Published in final edited form as:

J Proteome Res. 2017 September 01; 16(9): 3433–3442. doi:10.1021/acs.jproteome.7b00448.

Elucidating Protein-DNA Interactions in Human Alphoid Chromatin via Hybridization Capture and Mass Spectrometry

Katherine E. Buxton¹, Julia Kennedy-Darling¹, Michael R. Shortreed¹, Nur Zafirah Zaidan², Michael Olivier³, Mark Scalf¹, Rupa Sridharan^{2,4}, and Lloyd M. Smith^{1,5,*}

¹Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, United States

²Epigenetics Theme, Wisconsin Institute for Discovery, University of Wisconsin, Madison, Wisconsin 53715, United States

³Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas 78227, United States

⁴Department of Cell and Regenerative Biology, University of Wisconsin, Madison, Wisconsin 53715, United States

⁵Genome Center of Wisconsin, University of Wisconsin, Madison, Wisconsin 53706, United States

Abstract

The centromere is the chromosomal locus where the kinetochore forms and is critical for ensuring proper segregation of sister chromatids during cell division. A substantial amount of effort has been devoted to understanding the characteristic features and roles of the centromere, yet some fundamental aspects of the centromere, such as the complete list of elements that define it, remain obscure. It is well-known that human centromeres include a highly repetitive class of DNA known as alpha satellite, or alphoid, DNA. We present here the first DNA-centric examination of human protein-alpha satellite interactions, employing an approach known as HyCCAPP (hybridization capture of chromatin-associated proteins for proteomics) to identify the protein components of alphoid chromatin in a human cell line. Using HyCCAPP, cross-linked alpha satellite chromatin was isolated from cell lysate, and captured proteins were analyzed via mass spectrometry. After being compared to proteins identified in control pulldown experiments, 90 proteins were identified as enriched at alphoid DNA. This list included many known centromere-binding proteins in addition to multiple novel alpha satellite-binding proteins, such as LRIF1, a heterochromatin-associated protein. The ability of HyCCAPP to reveal both known as well as novel alphoid DNA-interacting proteins highlights the validity and utility of this approach.

*Corresponding author: Phone: (608) 263-2594. Fax: (608) 265-6780. smith@chem.wisc.edu.

Supplementary information is available free of charge at the ACS website (<http://pubs.acs.org>).

Conflict of interest disclosure: The authors declare no competing financial interest.

Author contributions: K.E.B., J.K.D., M.O., and L.M.S. conceived of HyCCAPP experiments. K.E.B. and J.K.D. developed experimental aspects of HyCCAPP, and K.E.B. performed HyCCAPP experiments. M.S. and K.E.B. analyzed captured proteins via MS. K.E.B. and M.R.S. analyzed MS data. K.E.B., N.Z.Z., and R.S. provided biological insight into proteins identified as enriched at alpha satellite DNA via HyCCAPP. K.E.B., N.Z.Z., and R.S. conceived of ChIP validation experiments, and K.E.B. executed these experiments. K.E.B. and L.M.S. prepared the manuscript. All authors edited and reviewed the manuscript.

Graphical abstract

Keywords

HyCCAPP; chromatin; hybridization capture; DNA-binding proteins; protein-DNA interactions; alpha satellite DNA; centromere; LRIF1

Introduction

The centromere is the chromosomal region that is responsible for linking replicated sister chromatids, appearing as the primary constriction in condensed, mitotic chromosomes. The centromere is also where the kinetochore, the proteinaceous structure that serves as the attachment point for spindle microtubules, forms and is therefore critical for ensuring proper chromosomal segregation during cell division. Human centromeres and their surrounding pericentromeric regions include a highly repetitive class of DNA known as alpha satellite, or alphoid, DNA.¹ The repeating unit of alpha satellite DNA is 171 base pairs in length, and these monomeric units are arranged tandemly in a head-to-tail fashion within the centromeric region.² An integral number of repeated alpha satellite monomers make up a structure known as a higher-order repeat (HOR), which is itself tandemly repeated so that the alpha satellite array may extend for ~0.2–5 Mb (Fig. 1).² Because this large block of alpha satellite DNA is found within the centromeric region of every chromosome, it is estimated that alphoid DNA makes up as much as ~3–5% of the human genome as a whole.¹ Individual alpha satellite monomers are relatively heterogeneous in nature, varying in sequence by as much as ~30–50%, but the HORs that make up an array show much more sequence similarity, varying by less than ~2%.^{2,3} Flanking the HOR array, the pericentromeric region contains disordered alpha satellite monomers, which become more heterogeneous in sequence further away from the centromeric core and gradually lead into nonalphoid genomic DNA.^{2,4}

Highly repetitive centromeric DNA is not unique to human cells. In fact, with few exceptions including the point centromeres of budding yeast and the holocentric chromosomes of *Caenorhabditis elegans*, repetitive centromeric sequences have been identified in most other animal, plant, and fungal species studied to date.^{3,5} This suggests that such repetitive sequences may be critical to centromere identity or kinetochore function, a hypothesis supported by the finding that both alpha satellite DNA and the CENP-B box, a 17 bp sequence found within a subset of alpha satellite repeats to which centromere protein B binds,⁶ are required for the *de novo* formation of functional human artificial chromosomes

(HACs).⁷ These results suggest that alpha satellite DNA may have intrinsic properties without which a functional human centromere cannot form.

However, the conclusion that repetitive DNA is required for normal centromere function is called into question when one considers the phenomenon of neocentromeres.

Neocentromeres, first identified by Voullaire et al. in 1993,⁸ are novel, ectopic centromeres that may form after the disruption or inactivation of a natural centromere.⁹ Unlike natural centromeres, neocentromeres form at diverse, nonalphoid genomic loci.^{9,10} These neocentromeres appear to be functional, supporting chromosomal segregation during cell division,^{8,10} which refutes the idea that alpha satellite DNA is required for centromere formation.

Together, the observations outlined above raise questions about what role, if any, alpha satellite DNA actually plays in centromeric identity or function. The discovery of neocentromeres at nonalphoid sequences indicates that the human centromere is not defined solely by a particular DNA sequence, suggesting that the specification of the centromeric locus may involve other factors.

Much work has already been completed to shed light on the nongenetic elements, particularly proteins and associated post-translational modifications (PTMs), that may be important for centromere specification.^{4,11–13} Centromere protein A, or CENP-A, is just one example of a protein that appears to be important for centromere function. CENP-A is a histone H3 variant that is found in some centromeric nucleosomes,^{14–16} and homologues of CENP-A have been shown to be important for viability in budding yeast, mice, and *Drosophila*.^{17–19} Furthermore, extensive affinity purification experiments in human HeLa cells have revealed that CENP-A nucleosomes recruit a set of six proteins (CENP-C, -H, -M, -N, -T, and -U), collectively referred to as the CENP-A nucleosome-associated complex (NAC).²⁰ The CENP-A NAC, in turn, recruits a set of seven additional proteins (CENP-K, -L, -O, -P, -Q, -R, and -S), which are collectively referred to as the CENP-A distal (CAD) complex, and it has been observed that the disruption of CENP-A NAC formation results in mitotic errors.²⁰ However, despite its clearly important role, it does not appear that the presence of CENP-A alone is sufficient for centromere specification, as CENP-A nucleosomes are also found at ectopic sites, where functional centromeres do not form.²¹

CENP-A and its associated NAC and CAD proteins are just a few examples of the many players that have been identified as important for centromere function and kinetochore recruitment. Additionally, the transcriptional status of the centromeric region appears to be critical for proper centromere and kinetochore performance, requiring a balance between open, euchromatic chromatin and closed, heterochromatic chromatin.²² Still, with all of this information we do not yet have conclusive answers to questions about why human centromeres almost invariably form on repetitive, alpha satellite DNA, when other genomic loci are clearly also capable of performing this task.

In this study, we seek to further investigate the protein components of alpha satellite chromatin to reveal novel protein-alpha satellite interactions that may be important for centromere specification or function. We have used an approach known as HyCCAPP, or

hybridization capture of chromatin-associated proteins for proteomics.^{23,24} Briefly, human K562 cells are cross-linked using formaldehyde to fix protein-DNA interactions, then cells are lysed and chromatin is sheared. The lysate is incubated with biotinylated capture oligonucleotides, whose sequences are designed to complement the human alpha satellite consensus sequence,²⁵ and hybridized complexes are isolated using streptavidin-coated magnetic beads. Following washing steps, the hybridized complexes are released from the beads, and associated proteins are identified via bottom-up mass spectrometry.²⁶ The HyCCAPP approach is distinct from other strategies commonly used to study protein-DNA interactions, such as affinity purification or immunoprecipitation, in that it is DNA- rather than protein-centric. HyCCAPP has proven to be a useful tool, revealing both known as well as novel protein-DNA interactions in yeast.^{23,24} Here, we apply HyCCAPP to the human genome for the first time, providing a DNA-centric lens through which to view protein-alpha satellite DNA interactions. Using this approach, 90 proteins were identified as enriched at the alpha satellite repeats, including many known centromere-binding proteins in addition to a number of novel alpha satellite-binding proteins, which may provide new insights into centromere structure or function.

Methods

A detailed account of all experimental procedures employed in this work can be found in the Supplementary Information. Brief summaries of these procedures are provided here.

HyCCAPP

The HyCCAPP procedure was performed essentially as described previously²³ with relatively minor adjustments. Human K562 cells were cross-linked for 30 min at room temperature via gentle shaking with a final concentration of 3% (w/v) formaldehyde, and excess formaldehyde was quenched with the addition of Tris buffer pH = 8. Cells were washed, pelleted, and stored at -45°C for up to one month prior to use. For each HyCCAPP experiment, $\sim 3.3 \times 10^9$ cells were resuspended in lysis buffer containing 200 mM NaCl, 20 mM EDTA pH = 8, 50 mM Tris pH = 7, and protease inhibitors diluted 200X from a concentrated cocktail (Sigma-Aldrich #P8340). The cells were lysed using a Constant Systems TS Series cell disruptor at 18 kpsi, and then SDS was added to a final concentration of 1% and the lysate was heated in a 65°C water bath for 8 min to aid in chromatin solubilization. The chromatin was sheared via sonication to a median fragment size of ~3 kb (Misonix Ultrasonic Processor S4000) and the lysate was cleared of insoluble cellular debris via centrifugation. The supernatant was diluted five-fold with lysis buffer and RNA was digested via incubation with RNase A for 1 h at 37°C. The lysate was centrifuged again to remove any remaining particulates, and streptavidin-coated magnetic beads (Fisher Scientific #09-981-140) were added to the supernatant to clear endogenously biotinylated moieties. The lysate was incubated with the beads for 1 h at room temperature, after which the beads were removed and discarded. At this point, a 1 mL aliquot of the lysate was removed and stored at -20°C for eventual qPCR analysis.

Next, biotinylated alpha satellite capture oligonucleotides were added to the lysate. Eight capture oligonucleotides were designed to complement the human alpha satellite consensus

sequence²⁵ and added at equimolar concentrations. As a control, a biotinylated scrambled sequence capture oligonucleotide was also added to the lysate at this point. The sequences of all nine capture oligonucleotides used in HyCCAPP experiments are listed in Supplementary Table S-1. The lysate was incubated with the oligonucleotides for 3 h at 37°C, then allowed to cool to room temperature. To capture the hybridized complexes, streptavidin-coated magnetic beads were added and the lysate was incubated at room temperature for 1 h. The beads were isolated from the lysate using a magnet stand.

The beads were washed four times for 5 min each with wash buffer containing 50 mM Tris pH = 8, 200 mM NaCl, and 0.2% SDS. The beads were then concentrated into a smaller volume and washed once for 5 min and once for 1 h. To release the hybridized complexes, a toehold-mediated strand displacement strategy was employed (see Supplementary Information for details). This strategy was developed and demonstrated for the selective release of individual target loci from cross-linked yeast chromatin,^{27,28} and here we extended the strategy to cross-linked human chromatin. First, alpha satellite release solution containing equimolar concentrations of the eight alpha satellite release oligonucleotides was added to the beads, and the resultant bead slurry was gently mixed for 15 min at room temperature. The release solution was then removed and stored at 4°C, and the beads were washed twice for 5 min each. Next, scrambled release solution containing the scrambled release oligonucleotide was added to the beads, and the slurry was again gently mixed for 15 min at room temperature. The scrambled release solution was removed and stored at 4°C, and the beads were discarded. The sequences of all nine release oligonucleotides used in HyCCAPP experiments are listed in Supplementary Table S-1.

After removing a small aliquot of each release solution for eventual qPCR analysis, the proteins in both the alpha satellite and scrambled release solutions were precipitated using cold trichloroacetic acid (TCA). An aliquot of lysate material removed after hybridization capture was also precipitated at this time. The three pellets were resuspended in buffer containing 8 M urea and 0.1% deoxycholic acid, and the proteins were prepared for mass spectrometric analysis using the eFASP method.²⁹ Note that an explicit cross-link reversal step was not employed during the preparation of peptides for mass spectrometric analysis as, given the known rates of spontaneous formaldehyde cross-link reversal,³⁰ such a step was not deemed necessary. After removal of surfactant via extraction with ethyl acetate, the peptides in the aqueous phase were dried down and resuspended in 0.1% trifluoroacetic acid. The samples were then desalted using C18 solid-phase extraction pipet tips (Agilent Technologies #A57003100K), and the desalted peptides were dried down and resuspended in 95:5 water/acetonitrile with 0.2% formic acid.

Mass spectrometric analysis of HyCCAPP samples

Three replicates of the HyCCAPP experiment were carried out as described. Samples were analyzed via HPLC-ESI-MS/MS using a system consisting of a high performance liquid chromatograph (nanoAcquity, Waters) connected to an electrospray ionization (ESI) orbitrap mass spectrometer (LTQ Velos, Thermo Fisher Scientific). HPLC separation employed a 100 × 365 µm fused silica capillary microcolumn packed with 20 cm of 1.7 µm diameter, 130 Å pore size C18 beads (Waters BEH), with an emitter tip pulled to approximately 1 µm using a

laser puller (Sutter instruments). Peptides were loaded on-column at a flow-rate of 400 nL/min for 30 min, then eluted over 125 min at a flow-rate of 300 nL/min with a gradient from 2% to 30% acetonitrile in 0.1% formic acid. Full-mass profile scans were performed in the orbitrap between 300 and 1500 m/z at a resolution of 60000, followed by MS/MS HCD scans of the ten highest intensity parent ions with $z > 1$ at 42% relative collision energy and 7500 resolution, with a mass range starting at 100 m/z . Dynamic exclusion was enabled with a repeat count of two over a duration of 30 s and an exclusion window of 120 s. All raw data files are available from the MassIVE repository under the name “AlphaSatelliteHyCCAPP”.

The raw mass spectrometric data were analyzed using MaxQuant (version 1.5.3.30).³¹ The detected features were searched against the human canonical protein database from UniProt (downloaded on April 22, 2016) and a list of potential contaminants. A false discovery rate (FDR) of 1% at both the peptide and protein levels was allowed. Relative quantification was performed for each capture sample (alpha satellite and scrambled) of each HyCCAPP experiment (1–3) using the protein intensities reported in the “proteinGroups.txt” file produced by MaxQuant. Individual protein intensities were normalized to the median protein intensity of the corresponding capture sample after removing potential contaminants, and these normalized protein intensities were loaded into Perseus (version 1.5.3.2)³² for downstream statistical analysis. Two-sided Welch’s t tests were performed to compare protein abundances between groups (alpha satellite capture versus scrambled capture). The statistical tests were corrected for multiple hypothesis testing using a permutation-based FDR cutoff of 5% ($S_0 = 2$) (see Supplementary Table S-2 for output of Perseus analysis).

Quantitative real-time PCR analysis of HyCCAPP samples

Quantitative real-time PCR (qPCR) was performed to assess the efficiency and specificity with which the alpha satellite repeats were captured in HyCCAPP experiments. Because of the heterogeneity of alpha satellite monomeric sequences, it was not feasible to design qPCR assays to target every individual alpha satellite monomer. Instead, a TaqMan assay was designed to target a single alpha satellite sequence, and the capture efficiency of this sequence was used as a proxy for the capture efficiency of alpha satellite DNA as a whole. Additionally, a qPCR assay for an off-target sequence (28S rDNA) was designed to assess capture specificity. These assays were ordered from IDT, and the sequences of the primers and probes are listed in Supplementary Table S-3. qPCR was performed essentially as described previously²³ to measure the concentrations of the target alpha satellite sequence and the off-target 28S rDNA sequence in the alpha satellite release, scrambled release, and lysate (input) samples. Capture efficiency was then calculated by dividing the amount of alpha satellite DNA or 28S rDNA present in the release solution of interest by the amount of that sequence present in the lysate sample (prior to capture).

Chromatin immunoprecipitation with qPCR (ChIP-qPCR)

Human K562 cells were cross-linked for 10 min at room temperature via gentle shaking with a final concentration of 1% (w/v) formaldehyde, and excess formaldehyde was quenched with the addition of Tris buffer pH = 8. Cells were washed, pelleted, and stored at -80°C prior to use. To prepare chromatin for ChIP experiments, cells were resuspended in ice-cold buffer containing 50 mM HEPES pH = 7.9, 140 mM NaCl, 1 mM EDTA, 10%

glycerol, 0.5% NP-40 substitute (Amresco #M158), 0.25% Triton X-100, and protease/phosphatase inhibitors diluted 100X from a concentrated cocktail (Thermo Fisher Scientific #1861281). The resuspended cells were incubated on ice for 20 min, and the nuclear pellet was collected via centrifugation and washed. The pellet was resuspended at a final concentration of $\sim 7.4 \times 10^6$ nuclei/mL in ice-cold shearing buffer containing 0.1% SDS, 1 mM EDTA, 10 mM Tris pH = 8.1, and protease/phosphatase inhibitors, and the chromatin was sheared to ~200–500 bp using a Covaris S220 water bath sonicator. The sheared chromatin was centrifuged to remove insoluble debris, and the supernatant was aliquoted into 10- μ g and 1- μ g portions (based on Nanodrop measurement of DNA concentration) and stored with 10% glycerol at -80°C for up to two months prior to use.

For each ChIP experiment, a 10- μ g aliquot of chromatin was thawed on ice and diluted 10-fold in ice-cold buffer containing 16.7 mM Tris pH = 8, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 333 mM NaCl, and protease/phosphatase inhibitors. Then, 2 μ g of either CENP-B antibody (abcam #ab25734), LRIF1 antibody (Atlas Antibodies #HPA044515), or IgG (Jackson ImmunoResearch #312-005-003) were added to the chromatin and incubated overnight at 4°C with gentle mixing. Next, 10 μ L each of protein A- and protein G-coated magnetic beads (Thermo Fisher Scientific #10002D and #10004D) were washed, then added to the sample and gently mixed for 2 h at 4°C . The beads were isolated using a magnet stand and the supernatant was discarded. The beads were then washed thoroughly to remove nonspecifically-bound proteins/DNA, and the captured chromatin was eluted from the beads by heating at 65°C for 10 min in 100 μ L of elution buffer containing 50 mM Tris pH = 8, 1 mM EDTA, 1% SDS, and protease/phosphatase inhibitors. Next, 80 μ L of TE buffer containing 0.67% SDS and protease/phosphatase inhibitors were added to the sample, and the sample was heated overnight at 65°C . At this time, a 1- μ g aliquot of chromatin (“10% input sample”) was thawed on ice and diluted with 100 μ L of elution buffer and 80 μ L of TE buffer containing 0.67% SDS and protease/phosphatase inhibitors and heated overnight at 65°C .

The ChIP and 10% input samples were treated with RNase A and proteinase K, and the beads were removed from the ChIP sample using a magnet stand. The DNA in both samples was then purified via phenol-chloroform extraction and ethanol precipitation, and the DNA pellets were resuspended in 10 mM Tris pH = 7. The DNA was then analyzed via qPCR using two assays designed to target alpha satellite DNA and one assay designed to target 28S rDNA (negative control sequence). These assays were ordered from IDT, and the sequences of the primers and probes are listed in Supplementary Table S-3. For each assay, a standard curve was generated by preparing serial dilutions of the 10% input sample, and each standard/ChIP sample was analyzed in duplicate on a 96-well plate using a Roche 480 LightCycler.

Results and Discussion

Evaluation of hybridization capture efficiency and specificity

Quantitative real-time PCR was employed to evaluate the efficiency and specificity of hybridization capture in HyCCAPP. Here, capture efficiency was defined as the percentage of input alpha satellite DNA present in the final capture sample. Using a qPCR assay

designed to target an alpha satellite sequence, the efficiency with which alphoid DNA was captured in the alpha satellite pulldown experiments was determined to be ~0.22%, which is comparable to capture efficiencies reported previously for HyCCAPP in yeast.²³ The efficiency with which alphoid DNA was captured in the scrambled pulldown experiments was approximately 44-times lower, or ~0.0049% (Fig. 2). In small-scale HyCCAPP experiments, the effect of release order (alpha satellite first followed scrambled or vice versa) on alpha satellite capture efficiency was evaluated, and no dramatic differences were apparent (Supplementary Figure S-1).

To evaluate the specificity of capture, the capture efficiency of an off-target sequence (28S rDNA) was also calculated for each sample. Neither the alpha satellite nor the scrambled capture oligonucleotides captured the off-target sequence with appreciable efficiency (Fig. 2). Additionally, the enrichment of alpha satellite DNA after hybridization capture was determined by comparing the ratio of alpha satellite DNA to 28S rDNA in whole cell lysate (precapture) to the ratio in the alpha satellite capture sample. This gave an average enrichment factor of ~140-fold, which indicated that the hybridization capture step effectively enriched alphoid DNA.

The success of a HyCCAPP experiment hinges on its ability to specifically isolate the target genomic locus from cross-linked chromatin in sufficient quantities that its associated proteins are detectable via mass spectrometry. Because of the exceedingly high copy number of alpha satellite DNA, a capture efficiency of ~0.22% provided enough material to meet this criterion, allowing for the detection of locus-specific proteins by standard bottom-up proteomics. Given that we expect capture efficiency in HyCCAPP to be related to the presence of single-stranded regions within the input chromatin, it is interesting to consider the mechanism(s) by which single-stranded regions are available in the first place. An earlier generation of the HyCCAPP technology, known as GENECAPP (global exonuclease-based enrichment of chromatin-associated proteins for proteomics), employed exonuclease treatment to produce single-stranded regions amenable to capture. This approach was successful in *in vitro* studies on a model system,³³ but transitioning to the capture of chromatin fragments from whole cell lysate proved challenging. Initially, a strategy involving restriction enzyme digestion followed by exonuclease treatment was employed to generate single-stranded regions in a sequence-specific manner. However, the restriction enzyme digestion step proved inefficient in the context of whole cell lysate, and sonication was instead selected as the means of chromatin fragmentation. It was assumed that exonuclease digestion would still be necessary to generate single-stranded regions in the chromatin, but control experiments showed that hybridization occurred equally well either with or without exonuclease treatment. This observation led to the development of HyCCAPP, which does not rely on enzymatic digestions, thereby reducing the cost and complexity of the procedure. Still, the mechanism by which chromatin is available for capture in HyCCAPP is not yet completely clear. One hypothesis is that the sonication step employed early in the HyCCAPP experiment serves not only to shear DNA into smaller pieces, but also to introduce single-stranded regions^{34,35} and overhanging ends, which could be amenable to hybridization capture. Additional work will be required to fully understand the mechanism(s) by which single-stranded regions of chromatin are available in HyCCAPP and to take advantage of these mechanisms to increase capture efficiency.

Identification of proteins enriched at the alpha satellite repeats

Identification and extracted ion chromatogram-based label-free quantification of the proteins present in the alpha satellite capture and scrambled capture samples were performed using MaxQuant software. Perseus software was then used to perform *t* tests to determine which proteins were enriched in the alpha satellite capture samples as compared to the scrambled capture samples. Comparison to the scrambled oligonucleotide pulldown allows one to control for background proteins which may be carried through to the final alpha satellite capture sample, perhaps due to their high abundance or through nonspecific binding to the beads, as these proteins are presumably present in the alpha satellite and scrambled capture samples in similar amounts. The comparison between the alpha satellite capture and scrambled capture samples yielded 90 proteins which were significantly enriched at the alpha satellite repeats, with an FDR of 5% and a minimum fold-change of 5.8 (Fig. 3). This list of 90 proteins will henceforth be referred to simply as “alpha satellite-enriched proteins” (a complete list of these proteins can be found in Supplementary Table S-4).

Analysis of the alpha satellite-enriched protein list

To ascertain whether the final list of 90 alpha satellite-enriched proteins was reasonable, the list was analyzed via the PANTHER over-representation test (released on July 15, 2016).^{36,37} This test finds GO terms that are more abundant among input proteins than would be expected based on a random list of proteins from the reference *Homo sapiens* database. Three over-representation tests were performed to find 65 enriched cellular component, 15 enriched molecular function, and 100 enriched biological process GO terms, all with Bonferroni-corrected *p*-values less than 0.05. The list of enriched GO terms included terms like “chromosome, centromeric region”, “mitotic cell cycle”, “sister chromatid cohesion”, “kinetochore”, “CENP-A containing chromatin organization”, and “pericentric heterochromatin”. These terms make sense based on what is known about the alpha satellite repeats such as their localization at the centromeric/pericentromeric region of human chromosomes and the centromere’s relationship to the kinetochore and cell cycle. This GO term analysis thus served as an initial indication that the HyCCAPP experiments and subsequent data analysis steps were working as expected to identify true alpha satellite-interacting proteins. Table 1 lists the 30 enriched GO terms with the smallest corrected *p*-values (see Supplementary Table S-5 for a complete list of enriched GO terms). Because the HyCCAPP experiments were performed using asynchronous cells, we expect the set of enriched proteins to reflect activities of the centromeric region at various points of the cell cycle.

As another way to validate that the final list of alpha satellite-enriched proteins was nonrandom and reflected real associations existing within the cell, we used STRING (version 10.0)³⁸ to identify interactions between these 90 proteins (Fig. 4). The resultant network consisted of 464 edges, with an average of 10.3 interactions per protein. The expected number of edges for a random network of this size is 76, meaning that this collection of proteins is very likely nonrandom.

HyCCAPP identifies known centromeric proteins as enriched at the alpha satellite repeats

After the list of alpha satellite-enriched proteins was evaluated as a whole, individual proteins on the list were explored further. It was immediately obvious that HyCCAPP had identified a number of known alpha satellite-binding proteins as enriched at alphoid DNA (~30–50, depending on the criteria used for classification). For example, all six members of the CENP-A nucleosome-associated complex (CENP-C, -H, -M, -N, -T, and -U) and six out of seven members of the CENP-A distal complex (CENP-K, -O, -P, -Q, -R, and -S) were identified as enriched. Interestingly, despite the presence of all of these proteins which are, at least to some extent, recruited by CENP-A,²⁰ CENP-A itself was not identified by MaxQuant in any of the samples analyzed.

Although it is well-known that CENP-A is enriched at the centromere, it is not unreasonable that HyCCAPP did not identify this protein as one of the 90 alpha satellite-enriched proteins. It is important to understand that the HyCCAPP technology is fundamentally limited by the sensitivity of the mass spectrometric analysis. The detection limit for a given peptide in bottom-up proteomics is dependent on factors such as its solubility and ionization efficiency, and if none of the tryptic peptides corresponding to a protein are present at levels above their detection limits, the protein will not be detected. Additionally, even if a peptide is present at quantities above its detection limit, it still may not be detected if it coelutes from the LC column with other, more abundant peptides that are preferentially selected for fragmentation. It is also worth mentioning that although CENP-A is enriched at the centromere compared to other regions of the genome, it is estimated that CENP-A nucleosomes represent only ~4% of all centromeric nucleosomes and that a typical human centromere contains about 400 molecules of CENP-A.²¹ Thus, even within the centromere, histone H3 nucleosomes outnumber CENP-A nucleosomes by approximately 25:1, which means that HyCCAPP likely pulls down many more histone H3 than CENP-A nucleosomes (indeed, histone H3 was detected in all three alpha satellite and scrambled capture samples). On top of this, we must consider that histones are often decorated with a number of post-translational modifications, and indeed a search of CENP-A in the neXtProt database⁴⁰ reveals a number of annotated PTMs. The presence of multiple post-translationally modified forms of a base peptide sequence serves to “dilute” the signal of that peptide, thereby lowering the signal-to-noise ratio and making the peptide more difficult to detect. Taken together, these considerations present a plausible explanation for the absence of detected CENP-A peptides in HyCCAPP samples via mass spectrometry.

Besides CENP-A NAC and CAD proteins, many other known centromere-binding proteins were identified as enriched at alphoid DNA via HyCCAPP. Such proteins include CENP-B, -I, and -V. CENP-B in particular is a well-known alpha satellite-binding protein, and has been observed to bind to a specific 17 bp sequence, known as the CENP-B box, which is contained within a subset of alpha satellite repeats.⁶ Additionally, the four members of the chromosomal passenger complex (CPC), INCENP, CDCA8, BIRC5, and AURKB, were all enriched at the alpha satellite repeats, which makes sense given that the CPC is known to localize to the centromere during certain phases of mitosis.⁴¹ Multiple nuclear membrane-associated proteins, including LMNA, LMNB1, LMNB2, LBR, EMD, TMPO, and TOR1AIP1, were also identified as enriched at the alpha satellite repeats, which could be

explained by the observation that centromeres tend to localize near the nuclear periphery.⁴² Furthermore, two members of the heterochromatin protein 1 (HP1) family, HP1 α and HP1 β , were identified as enriched at the alpha satellite repeats, which validates prior evidence of interactions between these proteins and the centromere.^{43–45}

While it is clear from this discussion that the list of proteins identified by HyCCAPP as enriched at alphoid DNA reflects legitimate centromeric protein-DNA interactions, the results are further corroborated when compared to those from a related study performed in mouse embryonic stem cells.⁴⁶ In that study, a technique similar to HyCCAPP known as PICH (proteomics of isolated chromatin segments) was used to identify protein-DNA interactions occurring at mouse major satellite repeats. Similar to human cells, mouse centromeres are made up of repetitive satellite sequences. However, although alpha satellite DNA is found within both the centromeric and pericentromeric regions of human chromosomes, the mouse centromeric and pericentromeric regions are made up of two distinct types of satellite DNA. The 234 bp major satellite repeats make up the pericentromeric region, while the 120 bp minor satellite repeats make up the centromeric region.^{3,47} Using PICH, the authors identified 135 proteins as enriched at the mouse major satellites and divided those proteins into five main categories including “histone/DNA regulation and chromosome organization”, “DNA replication and cell cycle control”, “RNA processing/nucleolus”, “DNA damage”, and “lamina”. It is worth noting that the criteria used to classify a protein as enriched at the major satellites in that work were different than the criteria employed in this work. For the PICH experiments, proteins identified in major satellite pulldown experiments were compared to those identified in “input chromatin” samples. A protein was denoted as “pericentromere-enriched” if it was found only in the major satellite pulldown and not in the input chromatin sample in at least three of seven independent experiments or if, on average, three times more peptide counts were observed in the major satellite pulldown experiments than in input chromatin samples. This approach contrasts with the more rigorous statistical analysis of extracted ion chromatogram-based label-free quantification data described herein to identify alpha satellite-enriched proteins from HyCCAPP experiments.

Despite the differences in these two approaches, of the 135 proteins identified as enriched at the mouse major satellites, the human homologues of 39 were found in the list of 90 proteins identified as enriched at alphoid DNA via HyCCAPP. It was particularly interesting to note that two proteins, SMCHD1 and MKI67, whose interactions with the major satellites were suggested by PICH experiments and validated via immuno-FISH experiments, were also found in the list of alpha satellite-enriched proteins. The overlap between proteins identified at mouse major satellites by PICH and those identified at human alphoid DNA using HyCCAPP provides corroborating evidence for and strengthens confidence in the existence of these protein-DNA interactions. However, it is also interesting to consider proteins that were only found in one of these two experiments. For example, it was intriguing to find that, though HyCCAPP identified many CENP-A NAC and CAD proteins as enriched at the alpha satellite repeats, these proteins were entirely missing from the list of proteins identified as enriched at the mouse major satellites. This is likely due to the fact that the major satellites make up the pericentromeric region of mouse chromosomes, while the centromeric region is composed of minor satellites. Because NAC and CAD proteins are

recruited by CENP-A, which is primarily localized to the minor satellites in mouse cells, the absence of these proteins at the major satellites is expected. In human cells, however, where alphoid DNA is found both within the core, CENP-A-containing centromere as well as within the pericentromeric region, the identification of these proteins by HyCCAPP makes sense.

HyCCAPP identifies novel alpha satellite-binding proteins

In addition to many well-known centromere-binding proteins, HyCCAPP also identified a number of proteins as enriched at the alpha satellite repeats for which there is not, to our knowledge, prior evidence suggesting an interaction with the centromere (~20–40, depending on the criteria used for classification). These novel protein hits are exciting because their interactions with alpha satellite DNA may provide insight into additional, previously unsuspected factors important for centromere structure or function. Here, we discuss one such protein—LRIF1.

LRIF1 (ligand-dependent nuclear receptor-interacting factor 1) was found to be enriched at alpha satellite DNA with a fold-change of ~63 and a permutation-based *q*-value of 0. The LRIF1 protein had previously been identified in a yeast two-hybrid screen for potential retinoic acid receptor (RAR)-interacting proteins.⁴⁸ RARs are members of the steroid/nuclear receptor (NR) family of ligand-dependent transcription factors, which regulate genes important for a variety of essential biological processes.⁴⁸ In their screen for RAR-interacting proteins, Li et al. used RAR α as a bait protein to pull down LRIF1 and subsequently showed that LRIF1 also interacts with a number of other members of the NR family of transcription factors.⁴⁸ They further showed that LRIF1 is a nuclear protein, and that it represses RAR α -mediated transcriptional activation in a ligand-dependent manner, possibly through the recruitment of histone deacetylases. Additional studies showed that LRIF1 associates with HP1 α , HP1 β , HP1 γ , and SMCHD1 in human T-REx 293 cells and is concentrated on the inactive X chromosome in female human fibroblasts.⁴⁹ The inactive X chromosome (Xi), or Barr body, is an example of facultative heterochromatin and is transcriptionally silenced in mammalian female cells. LRIF1 has been observed to play an important role in Xi compaction, by linking the H3K9me3 and XIST-associated H3K27me3 regions of Xi via interactions with HP1 and SMCHD1, respectively.⁴⁹ Furthermore, a recent study in mouse cells showed that LRIF1, in conjunction with HP1 γ , can mediate the interaction between SMCHD1 and H3K9me3, potentially facilitating transcriptional repression.⁵⁰ Taken together, these studies provide evidence that LRIF1 may play a role in the formation and/or maintenance of heterochromatin.

It has been well-established that human centromeres are embedded within constitutive heterochromatin, and therefore the finding that LRIF1 is enriched at alpha satellite DNA is intriguing. On the basis of what is already known about LRIF1, it seems plausible that its enrichment on centromeric/pericentromeric alphoid DNA indicates a role for LRIF1 in the maintenance or formation of centromeric heterochromatin. This hypothesis is supported by the observation that HP1 α , HP1 β , and SMCHD1, proteins which appear to play a role in transcriptional silencing by LRIF1, were also found to be enriched at alpha satellite DNA via HyCCAPP. To validate the LRIF1-alpha satellite interaction suggested by HyCCAPP,

ChIP-qPCR experiments were performed to see whether alpha satellite DNA was enriched in LRIF1 immunoprecipitations compared to control IgG pulldowns. The ChIP-qPCR experiments revealed that approximately twice as much alpha satellite DNA was precipitated using an LRIF1 antibody compared to IgG (Fig. 5B). Such enrichment was not observed for 28S rDNA, a nonalphoid sequence, suggesting that LRIF1 may indeed preferentially localize to alpha satellite DNA (Fig. 5B). As a positive control, ChIP-qPCR experiments also showed substantial enrichment of alpha satellite DNA in CENP-B pulldown experiments as compared to the IgG control, but no enrichment of 28S rDNA, a sequence that CENP-B is not expected to bind (Fig. 5A).

It is worth noting that we attempted to validate a second novel alpha satellite-interacting protein suggested by HyCCAPP experiments, NASP, via ChIP-qPCR, but did not see enrichment of alphoid DNA in the NASP pulldown as compared to the IgG control (i.e., both NASP and IgG pulled down approximately the same amount of alpha satellite DNA; data not shown). Although we were unable to validate the NASP-alpha satellite interaction, these NASP ChIP-qPCR experiments did prove informative, as they showed that not all proteins show enrichment of alphoid DNA as compared to the IgG control. This provides support for our interpretation that the enrichment of alpha satellite DNA seen in the CENP-B and LRIF1 ChIP experiments reflects genuine protein-alpha satellite interactions.

Conclusion

HyCCAPP was used in this study to identify protein-alpha satellite interactions. This constitutes the first application of the HyCCAPP technology in mammalian cells and, to our knowledge, is the first DNA-centric study of human protein-alpha satellite interactions. Many known centromere-binding proteins were identified as enriched at the alpha satellite repeats, validating the approach and providing confidence in the results obtained. HyCCAPP also identified a number of novel alpha satellite-binding proteins, including LRIF1, a heterochromatin-associated protein. Further study will be required to elucidate the roles that the novel alpha satellite-binding proteins revealed in this work may play in centromere structure and/or function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Professor Emery H. Bresnick and members of his group for allowing us to use their cell culture facilities and for providing a stock of human K562 cells, Professor Ronald T. Raines for use of the cell disruptor, and Professor Eric R. Strieter for use of the Misonix sonicator. We are grateful to Yunxiang Dai for helpful conversations about HyCCAPP experiments.

Funding sources: This work was supported by the National Institutes of Health Center of Excellence in Genomic Science grant No. 1P50HG004952. K.E.B. was supported by the National Human Genome Research Institute grant to the Genomic Sciences Training Program, grant No. 5T32HG002760.

References

1. Willard HF. Evolution of alpha satellite. *Curr Opin Genet Dev.* 1991; 1:509–514. [PubMed: 1822282]
2. Aldrup-MacDonald ME, Sullivan BA. The past, present, and future of human centromere genomics. *Genes.* 2014; 5:33–50. [PubMed: 24683489]
3. Choo, KH. *The Centromere.* Oxford University Press; Oxford: 1997.
4. McKinley KL, Cheeseman IM. The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol.* 2015; 17:16–29. [PubMed: 26601620]
5. Eichler EE. Repetitive conundrums of centromere structure and function. *Hum Mol Genet.* 1999; 8:151–155. [PubMed: 9931322]
6. Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol.* 1989; 109:1963–1973. [PubMed: 2808515]
7. Ohzeki J, Nakano M, Okada T, Masumoto H. CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *J Cell Biol.* 2002; 159:765–775. [PubMed: 12460987]
8. Voullaire LE, Slater HR, Petrovic V, Choo KHA. A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *Am J Hum Genet.* 1993; 52:1153–1163. [PubMed: 7684888]
9. Fukagawa T, Earnshaw WC. The centromere: chromatin foundation for the kinetochore machinery. *Dev Cell.* 2014; 30:496–508. [PubMed: 25203206]
10. Burrack LS, Berman J. Neocentromeres and epigenetically inherited features of centromeres. *Chromosome Res.* 2012; 20:607–619. [PubMed: 22723125]
11. Sullivan BA, Karpen GH. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol.* 2004; 11:1076–1083. [PubMed: 15475964]
12. McAinsh AD, Meraldi P. The CCAN complex: linking centromere specification to control of kinetochore-microtubule dynamics. *Semin Cell Dev Biol.* 2011; 22:946–952. [PubMed: 22027615]
13. Bailey AO, Panchenko T, Shabanowitz J, Lehman SM, Bai DL, Hunt DF, Black BE, Foltz DR. Identification of the post-translational modifications present in centromeric chromatin. *Mol Cell Proteomics.* 2016; 15:918–931. [PubMed: 26685127]
14. Palmer DK, O'Day K, Wener MH, Andrews BS, Margolis RL. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J Cell Biol.* 1987; 104:805–815. [PubMed: 3558482]
15. Palmer DK, O'Day K, Trong HL, Charbonneau H, Margolis RL. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc Natl Acad Sci U S A.* 1991; 88:3734–3738. [PubMed: 2023923]
16. Yoda K, Ando S, Morishita S, Houmura K, Hashimoto K, Takeyasu K, Okazaki T. Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution *in vitro*. *Proc Natl Acad Sci U S A.* 2000; 97:7266–7271. [PubMed: 10840064]
17. Stoler S, Keith KC, Curnick KE, Fitzgerald-Hayes M. A mutation in *CSE4*, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev.* 1995; 9:573–586. [PubMed: 7698647]
18. Howman EV, Fowler KJ, Newson AJ, Redward S, MacDonald AC, Kalitsis P, Choo KHA. Early disruption of centromeric chromatin organization in centromere protein A (*Cenpa*) null mice. *Proc Natl Acad Sci U S A.* 2000; 97:1148–1153. [PubMed: 10655499]
19. Blower MD, Karpen GH. The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat Cell Biol.* 2001; 3:730–739. [PubMed: 11483958]
20. Foltz DR, Jansen LET, Black BE, Bailey AO, Yates JR III, Cleveland DW. The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol.* 2006; 8:458–469. [PubMed: 16622419]

21. Bodor DL, Mata JF, Sergeev M, David AF, Salimian KJ, Panchenko T, Cleveland DW, Black BE, Shah JV, Jansen LET. The quantitative architecture of centromeric chromatin. *eLife*. 2014; 3:e02137. [PubMed: 25027692]
22. Nakano M, Cardinale S, Noskov VN, Gassmann R, Vagnarelli P, Kandels-Lewis S, Larionov V, Earnshaw WC, Masumoto H. Inactivation of a human kinetochore by specific targeting of chromatin modifiers. *Dev Cell*. 2008; 14:507–522. [PubMed: 18410728]
23. Kennedy-Darling J, Guillen-Ahlers H, Shortreed MR, Scalf M, Frey BL, Kendzierski C, Olivier M, Gasch AP, Smith LM. Discovery of chromatin-associated proteins via sequence-specific capture and mass spectrometric protein identification in *Saccharomyces cerevisiae*. *J Proteome Res*. 2014; 13:3810–3825. [PubMed: 24999558]
24. Guillen-Ahlers H, Rao PK, Levenstein ME, Kennedy-Darling J, Perumalla DS, Jadhav AYL, Glenn JP, Ludwig-Kubinski A, Drigalenko E, Montoya MJ, Göring HH, Anderson CD, Scalf M, Gildersleeve HIS, Cole R, Greene AM, Oduro AK, Lazarova K, Cesnik AJ, Barfknecht J, Cirillo LA, Gasch AP, Shortreed MR, Smith LM, Olivier M. HyCCAPP as a tool to characterize promoter DNA-protein interactions in *Saccharomyces cerevisiae*. *Genomics*. 2016; 107:267–273. [PubMed: 27184763]
25. Choo KH, Vissel B, Nagy A, Earle E, Kalitsis P. A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. *Nucleic Acids Res*. 1991; 19:1179–1182. [PubMed: 2030938]
26. Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates JR III. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev*. 2013; 113:2343–2394. [PubMed: 23438204]
27. Kennedy-Darling J, Holden MT, Shortreed MR, Smith LM. Multiplexed programmable release of captured DNA. *ChemBioChem*. 2014; 15:2353–2356. [PubMed: 25157426]
28. Dai Y, Kennedy-Darling J, Shortreed MR, Scalf M, Gasch AP, Smith LM. Multiplexed sequence-specific capture of chromatin and mass spectrometric discovery of associated proteins. *Anal Chem*. 2017; doi: 10.1021/acs.analchem.7b01784
29. Erde J, Loo RRO, Loo JA. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J Proteome Res*. 2014; 13:1885–1895. [PubMed: 24552128]
30. Kennedy-Darling J, Smith LM. Measuring the formaldehyde protein-DNA cross-link reversal rate. *Anal Chem*. 2014; 86:5678–5681. [PubMed: 24848408]
31. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008; 26:1367–1372. [PubMed: 19029910]
32. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016; 13:731–740. [PubMed: 27348712]
33. Wu C-H, Chen S, Shortreed MR, Kreitinger GM, Yuan Y, Frey BL, Zhang Y, Mirza S, Cirillo LA, Olivier M, Smith LM. Sequence-specific capture of protein-DNA complexes for mass spectrometric protein identification. *PLoS One*. 2011; 6:e26217. [PubMed: 22028835]
34. Pinamonti S, Caruso A, Mazzeo V, Zebini E, Rossi A. DNA damage from pulsed sonication of human leukocytes *in vitro*. *IEEE Trans Ultrason Ferroelectr Freq Control*. 1986; 33:179–185. [PubMed: 18291769]
35. Elsner HI, Lindblad EB. Ultrasonic degradation of DNA. *DNA*. 1989; 8:697–701. [PubMed: 2693020]
36. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*. 2013; 8:1551–1566. [PubMed: 23868073]
37. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. 2017; 45:D183. [PubMed: 27899595]
38. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015; 43:D447. [PubMed: 25352553]

39. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13:2498–2504. [PubMed: 14597658]
40. Gaudet P, Michel PA, Zahn-Zabal M, Britan A, Cusin I, Domagalski M, Duek PD, Gateau A, Gleizes A, Hinard V, Rech de Laval V, Lin JJ, Nikitin F, Schaeffer M, Teixeira D, Lane L, Bairoch A. The neXtProt knowledgebase on human proteins: 2017 update. *Nucleic Acids Res.* 2017; 45:D177. [PubMed: 27899619]
41. Carmena M, Wheelock M, Funabiki H, Earnshaw WC. The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. *Nat Rev Mol Cell Biol.* 2012; 13:789–803. [PubMed: 23175282]
42. Weierich C, Brero A, Stein S, von Hase J, Cremer C, Cremer T, Solovei I. Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes. *Chromosome Res.* 2003; 11:485–502. [PubMed: 12971724]
43. Wreggett KA, Hill F, James PS, Hutchings A, Butcher GW, Singh PB. A mammalian homologue of *Drosophila* heterochromatin protein 1 (HP1) is a component of constitutive heterochromatin. *Cytogenet Genome Res.* 1994; 66:99–103.
44. Minc E, Allory Y, Worman HJ, Courvalin J-C, Buendia B. Localization and phosphorylation of HP1 proteins during the cell cycle in mammalian cells. *Chromosoma.* 1999; 108:220–234. [PubMed: 10460410]
45. Minc E, Courvalin J-C, Buendia B. HP1 γ associates with euchromatin and heterochromatin in mammalian nuclei and chromosomes. *Cytogenet Genome Res.* 2000; 90:279–284.
46. Saksouk N, Barth TK, Ziegler-Birling C, Olova N, Nowak A, Rey E, Mateos-Langerak J, Urbach S, Reik W, Torres-Padilla M-E, Imhof A, Déjardin J. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Mol Cell.* 2014; 56:580–594. [PubMed: 25457167]
47. Joseph A, Mitchell AR, Miller OJ. The organization of the mouse satellite DNA at centromeres. *Exp Cell Res.* 1989; 183:494–500. [PubMed: 2767161]
48. Li HJ, Haque ZK, Chen A, Mendelsohn M. RIF-1, a novel nuclear receptor corepressor that associates with the nuclear matrix. *J Cell Biochem.* 2007; 102:1021–1035. [PubMed: 17455211]
49. Nozawa R-S, Nagao K, Igami K-T, Shibata S, Shirai N, Nozaki N, Sado T, Kimura H, Obuse C. Human inactive $\times$ chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nat Struct Mol Biol.* 2013; 20:566–573. [PubMed: 23542155]
50. Brideau NJ, Coker H, Gendrel A-V, Siebert CA, Bezstarosti K, Demmers J, Poot RA, Nesterova TB, Brockdorff N. Independent mechanisms target SMCHD1 to trimethylated histone H3-modified chromatin and the inactive $\times$ chromosome. *Mol Cell Biol.* 2015; 35:4053–4068. [PubMed: 26391951]

Fig. 1.

Alpha satellite repeats are found in the centromeric/pericentromeric region of every human chromosome. Individual alpha satellite monomers (depicted here as small arrows) are 171 bp in length and are relatively heterogeneous in sequence (heterogeneity is represented by different colors). These monomers are arranged in a tandem, head-to-tail fashion within the centromere, forming higher-order repeat (HOR) structures (depicted here as large arrows). HORs show much more sequence homogeneity than their individual monomeric constituents and are themselves repeated to form large alpha satellite arrays within the centromeric region. The HOR array is flanked by a pericentromeric region containing disordered alpha satellite monomers. HOR and HOR array size estimates for this figure are from Aldrup-MacDonald and Sullivan.²

Fig. 2.

qPCR was used to measure the capture efficiencies of alphoid DNA and an off-target sequence (28S rDNA) in both the alpha satellite capture and scrambled capture samples. These measurements were made for all three HyCCAPP experiments, and the labels denote average values. Error bars represent +1 standard deviation of the three experimental replicates. Note that the average capture efficiency of 28S rDNA in the scrambled capture samples could not be calculated because the concentration of 28S rDNA in scrambled capture samples was too low to be accurately quantified (NQ = not quantified).

Fig. 3.

Volcano plot showing the results of *t* tests performed to compare protein abundances between the alpha satellite capture and scrambled capture samples. Data points represent individual proteins. Red points indicate those proteins that were found to be significantly enriched either in the alpha satellite or scrambled capture samples, with a permutation-based FDR = 5% and a minimum fold-change of 5.8. Subject to these criteria, 90 proteins were found to be significantly enriched at the alpha satellite repeats. The data used to generate this volcano plot, including the identity of each protein and its associated *p*-value, permutation-based *q*-value, and fold-change, can be found in Supplementary Table S-2.

Fig. 4.

Interaction network for the 90 alpha satellite-enriched proteins identified through HyCCAPP experiments. The network was created in Cytoscape (version 3.5.1),³⁹ using interactions identified by STRING (version 10.0).³⁸ Textmining, experiments, and databases were used as sources of evidence. The network reflects interactions that STRING identified with at least medium confidence, and the thickness of the line connecting two nodes reflects the strength of the data supporting that interaction. Proteins with selected GO terms are outlined to aid in interpretation of the network, and associated Bonferroni-corrected p -values for GO term enrichment are indicated.

Fig. 5.

ChIP-qPCR experiments were performed using antibodies against CENP-B, a known alpha satellite-binding protein (panel A), as well as LRIF1, a novel alpha satellite interactor suggested by HyCCAPP (panel B). IgG was used as a control for the background level of the immunoprecipitation. The DNA recovered from each ChIP experiment was analyzed using two qPCR assays designed to target alpha satellite DNA (α -SL 1 and 2) and one qPCR assay designed to target a nonalphoid sequence (28S rDNA). The amount of each sequence recovered is plotted as a percentage of the input chromatin material. Average values of two experimental replicates are plotted, and error bars represent ± 1 standard deviation.

Table 1

Selection of GO terms over-represented in the list of 90 alpha satellite-enriched proteins identified by HyCCAPP. The PANTHER over-representation tests generated a list of 180 over-represented GO terms with Bonferroni-corrected p -values less than 0.05. This table shows the 30 GO terms from that list with corrected p -values less than 3.5×10^{-24} .

GO Term	p -value	GO Term	p -value
Chromosome	1.53×10^{-64}	Organelle organization	7.15×10^{-31}
Chromosomal region	7.25×10^{-57}	Cell cycle	1.12×10^{-30}
Chromosomal part	7.26×10^{-53}	Condensed chromosome	1.12×10^{-30}
Chromosome organization	6.39×10^{-48}	Nuclear chromosome	9.51×10^{-30}
Nuclear part	2.20×10^{-41}	Nucleus	3.21×10^{-28}
Nuclear lumen	1.26×10^{-38}	Sister chromatid segregation	4.01×10^{-28}
Intracellular non-membrane-bounded organelle	6.21×10^{-38}	Nuclear chromosome part	6.42×10^{-28}
Non-membrane-bounded organelle	6.21×10^{-38}	Protein-DNA complex assembly	2.72×10^{-26}
DNA conformation change	9.82×10^{-36}	Sister chromatid cohesion	4.60×10^{-26}
Chromosome, centromeric region	2.23×10^{-35}	Nuclear chromosome segregation	4.75×10^{-26}
Organelle lumen	8.22×10^{-34}	DNA packaging	9.19×10^{-26}
Intracellular organelle lumen	8.22×10^{-34}	Intracellular organelle part	4.28×10^{-25}
Membrane-enclosed lumen	8.22×10^{-34}	Protein-DNA complex subunit organization	8.32×10^{-25}
Cell cycle process	3.78×10^{-32}	Chromosome segregation	2.39×10^{-24}
Nucleoplasm	6.28×10^{-31}	Organelle part	3.04×10^{-24}