Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere

Anderly C. Chueh, Lee H. Wong, Nicholas Wong and K.H. Andy Choo*

Chromosome Research Laboratory, Murdoch Childrens Research Institute, Melbourne University Department of Paediatrics, Royal Children’s Hospital, Parkville 3052, Australia

Received August 25, 2004; Revised October 13, 2004; Accepted October 25, 2004

Human neocentromeres are fully functional centromeres that arise epigenetically from non-centromeric precursor sequences that are devoid of α-satellite DNA. Using chromatin immunoprecipitation (ChIP) and BAC-array analysis, we have previously described a 330 kb binding domain for CENP-A (a histone H3 variant that confers centromere-specific nucleosomal property) at the 10q25 neocentromere found on a chromosome 10-derived marker chromosome mardel(10). For the further detailed analysis of the CENP-A-associated chromatin, we have generated a high-resolution genomic array consisting of PCR fragments with an average size of 8 kb, providing an ~20-fold increment in analytical resolution. ChIP and PCR-array analysis reveals seven distinct CENP-A-binding clusters within the 330 kb domain, demonstrating the interspersion of CENP-A-associated nucleosomal blocks within the neocentromeric chromatin. Independent ChIP–PCR analysis verified this distribution profile and indicated that histone H3-containing nucleosomes directly intervene the CENP-A-binding clusters. The CENP-A-binding clusters are uneven in size, with the central cluster (>50 kb) being significantly larger than the flanking ones (10–30 kb), and the flanking clusters arranged in an interesting hierarchical and symmetrical configuration of alternating larger and smaller sizes around the central cluster. In silico sequence analysis indicates an ~2.5-fold increase in the prevalence of L1 retroelements within the CENP-A-binding clusters when compared with the non-CENP-A-binding regions. These results provide insight into the possible role of retroelements in determining the positioning of CENP-A binding at human neocentromeres, and that a hierarchical and symmetrical arrangement of CENP-A-binding clusters of varying sizes may be an important structural requirement for mammalian kinetochore assembly and/or to provide stability to withstand polar microtubule forces.

INTRODUCTION

The functional role of the centromere in mitotic and meiotic cell divisions is evolutionarily conserved. However, centromeric DNA sequences are highly variable across the phylogeny with no obvious conservation even among closely related species (1). The specific requirements that underpin the different types of DNA sequences that provide the templates for centromere formation remains a puzzle but point to the importance of epigenetic factors (2). The ability of non-centromeric sequences to acquire centromere activity de novo as demonstrated in both humans and Drosophila melanogaster adds direct support to the epigenetic model (3–5). Despite the lack of obvious nucleotide sequence conservation, several proteins involved in centromere function have been found to be conserved in highly divergent eukaryotic species, including Saccharomyces cerevisiae, Caenorhabditis elegans, D. melanogaster, humans and plants (6,7). These proteins are critical to the organization of the eukaryotic centromeres and pericentromeric regions into distinct conserved functional and structural domains (8,9). Centromere protein CENP-A is directly involved in the establishment of a centromere-specific chromatin structure by replacing normal histone H3 at the nucleosomal level (10–12). Defects in this protein seriously impair centromere function and result in mitotic disarray and cell death (13,14). On
human chromosomes, CENP-A has been shown to bind specifically to a subset of the tandemly repeating centromeric α-satellite sequences (15,16).

Neocentromeres are fully functional centromeres that arise from non-tandemly repetitive chromosomal sites that have not previously been shown to exhibit any centromere activity (5). This class of centromeres represents a sequence-independent epigenetic model for centromerization and provides a useful tool for mapping important centromeric chromatin domains, which has been previously hampered by the repetitive nature of mammalian centromeric DNA (11,17–19). Using a BAC-based chromatin immunoprecipitation (ChIP) and genomic array analysis, the CENP-A-binding domain has previously been mapped to a genomic segment spanning 330 kb along the 10q25 neocentromere of the mardel(10) marker chromosome (11). However, the low resolution associated with the BAC-based mapping approach (average BAC insert size of 169 kb, ranging from 71 to 372 kb) does not allow us to determine the finer structural organization of CENP-A-associated chromatin at the neocentromere.

In this study, we have developed a high-resolution ChIP and genomic array analysis strategy through the use of the genomic arrays consisting of PCR fragments with an average size of 8 kb, spanning 1 Mb across the previously characterized CENP-A domain and flanking regions. In silico sequence analysis was also performed on the data obtained from the high-resolution mapping study to search for sequence properties that may be of potential importance in establishing the specialized chromatin structure at the neocentromere.

RESULTS

Construction of 10q25 genomic PCR array

We have previously used BAC-based chromatin immunoprecipitation and array (CIA) analysis to define the protein-binding domains for CENP-A, CENP-H and HP1α along the chromatin of the 10q25 mardel(10) neocentromere (11,18). The CENP-A-binding domain was localized to a genomic region of ~330 kb on this neocentromere. However, the resolution of the genomic BACs used (average BAC insert size of 169 kb, ranging from 71 to 372 kb) was insufficient to provide further information regarding the more detailed structural arrangement of the CENP-A-containing nucleosomes within the broader CENP-A-binding domain; that is, whether the CENP-A-binding nucleosomes were present in a continuous manner and, if not, what their pattern of organization was. In order to address these questions, we constructed a genomic array consisting of overlapping PCR fragments with an average size of ~8 kb, ranging from 2 to 16 kb (Fig. 1A and B). A total of 125 fragments were generated by long-range PCR amplification, with 95% coverage of the ~1.0 Mb chromosomal region corresponding to the previously mapped CENP-A-binding and immediately flanking domains of the 10q25 neocentromere.

Variable-sized CENP-A-binding clusters within neocentromeric chromatin

ChIP was performed on Chinese hamster ovary (CHO)-human somatic hybrid cell lines CHOM10 (test) and CHON10 (control), containing the mardel(10) chromosome and normal chromosome 10, respectively, using an anti-CENP-A antibody. The immunoprecipitated chromatin was isolated using protein A–Sepharose. The DNA extracted from both input and bound fractions was randomly amplified using degenerate oligonucleotide primed-polymerase chain reaction (DOP-PCR) as probes in the dot-blotted genomic PCR array. The relative hybridization signals for the bound and input fractions in CHOM10 and CHON10 were determined and subsequently used to generate the CENP-A-binding profile across the PCR array (Fig. 2A; Supplementary Material, Table S1). A total of seven CENP-A-binding clusters were found (P < 0.05), with each of these clusters being defined by two or more PCR fragments. The difference in the ratio of hybridization signal enhancement ranges from 250 to 500%. All the CENP-A-binding clusters mapped tightly within the 330 kb CENP-A-binding domain previously localized using BACs RP11-87P3, RP11-153G5 and RP11-87E14 (11), providing independent verification for the reproducibility of the two different methods of mapping (Figs 1B and 2A).

As the array hybridization was performed in the presence of Cot-1, could the periodic pattern observed merely reflect Cot-1 suppression of signals in the ‘trough’ regions because these regions were more repetitive? Our in silico analysis (discussed later) showed no significant difference (P > 0.05) in term of the total interspersed repeat content between the CENP-A-binding clusters (66.8%) and the troughs (56.1%). If anything, the CENP-A-binding clusters contained a slightly higher overall amount of interspersed repeats (due to the L1 repeats) when compared with the trough regions that should make any Cot-1 suppression more apparent at the CENP-A-binding clusters. Thus, the periodic CENP-A distribution pattern is unlikely a result of Cot-1 suppression of signal at the regions intervening the CENP-A-binding clusters.

The estimated size for each of the CENP-A-binding clusters was shown to vary from 10.5 to 51.8 kb (Fig. 1C). The distribution profile indicated that the central cluster, spanning position from 495 to 540 kb on the 10q25 PCR array, was noticeably larger than any of the remaining clusters. Interestingly, these remaining clusters appeared to be distributed in an alternating pattern of larger and smaller blocks towards the outer edges of the overall CENP-A-binding domain. The intervening regions between the CENP-A clusters ranged in size from 20.5 to 49.5 kb, and showed no discernible pattern of size distribution. These results indicate that CENP-A does not bind uniformly across the previously defined 330 kb binding domain but binds as a series of discontinuous blocks with what appears to be a hierarchical pattern of size distribution.

Histone H3-containing nucleosomes intervene the CENP-A-binding clusters

We next investigated the nature of the chromatin that is present between the CENP-A-binding clusters, in particular the status of histone H3. Thirty-eight PCR primer sets were

86 Human Molecular Genetics, 2005, Vol. 14, No. 1
designed across the CENP-A-binding and intervening regions, and used for the semi-quantitative PCR analysis of CHOM10 DNA following ChIP with either anti-CENP-A or anti-H3 antibody (Fig. 2B). Use of the anti-CENP-A antibody resulted in the detection of the expected PCR fragments only within the CENP-A-binding clusters but not the intervening regions. Conversely, use of the anti-histone H3 antibody gave the expected PCR fragments only within the intervening regions but not the CENP-A-binding clusters. These results provide independent verification of the profile obtained in the PCR-array analysis. Additionally, they indicate that histone H3-containing nucleosomes directly constitute the regions that intervene the CENP-A-binding clusters.

High L1 content associated with the CENP-A-binding clusters

The sequence properties of the individual genomic fragment within the 10q25 PCR array were analysed in silico using the RepeatMasker program. On the basis of the results obtained from the high-resolution PCR-array CIA analysis, the output data were classified into two groups—for fragments corresponding to the CENP-A-binding clusters and for those corresponding to the non-binding regions (Fig. 3). The CENP-A-binding clusters as a group was shown to have an average AT content of 65.3% compared with the combined non-CENP-A-binding regions of 63.4%.
within the PCR array. Overall, the 10q25 genomic position and its immediately flanking regions to which CENP-A binds have a higher AT content compared with that of the rest of the genome (59.0%).

A search for GATA-like sequences was also performed across the 1.0 Mb region spanning the CENP-A-binding domain given that the consensus recognition core sequences (HGATAR; \(\text{H} = \text{T}/\text{A}/\text{C}, \text{R} = \text{A}/\text{G}\) for GATA-type zinc finger proteins (20) was recently found to be plentiful in the centromeric sequences of *Schizosaccharomyces pombe* and implicated in the regulation of CENP-A localization in this species (21). However, we did not observe any significant difference in the frequency of these sequences between the CENP-A-binding clusters (~0.001333 occurrence/bp) and non-binding regions (~0.001323 occurrence/bp). The values obtained from both groups are not significantly different to the calculated theoretical expected frequency (~0.001465 occurrence/bp).

When the composition of the interspersed repetitive elements was analysed, no significant difference was found between the CENP-A-binding clusters and the non-CENP-A-binding regions. A noted exception, however, was the L1 member of the long interspersed nuclear elements (LINEs). An average L1 density of 51.7% was observed for the CENP-A-binding clusters, which was shown to be significantly different \((P < 0.001)\) from that of the non-CENP-A-binding regions (18.6%) and the genome as a whole (16.9%). The average

![Figure 2. High-resolution mapping of CENP-A-binding domain at 10q25 neocentromere. (A) PCR-array CIA analysis. The y-axis represents the percentage difference between the normalized bound/input ratio of CHOM10 and CHON10. A mean of five independent experiments is shown, with the width of the histogram bars drawn to scale according to the size of each PCR fragment used in the array. P-values were calculated based on Student’s t-test in order to determine the statistical significance for the enrichment of individual PCR fragments. A total of seven CENP-A-binding clusters were found \((P < 0.05)\), with each of these clusters being defined by two or more PCR fragments. These clusters (denoted by the grey boxes and vertically shaded areas) are located within the 330 kb CENP-A-binding domain (indicated by a dotted line) previously defined using CIA on a BAC array. The estimated sizes of the CENP-A clusters range from 10.5 to 51.8 kb (detailed in Fig. 1C). Raw data including P-values and standard deviations are listed in Supplementary Material, Table S1. (B) ChIP–PCR analysis. Semi-quantitative PCR was performed using DNA extracted from CHOM10 chromatin immunoprecipitated using anti-CENP-A or anti-H3 antibody. A total of 38 PCR fragments corresponding to genomic segments within the CENP-A-binding clusters (shaded in grey) and the non-CENP-A-binding regions were analysed. CENP-A does not colocalize with histone H3 on any of the fragments analysed and H3-associated nucleosomes are present at the intervening region between the CENP-A-binding clusters.]
L1 densities for each of the seven CENP-A-binding clusters were as follows: 77.4, 51.7, 32.4, 57.7, 73.1, 25.5 and 48.5% (given in the order from nucleotide 0 to 1000 Mb). These results indicate that a significantly higher-than-genome-average prevalence of L1 is seen with the CENP-A-binding clusters, either collectively or individually. However, there is no obvious correlation between the size of a cluster and the density of the L1 (e.g. the largest cluster does not contain the highest density of L1).

DISCUSSION

Variable sizes of interspersed CENP-A-binding clusters arranged in a symmetrical and hierarchical distribution pattern

In contrast to the paucity of useful DNA sequence markers due to the highly repetitive nature of the normal mammalian centromeres, the fully sequenced human neocentromeres provide a useful system for delineating the detailed molecular organizational properties of the centromeric chromatin domains. CIA analysis has been successfully used for the mapping of the CENP-A-binding domains of a number of human neocentromeres, with the results showing the size range for such domains of 130 to 460 kb (11,17–19). However, the resolution of the reported CIA analyses using genomic BAC arrays provides no molecular information for possible finer subregional organizational properties for CENP-A binding at these neocentromeres. In this study, we have demonstrated the feasibility of using a genomic PCR array (containing sequence fragments with an average size of 8 kb) to provide a >20-fold increase in the mapping resolution for the CENP-A-binding domain of the 10q25 neocentromere.

Our results indicate that while the boundary of the outermost CENP-A-binding PCR fragments defines a domain of ~330 kb that closely matches the region previously defined using the BAC array (11), CENP-A itself does not bind uniformly across the entire 330 kb genomic segment. Instead, a finer structure consisting of multiple distinct CENP-A-binding clusters that are intervened by histone H3-containing nucleosomes is found. It is unlikely that the observed CENP-A clustering reflects heterogeneous binding between cells in the population, rather than a complex that is more or less identical for all the cells. The use of five biological replicates each representing the sum of a large number of cells used for chromatin preparation and immunoprecipitation should ensure the nullification of heterogeneous binding at the individual cell level. Additionally, our observed organizational profile is in close agreement with those reported for the CENP-A-binding region using immunofluorescence analysis of extended chromatin fibres at the normal centromeres of various species including humans, D. melanogaster and Zea mays (12,22). The estimated sizes of
the CENP-A-binding clusters along 10q25 neocentromere (~10.5–51.8 kb) are similar to those determined for the D. melanogaster CENP-A homologue CID (~15–40 kb) (12). These observations indicate a high degree of conservation for both the discontinuous nature and the size range of the CENP-A-binding clusters between the human and Drosophila centromeres.

A notable drawback of earlier studies based on repetitive DNA-based centromeres using immunofluorescence analysis of extended chromatin fibres is the difficulty in generating uninterrupted fibres to allow the definition of the distribution profile for all member-clusters as well as their intervening regions within a complete centromere. Using a neocentromere-based CIA strategy, we show here that such a study is possible and provide the first high-resolution and complete CENP-A-binding profile for a functional higher eukaryotic centromere. Our results reveal that the most centrally positioned CENP-A-binding cluster within the overall CENP-A-binding domain is significantly larger than each of the remaining flanking clusters. Furthermore, the flanking clusters appear to be arranged in a configuration of alternating larger and smaller sizes around the central cluster.

The organization of CENP-A nucleosomes into discontinuous blocks has been proposed to be essential for the packaging of the centromeric chromatin into a spiral or loop pattern (12,23,24). According to this model, CENP-A-binding segments are brought together through chromatin condensation into parallel register forming a disc or plate upon which the foundation kinetochore proteins can be assembled (9). In contrast, the intervening non-CENP-A-binding and histone H3-containing segments are folded inward to form the inner centromere domain sequestered between two sister kinetochores. The results of the present study indicate that a hierarchical and symmetrical arrangement of CENP-A-binding clusters of varying sizes may be an important structural requirement for mammalian kinetochore assembly and/or to provide stability to withstand polar microtubule forces (Fig. 4). This provides a useful basis for the further investigation of the CENP-A-binding properties of the repetitive DNA-based centromeres in humans and other species.

**Figure 4.** Organization of CENP-A-binding regions within the 10q25 neocentromeric chromatin. A schematic diagram showing a cylinder-like solenoid structure (12,23) of the previously defined 330 kb CENP-A-associated domain within a larger 3.5 Mb S/MAR (scaffold/matrix attachment region)-enriched domain at the 10q25 neocentromere (11,18). The 330 kb CENP-A-associated centromeric segments are organized in loops and form a plate or disc structure to allow the assembly of other kinetochore proteins (9). A hierarchical and symmetrical arrangement, of which the largest central CENP-A-binding cluster is flanked by smaller clusters arranged in an alternating larger and smaller size configuration, may be essential for optimal kinetochore stability and/or withstanding the tensile forces of the polar microtubules.

**Enrichment of L1 retroelements within CENP-A-binding domain**

Our previous sequence analysis of the 330 kb 10q25 CENP-A-binding domain revealed a higher AT content (65.4%) when compared with that of the genome average (59.0%), suggesting that AT-rich sequences may be the preferred sites for CENP-A-binding and neocentromere formation (11). However, in this study, *in silico* analysis shows that the average AT contents of both CENP-A-binding and non-CENP-A-binding clusters are similar despite both being higher than that of the genome average. These results suggest that higher AT content alone is probably not the primary factor in determining the differential binding of CENP-A across its overall CENP-A-binding domain. Also, none of the interspersed repeats, with the exception of L1, show significant difference in prevalence to make them likely candidates to have a role in determining CENP-A binding.

The L1 retroelements show a >2.5-fold increase in prevalence in the CENP-A-binding clusters when compared with non-CENP-A-binding regions or the genome average. At present, we can only speculate on the potential involvement, if any, of an enhanced level of such elements in the localization of CENP-A-associated nucleosomes and the organization of centromeric chromatin (32). One possibility could be that the heterochromatic nature associated with L1 makes the genomic sites enriched in these sequences more favourable for the nucleation of neocentromeric chromatin (25,26).

Alternatively, L1 may adopt a structural configuration that is more suitable for the loading of CENP-A nucleosomes as compared to other genomic sequences. The close correlation of mouse L1, L1-Md elements with the nuclear matrix anchorage regions (MARs) at the mouse IgH locus provides some support for this possibility (27). In addition, these repetitive elements are prone or predisposed to epigenetic modifications (28,29), making them potentially versatile adaptation sites for neocentromere formation. Of interest, the human X chromosome also has a 2-fold increase in L1 elements that have been proposed to account for the initiation of heterochromatin formation and the spreading of inactivation signal during X chromosome inactivation (30,31).

Additional evidence for the possibility that L1 may have a role in the formation and evolution of centromeric chromatin come from studies in plants (32). The presence of centromere-specific retrotransposons (CRs) in various plant centromeres (33,34) and the direct interaction of these elements with the CENP-A homologue in maizee are suggestive of their contributions to the organization of centromeric chromatin (35).
A better understanding of the role of the retroelements at these centromeres or human neocentromeres may come from direct study of possible transcriptional activities of the centromeric or neocentromeric retroelements or of the effects disrupting such activities may have.

**MATERIALS AND METHODS**

**Cell culture condition**

CHO-human somatic cell hybrid lines CHON10 and CHOM10, containing normal human chromosome 10 and the neocentromeric mardel(10) chromosome tagged with a zeocin-resistant gene, respectively, were generated by microcell-mediated transfer as previously described (36). Both cell lines were maintained in Ham’s Kao and Michayluk medium (KAO) supplemented with 12% FCS. For CHOM10, 100 µg/ml zeocin was added into the culture medium to maintain selection for mardel (10).

**10q25 genomic PCR array construction**

Primers were designed to amplify 125 genomic fragments spanning ~1.0 Mb within and surrounding the previously mapped CENP-A-binding domain of the 10q25 neocentromere. PCR primer lengths varied between 22 and 28 nucleotides, and the melting temperature was set between 56 and 60°C using the Jellyfish software (Lab Velocity, Burlingame, USA). The size of PCR fragments ranged from 2 to 16 kb with varying degrees of overlap between neighbouring fragments. A complete list of primer sequences is provided in Supplementary Material, Table S2.

The fragments were amplified using Expand Long Template PCR system (Roche) according to the manufacturer’s protocols. The BAC templates used for the PCR amplification were obtained from either Genome Therapeutics Corporation (Walthan, USA) or the Sanger Institute (Cambridge, UK). The cycling parameters used were as follows: an initial denaturation step for 2 min at 94°C followed by 30 cycles of denaturation for 10 s at 94°C, annealing for 30 s at 52–56°C (depending on the primer sets used) and elongation for 10 min at 68°C. The PCR products were purified using High Pure PCR Product Purification Kit (Promega). Twenty five nanograms of the purified PCR products was then immobilized onto Hybond N+ nylon membranes using a 96-well dot blotter (Minifold SRC-96, Schleicher & Schuell, Dassel, Germany).

**Nuclear extract preparation and ChIP**

ChIP was performed as previously described (11). Approximately 107 exponentially growing cells were resuspended in TBS (0.01 M Tris–HCl at pH 7.5, 3 mM CaCl2, 2 mM MgCl2, 0.1 mM PMSF, protease inhibitors) with 0.25% Tween-40 and subjected to constant shaking for 1 h at 4°C. The nuclei were purified by Dounce homogenization (30 strokes on ice, Wheaton, Millville, USA), followed by centrifugation at 400 g for 20 min at 4°C through a 25%/50% discontinuous sucrose gradient. In order to obtain short polynucleosomes, ranging from 100 to 1000 bp, the purified nuclei were digested with micrococcal nuclease (USB Corp., Cleveland, USA) in digestion buffer (0.32 M sucrose, 50 mM Tris–HCl at pH 7.5, 4 mM MgCl2, 1 mM CaCl2, 0.1 mM PMSF, protease inhibitors) at a concentration of 64 U/mg of DNA for 6 min at 37°C. The reaction was stopped by adding EDTA to a final concentration of 5 mM followed by centrifugation at 14 000 g for 15 min at 4°C. The chromatin-containing supernatant was then kept on ice and the pellet fraction was further processed by incubating in lysis buffer (1 mM Tris–HCl at pH 7.5, 0.2 mM EDTA, 0.1 mM PMSF, protease inhibitors) for 1 h on ice, and centrifuged at 14 000 g for 20 min at 4°C. The two supernatant fractions were then pooled and mixed with equal volume of incubation buffer (50 mM NaCl, 20 mM Tris–HCl at pH 7.5, 5 mM EDTA, 0.1 mM PMSF, protease inhibitors). A portion of sample was set aside as input fraction, while the remaining was incubated with a polyclonal anti-mouse CENP-A antibody (Santa Cruz) at 1:500 dilution or polyclonal anti-histone H3 antibody (Santa Cruz) at 1:250 dilution overnight at 4°C. Five independent biological replicate ChIP experiments using anti-CENP-A antibody were performed on chromatin prepared from different flasks of cells. The immunocomplexes were recovered by incubation with 12.5% protein-A–Sepharose for 4 h at 4°C and washed in a stepwise manner with buffer A (10 mM EDTA, 50 mM Tris–HCl at pH 7.5) containing 50, 100 and 150 mM NaCl. Bound immunocomplexes were then eluted with two volumes of incubation buffer containing 1% SDS.

**DNA amplification, labelling and array hybridization**

DNA from both the input and the bound fractions were purified by standard phenol/chloroform extraction and isopropanol precipitation using glycogen (50 µg/sample) as a carrier (38). For generating the probes, DOP-PCR was performed in a reaction volume of 50 µl 1× PCR buffer (20 mM Tris–HCl, pH 8.4, 50 mM KCl) containing 200 ng immunoprecipitated DNA as template, 1 U of AmpliTaq DNA polymerase (Applied Biosystems), 0.2 mM dNTPs, 2 mM MgCl2 and 1 µM DOP primer (5'-CCAACCTCGAGNNNNNATTGTG-3'). The cycling conditions were performed as follows: an initial denaturation of 94°C for 2 min; followed by eight cycles of 94°C for 1 min, 30°C for 90 s, 72°C for 3 min with ramping rate at 50%; then 32 cycles of 94°C for 1 min, 54°C for 90 s and 72°C for 3 min; a final extension step was done at 72°C for 10 min.

Hybridization was performed using the genomic PCR arrays generated in this study. Each membrane was washed briefly in 2× SSC, and prehybridized with 50 µg of denatured salmon sperm DNA in prehybridization solution (5× SSC, 10× Denhardt’s, 0.05 M NaPO4 at pH 6.7, 5% dextran sulphate, 50% formamide) at 42°C overnight. DOP-PCR amplified DNA was 32P radioactively labelled using Rediprime II (Amersham Biosciences) according to the manufacturer’s instructions. A 500 ng aliquot of radioactively labelled probes was denatured with 10 µg of human Cot-I DNA (Invitrogen), 20 µg of human placenta DNA (Promega), 100 µg/ml denatured salmon sperm DNA in hybridization buffer (5× SSC, 1× Denhardt’s, 0.02 M NaPO4 at pH 6.7, 5% dextran sulphate, 50% formamide) and preannealed at 42°C for 1 h before hybridizing to the membranes overnight at 42°C. Sequential washes with 2× SSC/0.1% SDS, 1× SSC/0.1% SDS.
SDS and 0.1 x SSC/0.1% SDS were carried out at 65°C with constant agitations.

Genomic array data analysis
Quantification of radioactive signals was performed using a Phosphorimager system (Storm 860 Gel and Blot Imaging System, Molecular Dynamics) coupled with ImageQuaNT software (Molecular Dynamics). Signals obtained from each round DNA spot on the dot-blotted PCR array were compared with that on a duplicate blot hybridized with the input DNA. Areas without immobilized DNA on the same hybridized blot were used for background correction. The ratio of enhancement (R) was calculated by dividing the intensity of bound fraction by intensity of input fraction. For correcting intra- and inter-experimental variations, normalization was performed against the mean R value of seven control PCR fragments (Cel-35F/Cel-35R, Cel-36F/Cel-36R, Cel-37F/Cel-37R, Cel-38F/Cel-38R, Cel-39F/Cel-39R, Cel-40F/Cel-40R and Cel-41F/Cel-41R) that localized outside the CENP-A-binding domain. For each spot, normalized R (nR) value was then compared by expressing the difference in R values between the test (CHOM10) and control (CHON10) lines as a percentage of the R value.

ChIP-PCR analysis
PCR was performed on CHOM10 DNA isolated by ChIP using anti-CENP-A and anti-H3 antibodies. A list of the primer sequences is provided in Supplementary Material, Table S3. Amplification was achieved using Expand Long Template PCR system (Roche) with 100 ng of input or immuno-precipitated DNA according to the manufacturer’s protocols. The cycling parameters used were as follows: an initial denaturation step for 2 min at 94°C followed by 40 cycles of denaturation for 10 s at 94°C, annealing for 20 s at 52°C and elongation for 30 min at 68°C. The PCR products were visualized by 2% agarose gel electrophoresis.

In silico sequence analysis
All DNA sequences and their genomic positions were obtained from the Jul 2003 freeze version of the UCSC human genome database (http://genome.ucsc.edu/). The accession numbers and sequence coordinates of the BACs used for mapping the CENP-A-binding clusters in this study are as follows: 116624033–116638762, 14729 bp; 116659298–116686350, 2702 bp; 116697079–117052058. The sequence coordinates for the newly defined CENP-A-binding clusters in this study are as follows: 116642033–116683762, 14729 bp; 116697091–116900844, 29853 bp and 116950436–116960927, 10491 bp. RepeatMasker program (http://www.repeatmasker.org/) was used to analyse the base composition and the interspersed repeat content for all genomic fragments included in 10q25 PCR array. The raw RepeatMasker output data derived from each fragment were summarized in Supplementary Material, Table S4.

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS
We thank P. Kalitsis for anti-CENP-A antibody, M. Anderson for reading the manuscript, and all current laboratory members for their helpful discussions and technical advice. A.C.C. receives an Australian Postgraduate Award from the Commonwealth Government of Australia. This work was funded by NHMRC and NIGMS/NIH. K.H.A.C. is a senior principle research fellow of NHMRC.

REFERENCES
(CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. Genome Res., 11, 448–457.