Published in final edited form as:

Chromosome Res. 2016 December ; 24(4): 451–466. doi:10.1007/s10577-016-9536-6.

Human centromere repositioning within euchromatin after partial chromosome deletion

Lori L. Sullivan¹, Kristin A. Maloney^{1,*}, Aaron J. Towers^{2,#}, Simon G. Gregory^{3,4}, and Beth A. Sullivan^{1,4,5}

¹Department of Molecular Genetics and Microbiology, Duke University School of Medicine, DUMC 3054 Durham, NC 27710 USA

²University Program in Genetics and Genomics, Duke University School of Medicine, Durham, NC 27710 USA

³Duke Molecular Physiology Institute, Department of Medicine, 300 N. Duke Street, Durham, NC, 27701 USA

⁴Division of Human Genetics, Duke University School of Medicine, Durham, NC 27710

Abstract

Centromeres are defined by a specialized chromatin organization that includes nucleosomes that contain the centromeric histone variant CENP-A instead of canonical histone H3. Studies in various organisms have shown that centromeric chromatin (i.e. CENP-A chromatin or centrochromatin) exhibits plasticity, in that it can assemble on different types of DNA sequences. However, once established on a chromosome, the centromere is maintained at the same position. In humans, this location is the highly homogeneous repetitive DNA alpha satellite. Mislocalization of centromeric chromatin to atypical locations can lead to genome instability, indicating that restriction of centromeres to a distinct genomic position is important for cell and organism viability. Here, we describe a rearrangement of *Homo sapiens* chromosome 17 (HSA17) that has placed alpha satellite DNA next to euchromatin. We show that on this mutant chromosome, CENP-A chromatin has spread from the alpha satellite into the short arm of HSA17, establishing a ~700kb hybrid centromeric domain that spans both repetitive and unique sequences and changes the expression of at least one gene over which it spreads. Our results illustrate the plasticity of human centromeric chromatin and suggest that heterochromatin normally constrains CENP-A chromatin onto alpha satellite DNA. This work highlights that chromosome rearrangements, particularly those that remove the pericentromere, create opportunities for centromeric nucleosomes to move into non-traditional genomic locations, potentially changing the surrounding chromatin environment and altering gene expression.

Keywords

keywords: CENP-A; histone; alpha satellite; transposable element; transcription; neocentromere

⁵corresponding author (beth.sullivan@duke.edu).

*current affiliation: Division of Endocrinology, Diabetes, and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21201 USA

#current affiliation: Quintiles, 4820 Emperor Blvd., Durham, NC 27703

Introduction

Centromeres are specialized chromosomal loci that are essential for chromosome segregation during cell division. Human centromeres are genetically and functionally defined by alpha satellite DNA, a ~171bp repeat subunit, that is tandemly organized into long homogeneous arrays (Waye and Willard, 1986, Waye et al., 1987, Willard and Waye, 1987, Schueler et al., 2001, Rudd and Willard, 2004). On endogenous chromosomes, alpha satellite is organized into higher-order repeat units (HOR) that contain a defined number of monomers; the number of monomers in a HOR confer chromosome-specificity. Hundreds of copies of a HOR give rise to a homogenous array at each primary constriction. Higher order alpha satellite DNA is the site of centromere protein binding, including CENP-A, the centromere-specific histone H3 variant (Vafa and Sullivan, 1997, Mravinac et al., 2009). CENP-A contributes to a unique type of chromatin (centromeric chromatin or centrochromatin) that is the structural foundation for the three-dimensional kinetochore, linking centromeric DNA and additional protein components of the inner and outer kinetochore (Perpelescu and Fukagawa, 2011).

In addition to CENP-A nucleosomes, centromeric chromatin contains subdomains of nucleosomes that contain histone H3. Interspersion of CENP-A and H3 is important for proper centromere function, including recruitment of CCAN network proteins CENP-S, -T, -W, -X and maturation of the kinetochore (Hori et al., 2008, Prendergast et al., 2011, McKinley et al., 2015). The H3 within centromeric chromatin carries post-translational modifications, including methylation of lysine 4 (K4), lysine 36 (K36), and to a lesser extent lysine 27 (K27), that are needed for centromeric transcription, centromere protein recruitment, and establishment of distinct chromatin domains within the broader centromere region (Bergmann et al., 2011, Ohzeki et al., 2012, Ohzeki et al., 2016). Although chromosome-specific alpha satellite arrays extend for several megabases within a given human centromere region, centromeric chromatin and centromere proteins are positioned on only 30–45% of the large repetitive array (Zeng et al., 2004, Lam et al., 2006a, Mravinac et al., 2009, Sullivan et al., 2011). Moreover, the CENP-A chromatin domain appears to be relatively fixed at a similar position on a given chromosome-specific array in different individuals (Ross et al., 2016). Restriction of centromeric chromatin suggests that only a limited section of alpha satellite DNA is required for proper centromere function in humans. Such spatial confinement may also protect the genome from the accumulation of CENP-A at inappropriate genomic regions.

The molecular basis for CENP-A chromatin restriction is not well understood. It is clear that centromeric chromatin must be amassed at one locus per chromosome; when CENP-A is promiscuously loaded throughout the genome or tethered to multiple sites on the same chromosome, neocentromeres and dicentric chromosomes arise and lead to genome instability, developmental defects, and cancer (Heun et al., 2006, Marshall et al., 2008, Mendiburo et al., 2011, Chen et al., 2014, Garsed et al., 2014, Athwal et al., 2015). Overexpression of CENP-A or its chaperone HJURP, experimentally or as an inherent feature of cancer cells, leads to centromere expansion (Sullivan et al., 2011, Perpelescu et al., 2015), indicating that appropriately regulated levels of histone variants and their

chaperones coordinately constrain the CENP-A chromatin domain. Moreover, studies in model systems have shown that nearby chromatin, specifically heterochromatin, establishes and sustains CENP-A chromatin at the same chromosomal location (Ekwall et al., 1997, Williams et al., 1998, Partridge et al., 2000, Folco et al., 2008). Pericentromeric heterochromatin in mammalian cells may also play a similar “gatekeeper” role for centromeric chromatin in order to ensure chromosome stability (Taddei et al., 2001, Boyarchuk et al., 2014), although few studies have directly tested this model.

In this work, we describe a naturally-occurring human chromosome 17 mutant that asymmetrically lacks pericentromeric heterochromatin and a large proportion of alpha satellite DNA. Loss of pericentromeric heterochromatin results in repositioning of centromeric chromatin within euchromatin and changes in gene expression. The consequences of centromere spreading and CENP-A nucleosomes expansion is not unlike that described in model systems.

Materials and Methods

Cell culture

The fibroblast line GM8148 that contains a normal HSA17, del(17), and mar(17) (Coriell Institute for Medical Research) was grown in Minimal Essential Medium (MEM) alpha, supplemented with 10% fetal bovine serum (FBS; Cellgro) and 1X antibiotic-antimycotic solution (Gibco). Somatic cell hybrids containing the normal HSA17 (L65-14A) or del(17) (L65-13A) derived from a fibroblast cell line GM8148 as previously described (Wevrick et al., 1990) were grown in MEM alpha supplemented with 10% FBS, 1X antibiotic-antimycotic, and 1X hypoxanthine-aminopterin-thymidine (HAT; Gibco). All cells were cultured at 37°C in 5% CO₂.

FISH on metaphase chromosomes

Metaphase chromosomes were isolated from GM8148 using standard protocols with hypotonic treatment in either 75mM KCl (GM8148) or a 1:1:1 v/v/v mixture of 75mM KCl/0.8% sodium citrate/dH₂O. Cells were fixed in 3:1 methanol:acetic acid multiple times prior to immobilization on glass slides. Chromosome preparations that were less than a week old were used for FISH with a probe specific for D17Z1 (p17H8, a generous gift from H.F. Willard) that had been labeled with biotin-12-dUTP or directly labeled with Alexafluor 488- or Alexafluor 594-dUTP. Hybridization and washing was done at high stringency (68% formamide). Indirectly labeled biotin probes were detected with Cy3-avidin. Chromosomes were counterstained with 1ng/μL DAPI diluted in Vectashield (Vector Laboratories).

PCR walking using STS markers

Genomic DNA (50ng) from L65-13A [del(17)] and L65-14A (normal HSA17 control) was amplified in a 15μL reaction using primers to STS markers spanning the proximal short arm of chromosome 17 from 17p11–17p13 (Table 1). The entire PCR reaction was run on a 1.5% agarose gel that was stained with ethidium bromide and imaged on an AlphaImager documentation system.

Chromosome 17 tile path array comparative genomic hybridization (CGH) and analysis

High-resolution genome tilepath microarrays were constructed using large bacterial clones (bacterial artificial chromosome (BAC) clones) as previously described (Fiegler et al., 2003). The chromosome 17 tile path array included 923 individual BACs spanning HSA17. A chromosome 4 tile path BAC array containing 2303 BACs was used as a control CGH for normal copy number. Briefly, BAC DNA was purified by alkaline lysis and used as the template for three different, degenerate oligonucleotide primed PCR reactions, prior to a second round of amino linking PCR and printing onto Codelink slides (GE Bioscience, Piscataway, NJ, USA) using a Genetix Qarray2 (Genetix, Boston, MA, USA). A total of 450 ng of WT or MUT HSA17 DNA and control DNA, from human peripheral blood, were differentially labeled with Cy3-CTP and Cy5-CTP (BioPrime Labeling Kit; Invitrogen, Carlsbad, CA, USA), purified (Qiagen, Hilden, Germany) and hybridized to the tilepath arrays using the MAUI hybridization station (BioMicro Systems Inc, Salt Lake City, UT, USA). Image capture of the hybridized arrays for fluorescent intensity extraction was performed using a Genepix 4100A scanner (Molecular Devices, Sunnyvale, CA, USA). Bluefuse microarray software (BlueGnome http://www.cambridgebluegenome.com/bluefuse_micro) was used for data processing of the scanned images prior to porting into Nexus (BioDiscovery <http://www.biodiscovery.com/index/nexus>) for analysis.

Microarray data was preprocessed in Nexus by removing poor quality flagged spots (confidence <0.3 as defined by Bluefuse software), removing background by Lowess correction before normalization of log2 ratios and combining BAC replicates. BioDiscovery's rank segmentation algorithm (RSA), which is similar to circular binary segmentation (CBS) (Olshen et al., 2004), was used to identify genomic rearrangements. Briefly, the algorithm uses a normal distribution function for testing for change points as opposed to the non-parametric permutation based statistics used in the original CBS algorithm. It also uses the log ratio rank of each probe as oppose the log ratio value itself. The calling algorithm used cluster values and defined log2 thresholds of - 0.5 for one copy loss. We applied a conservative cut-off of three BAC clones showing the same copy number change trend to define genomic deletion or duplication with the significance threshold set at 0.005.

DNA sequencing

Genomic DNA extracted from cell lines L65-13A [mutant, MUT; del(17)] and L65-14A (wildtype, WT; normal HSA17 control) was sequenced on the Illumina HiSeq2500 using Illumina's library preparation kit. Single end 50bp reads were used for sequencing. Reads were first processed to remove low quality bases and adapter sequences from the 3' end using the TrimGalore! Toolkit (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Reads that were at least 20nt in length after trimming were then aligned to a custom genome that contained the full mouse genome (GRCm38 from ENSEMBL (Kersey et al., 2012)) as well as human chr17 (GRCh38 from ENSEMBL) by the Bowtie alignment algorithm (Langmead et al., 2009) (-m 500 --v3 --all --best --strata). The breakpoint region in the MUT samples was identified by averaging the ratio of input DNA in the WT samples to the MUT samples for the two factors in a sliding window across all of the replicates. A

5kb window was identified between two windows that had a greater than 10-fold change in the ratio.

Immunofluorescence-FISH on chromatin fibers

Chromatin fibers were prepared from somatic cell hybrids L65-14A and L65-13A as previously described (Sullivan, 2010, Sullivan et al., 2011, Maloney et al., 2012). Custom rabbit polyclonal antibodies to mouse Cenp-A (Maloney et al., 2012) were used at 1:50 dilution and incubated overnight at 4°C followed by washing in KCM buffer without agitation. Cenp-A antibodies were detected with Alexa Fluor 594-conjugated anti-rabbit secondary antibodies (Invitrogen) at room temperature for two hours, followed by KCM washes. Antibodies were cross-linked to chromatin fibers with 10% formalin (4% formaldehyde) for 10–15 minutes at room temperature. D17Z1 plasmid p17H8 was directly labeled with ChromaTide 488-dUTP using nick translation and 200ng was used for hybridization to chromatin fibers. Chromatin fibers were denatured in 70% formamide/2X SSC, pH 7.0 for 5 minutes and denatured probe resuspended in 68% formamide hybridization buffer was quickly applied. Hybridization was performed at 37°C overnight, followed by washes in 68% formamide/2X SSC/0.05% Tween-20 at 37°C.

Microscopy and image analysis

All images were acquired using an inverted Olympus IX-71 microscope connected to the Deltavision Elite imaging system (Applied Precision/GE Healthcare) equipped with a Photometrics CoolSNAP HQ² CCD camera and running the SoftWoRx imaging software. Fiber images were taken with a 100X objective (N.A. 1.40). Those extending through multiple fields of view were captured using the Panels option in the SoftWoRx Acquire 3D program and merged into single images using the “Stitch” function. The “Measure Distances” tool was used to measure lengths of fluorescent signals in micrometers. Micrometer distances were converted to kilobases using the D17Z1 fluorescent probe signal from each line as a normalizer for size. The molecular sizes of D17Z1 in L65-14A (3.7Mb) and L65-13A (0.7 Mb) had been previously calculated (Wevrick et al., 1990) and re-confirmed in our lab (Aldrup-MacDonald et al., *in press*). Cenp-A domain size was calculated from the ratio of the length of Cenp-A antibody signal over the length of the D17Z1 array signal (Sullivan et al., 2011, Ross et al., 2016)

ChIP-PCR with Cenp-A antibodies

Native chromatin containing oligonucleosomes (mono, di, and tri) was prepared by micrococcal nuclease digestion as previously described (Mravinac et al., 2009, Maloney et al., 2012) using a modification of the MAGnify ChIP system (Invitrogen) for native chromatin. One microgram of chromatin was incubated with 5µg of polyclonal mouse Cenp-A antibodies (AP601); ChIPs for each antibody were performed in duplicate per sample (WT or MUT). Two microliters of immunoprecipitation product (IP) was used for qPCR that was performed in triplicate for each IP sample. Primers for each genomic location and qPCR conditions are listed in Table 1. Enrichment for Cenp-A was calculated as percentage of input and presented as enrichment above D17Z1, the site on HSA17 where Cenp-A is normally highly concentrated.

RT-qPCR

RNA was isolated in duplicate for each cell line ($\sim 5 \times 10^6$ cells per reaction) using the RNeasy kit (Qiagen). RNA quality was confirmed by 1.1% general purpose agarose gel containing 1% household bleach. RNA was reverse transcribed using QuantiTect Reverse Transcription Kit (Qiagen). Both RT+ and RT- samples were generated for each duplicate. For each resulting cDNA sample, RT-qPCR was performed in triplicate (2uL template/reaction) using QuantiFast SYBR Green PCR kit (Qiagen). RT-qPCR runs included a target gene (ELAC2; F: 5' TCTTCAACTGTGGAGAAGGC 3'; R: 5' CTGTGTATGGGGATCTGGTA 3') and a reference gene (mouse beta actin; F: 5' GTGGGCCGCTCTAGGCACCA 3'; R: 5' CGGTTGGCCTTAGGGTTCAGGGGGG 3'). Relative expression was calculated by the C_t method using the normal 17 as the control. A Students t-test was used to determine significance between normal and mutant HSA17 gene expression, with p values less than 0.01 considered significant.

Results

Clarification of the del(17) structure and refinement of the breakpoint to a 5kb region

The *Homo sapiens* chromosome 17 (HSA17) has a complex centromere region, containing three, distinct, adjacent higher order alpha satellite arrays (D17Z1-B, D17Z1, D17Z1-C, moving from short arm toward the long arm) flanked by unordered monomeric alpha satellite DNA (Figure 1A) (Rudd and Willard, 2004, Maloney et al., 2012). Typically, D17Z1 is the location of centromere assembly, although in some individuals in the population, the centromere forms at D17Z1-B (Maloney et al., 2012). Centromere-functional alpha satellite arrays like D17Z1 are enriched for centromere proteins, euchromatic histone modifications such as H3 methylated at lysines 4 and 36 (H3K4me2, H3K36me2/3), and heterochromatic histone modifications H3 tri-methylated at lysines 9 and 27 (H3K9me3 and H3K27me3) (Sullivan and Karpen, 2004, Lam et al., 2006b, Bergmann et al., 2011). Inactive arrays such as D17Z1-B and D17Z1-C in the pericentromeric region lack CENP-A and other centromere proteins and instead are highly enriched for heterochromatic histone modifications (Mravinac et al., 2009, Maloney et al., 2012) (Figure 1A).

In this study, the HSA17 of interest has undergone a large structural change in which most of the centromere region and nearly half of the short arm (p) has been deleted. This resulted in a deleted chromosome [del(17)], and a small supernumerary marker chromosome [mar(17)] that retains the deleted portions of HSA17 (Howard-Peebles et al., 1985) (Figure 1B). The del(17) and the normal HSA17 from the patient were isolated into mouse-human somatic cell hybrids to facilitate comparative molecular analyses (Wevrick et al., 1990, Warburton et al., 1992). The del(17) breakpoint region was previously mapped by classical cytogenetic analysis to distal 17p11 (Howard-Peebles et al., 1985). However, because of the low resolution of metaphase analysis, we sought to refine the breakpoint region of the del(17) using a combination of approaches, including BAC CGH array, PCR, and DNA sequencing. By CGH array analysis, the breakpoints were narrowed to a 292kb region between two BACs, RP11-409O23 (hg38: 13086033-13254863) and RP11-196N4 (hg38: 13205501-13378336) (Figure 2A). The region was narrowed further by PCR with primers specific for several annotated STS markers (PCR walking) within 17p12. With this

approach, we mapped the breakpoint to a 28kb region (hg38: 13203706-13232342) between markers D17S2117 and D17S1808 (Figure 2B).

Illumina sequencing further defined the breakpoint to a 5kb region (hg38: 13206181-13211180) (Figure 2C). Structural variants such as segmental duplications or low copy repeats are known to drive recurrent and de novo chromosome rearrangements, and are especially enriched within the proximal short arm of HSA17 (Stankiewicz et al., 2001, Shchelochkov et al., 2010). When examining the area of breakage for structural variation within the genome assembly, no annotated segmental duplications or low copy repeats mapped to the breakpoint region. However, three retroelements, specifically L1s/LINEs (long interspersed nuclear elements), were present within the breakpoint region (Figure 2C), with one appearing as an insertion. The deletion, at least on the short arm side of the break, may have been driven by these mobile elements that are present within the breakpoint region and that may create an area of instability. Taking into account all of these experiments, we determined that the del(17) is missing 12.9Mb of HSA17 (15% of the chromosome), including 2.5Mb of D17Z1 alpha satellite, ~500kb of pericentromeric DNA, and 9.9Mb of the proximal short arm (Figure 2D).

Spreading of CENP-A chromatin into the short arm of HSA17

Previous analysis revealed that the alpha satellite binding protein CENP-B is present in reduced amounts on the del(17) (Wevrick et al., 1990). This observation is most likely explained by the decrease in D17Z1 alpha satellite array size and, by extension, the number of CENP-B boxes, i.e. the 17bp motif within alpha satellite DNA that is specifically recognized by CENP-B. Even though the del(17) lacks ~75% of D17Z1 and is associated with less CENP-B, it is mitotically and meiotically stable (Wevrick et al., 1990, Aldrup-MacDonald et al., 2016). Reduced or even absent CENP-B binding does not preclude recruitment of other centromere and kinetochore proteins. In fact, the stability of the del(17) and its inheritance through meiosis (Wevrick et al., 1990) suggest that this chromosome has a functional centromere and properly recruits other centromere proteins even though the D17Z1 alpha satellite array is extensively reduced in size.

We measured the amount of centromere proteins by optically mapping CENP-A chromatin on extended chromatin fibers from the del(17). We previously showed that CENP-A chromatin is normally assembled on 30–45% of alpha satellite DNA of any given human chromosome (Lam et al., 2006a, Mravinac et al., 2009, Sullivan et al., 2011). This ratio, and the chromatin organization within the centromere, is consistently preserved even when human chromosomes are transferred to rodent backgrounds (Mravinac et al., 2009, Sullivan et al., 2011, Maloney et al., 2012). Therefore, it appears that mouse Cenp-A assumes the location previously occupied by human CENP-A, and accurately reflects centromeric chromatin organization in the human background.

The normal HSA17 homolog from the same patient has a 3.7Mb alpha satellite D17Z1 array (Wevrick et al., 1990), and Cenp-A chromatin is assembled on ~40% of the array (1.45Mb $\pm$ 0.45Mb) (Figure 3A). The del(17) with its 700kb D17Z1 array exhibited a smaller Cenp-A chromatin domain. Surprisingly, Cenp-A only partially co-localized with D17Z1 (Figure 3A). The remainder of Cenp-A did not overlap with D17Z1 FISH signal. In fact, ~55% of

Cenp-A staining was located on D17Z1 (Figure 3B), and the remainder was associated with non-alpha satellite DNA. Using the length of the fluorescent signal for the D17Z1 array as a genomic size marker (i.e. 700kb), we measured the entire region of Cenp-A staining and estimated the genomic size of the Cenp-A chromatin domain on the del(17) to be ~677kb (± 170 kb) (Figure 3B'). Based on the percentages of Cenp-A staining on D17Z1 versus that extending beyond D17Z1, we estimated that only 360kb (± 104 kb) of the Cenp-A chromatin domain is located on D17Z1 (Figure 3C).

Cenp-A chromatin spreads several hundred kilobases into the HSA17 short arm euchromatin

The chromatin fiber calculations indicated that Cenp-A chromatin had spread ~300kb into non-centromeric DNA. The short arm of HSA17, rather than the long arm, seemed the most likely direction for CENP-A spreading, since the breakage events resulting in the del(17) placed D17Z1 immediately adjacent to the HSA17 short arm (i.e. 17p12). To verify the chromatin fiber results, we performed chromatin immunoprecipitation (ChIP) with antibodies to Cenp-A, followed by qPCR for specific genomic sites within 17p12, moving distal from the breakpoint. As expected, Cenp-A was concentrated on D17Z1 on both the normal HSA17 and del(17), although D17Z1 on the del(17) was less enriched for Cenp-A (Figure 4A). Five genomic sites within the HSA17 short arm all showed Cenp-A enrichment that matched or exceed that on D17Z1, the normal site of Cenp-A accumulation. A sixth location that was absent from the del(17) but present on the normal HSA17 was included as a control. The five interrogated locations spanned a 200kb region, confirming that Cenp-A chromatin had spread at least several hundred kilobases away from the 17p breakpoint. Overall, our findings demonstrate that a large-scale rearrangement of a human chromosome that places alpha satellite DNA next to non-centromeric DNA allows centromeric chromatin to spread into unique euchromatic regions, and centromeric chromatin is stably retained at its new genomic location.

Changes in gene expression within euchromatin covered by Cenp-A

The ChIP-PCR experiments revealed that Cenp-A was not equally enriched at each of the five interrogated genomic locations, possibly reflecting differences in the underlying DNA sequences. Analysis of the human genome assembly for this region of HSA17 indicated that Cenp-A chromatin on the del(17) had spread across a region that is not only rich in interspersed repeats/transposable elements but also includes at least one gene and several annotated RNA transcripts (Figure 4B). We tested if spreading of centromeric chromatin affected gene expression by measuring transcription of the ELAC2 gene (hg38 chr17:12,992,391-13,018,187), located within the repositioned centromeric chromatin domain (Figure 4A, genomic site 1). RT-qPCR showed that ELAC2 expression on the del(17) was reduced by more than 50% compared to the normal 17 (Figure 4C). These results demonstrate that although centromeric chromatin can readily spread into euchromatin/transcriptionally active chromatin, it has a notable repressive effect on gene transcription.

Discussion

Centromeric/CENP-A chromatin in humans is typically assembled on repetitive DNA sequences. However, it has been shown that CENP-A chromatin can form on unique sequences that are located outside of, and often far away from, the endogenous centromere. These neocentromeres were initially described in humans, but subsequently have been reported in a variety of organisms (Depinet et al., 1997, du Sart et al., 1997, Maggert and Karpen, 2001, Warburton, 2004, Heun et al., 2006, Ketel et al., 2009, Burrack and Berman, 2012, Shang et al., 2013, Thakur and Sanyal, 2013). Over 100 human neocentromeres have been described; for about a dozen, the centromere protein (CENP-A, CENP-C, CENP-H) binding domains have been defined in detail (Alonso et al., 2003, Chueh et al., 2005, Alonso et al., 2007, Hasson et al., 2013). Most of these CENP-A domains have been mapped to 300–500kb regions, suggesting that neocentromere-associated CENP-A chromatin domains are considerably smaller than endogenous CENP-A domains. Likewise, the CENP-A chromatin domain on the del(17) occupies half the amount of DNA compared to the CENP-A domain on a normal HSA17. These results suggest that CENP-A chromatin may not be able to spread as far within euchromatin. Alternatively, a smaller CENP-A chromatin domain simply correlates with the smaller D17Z1 array size, as has been observed for experimentally deleted human X chromosomes (Mravinac et al., 2009).

The single molecule chromatin fiber analyses did not reveal extensive variability in CENP-A spreading. In other words, we did not observe fibers on which CENP-A nucleosomes only associated D17Z1, nor did we observe those in which CENP-A chromatin assembled on only short arm material. These findings suggest a combined degree of rigidity and flexibility of centromeric chromatin expansion. CENP-A nucleosomes appear to spread only far enough to create a centromeric chromatin domain that will confer chromosome stability. It appears that once the centromeric chromatin domain was established on the del(17), it has remained largely static, suggesting that centromere formation on non-centromeric sequences follows similar rules of inheritance at endogenous centromeres (Ross et al., 2016).

Many naturally-occurring neocentromeres, particularly in humans, are formed at locations that are far, at least in linear distance, from the endogenous centromere. Conversely, neocentromere induction experiments in model organisms have shown that CENP-A is concentrated in genomic locations near or next to endogenous centromeres. In these instances, CENP-A chromatin does not spread per se, but the low to moderate levels of CENP-A nucleosomes within regions flanking centromeres nucleate a neocentromere when the endogenous centromere is deleted (Burrack and Berman, 2012, Shang et al., 2013, Thakur and Sanyal, 2013, Scott and Sullivan, 2014). The closest example of a neocentromere that is similar to the del(17) is the *Drosophila* minichromosome *Dp(1;f)1187*. This derivative chromosome was formed after successive rounds of gamma irradiation caused an X chromosome inversion followed by removal of all but the proximal and distal sequences within the inversion region (Karpen and Spradling, 1990). The 1.3Mb *Dp1187* minichromosome was used to molecularly define the first centromeric sequences in *Drosophila* (Murphy and Karpen, 1995, Sun et al., 1997, Sun et al., 2003). Multiple rounds of truncation of *Dp1187* produced a series of progressively deleted minichromosomes lacking the molecularly defined centromere. These derivatives were found to be stable

because the centromere had moved from satellite DNA into flanking euchromatin (Williams et al., 1998, Blower and Karpen, 2001, Maggert and Karpen, 2001), spreading over different genomic distances, depending on the minichromosome derivative (DE Pazin and BA Sullivan, unpublished). The del(17) mimics the structure and organization of these neocentromeric minichromosomes, in that the HSA17 rearrangement placed D17Z1 next to the short arm and centromeric chromatin then spread into euchromatin.

Together, our del(17) and the *Dp1187*-derived minichromosomes suggest that centromeric chromatin is normally constrained entirely or in part by the flanking pericentromere. Indeed, endogenous CENP-A chromatin domains are surrounded by constitutive heterochromatin that is enriched for H3K9me3 and H3K27me3 (Partridge et al., 2000, Choo, 2001, Mravinac et al., 2009). CENP-A overexpression studies have shown that CENP-A nucleosomes can readily spread into and nucleate at regions of open chromatin or at heterochromatin-euchromatin boundaries (Olszak et al., 2011, Athwal et al., 2015). In cancer genomes that inherently over-express CENP-A, excess CENP-A accumulates at promoters, transcription factor binding sites, and DNase hypersensitive (DNase HSS) sites, regions that are indicative of open or accessible chromatin (Crawford et al., 2004, Thurman et al., 2012, Lacoste et al., 2014, Athwal et al., 2015). We also observed many mapped DNase HSS sites within the 300kb euchromatic region on the del(17) occupied by CENP-A chromatin. Thus, our findings are in agreement with other reports claiming that CENP-A chromatin can readily invade regions of open and transcribed chromatin.

L1/LINE retroelements are structural and functional components of CENP-A domains at human neocentromeres (Chueh et al., 2005, Chueh et al., 2009). Retroelement transcription is also a feature of normal mammalian centromeres (Carone et al., 2009, Carone et al., 2013). We speculate that the transcriptional activity of the L1s/LINEs at the junction of the alpha satellite- short arm fusion point on the del(17) facilitated spreading of CENP-A nucleosomes into non-centromeric DNA. It is not clear if the euchromatic portion of the CENP-A domain on the del(17) is permissive for CENP-A spreading due solely to retroelement transcription or was additionally enabled by gene transcription within the region. Non-coding RNAs derived from satellites are components of centromere complexes and recruit or maintain kinetochore proteins (Topp et al., 2004, Wong et al., 2007, Du et al., 2010, Bergmann et al., 2011, Rosic et al., 2014). The role of transcription in centromere identity remains complex, in that transcripts are present, but too much or too little transcription is incompatible with centromere function (Hill and Bloom, 1987, Bloom et al., 1989, Cardinale et al., 2009, Bergmann et al., 2012, Shang et al., 2013). The del(17) also exhibited incompatibility of centromeric chromatin and gene expression, suggesting a difference between transcripts that maintain CENP-A chromatin and those produced by gene expression. The euchromatic nature of the region flanking the del(17) breakpoint may have permitted expansion of CENP-A chromatin, but stabilization of centromeric chromatin required a reduction in the transcription of ELAC2 and possibly of other genes located nearby.

Our studies highlight several directions for additional study, including RNA-sequencing to determine the broader effects on transcription throughout the ~300kb euchromatin region, and ChIP-seq to map chromatin organization within and around the new centromere

location. It would be interesting to determine if CENP-A nucleosomes are equally distributed over intergenic and coding regions within the short arm of HSA17, or instead are organized into large domains in which the H3 nucleosomes occupy the genic regions (Alonso et al., 2003, Alonso et al., 2007, Hasson et al., 2013). Furthermore, the use of genome engineering technologies to precisely delete pericentromere regions could identify key temporal aspects of CENP-A chromatin spreading immediately after flanking heterochromatin is removed. Such future studies could directly test the molecular mechanisms that control CENP-A invasion and repositioning after loss of heterochromatin, thereby increasing our understanding of normal centromere biology and disease states associated with inappropriate CENP-A incorporation.

Acknowledgments

We thank Huntington Willard for the generous gift of the GM08146 cell line and L65-13A [del(17)] and L65-14A (normal HSA17) somatic cell hybrids. We are grateful to Dr. David Corcoran in the Duke Genomics and Bioinformatics Analysis Core Facility for assistance with genomic informatics. This research was partially supported by grants from the March of Dimes Foundation (06-FY10-294 and 1-FY13-517 to B.A.S.) and the National Institutes of Health (GM069514 and GM098500 to B.A.S.).

References

- Aldrup-MacDonald ME, Kuo ME, Sullivan LL, Chew K, Sullivan BA. Genomic variation within alpha satellite DNA influences centromere location on human chromosomes with metastable epialleles. *Genome Res.* 2016; doi: 10.1101/gr.206706.116
- Alonso A, Fritz B, Hasson D, Abrusan G, Cheung F, Yoda K, Radlwimmer B, Ladurner A, Warburton P. Co-localization of CENP-C and CENP-H to discontinuous domains of CENP-A chromatin at human neocentromeres. *Genome Biol.* 2007; 8:R148. [PubMed: 17651496]
- Alonso A, Mahmood R, Li S, Cheung F, Yoda K, Warburton PE. Genomic microarray analysis reveals distinct locations for the CENP-A binding domains in three human chromosome 13q32 neocentromeres. *Human molecular genetics.* 2003; 12:2711–2721. [PubMed: 12928482]
- Athwal RK, Walkiewicz MP, Baek S, Fu S, Bui M, Camps J, Ried T, Sung MH, Dalal Y. CENP-A nucleosomes localize to transcription factor hotspots and subtelomeric sites in human cancer cells. *Epigenetics Chromatin.* 2015; 8:2. [PubMed: 25788983]
- Bergmann JH, Jakubsche JN, Martins NM, Kagansky A, Nakano M, Kimura H, Kelly DA, Turner BM, Masumoto H, Larionov V, Earnshaw WC. Epigenetic engineering: histone H3K9 acetylation is compatible with kinetochore structure and function. *J Cell Sci.* 2012; 125:411–421. [PubMed: 22331359]
- Bergmann JH, Rodriguez MG, Martins NM, Kimura H, Kelly DA, Masumoto H, Larionov V, Jansen LE, Earnshaw WC. Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *Embo J.* 2011; 30:328–340. [PubMed: 21157429]
- Bloom K, Hill A, Jones E. Conditional dicentric chromosomes in yeast. *Prog Clin Biol Res.* 1989; 318:149–158. [PubMed: 2696964]
- Blower MD, Karpen GH. The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nature cell biology.* 2001; 3:730–739. [PubMed: 11483958]
- Boyarchuk E, Filipescu D, Vassias I, Cantaloube S, Almouzni G. The histone variant composition of centromeres is controlled by the pericentric heterochromatin state during the cell cycle. *J Cell Sci.* 2014; 127:3347–3359. [PubMed: 24906798]
- Burrack LS, Berman J. Neocentromeres and epigenetically inherited features of centromeres. *Chromosome Res.* 2012; 20:607–619. [PubMed: 22723125]
- Cardinale S, Bergmann JH, Kelly D, Nakano M, Valdivia MM, Kimura H, Masumoto H, Larionov V, Earnshaw WC. Hierarchical inactivation of a synthetic human kinetochore by a chromatin modifier. *Mol Biol Cell.* 2009; 20:4194–4204. [PubMed: 19656847]

- Carone DM, Longo MS, Ferreri GC, Hall L, Harris M, Shook N, Bulazel KV, Carone BR, Obergfell C, O'Neill MJ, O'Neill RJ. A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. *Chromosoma*. 2009; 118:113–125. [PubMed: 18839199]
- Carone DM, Zhang C, Hall LE, Obergfell C, Carone BR, O'Neill MJ, O'Neill RJ. Hypermorphic expression of centromeric retroelement-encoded small RNAs impairs CENP-A loading. *Chromosome Res*. 2013; 21:49–62. [PubMed: 23392618]
- Chen CC, Dechassa ML, Bettini E, Ledoux MB, Belisario C, Heun P, Luger K, Mellone BG. CAL1 is the *Drosophila* CENP-A assembly factor. *J Cell Biol*. 2014; 204:313–329. [PubMed: 24469636]
- Choo K. Domain organization at the centromere and neocentromere. *Developmental cell*. 2001; 1:165–177. [PubMed: 11702777]
- Chueh AC, Northrop EL, Brettingham-Moore KH, Choo KHA, Wong LH. LINE Retrotransposon RNA Is an Essential Structural and Functional Epigenetic Component of a Core Neocentromeric Chromatin. *PLoS Genet*. 2009; 5:e1000354. [PubMed: 19180186]
- Chueh AC, Wong LH, Wong N, Choo KH. Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. *Human molecular genetics*. 2005; 14:85–93. [PubMed: 15537667]
- Crawford GE, Holt IE, Mullikin JC, Tai D, Blakesley R, Bouffard G, Young A, Masiello C, Green ED, Wolfsberg TG, Collins FS. National Institutes Of Health Intramural Sequencing C. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc Natl Acad Sci U S A*. 2004; 101:992–997. [PubMed: 14732688]
- Depinet TW, Zackowski JL, Earnshaw WC, Kaffe S, Sekhon GS, Stallard R, Sullivan BA, Vance GH, Van Dyke DL, Willard HF, Zinn AB, Schwartz S. Characterization of neo-centromeres in marker chromosomes lacking detectable alpha-satellite DNA. *Human molecular genetics*. 1997; 6:1195–1204. [PubMed: 9259264]
- du Sart D, Cancelli MR, Earle E, Mao JI, Saffery R, Tainton KM, Kalitsis P, Martyn J, Barry AE, Choo KH. A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. *Nat Genet*. 1997; 16:144–153. [PubMed: 9171825]
- Du Y, Topp CN, Dawe RK. DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet*. 2010; 6:e1000835. [PubMed: 20140237]
- Ekwall K, Olsson T, Turner BM, Cranston G, Allshire RC. Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. *Cell*. 1997; 91:1021–1032. [PubMed: 9428524]
- Fiegler H, Carr P, Douglas EJ, Burford DC, Hunt S, Scott CE, Smith J, Vetrie D, Gorman P, Tomlinson IP, Carter NP. DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer*. 2003; 36:361–374. [PubMed: 12619160]
- Folco HD, Pidoux AL, Urano T, Allshire RC. Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. *Science*. 2008; 319:94–97. [PubMed: 18174443]
- Garsed DW, Marshall OJ, Corbin VD, Hsu A, Di Stefano L, Schroder J, Li J, Feng ZP, Kim BW, Kowarsky M, Lansdell B, Brookwell R, Myklebost O, Meza-Zepeda L, Holloway AJ, Pedoutour F, Choo KH, Damore MA, Deans AJ, Papenfuss AT, Thomas DM. The architecture and evolution of cancer neochromosomes. *Cancer Cell*. 2014; 26:653–667. [PubMed: 25517748]
- Hasson D, Panchenko T, Salimian KJ, Salman MU, Sekulic N, Alonso A, Warburton PE, Black BE. The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nature structural & molecular biology*. 2013; 20:687–695.
- Heun P, Erhardt S, Blower MD, Weiss S, Skora AD, Karpen GH. Mislocalization of the *Drosophila* centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Developmental cell*. 2006; 10:303–315. [PubMed: 16516834]
- Hill A, Bloom K. Genetic manipulation of centromere function. *Mol Cell Biol*. 1987; 7:2397–2405. [PubMed: 3302676]
- Hori T, Amano M, Suzuki A, Backer CB, Welburn JP, Dong Y, McEwen BF, Shang WH, Suzuki E, Okawa K, Cheeseman IM, Fukagawa T. CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell*. 2008; 135:1039–1052. [PubMed: 19070575]

- Howard-Peebles PN, Friedman JM, Harrod MJE, Brookshire GS, Lockwood JE. A stable supernumerary chromosome derived from a deleted segment of 17p. *Am J Hum Genet.* 1985; 38:A97.
- Karpen GH, Spradling AC. Reduced DNA polytenization of a minichromosome region undergoing position-effect variegation in *Drosophila*. *Cell.* 1990; 63:97–107. [PubMed: 2208283]
- Kersey PJ, Staines DM, Lawson D, Kulesha E, Derwent P, Humphrey JC, Hughes DS, Keenan S, Kerhornou A, Koscielny G, Langridge N, McDowall MD, Megy K, Maheswari U, Nuhn M, Paulini M, Pedro H, Toneva I, Wilson D, Yates A, Birney E. Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species. *Nucleic Acids Res.* 2012; 40:D91–97. [PubMed: 22067447]
- Ketel C, Wang HS, McClellan M, Bouchonville K, Selmecki A, Lahav T, Gerami-Nejad M, Berman J. Neocentromeres form efficiently at multiple possible loci in *Candida albicans*. *PLoS Genet.* 2009; 5:e1000400. [PubMed: 19266018]
- Lacoste N, Woolfe A, Tachiwana H, Garea AV, Barth T, Cantaloube S, Kurumizaka H, Imhof A, Almouzni G. Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX. *Mol Cell.* 2014; 53:631–644. [PubMed: 24530302]
- Lam A, Boivin C, Bonney C, Rudd M, Sullivan B. Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *Proc Natl Acad Sci USA.* 2006a; 103:4186–4191. [PubMed: 16537506]
- Lam AL, Boivin CD, Bonney CF, Rudd MK, Sullivan BA. Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *Proc Natl Acad Sci U S A.* 2006b; 103:4186–4191. [PubMed: 16537506]
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10:R25. [PubMed: 19261174]
- Maggert KA, Karpen GH. The activation of a neocentromere in *Drosophila* requires proximity to an endogenous centromere. *Genetics.* 2001; 158:1615–1628. [PubMed: 11514450]
- Maloney KA, Sullivan LL, Matheny JE, Strome ED, Merrett SL, Ferris A, Sullivan BA. Functional epialleles at an endogenous human centromere. *Proc Natl Acad Sci U S A.* 2012; 109:13704–13709. [PubMed: 22847449]
- Marshall OJ, Chueh AC, Wong LH, Choo KH. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet.* 2008; 82:261–282. [PubMed: 18252209]
- McKinley KL, Sekulic N, Guo LY, Tsinman T, Black BE, Cheeseman IM. The CENP-L-N Complex Forms a Critical Node in an Integrated Meshwork of Interactions at the Centromere-Kinetochore Interface. *Mol Cell.* 2015; 60:886–898. [PubMed: 26698661]
- Mendiburo MJ, Padeken J, Fulop S, Schepers A, Heun P. *Drosophila* CENH3 is sufficient for centromere formation. *Science.* 2011; 334:686–690. [PubMed: 22053052]
- Mravinac B, Sullivan LL, Reeves JW, Yan CM, Kopf KS, Farr CJ, Schueler MG, Sullivan BA. Histone modifications within the human X centromere region. *PLoS One.* 2009; 4:e6602. [PubMed: 19672304]
- Murphy TD, Karpen GH. Localization of centromere function in a *Drosophila* minichromosome. *Cell.* 1995; 82:599–609. [PubMed: 7664339]
- Ohzeki J, Bergmann JH, Kouprina N, Noskov VN, Nakano M, Kimura H, Earnshaw WC, Larionov V, Masumoto H. Breaking the HAC Barrier: histone H3K9 acetyl/methyl balance regulates CENP-A assembly. *Embo J.* 2012; 31:2391–2402. [PubMed: 22473132]
- Ohzeki J, Shono N, Otake K, Martins NM, Kugou K, Kimura H, Nagase T, Larionov V, Earnshaw WC, Masumoto H. KAT7/HBO1/MYST2 Regulates CENP-A Chromatin Assembly by Antagonizing Suv39h1-Mediated Centromere Inactivation. *Developmental cell.* 2016; 37:413–427. [PubMed: 27270040]
- Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics.* 2004; 5:557–572. [PubMed: 15475419]
- Olszak AM, van Essen D, Pereira AJ, Diehl S, Manke T, Maiato H, Saccani S, Heun P. Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nature cell biology.* 2011; 13:799–808. [PubMed: 21685892]

- Partridge JF, Borgstrom B, Allshire RC. Distinct protein interaction domains and protein spreading in a complex centromere. *Genes Dev.* 2000; 14:783–791. [PubMed: 10766735]
- Perpelescu M, Fukagawa T. The ABCs of CENPs. *Chromosoma.* 2011; 120:425–446. [PubMed: 21751032]
- Perpelescu M, Hori T, Toyoda A, Misu S, Monma N, Ikeo K, Obuse C, Fujiyama A, Fukagawa T. HJURP is involved in the expansion of centromeric chromatin. *Mol Biol Cell.* 2015; 26:2742–2754. [PubMed: 26063729]
- Prendergast L, van Vuuren C, Kaczmarczyk A, Doering V, Hellwig D, Quinn N, Hoischen C, Diekmann S, Sullivan KF. Premitotic assembly of human CENPs -T and -W switches centromeric chromatin to a mitotic state. *PLoS Biol.* 2011; 9:e1001082. [PubMed: 21695110]
- Rosic S, Kohler F, Erhardt S. Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. *J Cell Biol.* 2014; 207:335–349. [PubMed: 25365994]
- Ross JE, Woodlief KS, Sullivan BA. Inheritance of the CENP-A chromatin domain is spatially and temporally constrained at human centromeres. *Epigenetics Chromatin.* 2016; 9:20. [PubMed: 27252782]
- Rudd M, Willard H. Analysis of the centromeric regions of the human genome assembly. *Trends Genet.* 2004; 20:529–533. [PubMed: 15475110]
- Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF. Genomic and genetic definition of a functional human centromere. *Science.* 2001; 294:109–115. [PubMed: 11588252]
- Scott KC, Sullivan BA. Neocentromeres: a place for everything and everything in its place. *Trends Genet.* 2014; 30:66–74. [PubMed: 24342629]
- Shang WH, Hori T, Martins NM, Toyoda A, Misu S, Monma N, Hiratani I, Maeshima K, Ikeo K, Fujiyama A, Kimura H, Earnshaw WC, Fukagawa T. Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. *Developmental cell.* 2013; 24:635–648. [PubMed: 23499358]
- Shchelochkov OA, Cheung SW, Lupski JR. Genomic and clinical characteristics of microduplications in chromosome 17. *Am J Med Genet A.* 2010; 152A:1101–1110. [PubMed: 20425816]
- Stankiewicz P, Parka SS, Holder SE, Waters CS, Palmer RW, Berend SA, Shaffer LG, Potocki L, Lupski JR. Trisomy 17p10-p12 resulting from a supernumerary marker chromosome derived from chromosome 17: molecular analysis and delineation of the phenotype. *Clin Genet.* 2001; 60:336–344. [PubMed: 11903333]
- Sullivan BA. Optical mapping of protein-DNA complexes on chromatin fibers. *Methods Mol Biol.* 2010; 659:99–115. [PubMed: 20809306]
- Sullivan BA, Karpen GH. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature structural & molecular biology.* 2004; 11:1076–1083.
- Sullivan LL, Boivin CD, Mravinac B, Song IY, Sullivan BA. Genomic size of CENP-A domain is proportional to total alpha satellite array size at human centromeres and expands in cancer cells. *Chromosome Res.* 2011; 19:457–470. [PubMed: 21484447]
- Sun X, Le HD, Wahlstrom JM, Karpen GH. Sequence analysis of a functional *Drosophila* centromere. *Genome Res.* 2003; 13:182–194. [PubMed: 12566396]
- Sun X, Wahlstrom J, Karpen G. Molecular structure of a functional *Drosophila* centromere. *Cell.* 1997; 91:1007–1019. [PubMed: 9428523]
- Taddei A, Maison C, Roche D, Almouzni G. Reversible disruption of pericentric heterochromatin and centromere function by inhibiting deacetylases. *Nature cell biology.* 2001; 3:114–120. [PubMed: 11175742]
- Thakur J, Sanyal K. Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function at native physical chromosomal loci in *Candida albicans*. *Genome Res.* 2013; 23:638–652. [PubMed: 23439889]
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutayavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang

- Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA. The accessible chromatin landscape of the human genome. *Nature*. 2012; 489:75–82. [PubMed: 22955617]
- Topp CN, Zhong CX, Dawe RK. Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc Natl Acad Sci U S A*. 2004; 101:15986–15991. [PubMed: 15514020]
- Vafa O, Sullivan KF. Chromatin containing CENP-A and alpha-satellite DNA is a major component of the inner kinetochore plate. *Curr Biol*. 1997; 7:897–900. [PubMed: 9382804]
- Warburton PE. Chromosomal dynamics of human neocentromere formation. *Chromosome Res*. 2004; 12:617–626. [PubMed: 15289667]
- Warburton PE, Wevrick R, Mahtani MM, Willard HF. Pulsed-field and two-dimensional gel electrophoresis of long arrays of tandemly repeated DNA : analysis of human centromeric alpha satellite. *Methods Mol Biol*. 1992; 12:299–317. [PubMed: 21409641]
- Waye JS, England SB, Willard HF. Genomic organization of alpha satellite DNA on human chromosome 7: evidence for two distinct alphoid domains on a single chromosome. *Mol Cell Biol*. 1987; 7:349–356. [PubMed: 3561394]
- Waye JS, Willard HF. Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome. *Mol Cell Biol*. 1986; 6:3156–3165. [PubMed: 3785225]
- Wevrick R, Earnshaw WC, Howard-Peebles PN, Willard HF. Partial deletion of alpha satellite DNA associated with reduced amounts of the centromere protein CENP-B in a mitotically stable human chromosome rearrangement. *Mol Cell Biol*. 1990; 10:6374–6380. [PubMed: 2247061]
- Willard HF, Waye JS. Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat. *J Mol Evol*. 1987; 25:207–214. [PubMed: 2822935]
- Williams BC, Murphy TD, Goldberg ML, Karpen GH. Neocentromere activity of structurally acentric mini-chromosomes in *Drosophila*. *Nat Genet*. 1998; 18:30–37. [PubMed: 9425896]
- Wong LH, Brettingham-Moore KH, Chan L, Quach JM, Anderson MA, Northrop EL, Hannan R, Saffery R, Shaw ML, Williams E, Choo KH. Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res*. 2007; 17:1146–1160. [PubMed: 17623812]
- Zeng K, de las Heras JI, Ross A, Yang J, Cooke H, Shen MH. Localisation of centromeric proteins to a fraction of mouse minor satellite DNA on a mini-chromosome in human, mouse and chicken cells. *Chromosoma*. 2004; 113:84–91. [PubMed: 15300445]

Figure 1. Molecular cytogenetic analysis of a patient cell line GM8148 containing a rearranged, mutant HSA17

(A) The HSA17 centromere region contains three higher order repeat alpha satellite arrays D17Z1-B (blue), D17Z1 (red), and D17Z1-C (purple). Centromeric chromatin defined by CENP-A nucleosomes and H3 nucleosomes methylated at lysines 4 (K4) and 36 (K36) (green circles) is typically assembled at D17Z1, on a portion of the array. Heterochromatin marked by H3K9me3 and H3K27me (orange circles) is assembled on the remainder of D17Z1, as well as on the flanking pericentromeric arrays D17Z1-B and D17Z1-C. (B) FISH with DNA probe p17H (red) that recognizes the D17Z1 alpha satellite array on HSA17 on the diploid patient cell line containing a normal 17 (nl17) the mutant deleted 17 (del17), and the supernumerary marker (mar17). Enlargement of each chromosome HSA17 from the patient showing the small amount of D17Z1 and the reduction in the short arm on the del(17). The del(17) lacks most of D17Z1 and all the short arm pericentromere that includes D17Z1-B. Chromosomes are counterstained with DAPI (blue). Scale bar equal 15 microns.

Figure 2. Molecular refinement of the extent of HSA17 short arm deletion and identification of the breakpoint region on the del(17)

(A) BAC CGH array narrowed the breakpoint to a 292kb region in band 17p12 between BACs RP11-409O23 and RP11-196N4. The gray line at 0 represent diploid copy number; the gray line below 0 represents deleted segments of the HSA17 short arm. Light blue circles denote individual BACs include in the array; arrows point to specific BACs. (B) Further definition of the breakpoint region by PCR using primers to STS markers. Genomic DNA isolated from somatic cell hybrids containing either the normal 17 or the del(17) was amplified with each primer set. PCR products run on agarose gels are shown as ethidium bromide DNA bands above the name of each STS marker. The white bands on dark background were inverted to dark bands on light background for presentation purposes only. The absence of bands in the del(17) lanes indicated that this region of HSA17 was deleted.

This type of analysis allowed us to narrow the breakpoint to a 28kb region between D17S2117 (light blue box) and D17S1808 (note the very faint band in the del(17) for D17S2117). (C) DNA sequencing further narrowed the breakpoint region (blue line) to a 5kb region in the middle of 17p12. The breakpoint region is rich in transposable elements, specifically L1/LINEs, and appears to be an area of mobile element insertion. (D) Summary schematic comparing the genomic structure of the mutant del(17) to normal HSA17. Yellow shading shows the 12.9Mb region of the centromere (green line) and short arm that is missing on the del(17).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. Mapping centromere position on del(17) using chromatin fiber immunostaining and FISH

(A) Representative images of chromatin fibers showing the centromere regions of the normal 17 and del(17) immunostained for CENP-A and hybridized with a DNA probe for D17Z1. Single panel images show the CENP-A and FISH signals separately, revealing that CENP-A (an indicator of CEN chromatin) localized to portion of the D17Z1 array on the normal 17. On the del(17), CENP-A was also localized to a portion of D17Z1, but some of the CENP-A staining did not overlap with D17Z1 suggesting that it was partially located within the nearby non-satellite DNA. Scale bars equal 5 microns. (B) The amount of overlap between CENP-A immunostaining and D17Z1 FISH signal was measured on individual chromatin fibers, revealing that 55% of CENP-A was associated with D17Z1. (B') Using the known size (700kb) of D17Z1 on the del(17), we estimated that the total CENP-A domain measured ~670kb in length, and that ~360kb co-localized with D17Z1. Thus, ~310kb of CENP-A chromatin is located on euchromatic/non-satellite DNA. (C) Schematic showing the HSA17 UCSC Genome Browser coordinates for HSA17 deletion (red box on ideogram), summarizing the chromatin fiber analysis of CENP-A chromatin (red bar) location to both short arm DNA and D17Z1 (green bar) on the del(17) compared to its localization only at D17Z1 on normal HSA17. The yellow bar represents the deleted portion of the HSA17 short arm on the del(17).

Figure 4. ChIP and gene expression analysis of Cenp-A spreading into euchromatin on del(17)
 (A) ChIP-qPCR showed that Cenp-A is enriched at D17Z1 on the normal HSA17, and to lesser extent on del(17). Five genomic sites distal to the breakpoint and a control genomic site present only on the normal HSA17 were interrogated by ChIP to measure the genomic extent of Cenp-A spreading. All five euchromatic HSA17 short arm sites were enriched from Cenp-A on the del(17) compared to the normal HSA17. Enrichment was determined based on % input compared to D17Z1 (the typical site of Cenp-A enrichment on HSA17). Error bars represent SEM. Significant differences between Cenp-A enrichment at each genomic site on the del(17) and normal 17 were determined using a Student's t-test. (B) Summary schematic of chromatin fiber and ChIP results showing the spreading of Cenp-A chromatin into ~300kb of HSA17 short arm euchromatin. Additional Genome Browser tracks are included below Cenp-A fiber and ChIP data to illustrate that Cenp-A chromatin spreads over

a genomic region that includes genes/RNA transcripts and transposable elements and is largely characterized by open chromatin (ie. DNase hypersensitive sites, DNase HSS). (C) RT-qPCR analysis of the ELAC2 gene that is present with the repositioned centromeric chromatin domain (genomic site 1, Figure 4A) revealed that ELAC2 transcription is reduced on the del(17) compared to the normal 17. RNA was extracted in duplicate from each cell line, and RT-qPCR was done in triplicate for each sample. Relative expression was calculated as a ratio of ELAC2 expression over beta actin expression, and the normal HSA17 was adjusted to 1. Error bars = SEM. A Student's t-test was used to determine that gene expression was significantly different between the normal 17 and del(17) ($p < 0.0001$).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

PCR					
STS Marker/ID	primer name	description	length (bp)	Position Assembly GRCh38/hg38	Ta
D17S2012/W19205	BS328/329	F: 5' CCCAGGGCCAAAGAGAAAG 3' R: 5' GCATGATGGGAGGGAATG 3'	121	chr17:9897857-9897977	60–61°C
D17S1525/W1-5843	BS326/327	F: 5' ATCACCTGGCAAAGTGATGG 3' R: 5' GACTAAAGGACCCCAAATGAGA 3'	105	chr17:11180020-11180124	58°C
D17S1692	BS342/343	F: 5' CACCAAGCCACCAACCCCTAC 3' R: 5' CATAATGCGTCCTTGGGCAC 3'	218	chr17:12338237-12338454	59°C
D17S1311	BS324/325	F: 5' AACACTCACCTCTCTGTGGGR: ATTTCCTGCGCTTGTGTCC 3'	152	chr17:12581372-12581523	52°C
D17S1477E	BS340/341	F: 5' AGTGTAAGTCACCCCTCCTC R: CAAATTGAGGCAATGACCTG 3'	96	chr17:12991442-12991537	59°C
11 kb distal to breakpoint	BS366/367	F: 5' GCAATGGGAAAGTGCTATGGTR: CTCATGGGCTCTTTTCATA 3'	155	chr17:13092677-13092831	59–60°C
63 kb distal to breakpoint	BS368/369	F: 5' TGGGTGAACTGGAGGGTAG 3' R: 5' TCTGCTCAACAATGGTGCTC 3'	114	chr17:13140670-13140783	59–60°C
45 kb distal to breakpoint	BS362/363	F: 5' CCAGACCTAGACCTGCAAGC 3' R: 5' AAATGGGGTAATGCAATGGA 3'	142	chr17:13159800-13159941	62°C
13 kb distal to breakpoint	BS364/365	F: 5' CAGAAAGGGCTCTCAACAG 3' R: 5' AGCCTTTCTCCTCCTCCTTG 3'	166	chr17:13190285-13190450	62°C
2.6 kb distal to breakpoint	BS358/359	F: 5' GGGAAAGGGGTGAGACTTTTC 3' R: 5' GTTGAGGTGCCCGTTTGTGTTT 3'	125	chr17:13201073-13201197	62°C
1 kb distal to breakpoint	BS360/361	F: 5' GCCACAAATTGCAACCTTTT 3' R: 5' TTATCTTGGCCACACACA 3'	213	chr17:13202647-13202860	59°C
D17S2117	BS338/339	F: 5' ACTCCGTGTCGAAAGAAAG 3' R: 5' TGGGAAGTGATAGAACGAAA 3'	100	chr17:13203706-13203805	56°C
D17S1808	BS336/337	F: 5' TCATTGTGTTTGACATGGGA 3' R: 5' GCATAAACCTCAGCACCTA 3'	278	chr17:13232065-13232342	53°C
D17S1550	BS334/335	F: 5' CCTCAGAGAAAATGTTCAATCG 3' R: 5' AGTAAAGGCCCTAGACCCTTCTT 3'	152	chr17:13557848-13557999	52–53°C
D17S1358	BS332/333	F: 5' CCTAATTACACAATACCTTTTGGGG 3' R: 5' GTTGGCTTACTCTAATACATCAGT 3'	181	chr17:14528891-14529071	57°C
D17S1356	BS322/323	F: GGCTTCTTGTGTTGTGAGGT R: TGACACTGTTCGATATACTGATTG	154	chr17:14991351-14991505	57°C
mouse minor satellite	BS591/592	F: 5' ACTCACTCTAATATGTTCTACAGTG 3' R: 5' AAAACACATTCGTTGGAACGGG 3'	ladder	mouse centromere regions	58°C
mouse major satellite	BS593/594	F: 5' CATATCCAGGTCTTTCAGTGTGC 3' R: 5' CACTTAGGACGTGAAATATGGCG 3'	ladder	mouse pericentromere regions	58°C

ChIP-qPCR

STS Marker/ID	primer name	description	length (bp)	Position Assembly GRCh38/hg38	Ta
D17S1692	BS342/343	F: 5' CACCAAGCCACCAACCCCTAC 3' R: 5' CATAATGCGTCCTTGGGCAC 3'	218	chr17:12338237-12338454	60°C
D17S1311	BS324/325	F: 5' AACACTCACCTCTCTGTGG 3' R: 5' ATTTCCTGCGCTTGTGTCC 3'	152	chr17:12581372-12581523	60°C
D17S1477E	BS340/341	F: 5' AGTGTAAGTCACCCCTCCTC 3' R: 5' CAAATTGAGGCAATGACCTG 3'	96	chr17:12991442-12991537	62°C
11 kb distal to breakpoint	BS366/367	F: 5' GCAATGGGAAAGTGCTATGGT 3' R: 5' CTCCAATGGGCTCTTTTCATA 3'	155	chr17:13092677-13092831	62°C

ChIP-qPCR						
STS Marker/ID	primer name	description	length (bp)	Position Assembly GRCh38/hg38	Ta	
63kb distal to breakpoint	BS368/369	F: 5' TGGGTTGAACCTGGAGGGTAG 3' R: 5' TCTGCTCAAAACATGGTGCTC 3'	114	chr17:13140670-13140783	62°C	
2.6kb distal to breakpoint	BS358/359	F: 5' GGGAAAGGGGTGAGACTTTTC 3' R: 5' GTTGAGGGTTGCCGTTTGTTT 3'	125	chr17:13045115-13045239	62°C	
D17S2117	BS338/339	F: 5' ACTCCGTGTCGAAAGAAAAG 3' R: 5' TGGGAAGTGATAGAACGAAA 3'	100	chr17:13203706-13203805	56°C	
D17S1550	BS334/335	F: 5' CCTCAGAGAAAAAATGTTTCATCG 3' R: 5' AGTAAAGGCCTTAGACCACCTTCTT 3'	152	chr17:13557848-13557999	62°C	
D17Z1	BS518/109	F: 5' AAAACTGCGCTCTCAAAAGG 3' R: 5' AATTTCAGCTGACTAAACA 3'	267	chr17:23195019-26566633	63°C	