Human Artificial Chromosome with Regulated Centromere: A Tool for Genome and Cancer Studies

Natalay Kouprina,*† Nikolai Petrov,‡ Oscar Molina,§ Mikhail Liskovykh,† Elisa Pesenti,† Jun-ichirou Ohzeki,† Hiroshi Masumoto,*† William C. Earnshaw,*§ and Vladimir Larionov*†

†Developmental Therapeutics Branch, National Cancer Institute, NIH, Bethesda, Maryland 20892, United States
‡Josep Carreras Leukaemia Research Institute, School of Medicine, University of Barcelona, Casanova 143, 08036 Barcelona, Spain
§Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, Scotland
¶Laboratory of Chromosome Engineering, Department of Frontier Research and Development, Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan

ABSTRACT: Since their description in the late 1990s, Human Artificial Chromosomes (HACs) bearing functional kinetochores have been considered as promising systems for gene delivery and expression. More recently a HAC assembled from a synthetic alphoid DNA array has been exploited in studies of centromeric chromatin and in assessing the impact of different epigenetic modifications on kinetochore structure and function in human cells. This HAC was termed the alphoidtetO-HAC, as the synthetic monomers each contained a tetO sequence in place of the CENP-B box that can be targeted specifically with tetR-fusion proteins. Studies in which the kinetochore chromatin of the alphoidtetO-HAC was specifically modified, revealed that heterochromatin is incompatible with centromere function and that centromeric transcription is important for centromere assembly and maintenance. In addition, the alphoidtetO-HAC was modified to carry large gene inserts that are expressed in target cells under conditions that recapitulate the physiological regulation of endogenous loci. Importantly, the phenotypes arising from stable gene expression can be reversed when cells are “cured” of the HAC by inactivating its kinetochore in proliferating cell populations, a feature that provides a control for phenotypic changes attributed to expression of HAC-encoded genes. AlphoidtetO-HAC-based technology has also been used to develop new drug screening and assessment strategies to manipulate the CIN phenotype in cancer cells. In summary, the alphoidtetO-HAC is proving to be a versatile tool for studying human chromosome transactions and structure as well as for genome and cancer studies.

KEYWORDS: human artificial chromosome, HAC, gene delivery vector, kinetochore, chromosome instability, CIN

Since their first description in the late 1990s, Human Artificial Chromosomes (HACs) carrying a functional kinetochore have been considered as a promising system for gene delivery and expression with the potential to overcome several problems caused by the use of viral-based gene transfer systems.1,6 HACs avoid the limited cloning capacity, lack of copy number control, and insertional mutagenesis during integration into host chromosomes that have hampered the use of viral vectors. Although it is not as routine as transfection or infection with a virus, HACs along with the genetic loci can be transferred from one cell to another in the laboratory. The ability of HACs to carry entire genomic loci with all regulatory elements allows them to faithfully mimic the normal pattern of natural gene expression. Moreover, not only single genes but also groups of genes encoding complex pathways can be carried on a single HAC.

Both top-down and the bottom-up approaches have been used to construct HACs. Top-down approaches are based on telomere-associated chromosome fragmentation in the homologous recombination-proficient chicken DT40 cell line. Using top-down approaches, linear minichromosomes ranging in size from 0.5 Mb to 10 Mb have been generated so far. Such derivatives of natural human chromosomes have been produced from chromosome X,9,10 chromosome Y,11–13 chromosome 22,14 chromosome 21,3,15 and chromosome 14.16 These minichromosomes retain a natural centromere and are mitotically stable in human cells during cell propagation with their size falling below ~300 Kb.11,13

This review focuses on the bottom-up approaches for de novo HAC construction. More precisely, we will focus on a de novo constructed synthetic HAC generated from an alphoid DNA array assembled from a 348 bp human centromeric repeat, and describe the multiple applications of this HAC for genome and cancer studies.

1. BOTTOM UP OR DE NOVO CONSTRUCTION OF HUMAN ARTIFICIAL CHROMOSOMES

1.1. Construction of Human Artificial Chromosomes from Natural Alphoid DNA. In the late nineties two groups independently reconstituted functional artificial human chromosomes. The Willard and Masumoto laboratories and their
respective coauthors were the first to show that alphoid DNA, the primary DNA satellite repeats in human centromeres, can "seed" formation of a functional kinetochore when transfected into the human fibrosarcoma HT1080 cell line.\textsuperscript{17,18} Subsequently, other groups have confirmed this observation and reported that natural higher-order repeat (HOR) arrays composed of 171 bp alpha-satellite monomer units containing CENP-B boxes,\textsuperscript{19} that are tandemly arranged in a directional head-to-tail fashion are sufficient for de novo HAC formation.\textsuperscript{20} These HACs ranged in size from 1 Mb to 10 Mb due to amplification of the input alphoid DNA during HAC establishment and were stably maintained as single copy episomes in the nucleus of transfected cells. HACs engineered by the bottom-up approach can be circular or linear if telomeric sequences are inserted. The resulting HACs are equally stable as both possess a functional centromere and therefore can autonomously replicate and segregate.\textsuperscript{1,2,4−8,21−30}

The first HACs were constructed from 50 to 100 kb alphoid DNA fragments identified in existing Yeast Artificial Chromosome (YAC) or Bacterial Artificial Chromosome (BAC) libraries. Using ligation-based reconstruction methods with alphoid DNA repetitive units, several studies proved that alphoid DNA bearing CENP-B boxes were required for de novo HAC formation.\textsuperscript{21,29,31} However, because the complete DNA sequence of these fragments was unknown, definitive studies of the structural requirements for de novo kinetochore formation were not feasible.

\subsection*{1.2. Construction of Human Artificial Chromosomes from Alphoid Synthetic Repeats.}

To address this problem, our group developed a method, RCA-TAR, to construct synthetic alphoid DNA arrays with the possibility to manipulate alphoid DNA arrays to introduce precisely defined DNA sequence variation.\textsuperscript{32,33} RCA-TAR involves two steps: rolling circle amplification (RCA) of alphoid DNA oligomers that may be as small as a dimer (348 bp) and subsequent assembly of the amplified fragments (1−3 kb) up to 140 kb by transformation-associated recombination (TAR) in yeast.\textsuperscript{34−38} Because the alphoid DNA repeat sequence can be altered before the amplification step, it is possible with this approach to introduce mutations, including defined deletions, insert recognition sites for DNA-binding proteins, or otherwise vary the alphoid DNA sequence and/or structure. Using the RCA-TAR method, synthetic alphoid DNA arrays from 50 kb to 140 kb have been generated from single alphoid repeats and used

![Figure 1. alphoid\textsuperscript{tetO}.HAC formation and detection. (a) de novo generation of the human artificial chromosome (HAC) via bottom−up approach using a synthetic alphoid DNA dimer. The first step includes amplification of the dimer by rolling circle amplification (RCA) up to 1−3 kb in size fragments. One monomer of the dimer derived from the chromosome 17 alphoid-type I 16-mer unit and contains a CENP-B box. The second monomer is a wholly synthetic sequence derived from alphoid DNA consensus, with sequences corresponding to the CENP-B box replaced by a 42 bp tetO motif. The second step includes assembly of the RCA-amplified fragments in yeast cells on the vector containing alphoid-specific hooks by transformation-associated recombination (TAR). End-to-end recombination of RCA-generated alphoid DNA fragments followed by interaction of the recombined fragments with the vector hooks results in the rescue of approximately a 50 kb synthetic alphoid array as a circular molecule in yeast. At the third step, HAC was formed in human cells by 50 kb transfected DNA multimerization up to 1.1 Mb in size. (b) Immunofluorescence on a metaphase chromosome spread of a cell containing the alphoid\textsuperscript{tetO}.HAC in human HT1080 cells. The HAC was stained with the centromeric CENP-C protein (red) and tetR-EYFP (green) to detect the HAC. (c) FISH analysis of the alphoid\textsuperscript{tetO}.HAC in hamster CHO cells. FISH analysis was performed using the PNA labeled probe for tetO sequences (green).]
for HAC formation. This accomplishment has made it possible to begin to analyze the genomic and proteomic requirements for de novo kinetochore formation and maintenance.

### 1.3. Construction of Synthetic Human Artificial Chromosome with a Conditional Centromere. A collaborative effort of three laboratories led to the generation of a circular HAC with a conditional centromere using the RCA-TAR technology (Figure 1). This HAC has been instrumental in resolving the role for various chromatin structures on kinetochore function.39–42 The HAC includes approximately 6000 copies of the tetracycline operator (tetO) sequence incorporated into a synthetic monomer synthesized according to the Choo consensus sequence43 and paired with a natural monomer from chromosome 17 containing a CENP-B box to make the 348 bp dimer unit.44 Ten copies of this dimer cloned into pBluescript were amplified by RCA-TAR up to 50 kb and used as input DNA for HAC formation after transfection into human HT1080 cells. This novel synthetic HAC was termed the alphoidtetO-HAC, as the synthetic monomers each contained a tetO sequence in place of the CENP-B box. Because tetO is bound with high affinity and specificity by the tet repressor (tetR), the final 1.1 Mb alphoid array containing tetO sequences in this HAC44 can be targeted specifically with tetR-fusion proteins.45 Importantly, the structure and functional domains of this HAC remain unchanged after several rounds of transfer into different host mammalian cells by microcell-mediated cell fusion.46 3D-CLEM confirmed that this HAC contains typical chromosome compartments (centromere, kinetochore, scaffold, and periphery).46 Knowledge of the alphoidtetO-HAC structure44 allowed us to monitor HAC integrity during different manipulations or its transfer from one cell line to another.

The synthetic HAC allows the targeted manipulation of chromatin within a single functional centromere without affecting the endogenous chromosomes of the host cell. This technology (termed epigenetic engineering) offers a unique approach to dissect the epigenetic factors that control centromere and kinetochore assembly and function to allow faithful chromosome segregation47–51 reviewed in refs 47,49, and 52. As described below, a modified version of the alphoidtetO-HAC with a “landing pad” for the insertion of genomic copies of genes is also useful as a full-length gene delivery vector for gene functional analyses.45,53–60

### 2. HUMAN ARTIFICIAL CHROMOSOME WITH REGULATED CENTROMERE FOR GENE FUNCTIONAL ANALYSES

#### 2.1. Conversion of AlphoidtetO-HAC into a Gene Delivery Vector. The ability of HACs to carry entire genomic loci with all regulatory elements should in principle allow them to faithfully mimic normal patterns of natural gene expression for inserted genomic loci. Several laboratories have constructed HACs with a single loxP gene loading site53,61–64 that was used for gene insertion and expression (reviewed in refs 1,2, and 4–8). However, the alphoidtetO-HAC with its conditional kinetochore allows unique controls for gene function analysis. Since its kinetochore can be specifically inactivated, this HAC provides the possibility to compare the phenotypes in target stable human human cell lines with and without a cloned genetic locus. In cells carrying the HAC, the consequences of expression or silencing of the cloned locus can be assessed. Then, by targeting the tetO sequences using chromatin modifiers that lead to centromere inactivation, populations are “cured” of the HAC as a result of loss during growth in culture. The return to the original basal phenotype in the same clone of cells is a rigorous control against unsuspected effects caused by the presence of the HAC, and thereby allows more confident interpretation of gene complementation and function studies.

To adopt the existing alphoidtetO-HAC for gene delivery and expression studies, the HAC was transferred from human HT1080 cells to recipient chicken DT40 cells by two rounds of Microcell-Mediated Chromosome Transfer (MMCT)65 using CHO hamster cells as an intermediate host. A Lox-P-S' HPRT-Hyg-TK cassette was targeted into the HAC by homologous recombination in DT40 cells. Clones with a single lox-P site were identified and the retrofitted HAC was transferred back to CHO cells.45 After all manipulations, the alphoidtetO-HAC with a single loxP site was mitotically stable in CHO cells. Cassette-associated transgenes, Hyg and TK, are stably expressed in CHO cells.45 In those cells, a gene of interest can be easily inserted into the loxP site of the HAC by Cre-mediated recombination. Because CHO cells form microcells at high frequency in response to Colcemid,66 the HAC can be easily moved from donor CHO cells into different recipient human or mouse cell lines via MMCT for complementation and function analysis.3,15,61–64 We also transferred the alphoidtetO-HAC containing an EGFP transgene inserted into the loxP site into mouse ES cells and assessed whether the presence of this extra chromosome affects their pluripotent properties. The alphoidtetO-HAC-bearing ES cells were indistinguishable from their wild-type counterparts: they retained self-renewal potential and full capacity for multilineage differentiation during mouse development, and the HAC itself was mitotically and transcriptionally stable during this process60 (Figure 2). Our data provided the first example of a fully synthetic chromosome behaving like a normal chromosome in cells of living animals, thereby opening new opportunities for functional genetic studies in laboratory animals as well as stem cell-based regenerative medicine.

#### 2.2. Re-engineering the alphoidtetO-HAC Vector To Allow a Unique Control for Gene Phenotypes by a
Simple Doxycycline Manipulation. As described above, the alphoidtetO-HAC can be easily eliminated from cells by inactivation of the HAC kinetochore via binding of chromatin modifiers, that induce either hypertranscription (tTS) or heterochromatinization (tTA) of its centromeric sequences (see also section 3.2). Induction of HAC loss provides the possibility to control for phenotypes induced by the presence of the HAC in a cell population and control more rigorously for the effects of genetic loci loaded onto the HAC. However, such inactivation of the HAC kinetochore initially required transfection of cells by lipofectamine or retroviral vectors in order to introduce exogenous DNAs expressing the tTS or tTA. Such transfections have the potential to cause insertional mutagenesis.

We therefore re-engineered the alphoidtetO-HAC vector to allow centromere inactivation without transfection of exogenous chromatin modulators. In this modified HAC vector, a cassette expressing the tTS was inserted into the loxP gene-loading site along with the genetic locus of interest. In the absence of doxycycline, expression of the tTS generates a self-regulating fluctuating heterochromatin state on the alphoidtetO-HAC. When the tTS binds to the tetO array, it induces heterochromatin formation that both inactivates the centromere and also inactivates expression of the tTS gene itself. The net result is a fast and strong silencing of the genetic locus being tested on the HAC without a significant effect on HAC segregation (the tTS silences itself before the heterochromatin is "deep" enough to inactivate the centromere). Silencing of the test transgene is reversible as its expression can be readily recovered by adding doxycycline, which blocks binding of the tTS to the tetO array. However, this system does not allow us to return cells to their "ground state" by curing them of the HAC. We therefore designed a second re-engineered alphoidtetO-HAC vector that allows fast and highly efficient inactivation of the HAC kinetochore. In this vector, a tTA(VP64) cassette carrying four tandem repeats of the VP16 domain is inserted into the single loxP gene-loading site along with the genetic locus of interest (Figure 3a). In medium plus doxycycline, tTA(VP64) binding to the alphoidtetO array is blocked. Thus, the genetic locus being tested is expressed under control of its own promoter and the HAC is stable (Figure 3b). In the absence of doxycycline, tTA(VP64) binds to the alphoidtetO array inducing a burst of transcription that leads to rapid disruption of kinetochore function followed by the HAC loss (Figure 3c,d).

These modified alphoidtetO-HAC-based vectors containing a single copy of the tTS or tTA(VP64) represent powerful tools for gene function studies because they allow us to control gene activity or to cure cells of the HAC by simple addition or removal of doxycycline.

2.3. Pericentromeric Gamma-Satellite DNA and tDNA Prevent Heterochromatin Spreading and Protect Gene Expression from AlphoidtetO-HAC. In the alphoidtetO-HAC vector, insertion of a gene-loading site into the alphoidtetO array created a domain that is permissive for transcription. However, the long-term stability of this transcriptionally active state within centromatin was unknown. It is widely known that transgenes inserted into ectopic sites in human cells tend to be silenced over time. Because the alphoidtetO expression domain is flanked by heterochromatin, which has a propensity to spread, we wondered whether chromatin barriers or insulator sequences might be essential for stable transgene expression in alphoidtetO-HAC centromatin.

So far, very few elements with such activity have been described (reviewed in refs 68 and 69). We therefore set out to screen for insulator sequences that would protect transgenes from epigenetic silencing in the alphoidtetO-HAC. Pericentromeric regions of mammalian chromosomes contain repetitive DNA sequences, including gamma-satellite DNA, that exhibit a high rate of evolutionary change. However, the exact role of these sequences with respect to kinetochore/heterochromatin structure and function remains unknown. We developed a system to study the function of gamma-satellite DNA in...
maintaining active chromatin domains. The system involved in vitro amplification of a 220 bp gamma-satellite repeat up to 3 kb, 9 kb, and 24 kb arrays by RCA.52 These arrays were then integrated into the RL5 locus on chromosome 4 in mouse erythroleukemia cells.70 Indeed, the human pericentromeric gamma-satellite DNA sequences promoted a transcriptionally permissive chromatin conformation in an adjacent transgene and protected the transgene from epigenetic silencing.71 In hematopoietic cells, the antisilencing and heterochromatin-arresting activities of gamma-satellite DNA require the binding of Ikaros, a protein that regulates hematopoiesis.72

The structural conservation of gamma-satellite in pericentromeric regions of most chromosomes in humans and nonhuman primates suggests that gamma-satellite arrays may have a structural and/or functional role in the centromere, possibly preventing the spreading of pericentric heterochromatin into chromosomal arms. As such, the gamma-satellite DNA resembles a barrier element.74 Together, these observations suggest that gamma-satellite DNA plays a role in separating specific domains of chromatin/heterochromatin and might therefore protect transgenes inserted into the HAC from silencing.

In budding and fission yeast, tRNA genes can function as chromatin barrier elements.73,74 However, until recently there was no experimental evidence that tRNA exhibits barrier activity in mammals. We therefore investigated whether tRNA genes can function as chromatin barrier elements. Indeed, we and others showed that functional copies of tRNA genes function as barrier insulators in mammalian cells.75,76

In a follow-up study, we compared the activity of different chromatin insulators on the HAC vector. We compared the effects of three chromatin insulators, cHS4,77 gamma-satellite DNA,71 and tDNA5,76 on the expression of an EGFP transgene inserted into the loxP site of the alphoidtetO-HAC vector. A tDNA insulator consisting of two functional copies of tRNA genes showed the highest barrier activity, followed by gamma-satellite DNA and cHS4.77 These results showed that proximity to centromochromatin does not protect genes lacking chromatin insulators from epigenetic silencing. Therefore, strategies for transgenesis using HAC vectors, including alphoidtetO-HAC, should include barrier elements such as tDNA and gamma-satellite DNA to prevent gene silencing (Figure 3).

2.4. An Improved Microcell-Mediated Chromosome Transfer (MMCT) Technique for HAC Transfer to Recipient Cells. Gene loading into the HAC is most easily performed in CHO cells, so MMCT transfer of alphoidtetO-HAC carrying a genetic locus to human gene-deficient cells is a key step for gene complementation and function analysis. Though the MMCT method was developed more than 40 years ago,55 two main limitations make the method tedious. First, the frequency of HAC transfer from donor CHO cells into recipient cells is very low. Second, MMCT is not applicable for all types of recipient cells, particularly those whose fusion with microcells is very inefficient. To minimize these problems, we have optimized the MMCT protocol.59 In the new protocol the following modifications have been made: (i) Colcemid, a microtubule inhibitor that arrests cells at metaphase, was replaced with TN-16 + Griseofulvin; (ii) Cytochalasin B, an actin inhibitor that induces actin cytoskeleton disassembling, was replaced with Latrunculin B. Such modifications in combination with a collagen/laminin surface coating, that improves adherence to the culture flask, increases the efficiency of HAC transfer to recipient cells at least 10 times (Figure 4). Moreover, the novel protocol is also less damaging to HAC than the standard MMCT method.59 The modified protocol was successfully applied to alphoidtetO-HAC transfer to several recipient cell lines, including human mesenchymal stem cells and mouse embryonic stem cells. Recently another group78 also developed a highly efficient chromosome transfer method, called retro-MMCT. This is based on Chinese hamster ovary cells (CHO) expressing envelope proteins derived from ecotropic or amphotropic murine leukemia viruses. Using this method, the HAC was transferred to mouse embryonic fibroblasts with 26.5 times greater efficiency than that obtained using conventional MMCT. Thus, both modified MMCT methods allow a significant improvement of HAC transfer to various types of target cell.

2.5. Expression of a Genomic Copy of Human Genes from alphoidtetO-HAC. The alphoidtetO-HAC has been used to deliver genomic loci carrying two human average-size cancer-associated genes, VHL (~25 kb) and NBS1 (~60 kb), and complement genetic deficiencies in cell lines derived from the patients with deficiencies in either VHL or NBS1 using the strategy summarized in Figure 5.53 Mutations in the VHL gene lead to von Hippel–Lindau syndrome (VHL; MIM 193300). Mutations in the NBS1 gene lead to Nijmegen breakage syndrome (NBS; MIM 251260). Functional expression of
pVHL and pNBS1 in recipient cells and rescue of the mutant phenotypes was demonstrated. Importantly, the mutant phenotypes were restored after specific elimination (“curing”) of the HAC from the cells following targeted inactivation of its kinetochore.

In another study, the alphoidtetO-HAC vector was used for delivery and expression of a 90 kb genomic copy of the BRCA1 gene into the gene-deficient human cells. BRCA1 is involved in many disparate cellular functions, including DNA damage repair, cell-cycle checkpoint activation, gene transcriptional regulation, DNA replication, centrosome function and others. However, no unifying mechanistic framework that links the reported biochemical activity of BRCA1 to its tumor suppressor function has yet been identified. After insertion of a full-length BRCA1 gene into the loxP site of the HAC and its transfer into BRCA1-deficient human cells, a battery of known functional tests was carried out to demonstrate functionality of the transgene. Then, specific experiments were performed to investigate a recently proposed role of BRCA1 in maintenance of global heterochromatin integrity. We demonstrated that BRCA1 deficiency results in an elevated level of transcription of diverged pericentromeric repeats forming constitutive heterochromatin as well as higher-order alpha-satellite repeats (HORs). Together, these could contribute to chromosome instability observed in the Brca1-deficient cells. Our data extended previous observations that BRCA1 may promote heterochromatin formation in a genomic locus-specific manner and support the hypothesis that epigenetic alterations of these regions initiated in the absence of BRCA1 could impact other gene(s) and nuclear structural interactions, leading to cell transformation.

All genes described above were isolated from the total human genome by a cloning technique that is based on transformation-associated recombination (TAR) in the yeast Saccharomyces cerevisiae. TAR cloning allows selective recovery of chromosome segments that are up to 300 kb in length from complex genomes. A modified CRISPR/TAR technology improved the efficiency of TAR cloning, with up to 32% of yeast transformants containing the gene of interest. The alphoidtetO-HAC-based delivery vector combined with TAR cloning can be widely used to characterize gene function and genome variation, including mutations and even polymorphic structural rearrangements in patient genomes.

2.6. Construction of AlphoidtetO-HAC with Multi-integration Site To Assemble Large Entire Genomic Loci and Engineer Synthetic Chromosomes with a Predetermined Set of Genes. The assembly of multiple genes or entire loci or transfer of multiple genes into desired cells using the HAC vector has multiple applications in functional genomics. A few years ago several laboratories suggested using artificial chromosomes (AC) to assemble large entire genomic loci or several genes on the same HAC molecule using a multi-integrase system. Therefore, construction of an alphoidtetO-HAC containing a multi-integration site was our next step. To reach this goal, we designed an iterative integration system (IIS) that utilizes three recombinases: Cre, ΦC31, and ΦBT1. This IIS-alphoidtetO-HAC system allows assembly of functional genes on the same HAC DNA molecule and has several notable advantages that set it apart from other artificial chromosome-based systems. These include the assembly of an unlimited number of genomic DNA segments and the opportunity to

Figure 5. (a) Transformation-associated recombination (TAR) cloning of a gene of interest from total genomic DNA with a TAR vector containing YAC and BAC cassettes and two unique targeting sequences (hook1 and hook2) (in green) homologous to the 5′ and 3′ ends of a gene. Genomic DNA and a linearized TAR vector are co-transformed into the yeast Saccharomyces cerevisiae cells. Recombination between targeting sequences in the vector and the targeted sequences in the genomic DNA fragment leads to rescue of a gene as a circular TAR/YAC/BAC molecule. (c) Transfer of the TAR-isolated molecules containing a region of interest from yeast cells to bacterial cells by electroporation with the followed isolation of BAC DNA by a standard procedure. (d) Loading of a TAR-isolated gene of interest into the single loxP site of the HAC by Cre-loxP-mediated recombination in hamster cultured cells. (e) The assembled HAC contains the rTA cassette and a gene of interest flanked by insulator sequences to protect the gene from epigenetic silencing.
remove mis-incorporated DNA segments. As for other alphoidtetO-HACs, the IIS-alphoidtetO-HAC can be “cured” from dividing cell populations, allowing target cells to revert to their pretransformed state. In future studies, the IIS-alphoidtetO-HAC may have the potential to engineer synthetic chromosomes with a predetermined set of genes, thereby allowing investigation of complex biomedical and gene regulation pathways.

3. USE OF THE ALPHOIDtetO-HAC FOR EPIGENETIC ENGINEERING OF CHROMATIN WITHIN A SINGLE FUNCTIONAL CENTROMERE

3.1. Human Centromeres: What We Know and What We Do Not. Centromeres define the site of the assembly of the kinetochore, a multiprotein complex that directs chromosome segregation by binding microtubules. In humans, endogenous centromeres typically form on chromosome-specific higher-order alphoid DNA arrays, that are composed of 171 bp alpha-satellite monomer units tandemly arranged in a directional head-to-tail fashion. The remarkable diversity in sequence composition of centromeres across species lies in contrast to their common function as a platform for kinetochore assembly. Indeed, evidence of centromere inactivation on stable dicentric chromosomes without the loss of the underlying DNA sequences led to the widespread belief that centromere specific is regulated by chromatin modifications (epigenetic). This was strongly supported by the discovery of stable chromosomes containing neocentromeres that completely lack alphoid DNA. The basis of this epigenetic regulation is still under investigation, but it was suggested that one key factor could be binding of the centromere-specific histone H3, CENP-A.

Microscopic investigation of stretched kinetochore fibers revealed that blocks of CENP-A nucleosomes are interspersed with H3 nucleosomes that contain transcription-associated modifications. This special chromatin, which has been termed “centrochromatin”, is flanked by constitutive heterochromatin and suggests a functional link between the local chromatin environment and kinetochore function. These observations raise several questions about the exact nature of the chromatin that specifies kinetochore assembly and propagation: (1) For centromeres, aside from CENP-A, what combination of histone modifications defines the elusive epigenetic state that is centrochromatin? (2) Can histone modifications be manipulated to turn normal chromatin into centrochromatin and to reverse—to inactivate—the established centromeres? (3) What barrier prevents heterochromatin spreading into centrochromatin? Answering these questions with native human chromosomes is extremely challenging. Aside from the fact that the sequence of centromeric DNA arrays remains largely uncharted territory, the use of RNAi or inhibitors to manipulate the composition of centrochromatin is fraught with difficulties beyond the usual specificity issues associated with both of those approaches.

Thus, despite their importance for proper chromosome segregation, the structure and function of human centromeres remain relatively poorly understood until now. This is at least in part because of the highly repetitive nature of centromeric DNA, which has made it difficult to clone, sequence, and study. Indeed, only the Y chromosome centromere has been recently sequenced in its entirety. The alphoidtetO-HAC, which contains tetO sequences that can be specifically targeted with chromatin modifiers fused to the tetR provides a unique tool for dissecting the chromatin (epigenetic) requirements for faithful chromosome segregation (reviewed in refs 49 and 52).

3.2. Centromere Chromatin State and Kinetochore Function are Interconnected. Since centromeres lie within constitutive heterochromatin regions, it was long assumed that centromeres themselves were a special form of heterochromatin. The first study using the alphoidtetO-HAC to manipulate the epigenetic state of centromeric chromatin examined the role of heterochromatin at centromeres. Tethering a heterochromatin-seeding transcription repressor (TS) into the alphoidtetO-HAC kinetochore resulted in a dramatic loss of kinetochore function over the course of several cell divisions. At a molecular level, heterochromatin formation caused a loss of CENP-A correlated with a decrease in H3K4me2 and an increase in H3K9me3 levels paralleled by recruitment of HP1 and chromosome segregation defects. Subsequent investigations using the alphoidtetO-HAC revealed that heterochromatin-associated loss of kinetochore structure was a hierarchical process, with CENP-C and CENP-H being rapidly displaced, and preceding a more gradual loss of CENP-A (Figure 7). In this study, HAC kinetochore function collapsed at a time when CENP-A was still present at the centromere, confirming that CENP-A alone is not sufficient for proper kinetochore function in an unfavorable chromatin environment. Together, these
unfavorable chromatin context. 

(a) A simplified structure of kinetochore in the alphoid\textsuperscript{tetO-HAC}. (b) A fusion of tet-repressor with the transcriptional organization. (a) A simplified scheme illustrates how protein targeting works to study kinetochore CENP-A is still present at the centromere con-

ning that HAC kinetochore function collapses at a time when kinetochore. In contrast, the CENP-A protein remains. It is worth noting that seeding heterochromatin within centrochromatin is detrimental to kinetochore function

studies provided the first conclusive evidence that centromere chromatin state and kinetochore function are tightly interconnected.

3.3. Human Kinetochores Have a Remarkable Plasticity That Tolerates Profound Changes to Their Chromatin Environment but Critically Sensitive to the Level of Centromeric Transcription. The demonstration that seeding heterochromatin within centromeratin is detrimental to kinetochore function fits well with previous observations linking the centromeratin domain to transcrip-
tional activity. The first of these was the discovery of H3K4me2 in centromeres.\textsuperscript{25} This modification is typically found at the $S^i$ region of poised or actively transcribed genes.\textsuperscript{27} Subsequent experiments (and references therein) revealed ongoing transcription at centromeres in various organisms, including humans and C. elegans.\textsuperscript{98,99} Remarkably, this transcription occurs during mitosis.\textsuperscript{98,100} The role of the transcripts is still debated.\textsuperscript{101–103}

We used the alphoid\textsuperscript{tetO-HAC} system to address specifically the question of how transcription and transcriptionally permissive chromatin influence kinetochore maintenance. Tethering a mild transcriptional activator (the minimal activation domain of NF-κB-p65) within the HAC centromere revealed a remarkable degree of plasticity of kinetochore toward the underlying chromatin. Despite the induction of local histone H3K9 hyper-acetylation and elevated RNA polymerase activity (a $\sim$10-fold elevation in transcript levels), levels of CENP-A or CENP-C at the HAC were virtually unaffected. Indeed, these engineered HACs retained a fully functional kinetochore and were stably propagated through subsequent cell divisions.\textsuperscript{37} In striking contrast, tethering the more potent transcriptional activator VP16 (herpes virus VP16 activation domain) allowed histone H3K9 hyperacetylation comparable to that seen after tethering p65 at the HAC, but resulted in an $\sim$150-fold elevation in transcripts, approaching the level of transcription of an endogenous housekeeping gene. VP16 rapidly inactivated the HAC kinetochore due both to a complete block of CENP-A loading, and to the specific and rapid stripping of preassembled CENP-A from the HAC kinetochore.\textsuperscript{47}

Detailed analysis confirmed that centromeratin of the HAC, like that of endogenous chromosomes, resembles domains found in the body of some actively transcribed housekeeping genes, containing H3K4me2 and H3K36me2. To study the functional link between this transcription-associated chromatin environment and kinetochore maintenance, we depleted H3K4me2 specifically from the alphoid\textsuperscript{tetO-HAC} centromere by tethering the H3K4me2-specific demethylase LSD1.\textsuperscript{39} Tethering LSD1 to the alphoid\textsuperscript{tetO-HAC} specifically and efficiently depleted H3K4me2 from its centromere, leaving other native centromeres untouched. H3K4me2 depletion caused a drop in transcription of the centromeric $\alpha$-satellite DNA accompanied by a loss of CENP-A. The latter was explained by a failure of the HAC centromere to efficiently recruit the CENP-A-specific chaperone HJURP. Kinetochores depleted of H3K4me2 were inactivated over the course of the next few days, probably because centromeres contain more CENP-A molecules than are required for kinetochore assembly.\textsuperscript{104,105}

Together, these results provide a functional link between centromeric chromatin, $\alpha$-satellite DNA transcription, maintenance of CENP-A levels, and kinetochore stability.

3.4. Histone H3K9 Acetyl/Methyl Balances Regulate CENP-A Assembly. The notion that the balance between heterochromatin and transcription are important in regulating \textit{de novo} CENP-A assembly and kinetochore formation on human centromeric alphoid DNA arrays was explored in a study inspired by previous experiments with the alphoid\textsuperscript{tetO-HAC}. This study started with the observation that HACs could be formed in HT1080 fibrosarcoma cells, but essentially never in other cell lines, such as HeLa. Careful analysis revealed that HT1080 appears to be hypomorphmic for the H3K9 methyltransferase Suv39H1. This led to experiments in
which it was shown that “open” chromatin induced by tethering histone acetyltransferases (HATs) to alphoid DNA arrays could induce assembly of CENP-A and other kinetochore proteins at ectopic alphoid<sup>tetO</sup> sites inserted into chromosome arms. These results confirmed that <em>de novo</em> CENP-A assembly requires an “open” chromatin state. Importantly, a brief pulse of histone H3 acetylation was observed <em>in vivo</em> at the end of mitosis, coincident with the onset of deposition of new CENP-A molecules at centromeres.<sup>106,107</sup> Importantly, transfection of alphoid<sup>tetO</sup> DNA arrays together with tetR-HAT fusions significantly increased the efficiency of HAC formation and <em>de novo</em> stable CENP-A assembly, even enabling HAC formation in HeLa cell, thereby breaking a barrier to HAC formation in cells other than HT1080.<sup>48</sup>

### 3.5. Centromeric Transcription and H3K9 Acetylation Keep Heterochromatin from Invading Centromatin.

To ask whether it was α-satellite transcription <em>per se</em>, or the chromatin environment generated as a result of transcription that is important for kinetochore maintenance, we recently designed a novel “<em>in situ</em> epistasis” assay in which the specific H3K4me2 demethylase LSD2 plus synthetic modules with competing activities were simultaneously targeted to the synthetic alphoid<sup>tetO</sup>-HAC centromere.<sup>41</sup> As in our previous study, H3K4me2 loss decreases centromeric transcription, CENP-A assembly, and HAC stability associated with spreading of the heterochromatin mark H3K9me3 across the HAC centromere (Figure 8). Surprisingly, cotethering of LSD2 plus the HAT subunit CENP-28/Eaf6 (which stimulated transcription associated with H4K12 hyperacetylation), did not rescue the phenotype observed after tethering LSD2 alone. In contrast, cotethering of LSD2 plus the p65 activation domain (which stimulated transcription associated with H3K9 hyperacetylation), did rescue kinetochore stability and function. These results suggest that H3K9 hyperacetylation might form a barrier to prevent heterochromatin invasion of human centromeres by blocking formation of H3K9me3 and by allowing proper H3.3 turnover, which in turn allows proper chromatin remodeling for <em>de novo</em> CENP-A deposition and long-term kinetochore maintenance.<sup>41</sup> Other recent studies showed that one HAT involved in centromere licensing for <em>de novo</em> CENP-A assembly is KAT7, which acetylates histone H3 specifically at lysine 14.<sup>50</sup> KAT7 recruits the chromatin remodeling factor RSF1 through chromatin acetylation, and these proteins prevent heterochromatin spreading by a histone eviction or turnover mechanism (Figure 9). Therefore,

### 3.6. CENP-C and CENP-I Are Key Connecting Factors for Kinetochore and CENP-A Assembly. Although it is generally accepted that CENP-A is an epigenetic mark that

![Figure 8](image1.png)  
**Figure 8.** Model of the effects observed after engineering the alphoid<sup>tetO</sup>-HAC by <em>in situ</em> epistasis assays. (a) Tethering LSD2 (H3K4 demethylase) and p65 (H3R9 HAT). Chromatin competent for RNAP II transcription and chromatin remodeling, allowing <em>de novo</em> CENP-A loading; (b) Tethering LSD2 (H3K4 demethylase) and CENP-28 (H4K12 HAT). Chromatin competent for RNAP II transcription but defective histone H3.3 turnover, abolishing new CENP-A loading, favoring heterochromatin spreading into centromatin and ultimately inactivating the HAC kinetochore.

![Figure 9](image2.png)  
**Figure 9.** A model of the balance between centromatin and heterochromatin. CENP-C binds to centromatin through CENP-A nucleosome and/or CENP-B. CENP-C, CENP-I, and M18BP1 interact with each other. M18BP1 recruits HJURP through interactions with Mis18α (α) and Mis18β (β). M18BP1 also interacts with acetyltransferase KAT7. RSF1 assembled on the acetylated chromatin and prevents heterochromatin spreading into centromatin through a histone eviction or turnover mechanism. This mechanism may proceed in parallel with new CENP-A deposition by HJURP.

epigenetic engineering studies using the alphoid<sup>tetO</sup> HAC allowed detailed analysis to understand the specific contributions of different histone modifications and transcription on kinetochore maintenance. Overall, they revealed that there is a balance between histone modifications and transcription that promote the proper epigenetic environment for kinetochore maintenance.
Figure 10. Next-generation of synthetic human artificial chromosomes. (a) Schematic representation of the genetic and epigenetic structure of the alphoid\textsuperscript{hybrid}-HAC. Green and orange arrows represent the types of high order repeats (HORs) used for synthesizing the HAC. (b) Oligo-FISH image of a metaphase spread containing the alphoid\textsuperscript{hybrid}-HAC. Oligonucleotides recognizing tetO (red) and lacO+gal4 sequences (green) were used.

specifies centromere identity, the pathways leading to the formation and maintenance of centromere chromatin remain unclear. We previously generated cell lines bearing the tetO alphoid array at ectopic integration sites on chromosomal arms—the alphoid\textsuperscript{tetO} integrations,	extsuperscript{18} and we have recently used these to examine the regulation of CENP-A assembly and maintenance at centromeres.\textsuperscript{41} The ability of the different factors fused to tetR to nucleate centromatin was assessed by their ability to assemble CENP-A at the ectopic integration sites. Many kinetochore structural components can induce de novo CENP-A assembly at the ectopic site. These components work by recruiting CENP-C and subsequently recruiting HJURP as part of the Mis18 complex.\textsuperscript{108,109} CENP-I can also recruit M18BP1 and therefore enhances centromeric M18BP1 assembly downstream of CENP-C (Figure 9). Thus, we suggest that CENP-C and CENP-I are key factors connecting the kinetochore itself to new CENP-A assembly. To summarize, the use of the alphoid\textsuperscript{tetO} arrays (either HAC or alphoidhybrid-HAC) offers the potential to further dissect the requirement of individual histone modifications and cellular processes involved in kinetochore formation and maintenance.

3.7. Future Direction: Next-Generation Human Artificial Chromosomes Containing Alphoid Arrays with Different Targeted Sequences. Although heterochromatin is frequently found in close proximity to centromeres, its role in chromosome segregation is not yet fully understood and may vary among species. S. pombe may provide an extreme example with its dependence on heterochromatin for de novo deposition of CENP-A and for proper cohesion dynamics.\textsuperscript{110,111} In plants, heterochromatin and chromosome segregation appear to be largely unrelated: chromosome transmission is impaired by mutations that affect cohesion but not by those that affect heterochromatin formation.\textsuperscript{112} The role of pericentromeric heterochromatin in the function of human kinetochores has been studied extensively but remains poorly understood despite suggestions that heterochromatin could be involved in maintaining genomic stability and that defects in heterochromatin assembly at centromeres are associated with chromosome segregation defects and tumorigenesis.\textsuperscript{113,114} as reviewed in ref 52.

One limitation of the alphoid\textsuperscript{tetO}-HAC is the impossibility of separately engineering the two centromeric domains (kinetochore and pericentromeric heterochromatin) and study their functional interactions. Using RCA-TAR technology, we have therefore constructed a novel synthetic HAC containing two centromeric domains, based on two different synthetic alphoid DNA arrays: the alphoid\textsuperscript{hybrid}-HAC (Figure 10). As a basis for the kinetochore, we used a dimeric repeat based on chromosome 21 HOR DNA with CENP-B boxes and tet operators in alternating repeats. As a basis for the heterochromatin, we constructed an array based on 11 monomers from the chromosome 21 monomeric alpha-satellite DNA lacking CENP-B boxes and containing lac operator (lacO) and yeast transcriptional factor Gal4. The lack of CENP-B boxes should preclude CENP-A assembly on this array.\textsuperscript{29,115} The two arrays were combined by recombination in the yeast S. cerevisiae. Transfection of the hybrid array into HT1080 cells resulted in formation of the alphoid\textsuperscript{hybrid}-HAC.\textsuperscript{116} Such hybrid HACs will allow to simultaneously target different centromeric compartments independently using different targeting systems.

4. HUMAN ARTIFICIAL CHROMOSOMES FOR MEASURING CHROMOSOME INSTABILITY (CIN) IN CANCER CELLS

4.1. Chromosome Instability (CIN) as a Driver of Tumorigenesis. Whole chromosomal instability (CIN), manifested as unequal chromosome distribution during cell division, is a distinguishing feature of most cancer types.\textsuperscript{117,118} CIN is generally considered to drive tumorigenesis, but a threshold level exists whereby further increases in CIN frequency actually hinder tumor growth. Therefore, CIN can be used as a strategy for cancer therapy.\textsuperscript{119−124} At present, the rate of chromosome mis-segregation is quantified by time-consuming techniques such as coupling clonal cell analysis with karyotyping, in vitro micronuclei (MN) assays or fluorescence in situ hybridization (FISH). In addition, while
CIN is appealing for therapeutic exploitation, drugs that increase CIN beyond the therapeutic threshold are currently few in number, and the clinical promise of targeting the CIN phenotype warrants new screening efforts.

4.2. HAC-Based “Loss of Signal” Assay for Measuring Chromosome Instability and Identification of Drugs that Elevate CIN in Cancer Cells. We have developed a new quantitative and sensitive assay for measuring CIN based on the use of the alphoid\textsuperscript{hybrid}-HAC carrying a constitutively expressed shRNA against the eGFP transgene integrated into the genome of the human HT1080 cells. Thus, cells that inherit the HAC do not display green fluorescence, while cells lacking the HAC do. It is expected that the control population of untreated cells should show uniform red fluorescence. A cell population that has lost the HAC after drug treatment should be highly variable in fluorescence. The actual number and percentage of cells with the HAC-shRNA can be measured by a scanning microscope. Thus, the drugs, which increase HAC loss and, therefore, increase spontaneous chromosome missegregation rates, may be identified. (HiTS stands for High-Throughput Screening.) Natural HAC loss is a parameter of cells that lose the HAC without drug treatment while cultured without selective antibiotics. It is usually 2–5\% of cells in population.

Figure 11. Scheme of the high-throughput assay using a fluorescence microtiter plate reader to characterize chemical libraries. In a new system, the alphoid\textsuperscript{hybrid}-HAC carries a constitutively expressed shRNA against the eGFP transgene integrated into the genome of the human HT1080 cells. Thus, cells that inherit the HAC do not display green fluorescence, while cells lacking the HAC do. It is expected that the control population of untreated cells should show uniform red fluorescence. A cell population that has lost the HAC after drug treatment should be highly variable in fluorescence. The actual number and percentage of cells with the HAC-shRNA can be measured by a scanning microscope. Thus, the drugs, which increase HAC loss and, therefore, increase spontaneous chromosome missegregation rates, may be identified. (HiTS stands for High-Throughput Screening.) Natural HAC loss is a parameter of cells that lose the HAC without drug treatment while cultured without selective antibiotics. It is usually 2–5\% of cells in population.
trigger CIN when overexpressed in human cells. The CIN genes identified in this work may reveal genes that cause CIN when overexpressed in cancer, which can then be leveraged through synthetic dosage lethality (SDL) to selectively target tumors.

5. CONCLUSIONS AND PROSPECTS

The alphoidtetO-HAC has proven to be a highly versatile reagent for studies of centromere epigenetics, for the permanent or temporary introduction of genetic loci (genes including both introns, exons, and their linked control elements) into wild-type and mutant cells as well as for screens for drugs and genetic alterations that induce chromosome instability. Advances in understanding chromatin determinants required for CENP-A deposition and kinetochore assembly offer opportunities to develop protocols for more efficient HAC formation in a wide variety of cell lines. Construction of new HACs containing different targeting sites in kinetochore chromatin and pericentromeric heterochromatin will open a unique opportunity to study functional interactions between these domains. The potential of the alphoidtetO-HAC and its derivatives in synthetic biology for cell and tissue engineering is only beginning to be explored. In the future, it will be very interesting to see how the alphoidtetO-HAC and its derivatives may be used to engineer novel biosynthetic pathways and novel synthetic chromosomes for potential gene therapy needs.

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: kouprina@mail.nih.gov. Tel: +1-240-760-7325.
*E-mail: larionov@mail.nih.gov. Tel: +1-240-760-7325.
*E-mail: masumoto@kazusa.or.jp. Tel: +81-438-52-395.
*E-mail: bill.earnshaw@ed.ac.uk. Tel: +44-(0)131-650-7101.

ORCID
Natalay Kouprina: 0000-0001-8690-5625
Oscar Molina: 0000-0001-7585-4519

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, USA (V.L. and N.K.), a core grant to the BBSRC/EPSRC/MRC Synthetic Biological Research Centre at Edinburgh (E.P.), a Wellcome Primary Research Fellowship [Grant No. 107022] (W.C.E.), MEXT KAKENHI grants [16H04747, 16H01414, and 8H04721] (H.M.), the Kazusa DNA Research Institute Foundation (H.M.), and the Beatriz de Pinós postdoctoral programme of the Government of Catalonia’s Secretariat for Universities and Research of the Ministry of Economy and Knowledge (O.M.).

■ REFERENCES


