

# Genomic microarray analysis reveals distinct locations for the CENP-A binding domains in three human chromosome 13q32 neocentromeres

Alicia Alonso<sup>1</sup>, Radma Mahmood<sup>1</sup>, Shulan Li<sup>1</sup>, Fanny Cheung<sup>1</sup>, Kinya Yoda<sup>2</sup> and Peter E. Warburton<sup>1,\*</sup>

<sup>1</sup>Department of Human Genetics, Mount Sinai School of Medicine, New York, NY 10029, USA and <sup>2</sup>Bioscience Center, Nagoya University, Nagoya 464-8601, Japan

Received July 9, 2003; Revised and Accepted August 13, 2003

Human neocentromeres are fully functional centromeres that provide mitotic stability to rearranged chromosomes that have separated from endogenous centromeres. A disproportionate number of neocentromeres has been observed in certain regions such as chromosome 3q ( $n=6$ ), 15q ( $n=9$ ) and 13q32 ( $n=7$ ), suggesting that these regions contain DNA sequences with a high propensity for neocentromere formation. Therefore, we have addressed the role of primary DNA sequence in neocentromere formation by asking whether multiple independent neocentromeres that were cytologically localized to chromosome 13q32 are in fact localized to the same underlying genomic DNA. Analysis of four independent 13q32 neocentromeres using simultaneous FISH with ordered YAC probes and immunofluorescence with antibodies to CENP-C have localized three neocentromeres to a distal ~7 Mb domain in chromosome 13q32, and one to an overlapping proximal domain of ~7 Mb. DNA was obtained from three of these neocentromeres by CENP-A chromatin immunoprecipitation (ChIP) and used to screen ordered BACs using both a slot-blotted BAC pool approach and a genomic microarray that contiguously spans 13q31.3–13q33.1. The CENP-A binding domains from each of these neocentromeres was identified to distinct genomic locations of ~130, 215 and 275 kb within an ~6.5 Mb region. Thus, the lack of coincidence of these neocentromeres to the same underlying DNA sequence refutes the idea of a DNA sequence based neocentromere 'hotspot' in 13q32 and further supports the sequence-independent epigenetic formation of human neocentromeres. The screening of genomic microarrays with ChIP DNA provides a powerful method to identify mammalian DNA sequences associated with particular functional chromatin states.

## INTRODUCTION

Centromeres are the critical chromosomal component responsible for the proper segregation of replicated chromatids to daughter cells during mitosis and meiosis (1,2). Normal endogenous human centromeres, and the proteinaceous kinetochores that form there, are found on megabase-sized arrays of tandemly repeated alpha satellite DNA (3). Formation of de novo centromeres on transfected alpha satellite DNA in human cells suggests a role for this DNA sequence in centromere formation (4–6). The inner kinetochore contains alpha satellite DNA in chromatin marked by Centromere Protein A (CENP-A), the centromere-specific histone H3 variant (7,8), and CENP-C, an additional DNA binding protein (9,10). The heterochromatin found between the sister kinetochores

also contains alpha satellite DNA, as well as CENP-B, Heterochromatin Protein 1, and specific covalent modifications of histone N-terminal tails (11). The large amounts and highly homologous repeat structure of alpha satellite DNA at human centromeres presents a major obstacle to the molecular analysis of human centromere structure and function (3).

A rare class of human variant centromeres, called neocentromeres, provide mitotic stability to rearranged chromosomes that have separated from endogenous centromeres (11–13). At least 60 cases of human neocentromere formation have been described thus far on derivatives of 19 different human chromosomes, identified primarily during routine clinical chromosomal analyses (12–14). Neocentromeres form fully functional kinetochores on apparently low-copy complex genomic DNA in the absence of detectable alpha satellite

\*To whom correspondence should be addressed at: Department of Human Genetics, Box 1498, Mount Sinai School of Medicine, 1425 Madison Ave, East Bldg 14-52A, New York, NY 10029, USA. Tel: +1 2126596728; Fax: +1 2128492508; Email: peter.warburton@mssm.edu

DNA, suggesting that centromere formation is an epigenetic process that may not be strictly dependent on primary DNA sequence (15–17). Nevertheless, a disproportionate number of neocentromeres have formed in certain chromosomal regions, such as 3q ( $n=6$ ), 15q ( $n=9$ ), and 13q ( $n=12$ ) (12–14,18), suggesting that these regions have an increased propensity for neocentromere formation. Thus, analysis of the genomic locations of these recurring neocentromeres permits investigation into the hypothesis that a particular underlying primary DNA sequence is responsible for the relatively high frequency of observed neocentromeres in these regions.

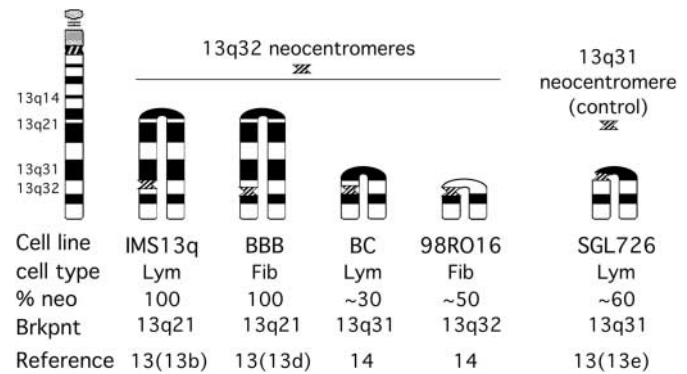
Here we present analysis of the chromosomal positions of four independent neocentromeres previously cytologically localized to chromosome 13q32. These cytological localizations were further refined using simultaneous FISH with physically mapped probes and immunofluorescence (IF) with antibodies to CENP-C. Three of the neocentromeres were localized to a  $\sim 7$  Mb domain spanning distal 13q32–13q33.1, and one neocentromere was localized to a  $\sim 7$  Mb domain spanning 13q31.3 to proximal 13q32. These domains overlapped in a  $\sim 3.0$  Mb region that also contained the inversion breakpoints of three neocentric invdup13q32 chromosomes. DNA obtained by CENP-A chromatin immunoprecipitation (ChIP) was used to screen ordered BACs using both a slot-blotted BAC pool approach and a unique genomic microarray contiguously spanning 14 Mb across 13q31.3–13q33.1. This analysis permitted the CENP-A binding domains of three neocentromeres to be localized to distinct domains of 130–275 kb within  $\sim 6.5$  Mb in chromosome 13q.

## RESULTS

We have previously described a total of 12 cell lines that each contain a supernumerary derivative 13q chromosome with a neocentromere, eleven of which are symmetrical inverted duplication chromosomes (invdups) with inversion breakpoints ranging from q14 to q32 (13,14), and one which is a small marker chromosome derived from an interstitial deletion of 13q21→22 (18). This collection of 12 patient-derived cell lines represent 21% of the 60 cases of neocentromeres thus far described, strongly suggesting that chromosome 13q represents a region with a high propensity to form neocentromeres (12–14). These neocentromeres were initially cytologically localized by positioning the neocentromere constriction to a chromosomal G or R band, with four localized in 13q21, one in 13q31, and seven in 13q32. Previous molecular cytogenetic analysis of each cell line confirmed the neocentromere by the absence of detectable alpha satellite DNA by FISH and the presence of kinetochore proteins including CENP-A, CENP-C, CENP-H and Mad2 (13,14,19). Four cell lines were chosen for further analysis because they had a derivative chromosome 13 with a 13q32 neocentromere in a relatively high proportion of cells, and because they were suitable for expansion to high numbers in culture (Fig. 1).

### Molecular cytogenetic localization of 13q32 neocentromeres

In order to further refine the cytological localization of 13q32 neocentromeres in these cell lines (Fig. 1), we performed FISH using YAC probes spanning 16 Mb across 13q31.3–13q33, and simultaneous IF using antibodies to CENP-C. These YAC FISH



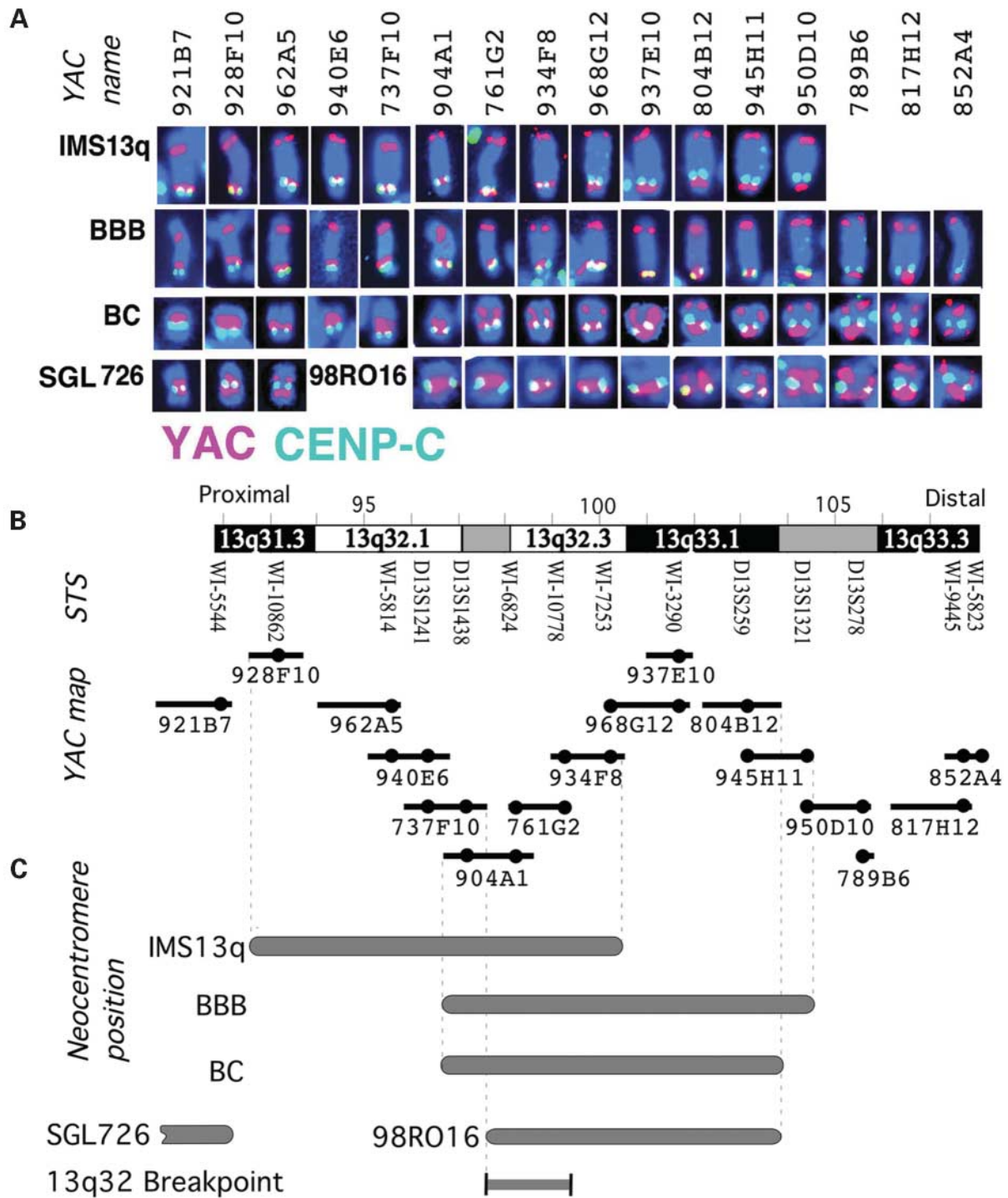
**Figure 1.** Ideograms of neocentric chromosomes examined. Ideograms for the invdup neocentric chromosomes in the five cell lines analyzed, shown relative to a normal chromosome 13 (left). Lym-lymphoblasts, Fib-fibroblasts. The name used for each cell line in the original references is given when necessary, e.g. ref 13 (cell line 13b).

probes hybridize to DNA sequences on both arms of the symmetrical inversion duplication chromosomes (Fig. 2A). Analysis of the morphologically similar invdup13q21 chromosomes from cell lines IMS13q and BBB showed that the CENP-C signals at these neocentromeres were in distinct locations relative to the YAC probes (e.g. compare YACs 928F10 or 937E10; Fig. 2A). The overlap between the CENP-C IF and the YAC FISH probes defined an  $\sim 7$  Mb proximal neocentromere domain for cell line IMS13q (YACs 928F10–934F8; Fig. 2B) and a distal  $\sim 7$  Mb neocentromere domain for cell line BBB (YACs 904A1–945H11; Fig. 2B), which overlapped by  $\sim 3$  Mb (Fig. 2C).

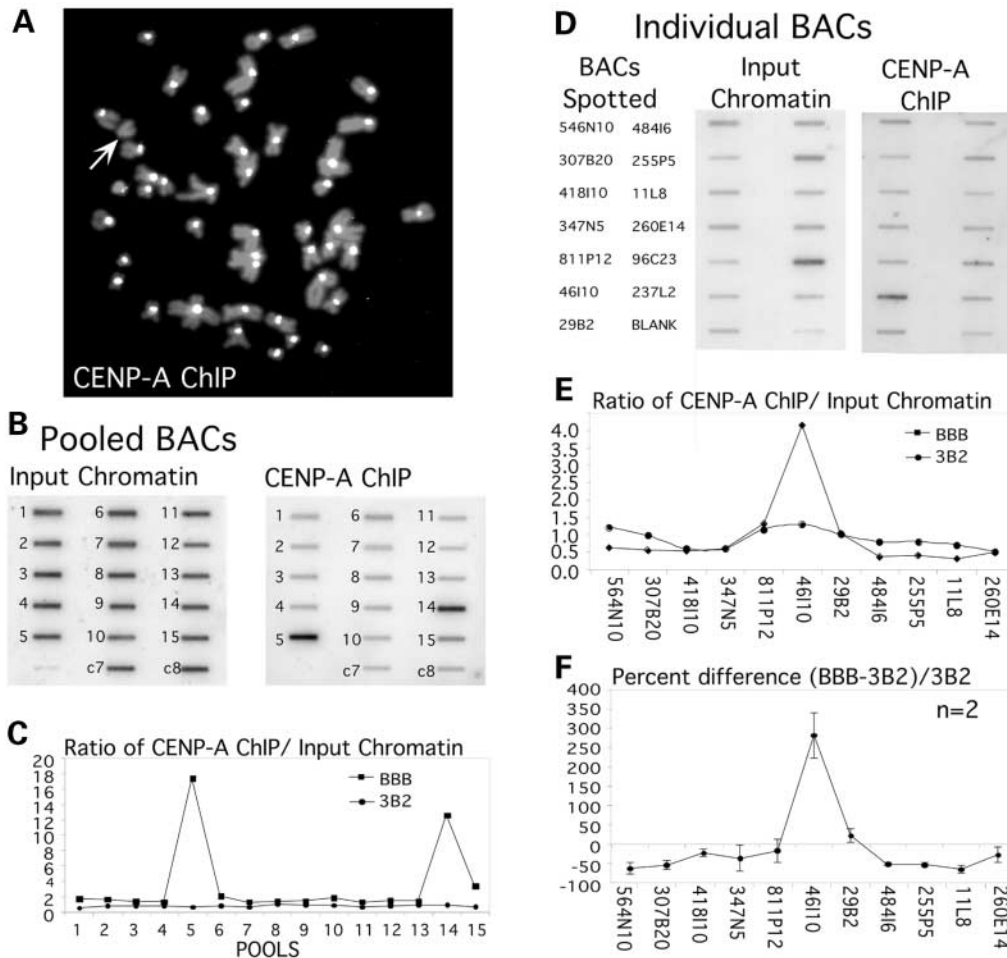
Similar analysis was performed on the morphologically similar invdup13q31 chromosomes in cell lines BC, which contains a 13q32 neocentromere, and cell line SGL726, which contains a 13q31 neocentromere included here as an outlying control (Fig. 1). YACs resolved from a single large FISH signal near the inversion breakpoints into separate signals on either chromosome arm for more proximal YACs in cell line SGL726 (YAC 962A5; Fig. 2A) than in cell line BC (YAC 904A1; Fig. 2A), confirming a more proximal breakpoint in cell line SGL726 (14). The CENP-C at the neocentromere in cell line SGL726 colocalized with YACs nearest to the inversion breakpoint (e.g. YAC 921B7; Fig. 2A), consistent with its cytologically localized 13q31 neocentromere. In contrast, the CENP-C at the neocentromere in cell line BC always appeared offset from YACs nearest the inversion breakpoint (e.g. YAC 921B7), and was localized to approximately the same distal neocentromere domain as seen in cell line BBB (Fig. 2C).

The inversion breakpoint in the invdup13q32 chromosome from cell line 98RO16 was defined by the absence of hybridization with YAC 737F10. The CENP-C at this neocentromere appeared to colocalize with YACs nearest the inversion breakpoint (e.g. YAC904A1; Fig. 2A), and was localized to approximately the same distal neocentromere domain as in cell line BBB (Fig. 2C). Notably, the inversion breakpoint in cell line 98RO16 was localized to the same  $\sim 2$  Mb region as the inversion breakpoints in two other previously reported invdup13q32 neocentric chromosomes (13,14).

This relatively low resolution molecular cytogenetic analysis was used to delineate the genomic regions, which extended beyond chromosome band 13q32, that we further examined



**Figure 2.** Molecular cytogenetic localization of 13q32 neocentromeres. (A) FISH and immunofluorescence analysis of neocentric chromosomes in each cell line as indicated. The YAC FISH probes (red) hybridized to either end of the invdup 13q chromosomes, and antibodies to CENP-C (green) mark the position of the neocentromeres. Representative pictures for each YAC for each cell line are shown. Cell line IMS13q- invdup13q21 chromosome, CENP-C colocalized with YACs 934F8 through YAC 928F10. Cell line BBB- invdup13q21 chromosome, CENP-C colocalized with YAC 945H11 through YAC 904A1. Cell line BC- invdup13q31 chromosome, YAC probes resolved into distinct signals on either chromosome arm at YAC 904A1. CENP-C colocalized with YACs 904A1 through 804B12. Cell line SGL726- invdup13q31 chromosome, YAC probes resolved into distinct signals on either chromosome arm at YAC 962A5. CENP-C was proximal (between) the FISH signals for YACs 928F10 and 962A5. Cell line 98RO16- invdup13q32 chromosome, YAC-737F10 did not hybridize, CENP-C colocalized with YACs 904A1 through 804B12. (B) Physical YAC map spanning 16.1 Mb from 13q31.3 to 13q33.3. STS markers indicate YAC overlaps. (C) The position of the proximal neocentromere domain in cell line IMS13q, and the distal neocentromere domain in cell lines BBB, BC and 98RO16 are shown. The common inversion breakpoint for cell lines 98RO16 and two other invdup13q32 chromosomes is shown. Vertical dotted lines indicate boundaries of the neocentromere domains on YAC map.



**Figure 3.** Hybridization of CENP-A ChIP DNA to slot-blotted BAC arrays. (A) Ligation-mediated PCR amplified CENP-A ChIP DNA hybridized to metaphase chromosomes from cell line BBB. All endogenous centromeres show a hybridization signal. The neocentromere (white arrow) does not show a hybridization signal. (B) Hybridization of input chromatin DNA and CENP-A ChIP DNA to slot blotted pooled BACs from 13q32/q33. BAC pools are as indicated in Table 1. Identical slot blots were hybridized with input chromatin DNA and CENP-A ChIP DNA from control cell line 3B2 (not shown). (C) Ratios of normalized CENP-A ChIP/ input chromatin signal intensities shown graphically for both BBB and 3B2. (D) Hybridization of input chromatin DNA and CENP-A ChIP DNA to slot blotted individual BAC DNA. 11 individual BACs centered on BAC46110 (underlined, Table 1) were slot-blotted, along with two control BACs not from contig. Identical slot blots were hybridized with input chromatin DNA and CENP-A ChIP DNA from control cell lines 3B2 and HeLa (not shown). (E) Ratios of normalized CENP-A ChIP/ input chromatin signal intensities shown graphically for both BBB (two independent experiments, mean  $\pm$  SE), and control cell line 3B2. (F) The percentage difference [(BBB-3B2)/3B2] in signal intensity, identifying BAC 46110 ( $\sim$ 300%) and BAC 29B2 ( $\sim$ 25%) as positive.

using more direct biochemical approaches to identify neocentromere DNA (15,16). Furthermore, the identification of the overlapping  $\sim$ 3 Mb region shared between the proximal and distal neocentromere domains suggested the possibility of a common neocentromere forming DNA sequence in that region.

#### ChIP of neocentromere DNA

In order to directly obtain neocentromere DNA, CENP-A ChIP was performed as described in the Materials and Methods. The quality of our CENP-A ChIP DNA was confirmed by sequencing random clones from independent CENP-A ChIP experiments from both neocentromere and control cell lines, which contained alpha satellite DNA in 38 to 60% of clones (7,20), while the remaining clones contained human repeats (e.g. LINES and SINES) or unique single copy sequences. We

performed ligation-mediated PCR amplification (LM-PCR) of the CENP-A ChIP DNA and input chromatin DNA for use as a hybridization probe. FISH using as probe LM-PCR amplified input chromatin DNA from cell line BBB painted human metaphase chromosomes (data not shown). FISH using LM-PCR amplified CENP-A ChIP DNA from cell line BBB showed intense hybridization to the centromeres of all human metaphase chromosomes (Fig. 3A), probably due to the Mb sized arrays of alpha satellite DNA at human centromeres acting as large targets for hybridization. The relatively homogeneous nature of these FISH signals at every endogenous human centromere suggests good enrichment for all centromeres in the CENP-A ChIP DNA. Nonetheless, a FISH signal was not seen at the neocentromere in cell line BBB under these hybridization conditions (Fig. 3A), presumably due to the relatively dilute neocentromere DNA in the FISH

**Table 1.** Slot-blotted BAC pools

	Pool 8	Pool 9	Pool 10	Pool 11	Pool 12	Pool 13	Pool 14	Pool 15
Pool 1	261P24	111L24	295B17	130L10	551M18	155N3	318G11	122A8
Pool 2	87L10	461N23	178C10	442I9	214F16	158C4	279D17	134O15
Pool 3	12G12	340C20	364F8	151A6	237L12	113J24	190K16	153O23
Pool 4	118F16	123G14	430M15	44I7	75C5	487A20	219L22	397O8
Pool 5	56G6	<u>564N10</u>	<u>307B20</u>	418I10	<u>347N5</u>	811P12	46I10	<u>29B2</u>
Pool 6	484I6	<u>255P5</u>	<u>11L8</u>	<u>260E 14</u>	113L12	<u>123H22</u>	357L1	485 E 1
Pool 7	<u>202O6</u>	<u>502J9</u>	<u>261F2</u>	<u>317H7</u>	78L19	233O7	262G14	17C16
Pool C7	100K21	166E2	566K4	128N18	165N12	96C23	237L2	
Pool C8	100K21	166E2	566K4	128N18	165N12	96C23	237L2	78J21

56 BACs spanning the BBB neocentromere domain (Fig. 2C) shown contiguously reading left to right across rows. Pools 1–7 (rows) contain eight contiguous BACs. Pools 8–15 (columns) contain seven non-contiguous BACs. Control pool C7 contains seven BACs. Control pool C8 contains eight BACs. BACs underlined were used in the individual BAC dot blot (Fig. 3D).

probe, which comprises at most ~2% of the total DNA (1 neocentromere/47 chromosomes).

#### Slot-blot hybridization of CENP-A ChIP DNA to ordered BAC clones

In order to determine whether the neocentromeres were in the region of overlap between the proximal and distal neocentromere domains (Fig. 2C), LM-PCR amplified CENP-A ChIP DNA from cell line BBB was used to screen the 56 ordered BACs spanning the 7 Mb distal neocentromere domain (including the region of overlap). We screened these BACs using a pooling approach, where identical molar quantities of BAC DNA were pooled into seven pools containing eight BACs each, and eight pools containing seven BACs each, such that each BAC was uniquely represented by two of the 15 pools (Table 1). This pooling approach permits screening of '*n*' BACs in twice the square root of *n* pools, significantly reducing the number of samples to be screened. These pools were immobilized on identical slot blots and hybridized with radiolabeled PCR amplified DNA from both the CENP-A ChIP and input chromatin from cell line BBB (Fig. 3B). Included on this slot blot were control pools (C7 and C8; Table 1) containing BACs from regions other than 13q32.

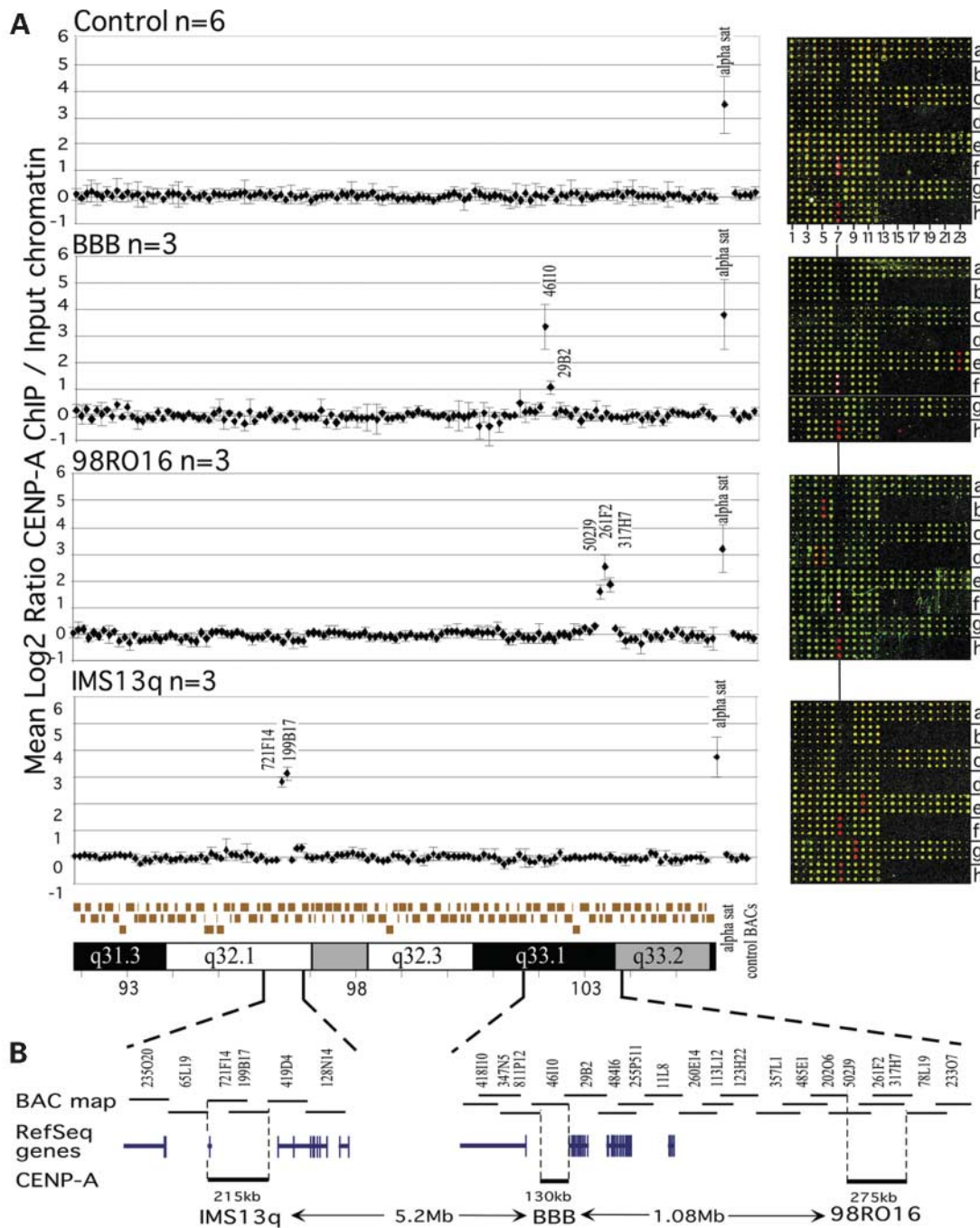
Quantitation of the hybridization signals was performed using a Phosphorimager following the described methods (16). On each blot, each hybridization signal was quantitated, local background subtracted and normalized to the control BACs on the blot, in order to control for differences in specific activity of the radiolabeled input chromatin and ChIP probes. The ratio of the hybridization signals from the CENP-A ChIP DNA to the input chromatin DNA was greatly increased for pools 5 and 14 (Fig. 3C). Thus, examination of Table 1 indicates that BAC 46I10 contains the CENP-A binding domain at the BBB neocentromere. BAC 46I10 was also uniquely identified when a different pool composition was screened using an independent CENP-A ChIP DNA as probe (data not shown). CENP-A ChIP was also performed on control cell lines, including HeLa cells and cell line 3B2, which was derived from an isolate of cell line BBB that had lost the neocentric chromosome. The ratio of the hybridization signals from the CENP-A ChIP DNA to the input chromatin DNA from these control cell lines to the

slot-blotted BAC pools shown in Table 1 revealed no increase in any pool (Fig. 3C and data not shown).

The individual BACs proximal and distal to BAC 46I10 in the contig (Table 1, underlined) were slot-blotted in equal molar quantities and hybridized with radiolabeled CENP-A ChIP DNA and input chromatin DNA from cell line BBB (Fig. 3D) and control cell lines 3B2 and HeLa (data not shown). The ratios of the hybridization signals were quantitated as performed for the BAC pools (Fig. 3E). This analysis confirmed BAC 46I10 as positive for CENP-A chromatin. A calculation of the percentage difference of the hybridization signal between cell line BBB and control cell lines showed a ~300% increase for BAC 46I10, and in addition an ~25% increase for BAC 29B2 (Fig. 3F). Notably, the CENP-A binding domain defined by these BACs was not within the ~3 Mb region of overlap with the proximal neocentromere domain (Fig. 2C), which demonstrated preliminarily that there were at least two distinct CENP-A binding domains for cell lines IMS13q and BBB.

#### Genomic microarray analysis of neocentromere position

In order to identify and compare the CENP-A binding domains from additional neocentromere cell lines, we developed a genomic microarray that spans both the proximal and distal neocentromere domains identified by FISH/IF. This microarray contained 126 contiguous BACs spanning 14.0 Mb of DNA across 13q31.3–13q33.1 (Fig. 4; <http://genome.cse.ucsc.edu>). These genomic microarrays show considerable technical advantages over the slot-blotted BAC approach shown above, including: (1) spotting each BAC in triplicate onto the microarray, ensuring reproducibility; and (2) performing a single simultaneous hybridization with Cy-5 labeled CENP-A ChIP DNA and Cy-3 labeled input chromatin DNA and identifying positive BACs by the Cy-5: Cy-3 intensity ratios. For each experiment in Fig. 4, the results of three independent CENP-A ChIP experiments for each cell line are shown. For control cell lines 3B2 and HeLa, all BACs showed background ratios ( $\log_2 < 0.26$ ; Fig. 4A), indicating no CENP-A binding in the region. An alpha satellite DNA plasmid included as a positive control in each microarray always showed highly



**Figure 4.** Genomic microarray analysis of neocentromere CENP-A binding domains. (A) Each microarray contains 126 contiguous BACs spanning 14 Mb from 13q31.3 to 13q33.2, shown across the x-axis at the bottom. Each microarray was hybridized simultaneously with CENP-A ChIP (Cy5-red) and input chromatin (Cy3-green). The mean Log<sub>2</sub> Cy-5: Cy-3 ratios and standard deviation for each BAC are shown plotted on the y-axis for multiple independent ChIP experiments from each cell line. A plasmid containing alpha satellite DNA was included as a positive control (log<sub>2</sub> ratio 3.5 ± 0.29). Control BACs (RP11-237L2, -96C23, -355P18, -78J21 and -128N18) from regions other than 13q31.3/13q33.2 (mean log<sub>2</sub> ratio 0.04 ± 0.1). A merged image of a representative microarray is shown to the right of each panel; BAC triplicates are spotted vertically (rows a-h), alpha satellite DNA in positions f7 and h7. The x-axis shows only unique sequenced regions for each BAC, spanning from proximal RP11-165N12 (AL159152) to distal RP11-358P11 (AL139379). BAC names and overlaps can be obtained from the April 2003 genome assembly. Control cell lines (top panel), n = 6 (three each for HeLa and 3B2), all BACs: mean log<sub>2</sub> ratio = 0.06 ± 0.09. Cell line BBB (second panel), n = 3, BAC RP11-46110: mean log<sub>2</sub> ratio = 3.36 ± 0.84 (position e23); BAC RP11-29B2: mean log<sub>2</sub> ratio = 1.08 ± 0.26 (position g23); all other BACs: mean log<sub>2</sub> ratio = 0.00 ± 0.15. Cell line 98RO16 (third panel), n = 3, BAC RP11-502J9: mean log<sub>2</sub> ratio = 1.62 ± 0.26 (position d4); BAC RP11-261F2: mean log<sub>2</sub> ratio = 2.55 ± 0.47 (position b5); BAC RP11-317H7: mean log<sub>2</sub> ratio = 1.90 ± 0.27 (position d5); all other BACs: mean log<sub>2</sub> ratio = -0.01 ± 0.13. Cell line IMS13q (bottom panel), n = 3, BAC RP11-721F14: mean log<sub>2</sub> ratio = 2.84 (position e10); BAC RP11-199B17: mean log<sub>2</sub> ratio = 3.15 (position g9); all other BACs: mean log<sub>2</sub> ratio = 0.02 ± 0.27. (B) The expanded areas show the 2.3 Mb region containing the CENP-A binding domain from BBB (130 kb) and 98RO16 (275 kb), and the ~1 Mb region containing the 215 kb CENP-A binding domain from IMS13q neocentromere. BAC overlaps are indicated. The genes in the area as seen by RefSeq are shown. The positions of RefSeq genes are shown approximately, please refer to UCSC Genome website for higher resolution. Genes shown near IMS13q CENP-A domain: HS6ST3, LOC283476, GPR80, MBNL2, RAP2A. Genes shown near BBB CENP-A domain: FGF14, TPP2, LOC93081, KDELC1, BIVM, ERCC5, SLC10A2.

significant increased ratios ( $\log_2 > 3.0$ ). The ability to identify neocentromere CENP-A binding regions using the genomic microarray approach was first validated by examination of cell line BBB, where all BACs showed background Cy-5: Cy-3 ratios ( $\log_2 < 0.47$ ), except for BACs RP11-46I10 and, more weakly, BAC RP11-29B2, in close agreement with results obtained using the BAC slot blot approach (Fig. 3). A CENP-A binding domain of  $\sim 130$  kb was defined for cell line BBB by removal of the sequence overlap between BAC RP11-46I10 with the neighboring proximal BAC, but inclusion of the overlap with BAC RP11-29B2 (Fig. 4B).

We then used the genomic microarray to define the CENP-A binding domain for cell line 98RO16, which was also in the  $\sim 7$  Mb distal neocentromere domain (Fig. 2). All BACs showed background ratios ( $\log_2 < 0.27$ ), except for three contiguous BACs RP11-502J9, -261F2 and -317H7 (Fig. 4A). A CENP-A binding domain of  $\sim 275$  kb was defined for cell line 98RO16 by removal of the sequence overlap with neighboring proximal and distal BACs (Fig. 4B). The relative intensity ratios and extent of overlap of the three positive BACs suggests an even smaller CENP-A binding domain centered on BAC RP11-261F2. This domain was  $\sim 1$  Mb distal to the domain from cell line BBB.

Analysis of cell line IMS13q showed background ratios for all BACs ( $\log_2 < 0.36$ ) except for two contiguous BACs RP11-721F4 and -199B17 (Fig. 4A). These BACs have minimal overlap (2000 bp) with the proximal and distal BACs, and define a CENP-A binding domain of  $\sim 215$  kb for cell line IMS13q, which was  $\sim 5$  Mb proximal to the domain from cell line BBB (Fig. 4B). Thus, the CENP-A binding domains for the three neocentromeres examined here all occupy distinct genomic locations (Fig. 4B), and are considerably smaller than those previously described by CENP-A ChIP analysis on chromosomes 10q25 and 20p12 (15,16).

### Sequence analysis of neocentromere DNA

The identification of these three new CENP-A binding domains using CENP-A ChIP provides an opportunity to perform DNA sequence comparisons between themselves and to the other reported neocentromere regions domains previously identified in chromosomes 10q25, 20p12 and 9p23 (15,16,21). Owing to the relatively large size of endogenous human centromeres, we also compared the 2 Mb of DNA flanking these CENP-A binding domains, in order to include regions potentially involved in formation of heterochromatin or other centromeric chromatin domains. Pairwise comparison in both orientations of these 4 Mb regions using AVID (22) and VISTA (23) (50 bp window, 50% homology) showed no significant homologies other than between known interspersed repeats. A BLASTn comparison of each of the three CENP-A domains identified here to the non-redundant DNA sequence data base with low stringency parameters (word size = 7, match = 1, mismatch = -1, open gap = -2 and extend gap = -1, human repeats masked) revealed no significant homologies to other regions of the genome. REPEATMASKER analysis of the CENP-A binding domains (Table 2) revealed that all neocentromere domains have a somewhat elevated AT richness and reduced percentage of SINES (Smit, AFA & Green, P

RepeatMasker at <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). The largest tandem repeat observed using Tandem Repeat Finder (24), located  $\sim 1.5$  Mb proximal to the BBB CENP-A domain, consisted of 19 copies (99% homologous) of a 56 bp 66% AT sequence, which showed no similarity to repeats in the other neocentromere regions. No internal duplications, multiple binding sites for CENP-B (TTCGNNNNANNCGGG), or homology to alpha satellite DNA were present in these regions. Analysis of the Known Genes and RefSeq tracks of the UCSC genome browser showed that all six neocentromere domains fall in relatively gene poor regions, with only the 10q25.3 CENP-A domain actually coinciding with a gene (KIAA0534), which is an unconfirmed partial coding sequence. Indeed, both the IMS13q and BBB CENP-A domains appear to fall directly between genes, while the 98RO16 CENP-A domain is located in a 2.5 Mb 'gene desert' (Fig. 4B).

### DISCUSSION

We have analyzed the genomic positions of four independent neocentromeres that had been cytologically localized to chromosome band 13q32 (Fig. 1). Simultaneous FISH with ordered YAC probes and immunofluorescence with antibodies to CENP-C permitted these neocentromere domains to be localized to overlapping  $\sim 7$  Mb chromosomal regions (Fig. 2). Screening contiguous BAC clones spanning these regions on slot blots (Fig. 3) and genomic microarrays (Fig. 4) by hybridization with DNA immunoprecipitated with CENP-A chromatin has revealed that at least three of these CENP-A domains occur in distinct genomic locations.

Human neocentromeres represent the epigenetic formation of a distinct functional chromatin structure, specifically the establishment of CENP-A chromatin and centromere function, in a previously non-centromeric location (8,17). Previous analysis of the CENP-A binding domains at two human neocentromeres failed to detect any significant conserved sequence features, although their formation on different chromosomes precluded the possibility that they were forming on precisely the same DNA sequence (15,16). Nevertheless, a role for the underlying primary DNA sequence in the establishment, choice or position of epigenetically determined chromatin structures is currently under investigation (25,26). Thus, the analysis performed here of multiple independent neocentromeres from chromosome 13q32 permitted the search for a DNA sequence within chromosome 13q32 that may play a role in neocentromere formation (a neocentromere hotspot), which could account for the relatively high frequency of observed neocentromeres in this region (1,12,13). Importantly, the identification of such a 'neocentromere hotspot' in the region would have been independent of our ability to recognize centromere-forming determinants in the primary DNA sequence. However, multiple distinct CENP-A binding domains in 13q32/q33.1 (Fig. 4) were identified, which suggested that there is no DNA sequence-based neocentromere hotspot in 13q32.

This report doubles the number of CENP-A binding domains available for analysis, adding the three identified in 13q32 (Fig. 4) to the two previously identified by CENP-A ChIP on

**Table 2.** Sequence analysis of human neocentromeres

Location	BBB	IMS13q	98RO16	10q25	20p12	9p23	Genome average
Basepairs	131 kb	214 kb	275 kb	329 kb	464 kb	437 kb	2.8 Gbp
%AT	61.2	59.9	65.0	65.4	61.1	64.7	59.0
Lines (%)	29.4	23.7	21.2	46.2	16.9	28.4	21.0
Sines (%)	7.7	9.2	6.5	6.4	6.3	4.5	13.6
LTRs (%)	12.3	8.6	8.0	10.1	8.2	19.3	8.6
CENP-B	0 <sup>a</sup>	1 <sup>a</sup>	0 <sup>a</sup>	0	0	0	

<sup>a</sup>Plus 2 Mb flanking sequences.

chromosomes 10q25.3 and 20p12 (15,16) and one identified by FISH and immunofluorescence on chromosome 9p23 (21). Extensive low-stringency pairwise sequence comparison of both the CENP-A binding regions as well as the 2 Mb of flanking DNA sequences failed to reveal any sequence characteristics shared even between any two of these regions, except for known human DNA interspersed repetitive elements. Thus, sequence comparisons of this relatively large set of neocentromere CENP-A binding domains strongly refute the possibility of DNA sequence preferences or 'desirable features' (1) that support neocentromere formation (12). Of course, it is possible that our analysis may fail to detect extremely subtle or poorly conserved sequence characteristics or secondary structural features in these regions that may be revealed by more sophisticated DNA sequence comparisons.

The position of two 15q neocentromeres has recently been roughly associated (to within several Mb) with genomic duplicons near inversion duplication breakpoints, which were in the general vicinity of an ancestral centromere observed in Old World Monkeys (27). However, no genomic duplicons or ancestral centromeres (from primates) have been found in the vicinity of the chromosome 13 neocentromeres in this study.

Current models of human centromeres suggest organization into distinct functional domains of chromatin, such as CENP-A containing chromatin in the kinetochore, and heterochromatin between the kinetochores (11). Recent evidence suggests that multiple domains of CENP-A chromatin are interspersed with Histone H3 chromatin at human and *Drosophila* centromeres (28). The CENP-A domains we have described here cover considerably less DNA (130–275 kb; Fig. 4B) than the previously described CENP-A domains (330 and 460 kb) (15,16). This may simply reflect a greater density of overlapping BACs in our contig, or may actually indicate quite significant variation in the size of functional CENP-A domains. Despite the small size of these CENP-A domains, these neocentric chromosomes are fully mitotically stable in these cell lines (Table 1). The apparent avoidance of active genes by CENP-A domains (Fig. 4B) may reflect a greater barrier for CENP-A chromatin formation onto active chromatin. It is not likely to be due to a negative selection for gene inactivation, as there are multiple other homologous loci in these cells that retain these active genes. It will be interesting to address whether and to what extent neocentromere formation inactivates nearby genes.

At endogenous centromeres, CENP-A has been shown to bind to some portion of the megabase-sized arrays of alpha satellite DNA, and to establish the inner kinetochore plate domain (7,28–30), while the remainder of the alpha satellite

DNA is presumably involved in the heterochromatin and/or sister cohesion functions (31). Thus, it is reasonable to assume that several megabases of DNA will also be involved in formation of the neocentromere. The CENP-A domains from cell lines BBB and 98RO16 are separated by ~1 Mb of DNA (Fig. 4B), which is within the range where they may have heterochromatin and/or cohesion domains in common. Thus, while there appears to be plasticity as to the position and size of the CENP-A domains, there may be preferred sites for heterochromatin or cohesion domains. It is unlikely, though, that the neocentromere from cell line IMS13q, which is ~5 Mb away, shares any domains with the BBB or 98RO16 neocentromeres. Using ChIP with additional antibodies (e.g. anti-Histone H3 K9-methyl or SMC3) coupled with our genomic microarray, the size and organization of the different chromatin domains at these neocentromeres can be mapped onto the underlying DNA sequence, which is not possible at endogenous centromeres due to the highly homologous alpha satellite DNA.

If it is not DNA sequence *per se* that specifies neocentromere location, what other factors may be playing a role in this determination? Endogenous centromeres replicate asynchronously and are not the latest DNA to replicate within the cell, indicating that centromere identity is not determined by temporal separation from bulk chromatin (32,33). Nevertheless, replication analysis of a normal chromosome 10 reveals that the ~450 kb region shown to contain the 10q25.3 neocentromere CENP-A binding domain was approximately the latest to replicate within the ~5 Mb examined, replicating in mid to late S phase. Upon neocentromere formation, an ~1.5 Mb region surrounding the CENP-A binding domain shifted its replication timing into mid to late S phase as well (15), suggesting that neocentromere formation is sensitive to and can affect local replication timing. However, treatment with TSA, a histone deacetylase inhibitor, can abolish this domain of delayed replication timing without affecting neocentromere function (34). These data suggest that the latest DNA to replicate within a localized chromosomal region may be predisposed to neocentromere formation (11,12), perhaps due to a localized depletion of histone H3 or other chromatin proteins. Our genomic microarray will facilitate comparison of the replication profile across chromosome 13q32/q33 to the positions of the multiple independent neocentromeres found there, which will help to clarify the role of local replication timing in neocentromere formation.

Neocentromere formation in *Drosophila* has been experimentally observed to occur by local *cis* spreading of endogenous centromeric chromatin to adjacent DNA sequences

during chromosomal rearrangement events (35,36). However, human neocentromeres are usually found far removed from endogenous centromere locations, most often on distal inverted duplication chromosomes (12). Interestingly, the three smallest invdup13q32 neocentric chromosomes reported, including the one from cell line 98RO16, appear to share a common inversion breakpoint (within ~2 Mb; Fig. 2) (13,14), suggesting the possibility of a chromosome region that may be prone to rearrangements and/or neocentromere formation. However, further analysis showed that none of the CENP-A binding domains identified in this study occurred in this region, and thus do not support a direct relationship between chromosome breakpoints and neocentromere positions. Indeed, the extreme rarity of observed neocentric chromosomes may in part be due to the low probability of two independent rare events: (1) a chromosomal rearrangement that creates an acentric chromosome; and (2) the formation of the neocentromere itself, to both occur on the same chromosome. Thus, mapping of these CENP-A binding domains within chromosome 13q represents a solid foundation for the characterization of human neocentromere structure and function and the factors that are important in their formation.

## MATERIALS AND METHODS

### Cell lines

HeLa cells, EBV transformed lymphoblasts (IMS13q, BC, SGL726) and fibroblast 98R016, BBB and 3B2 were grown in standard media (13,14). The control cell line 3B2 was obtained from isolated colonies of cell line BBB that had lost the neocentromere chromosome.

### 13q32 YAC and BAC contigs

The YAC contig (Fig. 2C) was assembled using the Genethon physical map (37). YACs were obtained from Research Genetics (Huntsville, AL, USA) and grown in selective media. The 14 Mb BAC contig was assembled by the UCSC Genome Bioinformatics group (<http://genome.cse.ucsc.edu>) and included all BACs from RP11-165N12 (AL159152) to RP11-358P11 (AL139379), which were obtained from the Wellcome Trust Sanger Institute (Cambridge, UK). BACs were grown in LB/kanamycin and purified using Qiagen columns (Qiagen Inc., Valencia, CA, USA). BACs and YACs were verified by FISH and STS content, and sized by PFGE. YAC and BAC maps were aligned by STS marker positions.

### FISH and immunofluorescence

FISH and immunofluorescence were performed essentially as described (8,13). Yeast genomic DNA containing YACs was nick translated with biotin-dUTP, denatured at 72°C and preannealed with human Cot-1 and herring sperm DNA at 37°C for at least 2 h. Images were collected on a Nikon E800 equipped with a SONY DKC 5000, using single-pass FITC, TRITC and DAPI filters (Chroma, Brattleboro, VT, USA) and merged using Adobe Photoshop. Correct alignment of FITC and TRITC images was confirmed by comparison to images collected on a FITC/TRITC double pass filter (Chroma). Metaphases were scored by eye and

by imaging six to 10 representative neocentric chromosomes for each YAC for each cell line.

### Chromatin immunoprecipitation

Soluble chromatin was obtained by micrococcal nuclease digestion of  $5 \times 10^7$  nuclei as described (20). For immunoprecipitation, the sample was made 0.1% NP40 and precleared for 20 min at 4°C with mouse IgG at 1:1000 and 2% blocked protein G (Amersham Pharmacia Biotech, Piscataway, NJ, USA). After centrifugation (250g  $\times$  5 min, 4°C), mouse monoclonal anti-CENP-A (20) was added to the supernatant (Input chromatin) at a 1:500 dilution and incubated for 2 h  $\times$  4°C. The immunocomplexes were recovered by incubation with 6% Protein G for 2 h  $\times$  4°C and centrifugation, and washed three times in WB buffer (20 mM Hepes pH 8.0, 20 mM KCl, 0.5 mM EDTA, 0.5 mM DTT), 0.3 M NaCl. Both input chromatin and immunocomplexes were digested with 500 µg/ml of Proteinase K (Roche, Indianapolis, IN, USA) in TE (10 mM Tris pH 8.0, 1 mM EDTA), 0.5% SDS for 4 h  $\times$  56°C, and DNA recovered by phenol extraction and ethanol precipitation. DNA was quantified using the DNA DipStick Kit (Invitrogen, Carlsbad, CA, USA). Approximately 30 ng of CENP-A ChIP DNA were treated with calf intestine phosphatase (CIP) (NEB, Beverly, MA, USA), and *Taq* DNA polymerase (Roche) to add dA tails, and ligated to pCRII-TOPO vector (Invitrogen). White colonies were grown and sequenced in an ABI Prism 3700 Capillary Array Sequencer, and the sequence identified using the NCBI BLASTn server ([www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST)).

### PCR amplification and labeling of chromatin DNA

Ligation-mediated PCR amplification was carried out as described (38), with slight modifications. Four nanogram each of input chromatin DNA and CENP-A ChIP DNA were CIP- and kinase-treated (NEB). They were ligated with 4 U of T4 DNA ligase (NEB) to linkers at a concentration of 0.1 µM, 4°C  $\times$  ON. Linkers were made by annealing a phosphorylated 24mer 5'pAGAGCGGCCGCTTCGAGCACTCAG to a 20mer 5'GTGCTCGAAGCGGCCGCTCT. Amplification was carried out using 1 ng of ligated product with 0.1 U of *Taq* polymerase (Roche) and 1 µM of the 20mer. The amplification cycles were one cycle of 94°C for 2 min; 35 cycles of 94°C for 1 min; 55°C for 1 min, 72°C for 3 min; and one cycle of 94°C for 1 min, 55°C for 1 min, 72°C for 10 min. Amplified products were purified using Qiagen columns (Qiagen Inc.) and quantified. For FISH probes, ~20 ng of amplified products were nick-translated with biotin dUTP (Roche). For slot blot hybridization, about 100 ng of PCR amplified product were random primed with the NEBlot kit (NEB), using high specific activity  $\alpha$ 32P-dATP (6000 Ci/mmol). For microarray hybridization, 1 ng of the ligated product was PCR amplified using 0.04 U/µl of *Taq* Pol (Roche), 1 µM of the 20mer, 250 µM of dATP, dCTP, dGTP, 50 µM of dTTP and 200 µM of aminoallyl-dUTP (Sigma, St. Louis, MO, USA). A 2 µg aliquot of the amplified product was dissolved in 5 µl of 0.2 M NaHCO<sub>3</sub> pH 9.0 and conjugated with 5 µl of 1.4 mM monofunctional reactive -Cy3 (Input chromatin) or -Cy5 (CENPA ChIP DNA) dye in DMSO (Amersham Biosciences UK Ltd,

Buckinghamshire, UK). Specific activity of the probes ranged from 70 to 150 nucleotides/dye.

### Hybridization to pooled BACs

Screening of the BAC contig was performed as described with modifications (16). Three picomoles of each of the 56 BACs (Table 1) were pooled into seven pools containing eight BACs each and eight pools containing seven BACs each, such that every BAC was uniquely represented by two of the 15 pools (Table 1). Control pools containing eight BACs (C8) or seven BACs (C7) from regions other than 13q32 were used to normalize each blot. The BAC pools were immobilized on Hybond-XL membranes (Amersham Pharmacia Biotech) using a Minifold II slot blot apparatus (Schleicher and Schuell, Keene, NH, USA). Blots were hybridized with  $2 \times 10^6$  cpm/ml of either input chromatin DNA or CENP-A ChIP DNA with specific activity of  $\sim 0.5\text{--}1 \times 10^9$  dpm/ $\mu\text{g}$ . The probe was preannealed for 2 h at 65°C with 300  $\mu\text{g}/\text{ml}$  of human placental DNA (Sigma). Hybridization was carried out ON at 65°C in 10% dextran sulfate, 1 M NaCl, 1% SDS and 50 mM Tris pH 7.5. Washes were performed at high stringency (0.1  $\times$  SSC/0.1% SDS, 65°C) and blots were quantified using a PhosphorImager system (Molecular Dynamics Storm 860) linked to ImageQuant software (Molecular Dynamics, CA, USA). Each hybridization signal was quantitated, local background subtracted, and normalized to the control BAC pools on the blot.

### BAC microarrays

Purified BAC DNAs were sonicated, resuspended in 50% DMSO at about 200 ng/ $\mu\text{l}$  and printed in triplicate onto aminosilane-coated glass slides (GAPS II, Corning, NY, USA), using a GMS 417 microarray pin and ring system (Affymetrix Inc., Santa Clara, CA, USA). Slides were pretreated with 0.145 M succinic borate pH 9.0 in 90% methylpyrrolidone for 30 min. Each array was prehybridized with 350  $\mu\text{g}$  of salmon sperm DNA, in 50% formamide, 2  $\times$  SSC, 10% dextran sulfate, and 4% SDS for 1 h at 42°C in an ArrayIt hybridization chamber (Telechem International Inc., Sunnyvale, CA, USA). A 500 ng aliquot of each Cy3- and Cy5-labeled probes were denatured with 7  $\mu\text{g}$  of Cot, 575  $\mu\text{g}$  of yeast tRNA, 10  $\mu\text{g}$  of sonicated *E. coli* DNA in 50% formamide, 2  $\times$  SSC and 4% SDS and preannealed at 42°C for 1 h before hybridizing to the microarray overnight at 42°C. The slides were washed for 5 min at room temperature in 1  $\times$  SSC, 0.1% SDS, followed by 0.2  $\times$  SSC, 0.1% SDS, and 0.2  $\times$  SSC. Arrays were imaged with an Affymetrix GMS 417 laser-based slide scanner (Affymetrix Inc.) and fluorescence intensity ratios (CENP-A ChIP-Cy5 to Input chromatin-Cy3) measured using Scan Array 2.0 (Perkin Elmer). Microarray images were false colored and merged in Adobe Photoshop for display purposes (Fig. 3A). For each microarray, the mean and standard deviation of the triplicate normalized ratios (Lowess) were calculated. Spots with a greater than 25% SD from the mean were rejected (less than 3% of total spots). For each microarray, the distribution of the intensity ratios was calculated, and positive BACs identified as those that were  $>3$  SD from the mean. BACs were identified as positive in three independent ChIP experiments.

### Sequence analysis

All DNA sequence positions in this report were obtained from the April 2003 freeze of the human genome sequence assembly (<http://genome.cse.ucsc.edu>). BAC clone RP11-46I10 is  $\sim 175$  kb, with basepairs 1–152 890 included in the genome assembly sequence as AL159155. RP11-46I10 overlaps proximal clone RP11-811P12 by 43 519 bp and overlaps distal clone RP11-29B2 by 21 649 bp. This overlap between RP11-46I10 and RP11-29B2 has been included in the putative CENP-A binding domain to account for the low but positive intensity ratio for BAC RP11-29B2. The overlapping BAC clones RP11-502J9 ( $\sim 200$  kb), RP11-261F2 ( $\sim 190$  kb) and RP11-317H7 ( $\sim 180$  kb) together cover  $\sim 380$  kb, with 275 241 bp included in the genome assembly sequence as AL356254, AL445226, AL358855, respectively. RP11-202O6 has an 86 783 bp overlap with RP11-502J9, which has a 56 145 bp overlap with RP11-261F2, which has a 135 913 bp overlap with RP11-317H7, which has a 20 242 bp overlap with RP11-78L19. The sequence coordinates and accession numbers for the CENP-A binding domains described here are as follows: cell line BBB-chr13: 101936717–102067737 (131 020 bp; AL159155, AL158063); cell line 98RO16-chr13: 103340031–103614973 (274 943 bp; AL356254, AL445226, AL358855); cell line IMS13q-chr13: 96473559–96687292 (213 733 bp; AL356486, AL359925). Sequence coordinates and accession numbers for the 3 other reported neocentromere domains are as follows: 10q25.3-chr10: 117112455–117441580 (329 125 bp; AL152304, AL357059 and AL356100, gene locus NCBI AB011106) (15); 20p12.2-chr20: 10663041–11127046 (464 005 bp; AL135937, 1050403, AL158042, 1050322, AL1078588) (16); 9p23-chr9: 10890577–11342074 (451 498 bp; AL162413, AL354952, 1354992, AL451129, AL355672) (21).

### ACKNOWLEDGEMENTS

The authors would like to thank Janet Partridge (Memphis) for comments on the manuscript. We gratefully acknowledge Tearina Chu and Yuexun Liu at the Mount Sinai Microarray Shared Resource Facility. We thank Andrew Dunham and the chromosome 13 group at the Wellcome Trust Sanger Institute (Cambridge, UK) for providing BACs and updated mapping information. This work was supported in part by a grant from the National Institutes of Health R01 GM061150.

### REFERENCES

- Mellone, B.G. and Allshire, R.C. (2003) Stretching it: putting the CEN(P-A) in centromere. *Curr. Opin. Genet. Dev.*, **13**, 191–198.
- Cleveland, D.W., Mao, Y. and Sullivan, K.F. (2003) Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell*, **112**, 407–421.
- Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K. and Willard, H.F. (2001) Genomic and genetic definition of a functional human centromere. *Science*, **294**, 109–115.
- Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H. and Masumoto, H. (1998) Creation of human artificial chromosomes by introduction of YACs retrofitted with human telomeric DNA. *Nat. Biotech.*, **16**, 431–439.
- Harrington, J.J., Bokkelen, G.V., Mays, R.W., Gustashaw, K. and Willard, H.F. (1997) Formation of de novo centromeres and construction

- of first-generation human artificial microchromosomes. *Nat. Genet.*, **15**, 345–355.
6. Grimes, B.R., Rhoades, A.A. and Willard, H.F. (2002) Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation. *Mol. Ther.*, **5**, 798–805.
  7. Vafa, O. and Sullivan, K.S. (1997) CENP-A is an inner kinetochore plate protein and is associated with alpha satellite DNA. *Curr. Biol.*, **7**, 897–900.
  8. Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B., Stetten, G., Gimelli, G., Warburton, D., Tyler-Smith, C., Sullivan, K.F. *et al.* (1997) Immunolocalization of CENP-A, a kinetochore-specific histone H3 variant, suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Curr. Biol.*, **7**, 901–904.
  9. Yang, C.H., Tomkiel, J., Saitoh, H., Johnson, D.H. and Earnshaw, W.C. (1996) Identification of overlapping DNA-binding and centromere-targeting domains in the human kinetochore protein CENP-C. *Mol. Cell Biol.*, **16**, 3576–3586.
  10. Politi, V., Perini, G., Trazzi, S., Pliss, A., Raska, I., Earnshaw, W.C. and Della Valle, G. (2002) CENP-C binds the alpha-satellite DNA *in vivo* at specific centromere domains. *J. Cell Sci.*, **115**, 2317–2327.
  11. Choo, K.H. (2001) Domain organization at the centromere and neocentromere. *Dev. Cell*, **1**, 165–177.
  12. Amor, D.J. and Choo, K.H. (2002) Neocentromeres: role in human disease, evolution, and centromere study. *Am. J. Hum. Genet.*, **71**, 695–714.
  13. Warburton, P.E., Dolled, M., Mahmood, R., Alonso, A., Li, S., Naritomi, K., Tohma, T., Nagai, T., Hasegawa, T., Ohashi, H. *et al.* (2000) Molecular cytogenetic analysis of eight inversion duplications of human chromosome 13q that each contain a neocentromere. *Am. J. Hum. Genet.*, **66**, 1794–1806.
  14. Li, S., Malafiej, P., Levy, B., Mahmood, R., Field, M., Hughes, T., Lockhart, L.H., Wu, Z., Huang, M., Hirschhorn, K. *et al.* (2002) Chromosome 13q neocentromeres: Molecular cytogenetic characterization of three additional cases and clinical spectrum. *Am. J. Med. Genet.*, **110**, 258–267.
  15. Lo, A.W., Craig, J.M., Saffery, R., Kalitsis, P., Irvine, D.V., Earle, E., Magliano, D.J. and Choo, K.H. (2001) A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. *EMBO J.*, **20**, 2087–2096.
  16. Lo, A.W., Magliano, D.J., Sibson, M.C., Kalitsis, P., Craig, J.M. and Choo, K.H. (2001) A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. *Genome Res.*, **11**, 448–457.
  17. Karpen, G.H. and Allshire, R.C. (1997) The case for epigenetic effects on centromere identity and function. *TIGS*, **13**, 489–496.
  18. Knegt, A.C., Li, S., Engelen, J.J.M., Bijlsma, E.K. and Warburton, P.E. (2002) Prenatal diagnosis of a karyotypically normal pregnancy in a mother with a supernumerary neocentric 13q21→13q22 chromosome and balancing reciprocal deletion. *Prenat. Diagn.*, **23**, 215–220.
  19. Sugata, N., Li, S., Earnshaw, W., Yen, T., Yoda, K., Masumoto, H., Munekata, E., Warburton, P. and Todokoro, K. (2000) Human CENP-H multimers colocalize with CENP-A and CENP-C at active centromere-kinetochore complexes. *Hum. Mol. Genet.*, **9**, 2919–2926.
  20. Ando, S., Yang, H., Nozaki, N., Okazaki, T. and 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.
  21. Satinover, D.L., Vance, G.H., VanDyke, D.L. and Schwartz, S. (2001) Cytogenetic analysis and construction of a BAC contig across a common neocentromeric region from 9p. *Chromosoma*, **110**, 275–283.
  22. Bray, N., Dubchak, I. and Pachter, L. (2003) AVID: a global alignment program. *Genome Res.*, **13**, 97–102.
  23. Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S. and Dubchak, I. (2000) VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics*, **16**, 1046–1047.
  24. Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucl. Acids Res.*, **27**, 573–580.
  25. Partridge, J.F., Scott, K.S., Bannister, A.J., Kouzarides, T. and Allshire, R.C. (2002) *cis*-Acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site. *Curr. Biol.*, **12**, 1652–1660.
  26. Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T. and Peters, A.H. (2003) Suv39h-mediated histone h3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.*, **13**, 1192–1200.
  27. Ventura, M., Mudge, J.M., Palumbo, V., Burn, S., Blennow, E., Pierluigi, M., Giorda, R., Zuffardi, O., Archidiacono, Jackson, M.S. *et al.* (2003) Neocentromeres in 15q24–26 map to duplicons which flank an ancestral centromere in 15q25. *Genome Res.*, **13**, 2059–2068.
  28. Blower, M.D., Sullivan, B.A. and Karpen, G.H. (2002) Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell*, **2**, 319–330.
  29. Blower, M.D. and Karpen, G.H. (2001) The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat. Cell Biol.*, **3**, 730–739.
  30. Spence, J.M., Critcher, R., Ebersole, T.A., Valdivia, M.M., Earnshaw, W.C., Fukagawa, T. and Farr, C.J. (2002) Co-localization of centromere activity, proteins and topoisomerase II within a subdomain of the major human X alpha-satellite array. *EMBO J.*, **21**, 5269–5280.
  31. Bernard, P. and Allshire, R. (2002) Centromeres become unstuck without heterochromatin. *Trends Cell Biol.*, **12**, 419.
  32. Shelby, R.D., Monier, K. and Sullivan, K.F. (2000) Chromatin assembly at kinetochores is uncoupled from DNA replication. *J. Cell Biol.*, **151**, 1113–1118.
  33. Sullivan, B. and Karpen, G. (2001) Centromere identity in *Drosophila* is not determined *in vivo* by replication timing. *J. Cell Biol.*, **154**, 683–690.
  34. Craig, J.M., Wong, L.H., Lo, A.W., Earle, E. and Choo, K.H. (2003) Centromeric chromatin pliability and memory at a human neocentromere. *EMBO J.*, **22**, 2495–2504.
  35. Williams, B.C., Murphy, T.D., Goldberg, M.L. and Karpen, G.H. (1998) Neocentromere activity of structurally acentric mini-chromosomes in *Drosophila*. *Nat. Genet.*, **18**, 30–37.
  36. Maggert, K.A. and Karpen, G.H. (2001) The activation of a neocentromere in *Drosophila* requires proximity to an endogenous centromere. *Genetics*, **158**, 1615–1628.
  37. Dib, C., Faure, S., Fizames, C., Samsom, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E. *et al.* (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, **380**, 152–154.
  38. Strutt, H., Cavalli, G. and Paro, R. (1997) Co-localization of polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.*, **16**, 3621–3632.