Chromosome Research **12**: 805–815, 2004.

Chromosome size and origin as determinants of the level of CENP-A incorporation into human centromeres

Danielle V. Irvine[1], David J. Amor[1], Jo Perry[1], Nicolas Sirvent[2], Florence Pedeutour[2], K. H. Andy Choo[1] & Richard Saffery[1*]

[1]*The Murdoch Childrens Research Institute, Royal Children's Hospital, Flemington Road, Parkville 3052, Australia; Department of Paediatrics, University of Melbourne, 3052, Australia;* [2]*Laboratoire de Genetique, Hopital de l'Archet, 151 route de Saint-Antoine de Ginestiere, 06202 Nice cedex 3, France*

**Correspondence: Tel: +61-3-83416255; Fax: +61-3-93481391; E-mail: andy.choo@mcri.edu.au*

Received 30 August 2004. Received in revised form and accepted for publication by Wendy Bickmore 25 September 2004

Key words: CENP-A, kinetochore, neocentromere

Abstract

We have expressed an EGFP-CENP-A fusion protein in human cells in order to quantitate the level of CENP-A incorporated into normal and variant human centromeres. The results revealed a 3.2-fold difference in the level of CENP-A incorporation into α-satellite repeat DNA-based centromeres, with the Y centromere showing the lowest level of all normal human chromosomes. Identification of individual chromosomes revealed a statistically significant, though not absolute, correlation between chromosome size and CENP-A incorporation. Analysis of three independent neocentromeres revealed a significantly reduced level of CENP-A compared to normal centromeres. Truncation of a neocentric marker chromosome to produce a minichromosome further reduced CENP-A levels, indicating a remodelling of centromeric chromatin. These results suggest a role for increased CENP-A incorporation in the faithful segregation of larger chromosomes and support a model of centromere evolution in which neocentromeres represent ancestral centromeres that, through adaptive evolution, acquire satellite repeats to facilitate the incorporation of higher numbers of CENP-A containing nucleosomes, thereby facilitating the assembly of larger kinetochore structures.

Introduction

Normal human centromeres are characterised by an abundance of 17l-bp α-satellite DNA repeat arrays (Choo 1997). These arrays are highly variable in size, ranging from ~240 kb on some Y chromosomes to >4000 kb at other centromeres. Variations are also common between homologous chromosomes. Such variation has raised questions concerning the functional organization of human centromeres. Centromeric kinetochores assemble on a subset of this DNA and mediate chromosome attachment to spindle microtubules to facilitate proper mitotic segregation. The primary link between centromeric DNA and the protein components of the kinetochore is the conserved histone H3-variant, CENP-A. CENP-A localises to the inner kinetochore plate of active centromeres (Warburton *et al.* 1997), co-purifies with nucleosomes (Palmer *et al.* 1987),

and can replace histone H3 in nucleosome reconstruction *in vitro* (Yoda *et al*. 2000) and *in vivo* (Lo *et al.* 2001a). Thus CENP-A forms specialized centromeric nucleosomes required for the recruitment of other kinetochore components.

The large arrays of centromeric repetitive DNA have greatly hindered the analysis of normal human centromere domain organization and have generally prevented detailed functional dissection. To date, only two normal human centromeres have undergone detailed molecular and functional analysis. The DNA sequences required for human Y centromere function have been mapped to a minimal region of ~200 kb of α-satellite and 300 kb of adjacent Yp sequences (Tyler-Smith *et al*. 1993) while analysis of the X chromosome centromeric DNA has defined the polymorphic size range of the α-satellite locus (DXZ1) of between ~2.2 and 3.7 Mb (Schueler *et al*. 2001). A single region of topoisomerase IIα cleavage activity is present at both the X and Y centromeres that tracks with kinetochore function (Floridia *et al*. 2000, Spence *et al*. 2002). Although these investigations have identified the general site of functional centromere activity they provide little insight into the size of the fully assembled kinetochore complex. Human neocentromeres lacking α-satellites have proven to be useful in dissecting the domain organisation of centromeres (reviewed in Amor & Choo 2002). Recent studies have defined distinct CENP-A binding domains of neocentromeres between 130 and 460 kb (Alonso *et al*. 2003, Lo *et al*. 2001a, b); however a corresponding dissection of CENP-A domains at normal human centromeres has not yet been undertaken. In the present study, we have used an EGFP-CENP-A fusion protein to quantify the incorporation of CENP-A into individual human centromeres and neocentromeres.

Methods

Construction of a human CENP-A/EGFP fusion protein

Forward (5′-GAGAGACTCGAGACCCTCTGCGGCGTGTCATGG-3′) and reverse primers (5′-GAGAGAGAATTCACTGGGTGCAGGAGCTCAGCC-3′) were used to amplify the entire coding region of the human CENP-A gene using clone ID#2853916 from the NCI_CGAP-Ut4 cDNA library as a template. PCR products were digested with *Xho*I and *Eco*RI (underlined) and cloned into pEGFP-Cl (Clontech, Palo Alto, CA, USA).

Cell culture and generation of stably transfected cell lines

Male HT1080 derivative cell lines containing mardel(10) or MiC5 were cultured as previously described (Saffery *et al*. 2001). Transformed female fibroblasts containing a neocentric inv dup(20p) marker chromosome were cultured in DMEM with 10% FCS (JRH Biosciences). The well-differentiated liposarcoma cell line 94778 (case 6b in Pedeutour *et al*. (l999) was cultured in RPMI with 20% FCS (JRH Biosciences). Transfections were performed by electroporation (Bio-Rad Gene Pulsar: one pulse at 400 V, 250 μF for HT1080 cells and liposarcoma line 94778; 2 pulses at 250 V, 250 μF and 250 V, 25 μF for fibroblasts). Linearized vector (2 μg) was electroporated and selection (450 μg/ml G418) applied 24 hours post-transfection. Cells were screened for EGFP-CENP-A expression in 96-weIl plates using an Olympus IX-70 Inverted Fluorescent Microscope. Those showing EGFP-CENP-A expression were scaled up for analysis. A single clone of each cell type exhibiting highly specific centromeric fluorescence with low background was used in subsequent quantification analysis.

Western blotting and immunofluorescence analysis

For Western analysis nuclear extracts were subjected to gel electrophoresis on a 4–12% gradient gel (NuPage, Invitrogen Corp.) followed by transfer onto Hybond-C membrane (Amersham Biosciences) using a Trans-Blot SD apparatus (Bio-Rad). Blocking of blots was carried out overnight followed by incubation with rabbit anti-CENP-A polyclonal antisera (1 : 200, Upstate Inc, Charlottsville, VA) overnight at 4°C. Following washing, blots were incubated with anti-rabbit horseradish peroxidase (1:1000, Silenius Laboratories, Melb, Aust) followed by chemoluminescence detection using a Bio-Rad ChemiDoc XRS system according to the manufacturer's instructions.

Dual immunofluorescence analysis with anti-GFP and anti-CENP-A was carried out as previously described (Saffery *et al*. 2000). Primary antibodies used in immunofluorescence were polyclonal goat anti-GFP (1:400 dilution, Rockland Immunochemicals Inc., Gilbertsville PA) and rabbit anti-CENP-A (1:100 dilution, Upstate Inc., Charlottsville, VA).

Analysis of EGFP-CENP-A expression at the kinetochore

EGFP-CENP-A fluorescence on 2% paraformaldehyde fixed, cytospun chromosomes was measured quantitatively using IPLab (Scanalytics) software as described previously (Craig *et al*. 2003). Background fluorescence was removed from each raw image prior to EGFP-CENP-A fluorescence quantitation. Fluorescence *in situ* hybridization (FISH) was performed as previously described (Saffery *et al*. 2000) to enable the identification of the 10q25 neocentromere of mardel(10) and MiC5, and the invdup(20p) neocentromere. The liposarcoma neocentromere was identified by chromosome morphology. Other chromosomes were identified by FISH with specific α-satellite probes or chromosome paints (Supplementary Table 1*)

Statistical analysis

Microsoft Excel was used for basic statistical analysis, including the linear regression. Confidence intervals (CI) for comparison of neocentromere CENP-A levels as a percentage of Y centromere and/or mean of all human chromosomes were calculated using the formula: CI (95%) = mean $\pm(1.96 \times \text{s.e.})$, where s.e. = s.d./$\sqrt{n}$. For comparison of the mardel(10) and MiC5 CENP-A levels the 95% confidence interval for the determination of differences between groups was tested using the formula: CI (95%) = $(\text{mean}_1\text{–mean}_0) - (z' \times \text{s.e.})$, to $(\text{mean}_1\text{–mean}_2) + (z' \times \text{s.e.})$, where $z' = 1.96$ and $\text{s.e.} = \sqrt{(\text{s.e.}_1{}^2 + \text{s.e.}_0{}^2)}$. The confidence interval provides a range of values that indicate a 95% probability of a difference in mean CENP-A levels between the two groups. The significance of this difference (expressed as a *p* value) is assessed using a *z*-test where $z = \text{mean}_1\text{–mean}_0/\text{s.e.}$ (Dytham, 2003).

Results

Faithful non-biased incorporation of EGFP-CENP-A into human kinetochores

We set out to examine CENP-A levels at human centromeres using a method involving the direct incorporation of exogenously expressed EGFP-tagged CENP-A.

Several previous studies have demonstrated the faithful incorporation of epitope tagged CENP-A variants into mammalian centromeres, including GFP fusion proteins (Blower *et al*. 2002, Kalitsis *et al*. 2003, Shelby *et al*. 1997, Sugimoto *et al*. 2000, Sullivan *et al*. 1994, Van Hooser *et al*. 2001). We produced several stably transfected human cell lines expressing EGFP-CENP-A fusion protein each of which contained a different neocentric marker chromosome as follows: a 58-Mb marker derived from chromosome 10 designated mardel (10) (Voullaire *et al*. 1993); a 0.7-Mb truncated derivative of mardel(10) designated MiC5 (Saffery *et al*. 2001); a 40-Mb inverted duplication of chromosome 20p designated invdup(20p) (Voullaire *et al*. 1999); and a 'giant rod' chromosome designated LGR identified in a liposarcoma (Pedeutour *et al*. 1999).

Production of stable cell lines carrying random integration of exogenously expressed genes invariably results in different expression levels of the introduced fusion protein between individual lines. In addition, although previous studies have demonstrated that expression of exogenous CENP-A at near physiological levels results in exclusive targeting to centromeres (Shelby *et al*. 1997, Sullivan *et al*. 1994), overexpression of this protein within cells can result in mistargeting to non-centromeric loci (Van Hooser *et al*. 2001). In light of this, we screened our stable cell lines for those that showed no localization of the fusion protein to metaphase chromosome arms. EGFP-CENP-A levels for these lines were further quantitated by Western blotting to determine the level of expression of the fusion protein in relation to that of the endogenous CENP-A protein (Figure 1). From this analysis, we identified several cell lines in which the overall level of EGFP-CENP-A was lower than that of the endogenous protein. Each of these cell lines showed

*Available online at http://www.kluweronline.com/issn/0967-3849

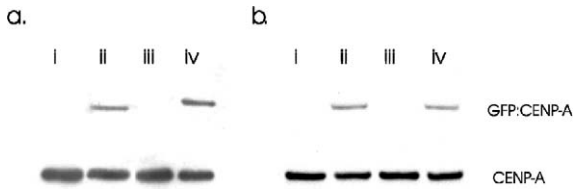

Figure 1. Low level expression of EGFP-CENP-A in stable cell lines. Examples of Western blot analysis of parental cell lines (Ai, Aiii, Bi, Biii) and stably transfected cell lines (Aii, Aiv, Bii, Biv) expressing the EGFP-CENP-A fusion protein using polyclonal anti-CENP-A. EGFP-CENP-A migrates slower than endogenous CENP-A due to increased size. A single band of CENP-A is present in untransfected parental cell lines while stable transfected lines contain both forms of CENP-A. Lower levels of the fusion protein are present than the endogenous CENP-A as evidenced by a fainter band following Western analysis. (**A**) 14ZBHT cell line containing mardel(10). (**B**) Liposarcoma cell line carrying neocentric giant rod chromosome.

specific centromere localization of the fusion protein on metaphase chromosomes (Figure 2).

Given that our analysis was aimed at determining the relative levels of CENP-A incorporation into centromeres, it was important to demonstrate equal relative incorporation of CENP-A and the EGFP-tagged CENP-A into individual centromeres. This was achieved using dual immunofluorescence analysis with antibodies to CENP-A (detecting both endogenous CENP-A and EGFP-CENP-A variant) and GFP (detecting only EGFP-CENP-A). The ratio of the different antisera signals for individual chromosomes was then quantitated and compared between chromosomes. No significant bias in the incorporation of the EGFP-CENP-A fusion protein into particular centromeres (including that of the Y-chromosome and neocentromeres) was detected within individual cells (an example is shown in Figure 3), indicating that each centromere incorporates the same relative level of EGFP-CENP-A in relation to the amount of endogenous CENP-A. In addition, we have previously shown that heterozygous CENP-A/EGFP-CENP-A mouse embryonic stem cell lines incorporate EGFP-CENP-A specifically at all the centromeres and support a normal growth phenotype (Kalitsis *et al.* 2003), suggesting that the simultaneous incorporation of EGFP-CENP-A and normal CENP-A into centromeres does not disrupt centromere functions in any detectable way and that any selection bias against GFP-CENP-A incorporation is unlikely. Furthermore, a *S. cerevisiae* Cse4p-GFP fusion protein can functionally replace human CENP-A at centromeres in human cells (Wieland *et al.* 2004).

Quantitation of CENP-A incorporation at normal human centromeres

In order to examine CENP-A incorporation into all 24 human centromeres, we quantitated EGFP-CENP-A levels in two male cell lines [mardel(10) and MiC5], Quantitative measurement was performed on EGFP-CENP-A fluorescence at all α-satellite-based centromeres in several metaphase spreads, using specific quantitation software (see Methods). The average variation in signal intensity for the normal chromosomes, calculated as the fold-difference between the lowest signal (which consistently corresponded to that of the Y chromosome, see below) and the highest signal present within a particular metaphase spread, was determined to be 3.18-fold [Confidence interval (95%); CI = 2.75 to 3.61], (Figure 4A–B). Therefore, the level of CENP-A incorporated into normal human centromeres varies over a 3-fold range within individual cells. FISH with specific probes was used to identify each of the 24 individual chromosomes within cells thereby allowing a comparison of CENP-A incorporation between centromeres on different chromosomes. Within each chromosome spread this was compared to the constant fluorescence signal present at the 10q25 neocentromere (see below) to give a relative fluorescence value for each of the 24 chromosome types in each metaphase spread. Comparison of these values for individual chromosomes (1–22, X and Y) between spreads demonstrated a relatively constant level of EGFP-CENP-A into each particular chromosome type (Supplementary Table 2*). Linear regression analysis also revealed a significant correlation between the physical size of chromosomes (as determined by

*Available online at http://www.kluweronline.com/issn/0967-3849

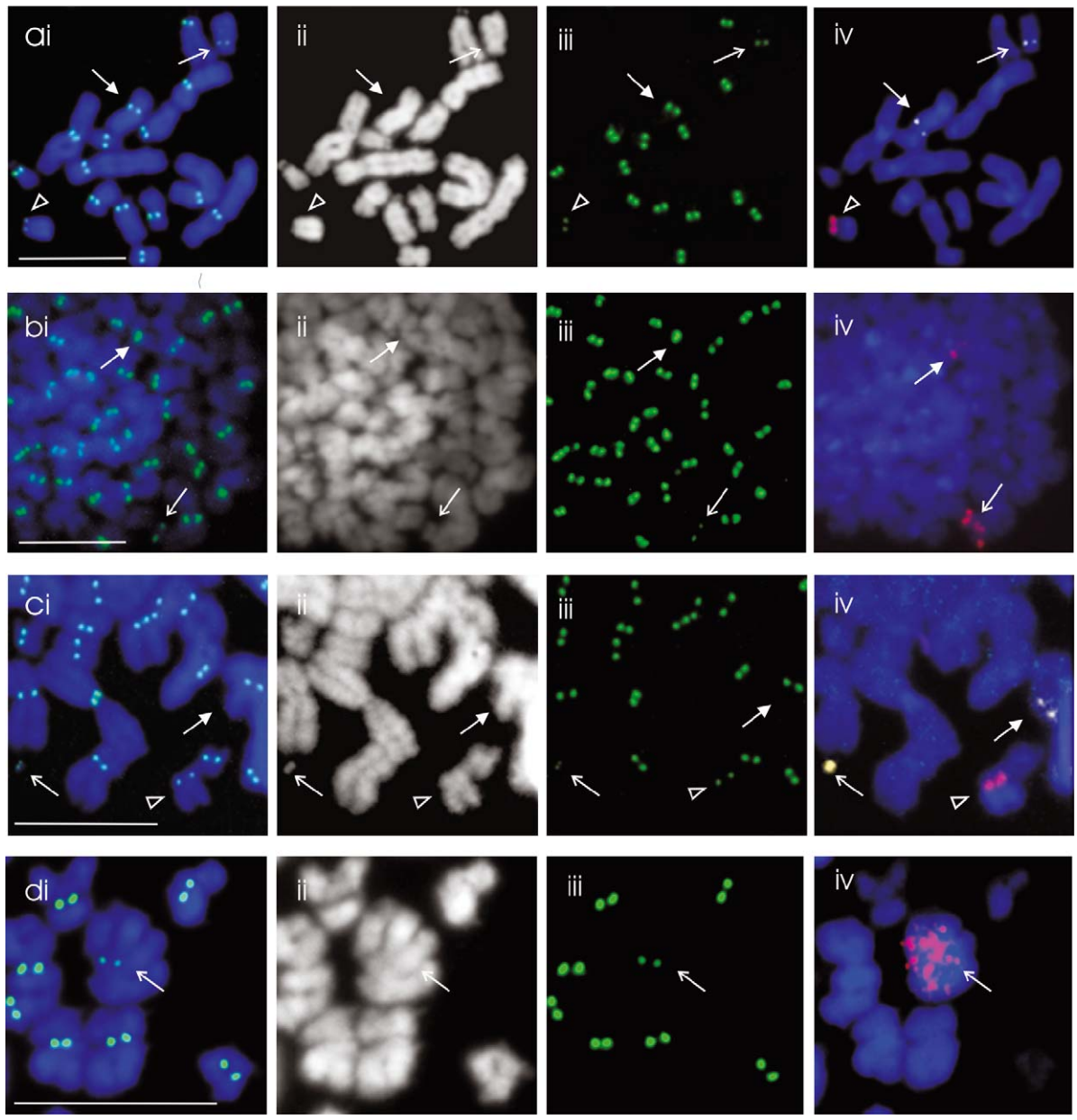

Figure 2. Centromere localization of EGFP-CENP-A in cell lines containing mardel(10), invdup(20), LGR or MiC5. Combined images showing EGFP fluorescence at the centromere of human chromosomes including mardel(10) (arrow; panels Ai–iv), invdup(20p) (arrow; panels Bi–iv), MiC5 (arrow; panels Ci–iv) and LGR (arrow; panels Di–iv), normal chromosome 10 (closed arrow; panels Ai–iv and Ci–iv), normal chromosome 20 (closed arrow; panels Bi–iv), and Y chromosome (arrowhead; panels Ai–iv and Ci–iv). Dual-colour FISH was subsequently performed with a 10q25 neocentromere BAC probe (A(iv), C(iv); green signal); a Y-chromosome α-satellite probe (A(iv),C(iv); red signal; a 20pl2 neocentromere BAC probe (B(iv); red signal), or a 12ql5-BAC (D(iv); red signal), were used to identify specific neocentromere containing chromosomes. DAPI counter staining is shown in panels A(ii), B(ii), C(ii) and D(ii) while EGFP-CENP-A signals are shown in panels A(iii), B(iii), C(iii) and D(iii).

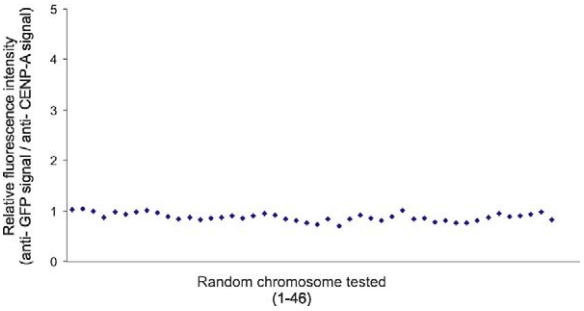

Figure 3. Example of relative level of EGFP-CENP-A incorporation into human centromeres. Dual immunofluorescence experiments were carried out with a goat anti-GFP antibody, detecting the EGFP-CENP-A fusion protein (anti-goat FITC detection), in combination with a rabbit anti-CENP-A antibody, detecting both EGFP-CENP-A and endogenous CENP-A (anti-rabbit Texas Red detection). The ratio of EGFP-CENP-A fluorescence to endogenous CENP-A fluorescence is expressed on the Y-axis for each of 46 randomly assayed centromeric signals (including that of the Y-chromosome and mardel(10) neocentromere) within a metaphase spread. Data for a single metaphase spread are presented. However no bias in loading for or against individual types of centromeres was detected in six individual metaphase spreads analysed (data not shown).

genomic content) and centromeric CENP-A incorporation, with larger chromosomes generally showing increasing centromeric CENP-A ($r^2 = 0.40$) (Figure 5). Of all normal centromeres, the Y centromere consistently showed the lowest level of CENP-A incorporation. Based on statistical analysis (see Methods) we can conclude that chromosome size is a determinant in the deposition of approximately 40% ($r^2 = 0.40$) of the level of CENP-A present at individual centromeres.

CENP-A incorporation into neocentromeres

We next examined the relative level of EGFP-CENP-A fluorescence at the neocentromeres of three different marker chromosomes and a truncation derivative of one of them. The mardel(10) neocentromere, derived from a complex rearrangement of chromosome 10 that resulted in the removal of the α-satellite centromere, contained an average of 61% (CI = 54% to 68%) of the level of EGFP-CENP-A fluorescence found at the Y chromosome, and 32% (CI = 26% to 38%) of the mean of all human chromosomes in the

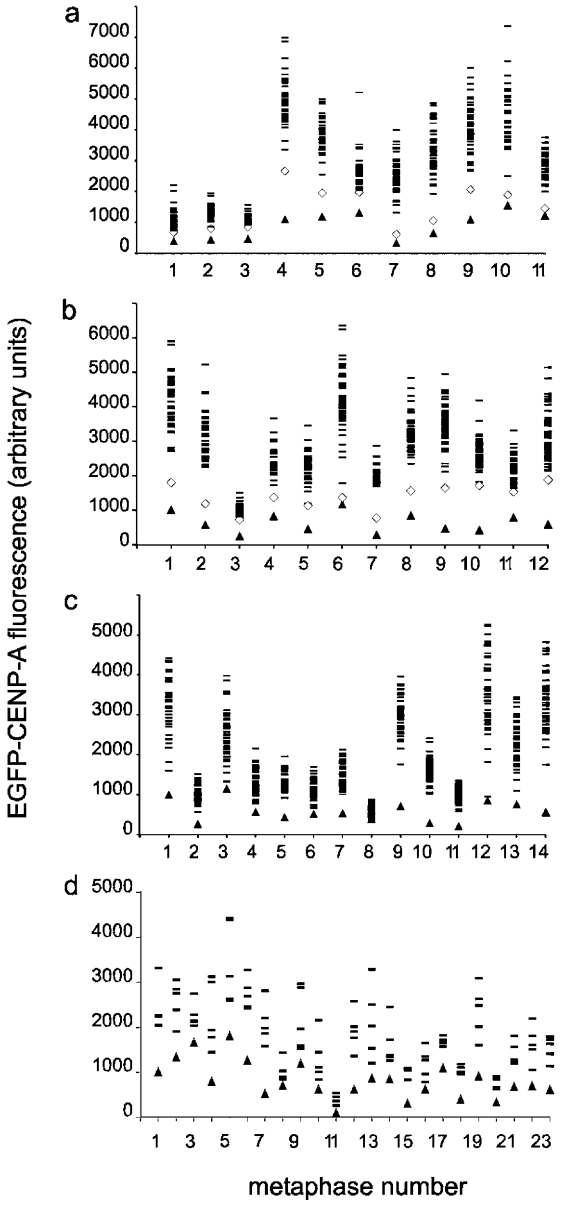

Figure 4. Range of EGFP-CENP-A fluorescence signals. Absolute centromeric fluorescence for individual chromosomes contained within particular metaphase spreads. The EGFP intensity of each chromosome within a mardel(10)-containing (**A**), MiC5-containing (**B**) and invdup(20p)-containing (**C**) cell is represented by a single data point, with all chromosomes within a single cell being scored. For the liposarcoma cell line 94778, the EGFP intensity of each of 6 random chromosomes immediately adjacent to and including the LGR chromosome (**D**) were scored. Neocentromere fluorescence intensity is represented by a closed triangle, while that of the Y chromosome is represented by an open diamond. Metaphase number refers to individual metaphase spreads scored in this analysis. Between 11 and 23 metaphase spreads were scored for each of the four cell lines used in this analysis.

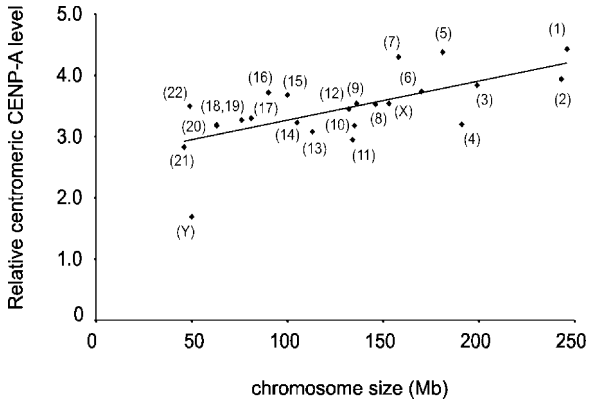

Figure 5. Relationship between kinetochore CENP-A level and chromosome size. Chromosome size is shown on the X-axis. The Y-axis shows CENP-A level expressed as a fluorescence ratio relative to that of the mardel(10) chromosome ($n = 10$–45 for individual chromosomes) in the 14ZBHT cell line. Using linear regression the proportion of CENP-A level variation attributable to chromosome size [CENP-A level $= 0.006 \times$ chromosome length $+ 2.73$] was determined to be 40% ($r^2 = 0.40$).

same metaphase spread (Figure 4A). Analysis of the invdup(20p) neocentromere in a female cell line produced similar results of 32% (CI = 26% to 38%) of the mean of all chromosomes (Figure 4C), whereas the neocentromere in a somatically-derived liposarcoma line contained an average of 45% (CI = 40% to 50%) of the mean level of GFP fluorescence of other human chromosomes (Figure 4D). These results supported our initial immunofluorescence findings demonstrating a constant but variable reduction in neocentromere CENP-A levels compared to normal centromeres (data not shown), and indicate that human neocentromere activation results in formation of kinetochores containing significantly less CENP-A than that typically assembled at human centromeres. This appears to be independent of chromosome size as the neocentric marker chromosome in the liposarcoma cell line is larger than all normal chromosomes (several hundred megabases) yet still demonstrates the lowest kinetochore incorporation of CENP-A in relation to all normal centromeres in metaphase spreads.

The neocentromere on minichromosome MiC5 (derived following truncation of the mardel(10) chromosome) was found to contain on average 46% (CI = 36% to 55%) of the amount of CENP-A present at the Y centromere and 23% (CI = 19% to 27%) of the mean of all human chromosomes (Figure 4B). This compares to the 61% and 32% values, respectively, obtained for the mardel(10) neocentromere, suggesting that the CENP-A-binding domain of the minichromosome kinetochore has contracted to approximately 72% of its original size as a result of minichromosome formation (95% CI = -0.27 to -0.034; $p = 0.006$).

Discussion

CENP-A incorporation shows 3.18-fold variation between normal human chromosomes

We have produced a series of stable cell lines expressing low levels of a EGFP-CENP-A variant that localizes faithfully to all human centromeres in a non-biased manner. We have used these lines to carry out the first quantitative comparison of native CENP-A at all normal human centromeres, identifying a greater than 3-fold variation in CENP-A levels between centromeres, with the Y chromosome centromere consistently incorporating the lowest level. Previous attempts at measuring aspects of kinetochore size have used autoimmune sera in immunofluorescence analysis. These studies have yielded conflicting results, with some studies suggesting little variation in fluorescence between the kinetochores of non-homologous chromosomes (Fantes *et al.* 1989, Schmitz *et al.* 1992), while others report an apparent heterogeneity in kinetochore fluorescence (Cherry & Johnston 1987, Peretti *et al.* 1986). One reason for these conflicting results may lie in the difficulty in interpretation of data obtained with ill-defined autoimmune sera that often contain antibodies to multiple centromere proteins (Earnshaw *et al.* 1989, Hudson *et al.* 1998).

Reduced CENP-A incorporation in neocentromeres

Measurements of the level of EGFP-CENP-A at three different neocentromeres have revealed the presence of significantly less CENP-A than that present at any of the normal centromeres, including that of the Y chromosome. This reduction in

CENP-A is unrelated to chromosome size as both small and large neocentric markers show a reduction in centromeric CENP-A level. This suggests that the process of neocentromere formation results in a 'minimal' centromere with a relatively low base-level of CENP-A incorporation, possibly reflecting an overall reduced kinetochore size than that found at repeat DNA-based centromeres (see below). This contrasts with previous results from two other neocentromeres using indirect immunofluorescence against a CENP-A-HA fusion protein, which detected no significant difference between normal centromeres and the neocentromeres (Warburton *et al.* 1997). However, in our own experience such an indirect immunofluorescence analysis is generally subject to high levels of variability between experiments possibly reflecting the increasing complexity of successive rounds of antibody detection required to detect antigen signals. This highlights the advantages of the direct EGFP-CENP-A incorporation technique used in the present study. Interestingly, immunofluorescence results using antisera for other kinetochore- and heterochromatin-associated proteins similarly show reduced signal intensity on neocentromeres (R. Saffery, unpublished data). It will be interesting in future to carry out similar analyses using fluorescent tagged variants of these and other centromere-associated proteins to see if the reduction in CENP-A levels at neocentromeres extends to other aspects of kinetochore structure.

The CENP-A associated domain of the 10q25 neocentromere has previously been defined using chromatin immunoprecipitation and DNA array analysis as a 330-kb region (Lo *et al.* 2001a). Taking into account relative CENP-A levels revealed in this study (Figure 4A), and assuming a consistent packaging density of CENP-A at human centromeres, we predict a minimal CENP-A-binding domain at normal human centromeres of between 540 and 1700 kb. This is consistent with recent chromatin fibre analysis of random human centromeres that identified the range over which CENP-A chromatin extends as between 500 and 1500 kb (Blower *et al.* 2002), although in this study specific chromosomes were not identified.

Analysis of the MiC5 minichromosome containing the 10q25 neocentromere has revealed a further significant reduction in the CENP-A binding domain compared to its progenitor neocentromere contained on the mardel(10) marker chromosome. This suggests a compaction or remodelling of centromeric chromatin on the minichromosome as a result of chromosomal truncation. This further highlights the emerging concept of plasticity of centromeric chromatin in response to an altered chromosomal environment (Craig *et al.* 2003, Floridia *et al.* 2000).

CENP-A incorporation correlates with chromosome size

We have identified a small but significant correlation between CENP-A levels and the size of the normal human chromosomes. However, the level of correlation ($r = 0.63$) indicates that chromosome size is not the sole determining factor in the overall level of CENP-A incorporation into human centromeres. Other possible contributing factors such as the quantity of α-satellite, the type of α-satellite subfamiliy (Choo 1997), the relative amounts of α-satellite, with and without CENP-B-box sequences (see below), the role of non-alphoid flanking repeats, or the amount and organization of pericentric heterochromatin, may be involved in specifying the level of CENP-A incorporation. At present however, the lack of comprehensive data on these variables does not allow proper discussion of their individual contribution to centromeric CENP-A levels.

Significance of variable CENP-A incorporation

Our data indicate that the level of CENP-A incorporation into centromeres is significantly correlated with chromosome size. Given that CENP-A forms the foundation of the active kinetochore, it is possible that variation in CENP-A levels also directly reflects a corresponding variation in the physical size of kinetochores. Thus, an increase in CENP-A incorporation might reflect the assembly of a larger kinetochore structure that may compensate for the mitotic segregation requirements of larger chromosomes namely to capture a greater number of microtubules to generate a greater 'pulling power' for chromosome movement and sister chromatid

separation. Previous studies in other species have suggested a link between chromosome and kinetochore size and have implied a relationship between kinetochore size and capacity to capture microtubules (Brinkley *et al.* 1984, Cherry *et al.* 1989, McEwen *et al.* 1998). Theoretically, such a compensatory mechanism would allow all chromosomes to be pulled to the spindle pole at a relatively constant rate regardless of total genomic content.

Our data supports the existence of an optimal range of CENP-A level within the α-satellite-based centromeres. Previous studies have similarly suggested that a minimum amount of kinetochore material may be necessary in order to maintain stable attachment to the mitotic spindle (Cherry *et al.* 1989). If this is the case, a significantly reduced CENP-A-binding domain may be expected to have consequences in terms of mitotic stability. Interestingly, an increased rate of mitotic instability has been reported for the Y chromosome compared to other human chromosomes (Nath *et al.* 1995), possibly consequent to the reduced CENP-A incorporation demonstrated in this study. We would similarly expect to see increased levels of mitotic instability for neocentromere-containing chromosomes because of their even lower levels of CENP-A incorporation. Although this may explain the observed mosaicism and instability for neocentric chromosomes in several patients (Amor & Choo 2002; Rivera *et al.* 1996), it cannot account for the apparent full mitotic stability of many other neocentromeres both in patient cells and in culture (Amor & Choo 2002), or the apparent stability of some neocentromere-based minichromosomes over prolonged periods of culture (Saffery *et al.* 2001). It is possible that smaller kinetochores assembled at sites of reduced CENP-A incorporation may associate with similar numbers of more tightly packed microtubules to achieve increased mitotic stability (McEwen *et al.* 1998), or that any instability due to reduced kinetochore size may be more pronounced in meiosis or be cell type specific.

The obvious difference in CENP-A levels observed for the Y chromosome and neocentromeres in relation to all other chromosomes may also reflect the generally lower levels of α-satellite DNA observed on the Y-chromosome (Tyler-Smith *et al.* 1993) and the absence of such arrays at neocentromeres. In addition, the α-satellite DNA of the Y-chromosome centromere specifically lacks the 17-bp CENP-B-box sequence motif to which the constitutive centromeric protein CENP-B binds. Although no definitive mitotic role of CENP-B has been identified, it is significant that *de novo* centromere formation is only observed on α-satellite DNA containing the CENP-B-box sequence motif (Ohzeki *et al.* 2002). The absence of CENP-B binding remains the only specific difference yet described between the Y centromere and all other centromeres in mammalian systems. Previous data have provided evidence for the specific association of CENP-A with type I α-satellite containing CENP-B boxes (Ando *et al.* 2002) although the existence of neocentromeres, and the lack of such CENP-B boxes on the Y chromosome, have suggested that such sequences are not absolutely required for CENP-A deposition into centromeric nucleosomes. Our data suggests that the presence of α-satellite DNA containing CENP-B boxes is associated with the incorporation of increasing levels of CENP-A at normal centromeres – an observation that appears to extend to the neocentromeres, which typically lack both α-satellite and CENP-B boxes, and incorporate reduced CENP-A compared to those of the normal centromeres. Interestingly, a direct interaction between CENP-B and another constitutive centromere protein CENP-C has also recently been demonstrated with the implication that the presence of CENP-B may be associated with increasing levels of (at least some) other kinetochore components in addition to CENP-A (Suzuki *et al.* 2004).

There is increasing evidence implicating centromere activation as a novel mechanism involved in karyotype evolution. It has been proposed that neocentromeres may represent the earliest stage in centromere evolution and that the acquisition of centromeric repeats occurs as a later event (Nagaki *et al.* 2004, Ventura *et al.* 2001). However, the higher order organization of the core CENP-A region of centromeres has yet to be fully elucidated. Several observations including the analysis of detached kinetochores and extended chromatin fibres (Blower *et al.* 2002, Zinkowski *et al.* 1991), centromeric fission in mammals (reviewed in Perry *et al.* 2004), and the phenomenon of neocentromere formation following

centromere spreading and fission in *Drosophila* (Maggert & Karpen 2001), have resulted in the development of a repeat subunit-based model of centromere structure. It remains to be seen whether an increasing level of CENP-A at indvidual centromeres equates to a larger number of such centromeric subunits. Our observation that human centromeres formed on α-satellite DNA contain higher CENP-A levels than those located on non-α-satellite DNA suggests a possible driving force for the accumulation of α-satellite repeats (in particular those containing the CENP-B-box motif) in centromere evolution to achieve the formation of larger and functionally more stable kinetochore structures.

Acknowledgments

We would like to thank Paul Kalitsis for the human CENP-A cDNA clone, Jeff Craig for his assistance with the quantification of centromeric GFP fluorescence, and the Clinical Epidemiology and Biostatistics Unit, MCRI, for help with statistical analysis. This work was supported by funding from NH&MRC and IGMS/NIH.

References

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

Amor DJ, Choo KH (2002) Neocentromeres: role in human disease, evolution, and centromere study. *Am J Hum Genet* **71**: 695–714.

Ando S, Yang H, Nozaki N, Okazaki T, Yoda K (2002) CENP-A, -B, and -C chromatin complex that contains the I-type alpha-satellite array constitutes the prekinetochore in HeLa cells. *Mol Cell Biol* **22**: 2229–2241.

Blower MD, Sullivan BA, Karpen GH (2002) Conserved organization of centromeric chromatin in flies and humans. *Dev* Cell **2**: 319–330.

Brinkley BR, Valdivia MM, Tousson A, Brenner SL (1984) Compound kinetochores of the Indian muntjac. Evolution by linear fusion of unit kinetochores. *Chromosoma* **91**: 1–11.

Cherry LM, Johnston DA (1987) Size variation in kinetochores of human chromosomes. *Hum Genet* **75**: 155–158.

Cherry LM, Faulkner AJ, Grossberg LA, Balczon R (1989) Kinetochore size variation in mammalian chromosomes: an image analysis study with evolutionary implications. *J Cell Sci* **92 (Pt 2)**: 281–289.

Choo KH (1997) *The Centromere*. Oxford University Press, New York, Tokyo.

Craig JM, Earle E, Canham P, Wong LH, Anderson M, Choo KH (2003) Analysis of mammalian proteins involved in chromatin modification reveals new metaphase centromeric proteins and distinct chromosomal distribution patterns. *Hum Mol Genet* **12**: 3109–3121.

Dytham C (2003) *Choosing and Using Statistics. A Biologists Guide*. Blackwell Publishing, Malden, Oxford, Melbourne, Berlin.

Earnshaw WC, Ratrie H, 3rd, Stetten G (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma* **98**: 1–12.

Fantes JA, Green DK, Malloy P, Sumner AT (1989) Flow cytometry measurements of human chromosome kinetochore labeling. *Cytometry* **10**: 134–142.

Floridia G, Zatterale A, Zuffardi O, Tyler-Smith C (2000) Mapping of a human centromere onto the DNA by topoisomerase II cleavage. *EMBO Rep* **1**: 489–493.

Hudson DF, Fowler KJ, Earle E *et al.* (1998) Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. *J Cell Biol* **141**: 309–319.

Kalitsis P, Fowler KJ, Earle E *et al.* (2003) Partially functional Cenpa-GFP fusion protein causes increased chromosome missegregation and apoptosis during mouse embryogenesis. *Chromosome Res* **11**: 345–357.

Lo AWI, Craig JM, Saffery R *et al.* (2001a) A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. *EMBO Journal* **20**: 2087–2096.

Lo AWI, Magliano DJ, Sibson NC, Kalitsis P, Craig JM, Choo KHA (2001b) A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A binding neocentromere DNA. *Genome Research* **11**: 448–457.

Maggert KA, Karpen GH (2001) The activation of a neocentromere in Drosophila requires proximity to an endogenous centromere. *Genetics* **158**: 1615–1628.

McEwen BF, Ding Y, Heagle AB (1998) Relevance of kinetochore size and microtubule-binding capacity for stable chromosome attachment during mitosis in PtKl cells. *Chromosome Res* **6**: 123–132.

Nagaki K, Cheng Z, Ouyang S *et al.* (2004) Sequencing of a rice centromere uncovers active genes. *Nat Genet* **36**: 138–145.

Nath J, Tucker JD, Hando JC (1995) Y chromosome aneuploidy, micronuclei, kinetochores and aging in men. *Chromosoma* **103**: 725–731.

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

Palmer DK, O'Day K, Wener MH, Andrews BS, Margolis RL (1987) A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J Cell Biol* **104**: 805–815.

Pedeutour F, Forus A, Coindre JM *et al.* (1999) Structure of the supernumerary ring and giant rod chromosomes in adipose tissue tumors. *Genes Chromosomes Cancer* **24**: 30–41.

Peretti D, Maraschio P, Lambiase S, Lo Curto F, Zuffardi O (1986) Indirect immunofluorescence of inactive centromeres as indicator of centromeric function. *Hum Genet* **73**: 12–16.

Perry J, Slater H, Choo KHA (2004) Centric fission-simple and complex mechanisms. *Chromosome Res* (in press)

Rivera H, Vassquez AI, Ayala-Madrigal ML, Ramirez-Duenas ML, Davalos IP (1996) Alphoidless centromere of a familial unstable inverted Y chromosome. *Ann Genet* **39**: 236–239.

Saffery R, Irvine DV, Griffiths B, Kalitsis P, Wordeman L, Choo KH (2000) Human centromeres and neocentromeres show identical distribution patterns of >20 functionally important kinetochore-associated proteins. *Hum Mol Genet* **9**: 175–185.

Saffery R, Wong LH, Irvine DV *et al.* (2001) Construction of neocentromere-based human minichromosomes by telomere-associated chromosomal truncation. *Proc Natl Acad Sci USA* **98**: 5705–5710.

Schmitz A, Guilly MN, Trask B, van den Engh G (1992) Immunofluorescent labeling of centromeres for flow cytometric analysis. *Cytometry* **13**: 478–484.

Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. *Science* **294**: 109–115.

Shelby RD, Vafa O, Sullivan KF (1997) Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. *J Cell Biol* **136**: 501–513.

Spence JM, Critcher R, Ebersole TH *et al.* (2002) Co-localisation of centromere activity, proteins and topoisomerase II within a subdomain of the major human X a-satellite array. *EMBO J* **21**: 5269–5280.

Sugimoto K, Fukuda R, Himeno M (2000) Centromere/kinetochore localization of human centromere protein A (CENP-A) exogenously expressed as a fusion to green fluorescent protein. *Cell Struct Funct* **25**: 253–261.

Sullivan KF, Hechenberger M, Masri K (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J Cell Biol* **127**: 581–592.

Suzuki N, Nakano M, Nozaki N, Egashira S, Okazaki T, Masumoto H (2004) CENP-B interacts with CENP-C domains containing Mif2 regions responsible for centromere localization. *J Biol Chem* **279**: 5934–5946.

Tyler-Smith C, Oakey RJ, Larin Z *et al.* (1993) Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. *Nat Genet* **5**: 368–375.

Van Hooser AA, Ouspenski II, Gregson HC *et al.* (2001) Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. *J Cell Sci* **114**: 3529–3542.

Ventura M, Archidiacono N, Rocchi M (2001) Centromere emergence in evolution. *Genome Res* **11**: 595–599.

Voullaire LE, Slater HR, Petrovic V, Choo KH (1993) A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *Am J Hum Genet* **52**: 1153–1163.

Voullaire L, Saffery R, Davies J *et al.* (1999) Trisomy 20p resulting from inverted duplication and neocentromere formation. *Am J Med Genet* **85**: 403–408.

Warburton PE, Cooke CA, Bourassa, S *et al.* (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Curr Biol* **7**: 901–904.

Wieland, G, Orthaus S, Ohndorf S, Diekmann S, Hemmerich P (2004) Functional complementation of human centromere protein A (CENP-A) by Cse4p from *Saccharomyces cerevisiae. Mol Cell Biol* **24**: 6620–6630.

Yoda K, Ando S, Morishita S *et al.* (2000) Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. *Proc Natl Acad Sci USA* **97**: 7266–7271.

Zinkowski RP, Meyne J, Brinkley BR (1991) The centromere-kinetochore complex: a repeat subunit model. *J Cell Biol* **113**: 1091–1110.