H04721] and the Kazusa DNA Research Institute Foundation (H.M.) and by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, USA (V.L.). M.A.A. and A.A.J. are supported by the Wellcome Trust through a Wellcome Senior Research Fellowship to A.A.J. (202811). We also thank Dr. Fernanda Cisneros Soberanis for her feedback on the manuscript and Dr. Nuno M. C. Martins and Dr. Oscar Molina for their help with data analysis.

## ■ REFERENCES

- (1) 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.
- (2) 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 Kinetochores by Specific Targeting of Chromatin Modifiers. *Dev. Cell* 14, 507.
- (3) Kouprina, N., Samoshkin, A., Erliandri, I., Nakano, M., Lee, H. S., Fu, H., Iida, Y., Aladjem, M., Oshimura, M., Masumoto, H., Earnshaw, W. C., and Larionov, V. (2012) Organization of Synthetic Alphoid DNA Array in Human Artificial Chromosome (HAC) with a Conditional Centromere. *ACS Synth. Biol.* 1, 590.
- (4) Pesenti, E., Kouprina, N., Liskovych, 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.
- (5) Lee, N. C. O., Kim, J. H., Petrov, N. S., Lee, H. S., Masumoto, H., Earnshaw, W. C., Larionov, V., and Kouprina, N. (2018) Method to Assemble Genomic DNA Fragments or Genes on Human Artificial Chromosome with Regulated Kinetochores Using a Multi-Integrase System. *ACS Synth. Biol.* 7, 63.
- (6) Cardinale, S., Bergmann, J. H., Kelly, D., Nakano, M., Valdivia, M. M., Kimura, H., Masumoto, H., Larionov, V., and Earnshaw, W. C. (2009) Hierarchical Inactivation of a Synthetic Human Kinetochores by a Chromatin Modifier. *Mol. Biol. Cell* 20 (19), 4194–4204.
- (7) Bergmann, J. H., Rodríguez, M. G., Martins, N. M. C., Kimura, H., Kelly, D. A., Masumoto, H., Larionov, V., Jansen, L. E. T., and Earnshaw, W. C. (2011) Epigenetic Engineering Shows H3K4me2 Is Required for HJURP Targeting and CENP-A Assembly on a Synthetic Human Kinetochores. *EMBO J.* 30, 328.
- (8) Bergmann, J. H., Jakubsche, J. N., Martins, N. M., Kagansky, A., Nakano, M., Kimura, H., Kelly, D. A., Turner, B. M., Masumoto, H., Larionov, V., and Earnshaw, W. C. (2012) Epigenetic Engineering: Histone H3K9 Acetylation Is Compatible with Kinetochores Structure and Function. *J. Cell Sci.* 125, 411.
- (9) Molina, O., Vargiu, G., Abad, M. A., Zhiteneva, A., Jeyaprakash, A. A., Masumoto, H., Kouprina, N., Larionov, V., and Earnshaw, W. C. (2016) Epigenetic Engineering Reveals a Balance between Histone Modifications and Transcription in Kinetochores Maintenance. *Nat. Commun.*, DOI: 10.1038/ncomms13334.
- (10) Ohzeki, J. I., 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.
- (11) Kim, J. H., Kononenko, A., Erliandri, I., Kim, T. A., Nakano, M., Iida, Y., Barrett, J. C., Oshimura, M., Masumoto, H., Earnshaw, W. C., Larionov, V., and Kouprina, N. (2011) Human Artificial Chromosome (HAC) Vector with a Conditional Centromere for Correction of Genetic Deficiencies in Human Cells. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20048.

- (12) Kononenko, A. V., Bansal, R., Lee, N. C. O., Grimes, B. R., Masumoto, H., Earnshaw, W. C., Larionov, V., and Kouprina, N. (2014) A Portable BRCA1-HAC (Human Artificial Chromosome) Module for Analysis of BRCA1 Tumor Suppressor Function. *Nucleic Acids Res.* 42, e164.
- (13) Kouprina, N., Tomilin, A. N., Masumoto, H., Earnshaw, W. C., and Larionov, V. (2014) Human Artificial Chromosome-Based Gene Delivery Vectors for Biomedicine and Biotechnology. *Expert Opin. Drug Delivery* 11, 517.
- (14) Kononenko, A. V., Lee, N. C. O., Liskovych, M., Masumoto, H., Earnshaw, W. C., Larionov, V., and Kouprina, N. (2015) Generation of a Conditionally Self-Eliminating HAC Gene Delivery Vector through Incorporation of a TTAVP64 Expression Cassette. *Nucleic Acids Res.* 43, e57.
- (15) Liskovych, M., Ponomartsev, S., Popova, E., Bader, M., Kouprina, N., Larionov, V., Alenina, N., and Tomilin, A. (2015) Stable Maintenance of de Novo Assembled Human Artificial Chromosomes in Embryonic Stem Cells and Their Differentiated Progeny in Mice. *Cell Cycle* 14, 1268.
- (16) Ponomartsev, S. V., Sinenko, S. A., Skvortsova, E. V., Liskovych, M. A., Voropaev, I. N., Savina, M. M., Kuzmin, A. A., Kuzmina, E. Y., Kondrashkina, A. M., Larionov, V., Kouprina, N., and Tomilin, A. N. (2020) Human AlphoidtetO Artificial Chromosome as a Gene Therapy Vector for the Developing Hemophilia A Model in Mice. *Cells* 9, 879.
- (17) Lee, H. S., Lee, N. C. O., Kouprina, N., Kim, J. H., Kagansky, A., Bates, S., Trepel, J. B., Pommier, Y., Sackett, D., and Larionov, V. (2016) Effects of Anticancer Drugs on Chromosome Instability and New Clinical Implications for Tumor-Suppressing Therapies. *Cancer Res.* 76, 902.
- (18) Lee, H. S., Carmena, M., Liskovych, M., Peat, E., Kim, J. H., Oshimura, M., Masumoto, H., Teulade-Fichou, M. P., Pommier, Y., Earnshaw, W. C., Larionov, V., and Kouprina, N. (2018) Systematic Analysis of Compounds Specifically Targeting Telomeres and Telomerase for Clinical Implications in Cancer Therapy. *Cancer Res.* 78, 6282.
- (19) Liskovych, M., Goncharov, N. V., Petrov, N., Aksenova, V., Pegoraro, G., Ozbun, L. L., Reinhold, W. C., Varma, S., Dasso, M., Kumeiko, V., Masumoto, H., Earnshaw, W. C., Larionov, V., and Kouprina, N. (2019) A Novel Assay to Screen siRNA Libraries Identifies Protein Kinases Required for Chromosome Transmission. *Genome Res.* 29, 1719.
- (20) Ikeno, M., and Suzuki, N. (2011) Construction and Use of a Bottom-up HAC Vector for Transgene Expression. *Methods Mol. Biol.* 738, 101.
- (21) 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.
- (22) Logsdon, G. A., Gambogi, C. W., Liskovych, M. A., Barrey, E. J., Larionov, V., Miga, K. H., Heun, P., and Black, B. E. (2019) Human Artificial Chromosomes That Bypass Centromeric DNA. *Cell* 178, 624.
- (23) Farr, C. J., Bayne, R. A., Kipling, D., Mills, W., Critcher, R., and Cooke, H. J. (1995) Generation of a Human X-Derived Minichromosome Using Telomere-Associated Chromosome Fragmentation. *EMBO J.* 14, 5444.
- (24) 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.
- (25) Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., and Okazaki, T. (1989) A Human Centromere Antigen (CENP-B) Interacts with a Short Specific Sequence in Alphoid DNA, a Human Centromeric Satellite. *J. Cell Biol.* 109, 1963.
- (26) Earnshaw, W. C., and Rothfield, N. (1985) Identification of a Family of Human Centromere Proteins Using Autoimmune Sera from Patients with Scleroderma. *Chromosoma* 91, 313.
- (27) Okada, T., Ohzeki, J. i., 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.
- (28) Masumoto, H., Nakano, M., and Ohzeki, J. I. (2004) The Role of CENP-B and  $\alpha$ -Satellite DNA: De Novo Assembly and Epigenetic Maintenance of Human Centromeres. *Chromosome Res.* 12, 543.
- (29) Ohzeki, J. i., 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.
- (30) Stephens, P. J., Greenman, C. D., Fu, B., Yang, F., Bignell, G. R., Mudie, L. J., Pleasance, E. D., Lau, K. W., Beare, D., Stebbings, L. A., McLaren, S., Lin, M. L., McBride, D. J., Varela, I., Nik-Zainal, S., Leroy, C., Jia, M., Menzies, A., Butler, A. P., Teague, J. W., Quail, M. A., Burton, J., Swerdlow, H., Carter, N. P., Morsberger, L. A., Iacobuzio-Donahue, C., Follows, G. A., Green, A. R., Flanagan, A. M., Stratton, M. R., Futreal, P. A., and Campbell, P. J. (2011) Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* 144, 27.
- (31) Liu, P., Erez, A., Nagamani, S. C. S., Dhar, S. U., Kołodziejaska, K. E., Dharmadhikari, A. V., Cooper, M. L., Wiszniewska, J., Zhang, F., Withers, M. A., Bacino, C. A., Campos-Acevedo, L. D., Delgado, M. R., Freedenberg, D., Garnica, A., Grebe, T. A., Hernández-Almaguer, D., Immken, L., Lalani, S. R., McLean, S. D., Northrup, H., Scaglia, F., Strathearn, L., Trapane, P., Kang, S. H. L., Patel, A., Cheung, S. W., Hastings, P. J., Stankiewicz, P., Lupski, J. R., and Bi, W. (2011) Chromosome Catastrophes Involve Replication Mechanisms Generating Complex Genomic Rearrangements. *Cell* 146, 889.
- (32) Crasta, K., Ganem, N. J., Dagher, R., Lantermann, A. B., Ivanova, E. V., Pan, Y., Nezi, L., Protopopov, A., Chowdhury, D., and Pellman, D. (2012) DNA Breaks and Chromosome Pulverization from Errors in Mitosis. *Nature* 482, 53.
- (33) Ebersole, T., Okamoto, Y., Noskov, V. N., Kouprina, N., Kim, J. H., Leem, S. H., Barrett, J. C., Masumoto, H., and Larionov, V. (2005) Rapid Generation of Long Synthetic Tandem Repeats and Its Application for Analysis in Human Artificial Chromosome Formation. *Nucleic Acids Res.* 33, e130.
- (34) Ohzeki, J. i., Larionov, V., Earnshaw, W. C., and Masumoto, H. (2015) Genetic and Epigenetic Regulation of Centromeres: A Look at HAC Formation. *Chromosome Res.* 23, 87.
- (35) 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.
- (36) Liskovych, M., Lee, N. C., Larionov, V., and Kouprina, N. (2016) Moving toward a Higher Efficiency of Microcell-Mediated Chromosome Transfer. *Mol. Ther.–Methods Clin. Dev.* 3, 16043.
- (37) Kimura, H., Hayashi-Takanaka, Y., Goto, Y., Takizawa, N., and Nozaki, N. (2008) The Organization of Histone H3 Modifications as Revealed by a Panel of Specific Monoclonal Antibodies. *Cell Struct. Funct.* 33, 61.
- (38) Schultz, D. C., Friedman, J. R., and Rauscher, F. J. (2001) Targeting Histone Deacetylase Complexes via KRAB-Zinc Finger Proteins: The PHD and Bromodomains of KAP-1 Form a Cooperative Unit That Recruits a Novel Isoform of the Mi-2 $\alpha$  Subunit of NuRD. *Genes Dev.* 15, 428.
- (39) Cardinale, S., Bergmann, J. H., Kelly, D., Nakano, M., Valdivia, M. M., Kimura, H., Masumoto, H., Larionov, V., and Earnshaw, W. C. (2009) Hierarchical Inactivation of a Synthetic Human Kinetochores by a Chromatin Modifier. *Mol. Biol. Cell* 20, 4194.
- (40) Martins, N. M. C., Cisneros-Soberanis, F., Pesenti, E., Kochanova, N. Y., Shang, W. H., Hori, T., Nagase, T., Kimura, H., Larionov, V., Masumoto, H., Fukagawa, T., and Earnshaw, W. C. (2020) H3K9me3 Maintenance on a Human Artificial Chromosome Is Required for Segregation but Not Centromere Epigenetic Memory. *J. Cell Sci.* 133, jcs242610.
- (41) Otake, K., Ohzeki, J. I., Shono, N., Kugou, K., Okazaki, K., Nagase, T., Yamakawa, H., Kouprina, N., Larionov, V., Kimura, H., Earnshaw, W. C., and Masumoto, H. (2020) CENP-B Creates



Alternative Epigenetic Chromatin States Permissive for CENP-A or Heterochromatin Assembly. *J. Cell Sci.* 133, jcs243303.

(42) Sullivan, B. A., and Karpen, G. H. (2004) Centromeric Chromatin Exhibits a Histone Modification Pattern That Is Distinct from Both Euchromatin and Heterochromatin. *Nat. Struct. Mol. Biol.* 11, 1076.

(43) Okada, T., Ohzeki, J. i., 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.

(44) Wake, C. T., Gudewicz, T., Porter, T., White, A., and Wilson, J. H. (1984) How Damaged Is the Biologically Active Subpopulation of Transfected DNA? *Mol. Cell. Biol.* 4, 387.

(45) Würtele, H., Little, K. C. E., and Chartrand, P. (2003) Illegitimate DNA Integration in Mammalian Cells. *Gene Ther.* 10, 1791.

(46) Dellaire, G., Yan, J., Little, K. C. E., Drouin, R., and Chartrand, P. (2002) Evidence That Extrachromosomal Double-Strand Break Repair Can Be Coupled to the Repair of Chromosomal Double-Strand Breaks in Mammalian Cells. *Chromosoma* 111, 304.

(47) Calos, M. P., Lebkowski, J. S., and Botchan, M. R. (1983) High Mutation Frequency in DNA Transfected into Mammalian Cells. *Proc. Natl. Acad. Sci. U. S. A.* 80, 3015.

(48) Razzaque, A., Mizusawa, H., and Seidman, M. M. (1983) Rearrangement and Mutagenesis of a Shuttle Vector Plasmid after Passage in Mammalian Cells. *Proc. Natl. Acad. Sci. U. S. A.* 80, 3010.

(49) Stepanenko, A. A., and Heng, H. H. (2017) Transient and Stable Vector Transfection: Pitfalls, off-Target Effects, Artifacts. *Mutat. Res., Rev. Mutat. Res.* 773, 91.

(50) Razzaque, A., Chakrabarti, S., Joffe, S., and Seidman, M. (1984) Mutagenesis of a Shuttle Vector Plasmid in Mammalian Cells. *Mol. Cell. Biol.* 4, 435.

(51) Semenova, N., Bosnjak, M., Markelc, B., Znidar, K., Cemazar, M., and Heller, L. (2019) Multiple Cytosolic DNA Sensors Bind Plasmid DNA after Transfection. *Nucleic Acids Res.* 47, 10235.

(52) Andreeva, L., Hiller, B., Kostrewa, D., Lässig, C., De Oliveira Mann, C. C., Jan Drexler, D., Maiser, A., Gaidt, M., Leonhardt, H., Hornung, V., and Hopfner, K. P. (2017) CGAS Senses Long and HMGB/TFAM-Bound U-Turn DNA by Forming Protein-DNA Ladders. *Nature* 549, 394.

(53) Chen, D., Murphy, B., Sung, R., and Bromberg, J. S. (2003) Adaptive and Innate Immune Responses to Gene Transfer Vectors: Role of Cytokines and Chemokines in Vector Function. *Gene Ther.* 10, 991.

(54) Liu, H., Zhang, H., Wu, X., Ma, D., Wu, J., Wang, L., Jiang, Y., Fei, Y., Zhu, C., Tan, R., Jungblut, P., Pei, G., Dorhoi, A., Yan, Q., Zhang, F., Zheng, R., Liu, S., Liang, H., Liu, Z., Yang, H., Chen, J., Wang, P., Tang, T., Peng, W., Hu, Z., Xu, Z., Huang, X., Wang, J., Li, H., Zhou, Y., Liu, F., Yan, D., Kaufmann, S. H. E., Chen, C., Mao, Z., and Ge, B. (2018) Nuclear CGAS Suppresses DNA Repair and Promotes Tumorigenesis. *Nature* 563, 131.

(55) Gentili, M., Lahaye, X., Nadalin, F., Nader, G. F. P., Puig Lombardi, E., Herve, S., De Silva, N. S., Rookhuizen, D. C., Zueva, E., Goudot, C., Maurin, M., Bochnakian, A., Amigorena, S., Piel, M., Fachinetti, D., Londoño-Vallejo, A., and Manel, N. (2019) The N-Terminal Domain of CGAS Determines Preferential Association with Centromeric DNA and Innate Immune Activation in the Nucleus. *Cell Rep.* 26, 2377.

(56) Fenech, M., Kirsch-Volders, M., Natarajan, A. T., Surralles, J., Crott, J. W., Parry, J., Norppa, H., Eastmond, D. A., Tucker, J. D., and Thomas, P. (2011) Molecular Mechanisms of Micronucleus, Nucleoplasmic Bridge and Nuclear Bud Formation in Mammalian and Human Cells. *Mutagenesis* 26, 125.

(57) Umbreit, N. T., Zhang, C. Z., Lynch, L. D., Blaine, L. J., Cheng, A. M., Tourdot, R., Sun, L., Almubarak, H. F., Judge, K., Mitchell, T. J., Spektor, A., and Pellman, D. (2020) Mechanisms Generating Cancer Genome Complexity from a Single Cell Division Error. *Science (Washington, DC, U. S.)* 368, eaba0712.

(58) Zhang, C. Z., Spektor, A., Cornils, H., Francis, J. M., Jackson, E. K., Liu, S., Meyerson, M., and Pellman, D. (2015) Chromothripsis from DNA Damage in Micronuclei. *Nature* 522, 179.

(59) MacKenzie, K. J., Carroll, P., Martin, C. A., Murina, O., Fluteau, A., Simpson, D. J., Olova, N., Sutcliffe, H., Rainger, J. K., Leitch, A., Osborn, R. T., Wheeler, A. P., Nowotny, M., Gilbert, N., Chandra, T., Reijns, M. A. M., and Jackson, A. P. (2017) CGAS Surveillance of Micronuclei Links Genome Instability to Innate Immunity. *Nature* 548, 461.

(60) Mao, Z., Bozzella, M., Seluanov, A., and Gorbunova, V. (2008) DNA Repair by Nonhomologous End Joining and Homologous Recombination during Cell Cycle in Human Cells. *Cell Cycle* 7, 2902.

(61) Ostrov, N., Beal, J., Ellis, T., Benjamin Gordon, D., Karas, B. J., Lee, H. H., Lenaghan, S. C., Schloss, J. A., Stracquadanio, G., Trefzer, A., Bader, J. S., Church, G. M., Coelho, C. M., William Efcavitch, J., Güell, M., Mitchell, L. A., Nielsen, A. A. K., Peck, B., Smith, A. C., Neal Stewart, C., and Tekotte, H. (2019) Technological Challenges and Milestones for Writing Genomes. *Science* 366, 310.

(62) 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., Cai, Y., Zeller, K., Agmon, N., Han, J. S., Hadjithomas, M., Tullman, J., Caravelli, K., Cirelli, K., Guo, Z., London, V., Yeluru, A., Murugan, S., Kandavelou, K., Agier, N., Fischer, G., Yang, K., Martin, J. A., Bilgel, M., Bohutski, P., Boulier, K. M., Capaldo, B. J., Chang, J., Charoen, K., Choi, W. J., Deng, P., DiCarlo, J. E., Doong, J., Dunn, J., Feinberg, J. I., Fernandez, C., Floria, C. E., Gladowski, D., Hadidi, P., Ishizuka, I., Jabbari, J., Lau, C. Y. L., Lee, P. A., Li, S., Lin, D., Linder, M. E., Ling, J., Liu, J., Liu, J., London, M., Henry, M., Mao, J., McDade, J. E., McMillan, A., Moore, A. M., Oh, W. C., Ouyang, Y., Patel, R., Paul, M., Paulsen, L. C., Qiu, J., Rhee, A., Rubashkin, M. G., Soh, I. Y., Sotuyo, N. E., Srinivas, V., Suarez, A., Wong, A., Wong, R., Xie, W. R., Xu, Y., Yu, A. T., Koszul, R., Bader, J. S., Boeke, J. D., and Chandrasegaran, S. (2014) Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science (Washington, DC, U. S.)* 344, 55.

(63) 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., Boeke, J. D., and Bader, J. S. (2017) Design of a Synthetic Yeast Genome. *Science (Washington, DC, U. S.)* 355, 1040.

(64) Pretorius, I. S., and Boeke, J. D. (2018) Yeast 2.0-Connecting the Dots in the Construction of the World's First Functional Synthetic Eukaryotic Genome. *FEMS Yeast Research*, DOI: 10.1093/femsyr/foy032.

(65) Pluta, A. F., Mackay, A. M., Ainsztein, A. M., Goldberg, I. G., and Earnshaw, W. C. (1995) The Centromere: Hub of Chromosomal Activities. *Science (Washington, DC, U. S.)* 270, 1591.

(66) Earnshaw, W. C., and Migeon, B. R. (1985) Three Related Centromere Proteins Are Absent from the Inactive Centromere of a Stable Isodiscentric Chromosome. *Chromosoma* 92, 290.

(67) Karpen, G. H., and Allshire, R. C. (1997) The Case for Epigenetic Effects on Centromere Identity and Function. *Trends Genet.* 13, 489.

(68) Allshire, R. C., and Karpen, G. H. (2008) Epigenetic Regulation of Centromeric Chromatin: Old Dogs, New Tricks? *Nat. Rev. Genet.* 9, 923.

(69) Ohzeki, J., Larionov, V., Earnshaw, W. C., and Masumoto, H. (2019) De Novo Formation and Epigenetic Maintenance of Centromere Chromatin. *Curr. Opin. Cell Biol.* 58, 15.

(70) Musacchio, A., and Desai, A. (2017) A Molecular View of Kinetochore Assembly and Function. *Biology* 6, 5.

(71) Fukagawa, T., and Earnshaw, W. C. (2014) The Centromere: Chromatin Foundation for the Kinetochore Machinery. *Dev. Cell* 30, 496.

(72) Lee, N. C. O., Larionov, V., and Kouprina, N. (2015) Highly Efficient CRISPR/Cas9-Mediated TAR Cloning of Genes and Chromosomal Loci from Complex Genomes in Yeast. *Nucleic Acids Res.* 43, e55.



(73) Okamoto, Y., Nakano, M., Ohzeki, J. I., 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.

(74) Bodor, D. L., Mata, J. F., Sergeev, M., David, A. F., Salimian, K. J., Panchenko, T., Cleveland, D. W., Black, B. E., Shah, J. V., and Jansen, L. E. T. (2014) The Quantitative Architecture of Centromeric Chromatin. *eLife*, DOI: 10.7554/eLife.02137.