A long non-coding RNA is required for targeting centromeric protein A to the human centromere

Delphine Quénet and Yamini Dalal*

Laboratory of Receptor Biology & Gene Expression,
Center for Cancer Research,
National Cancer Institute,
National Institutes of Health,
Building 41, Room D804B, 41 Library Drive MSC 5055, Bethesda, MD 20892,
USA

*Corresponding author: Yamini Dalal

Phone: 301-496-9247
Fax: 301-496-4951
Email: dalaly@mail.nih.gov

Keywords: CENP-A, HJURP, transcription, centromere, centromeric RNAs

Abbreviations: CENP-A, Centromeric Protein A; HJURP, Holliday Junction Recognition Protein; ncRNA, non-coding RNA; RNAPII, RNA Polymerase II; TBP: TATA-box Binding Protein.
Abstract

The centromere is a specialized chromatin region marked by the histone H3 variant CENP-A. Although active centromeric transcription has been documented for over a decade, the role of centromeric transcription or transcripts has been elusive. Here, we report that centromeric α-satellite transcription is dependent on RNA Polymerase II and occurs at late mitosis into early G1, concurrent with the timing of new CENP-A assembly. Inhibition of RNA Polymerase II-dependent transcription abrogates the recruitment of CENP-A and its chaperone HJURP to native human centromeres. Biochemical characterization of CENP-A associated RNAs reveals a 1.3 kb molecule which originates from centromeres, which physically interacts with the soluble pre-assembly HJURP/CENP-A complex in vivo, and whose down-regulation leads to the loss of CENP-A and HJURP at centromeres. This study describes a novel function for human centromeric long non-coding RNAs in the recruitment of HJURP and CENP-A, implicating RNA-based chaperone targeting in histone variant assembly.
Introduction

Specialized chromatin domains called centromeres play an essential role in chromosome segregation, serving as a platform for kinetochore complex formation, which in turn binds spindle microtubules at mitosis [1]. Although centromeric DNA sequence is not uniform across species, centromere function is conserved [2]. In humans, AT-rich 171 bp α-satellite repeats lacking any known genes are the primary DNA component of the centromere [3]. Centromeres are characterized by the presence of the centromeric histone H3 variant (CENH3/CENP-A in human). Centromeric chromatin has long been considered heterochromatic, despite exhibiting a bivalent organization with heterochromatin-like post-translational modifications (PTMs), such as H3 and H4 hypo-acetylation, and transcription-coupled PTMs, including dimethylation on H3 lysine 4 (H3K4me2) [4-6]. The function of such bivalent modifications has remained mysterious. Indeed, despite the common assumption that centromeres are largely dormant, a number of recent studies have pointed to the importance of transcription at centromeres in multiple organisms, which appears to be essential for the maintenance of centromere integrity [7]. Chemical inhibition of either RNA Polymerase I (RNAPI), or RNA Polymerase II (RNAPII), results in loss of the inner kinetochore protein CENP-C, and in chromosome mis-segregation [8,9]. Centromeric RNA components also seem to contribute to the structural integrity of the mitotic centromere [8]. However, the exact timing of centromeric transcription, the polymerase involved, the identity of centromeric RNAs and their...
precise role in maintaining native centromere integrity in human cells has been elusive.

In this study, we report that centromeric RNAs play a critical role in the maintenance of the human centromere \textit{in vivo}. Using chromatin immunoprecipitation (IP), and immuno-fluorescence (IF) on chromatin fibers, we find that RNAPII, in conjunction with TATA-box Binding Protein (TBP) localizes to, and actively transcribes native human centromeres from late mitosis to early G1 (eG1). Biochemical purification and sequencing of the RNA associated with human centromeric chromatin at eG1 reveals a 1.3 kb long transcript. This RNA physically interacts with CENP-A and its chaperone HJURP (Holliday junction recognition protein) in the pre-assembly soluble state \textit{in vivo}. Targeted sequence-specific knockdown of the transcript results in the formation of multipolar spindles and lagging chromosomes in subsequent mitoses, leading to chromosome instability. IF analysis of centromeric chromatin fibers demonstrates that these cellular and nuclear phenotypes arise specifically from the abrogation of CENP-A and HJURP localization to the centromere. Together, our data describe a direct physical role for a centromeric long non-coding RNA (lncRNA) in HJURP targeting, subsequent CENP-A loading, and the maintenance of centromere integrity. Our study supports the possibility that an lncRNA-based mechanism is involved in targeting CENP-A and its chaperone HJURP to the centromere.

RESULTS
RNAPII is associated with native human centromeres at eG1.

Centromeric transcription has been previously described in human cells, and RNAPII has been implicated in this process [8-11]. To investigate the timing of centromeric transcription, we used synchronized HeLa cells at G2, eG1 and G1/S to track the activated form of RNAPII (i.e., serine 2 phosphorylated, RNAPII\(^{S2P}\)) on centromeric chromatin fibers throughout the cell cycle by IF (Figure 1 - figure supplement 1). RNAPII\(^{S2P}\) co-localizes with the inner kinetochore protein CENP-B and centromeric \(\alpha\)-satellite DNA specifically at eG1 (Figure 1A and Figure 1 - figure supplement 2A). We also noted that TBP, a partner of RNAPII normally involved in transcription initiation [12], is localized on eG1 CENP-A-rich fibers (Figure 1A). These data suggest that centromeres are actively transcribed by RNAPII machinery at eG1.

We next sought to establish whether there was a physical interaction between CENP-A chromatin and RNAPII. In order to achieve this, we extracted chromatin from non-synchronized cells after a short MNase digestion, to obtain long chromatin arrays that are rich in tri-, tetra-, and penta-nucleosomes (Figure 1 - figure supplement 2B). From this input chromatin, centromeric chromatin was immuno-precipitated with specific antibodies against either CENP-A, or the inner kinetochore protein CENP-C or no antibody (mock IP). The mock IP control shows no enrichment of any of the centromeric proteins tested (Figure 1 - figure supplement 2C). As expected, Western blots revealed reciprocal co-purification of CENP-A and CENP-C (Figure 1B, left and middle panels). RNAPII and its partner TBP also co-purified with CENP-A and CENP-C (Figure 1B, left and
middle panels). To further establish an interaction between RNAPII and centromeric proteins, we performed the reciprocal experiment, precipitating RNAPII$^{S2P}$ from solubilized chromatin, and testing for centromeric partners. While Western blots revealed little or no interaction of RNAPII with CENP-C, a robust and reproducible binding of RNAPII to TBP and CENP-A was observed (Figure 1B, right panel). Thus, active RNAPII machinery is physically associated with CENP-A chromatin at eG1.

**RNAPII transcribes native human centromeres at late mitosis-eG1.**

Recent studies have indicated that RNAPII transcribes centromeres at mitosis [9]. However, our RNAPII localization data above showed RNAPII enrichment occurs primarily at eG1 (Figure 1A). To examine the consequence of eG1 RNAPII localization at centromeres, HeLa cells were synchronized at G2, mitosis and eG1, and briefly treated (2 hrs) with drugs to specifically block either RNAPI (actinomycin D) or RNAPII (α-amanitin) (Figure 1 - figure supplement 1) [13]. After RNA extraction and retro-transcription, we determined the expression of control genes and centromeric α-satellite repeats by semi-quantitative PCR. As expected, actinomycin D and α-amanitin inhibited transcription of target genes of RNAPI (e.g., 18S rRNA) or RNAPII (e.g., GAPDH), respectively (Figure 2, left and middle graphs). Compared to non-treated (NT) conditions, actinomycin D treatment or α-amanitin treatment in G2 phase had no impact on centromeric α-satellite expression (Figure 2, right graph). Consistent with a previous study [9], RNAPII inhibition in mitotic cells revealed a decrease (15.6%) in centromeric
transcripts (Figure 2, right graph). However, when RNAP II was blocked in eG1, a larger reduction (35.1%) was observed (Figure 2, right graph). These results suggest that RNAP II transcribes centromeres not solely at mitosis, but also throughout eG1.

Centromeric transcription at eG1 is required for HJURP and CENP-A targeting to the centromere.

The synchrony of centromeric transcription and CENP-A recruitment onto centromeres at late mitosis-eG1 led us to examine whether active transcription is required for CENP-A loading. To test this hypothesis, we briefly treated eG1-synchronized cells with α-amanitin (2 hrs) to block RNAP II activity as above, and quantified potential changes in intensity for CENP-A or CENP-B IF signal using ImageJ. Consistent with its role as a constitutive centromeric DNA-binding protein [1], CENP-B staining intensity was heterogeneous (Figure 3B, Figure 3-figure supplement 1A), but identical in both NT and α-amanitin treated cells (Figures 3A, 3B and Supplemental File 1), demonstrating that its localization is independent of centromeric transcription. Whereas punctate CENP-A spots can be seen under both conditions, when RNAP II was blocked, the intensity of the CENP-A signal was decreased by ~50% in cells (Figures 3A, 3B and Supplemental File 1). To ensure that this decrease was not due to reduced expression of CENP-A or its chaperone HJURP, total levels of both proteins were quantified by Western blot. No noticeable changes in protein levels were detected upon α-amanitin treatment (Figure 3 - figure supplement 1B). These
data indicate that equal amounts of CENP-A and HJURP were available at eG1, but potentially unable to load at the centromere. The decrease of CENP-A signal at centromeres during eG1 after RNAPII inhibition might be due to either the loss of pre-existing CENP-A, or a defect in targeting of newly synthesized CENP-A by its chaperone HJURP. To discriminate between these two hypotheses, we analyzed the localization of CENP-A, CENP-B, RNAPII S2P and HJURP on chromatin fibers, with or without α-amanitin treatment at eG1. Consistent with the known effect of α-amanitin on RNAPII (i.e., blocking RNAPII elongation without inducing the release of RNAPII) [13,14], inhibition of transcription did not affect RNAPII S2P localization onto the centromeric fiber (Figure 3C, first top panel). Similarly, CENP-B localization onto centromeric chromatin fibers remained unaffected as well (Figure 3C, first panel). Consistent with the whole cell data presented above (Figures 3A and 3B), after RNAPII inhibition, not only were CENP-A signals reduced on chromatin fibers (Figure 3C, second panel), HJURP was almost completely lost (Figure 3C, second panel). The loss of HJURP (Figure 3C) combined with the ~50% decrease of CENP-A signal after α-amanitin treatment (Figure 3B), suggests that RNAPII-dependent transcription is required for the targeting of HJURP, and for the subsequent loading of new CENP-A to the centromeric chromatin fiber at eG1.

A 1.3 kb centromeric RNA binds the soluble HJURP/CENP-A pre-assembly complex at eG1.
Previous data have documented the existence of ncRNA at centromeres in multiple species [7]. In humans, no genes have been annotated within native centromeres, suggesting a transcription event at α-satellite DNA repeats most likely leads to the synthesis of ncRNAs. In order to characterize potential centromeric transcripts, we sought to purify them biochemically. Total RNAs were extracted from cells, DNase I treated to remove genomic contamination, separated on denaturing gels, transferred to Northern blots, and subjected to hybridization with radiolabelled centromeric α-satellite probes, in order to reveal potential complementary transcripts. Northern blots revealed a unique centromeric RNA species migrating at approximately ~1.3 kb (Figure 4A and Figure 4 - figure supplement 1A). Control experiments were performed to exclude the possibility of trace genomic DNA contamination contributing to the 1.3 kb band. Treatment of RNA samples with RNase A (Figure 4 - figure supplement 1B), or purification of RNA from cells treated with α-amanitin (Figure 4 - figure supplement 1C), both demonstrated absence of the 1.3 kb band on Northern blots, supporting the interpretation that the 1.3 kb band derives solely from an RNA species.

The inhibition of transcription was accompanied by loss of CENP-A and HJURP at the centromere during eG1 (Figure 3C), and our results above (Figure 4A) supported the possibility of a unique RNA species present at centromeres in eG1. A logical prediction arising from these data is that centromeric transcripts might physically associate with the soluble pre-assembly HJURP/CENP-A complex \textit{in vivo}. Indeed, computational RNA-binding prediction algorithms
revealed potential RNA binding residues in both HJURP and CENP-A (Figure 4 - figure supplement 2) [15]. Thus, to further test this hypothesis, we probed for physical interactions between CENP-A and its chaperone HJURP with centromeric α-satellite transcripts. After a brief MNase digestion of eG1-synchronized cells, we immuno-precipitated CENP-A and HJURP from both, the soluble fraction (composed of free histones and nuclear factors, SF), and the chromatin fraction (composed of chromatin and associated complexes, CF) (Experimental Scheme, Figure 4 – figure supplement 3). CENP-A and HJURP complexes were immuno-precipitated from SF and CF. Mock IPs pulled down neither CENP-A nor HJURP (Figure 4 - figure supplement 4A). Consistent with HJURP chaperoning CENP-A at eG1 [16-18], these proteins co-purified from both SF and CF (Figure 4 - figure supplement 4A). From these IPs, RNAs were purified, electrophoresed, transferred to Northern blots, and subsequently hybridized to the same radiolabelled centromeric α-satellite probes as above (Figure 4A). These Northern blots revealed no RNA signal in the mock IP (Figure 4 – figure supplement 4B). In contrast, the 1.3 kb RNA is physically associated with CENP-A in both SF and CF (Figure 4B), and interacts with HJURP only in the SF (Figure 4B). These data provide evidence that the 1.3 kb centromeric RNA physically associates with the soluble HJURP/CENP-A pre-assembly complex at eG1.

**CENP-A associated RNA localizes to centromeres.**
We next sought to purify, clone and sequence CENP-A-associated RNA (Experimental Scheme, Figure 4 – figure supplement 3). This sequencing approach was moderately successful, yielding one sequence of ~665 nucleotides (cenRNA#1, Figure 5 - figure supplement 1). This RNA sequence is unique and contains four semi-regular spaced 28 bp repeats with a weak homology to the canonical CENP-B box (data not shown), but does not map to the currently annotated human genome sequence, to any other organisms, or to plasmids. Since most human centromeres remain intractable to complete sequencing [19], this outcome is not entirely surprising. We were unable to detect cenRNA#1 directly on Northern blots, presumably due to its low abundance or due to instability. Therefore, to directly reveal the genomic origin of the sequenced RNA, we converted it to a DNA FISH probe (cenRNA#1DNA). We next either performed control IF/FISH using the Xist locus (XistDNA, on the q-arm of the X-chromosome) and CENP-A, IF/FISH with cenRNA#1DNA and CENP-A, or IF/double-FISH with CENP-A, cenRNA#1DNA and centromeric α-satellite probes, on chromatin fibers. As expected, control IF/FISH between XistDNA and CENP-A yielded no appreciable signal (Figure 5A). In contrast, about half the CENP-A, or centromeric α-satellite stained fibers, contain the cenRNA#1DNA signal (Figure 5B, left versus right, quantification below panel). This result is consistent with the fact that centromeric α-satellite DNA sequences are not strictly conserved across all human centromeres [19]. Significantly, 100% of the cenRNA#1DNA was found associated with CENP-A (Figure 5B, top right panel), and with centromeric α-satellite DNA probe (Figure 5B, lower right panel), demonstrating robust
association on centromeric chromatin fibers. Thus, these data (Figure 5) indicate that the sequenced lncRNA (Figure 5 - figure supplement 1), which is associated with CENP-A at eG1 (Figure 4B), indeed has a centromeric origin.

Next, we assessed the functional role of cenRNA#1 by an shRNA strategy to down-regulate specifically its expression (Figure 5 – figure supplement 1). Cells were transfected with control scrambled (shRNA\textsuperscript{scram}) or shRNA\textsuperscript{cenRNA#1} constructs, and selected with puromycin. Cells transfected with shRNA\textsuperscript{scram} displayed no changes in morphology and density, whereas cells treated with shRNA\textsuperscript{cenRNA#1} displayed significant loss of cell density (down by ~70%, relative to control), and presented aberrant morphology (Figure 5 – figure supplement 2A and 2B). Altogether these data suggest that cenRNA#1 transcript significantly affected cell integrity.

Targeting destruction of centromeric $\alpha$-satellite transcripts results in severe mitotic defects.

We were curious whether it was the act of centromeric transcription alone, or the product of transcription (i.e., centromeric RNAs), that was necessary for HJURP and CENP-A targeting to the centromere at eG1. To distinguish between these hypotheses, we further examined functional consequences arising from the targeted loss of total centromeric $\alpha$-satellite RNA, without inhibiting RNAPII transcription.
Using the centromeric α-satellite consensus sequence [3], we designed two
shRNAs targeting α-satellite sequences (shRNA\textsuperscript{sat1} and shRNA\textsuperscript{sat2}) to destroy
centromeric transcripts (Figure 6 - supplement 1A).

At six days post-transfection with the control scrambled (shRNA\textsuperscript{scram}) or shRNA\textsuperscript{sat}
constructs and puromycin selection, the expression level of centromeric α-
satellite transcript was analyzed by qPCR. Compared to control shRNA\textsuperscript{scram},
cells transfected with shRNA\textsuperscript{sat} constructs displayed a significant decrease
(~70%) of the centromeric α-satellite transcript (Figure 6 - figure supplement
1B), confirming targeted destruction was accomplished.

Evaluation of cell morphology by phase contrast microscopy revealed that
shRNA\textsuperscript{sat} transfected cells were less dense (down by ~70%, relative to control)
and exhibited abnormal morphology (Figure 6A). Phenotypes included a large
and flat cytoplasm, cellular protrusions, and multinucleate cells (Figure 6A). To
better elucidate cell defects, we stained with β-actin, which revealed cells with
several nuclei and atypical shape (Figure 6B). Reduced cell density and
morphological abnormalities could result from cells exiting the replicative cell
cycle and undergoing senescence [20]. We performed β-galactosidase staining
to test for senescence [21]. Senescent BJ cells were used as positive control,
and displayed the expected blue color after the β-galactosidase assay (Figure 6 -
figure supplement 2). However, no significant increase in senescence was seen
in either shRNA\textsuperscript{scram} or shRNA\textsuperscript{sat} transfected cells (Figure 6 - figure supplement
2).
The kinds of cellular morphological changes observed above (Figures 6A and 6B) have previously been linked to defects in cell division, specifically in mitosis [22]. To test this alternative possibility, shRNA transfected cells were synchronized at mitosis, and at six days post-transfection, stained for markers of mitotic spindles (α-tubulin) and centromeres (CENP-B), respectively. shRNA\textsuperscript{scram} transfected cells displayed normal mitotic structures (Figure 6C). In contrast, almost half (42.2%) of shRNA\textsuperscript{sat} transfected cells presented abnormal mitoses, with multipolar spindles and lagging chromosomes (Figure 6C).

**Targeting destruction of centromeric α-satellite transcripts results in abrogation of CENP-A and HJURP targeting at eG1.**

A mechanistic explanation for the observed mitotic aberrances in the centromeric transcript depleted cells (Figure 6C) could be loss of centromere integrity, driven by deficient targeting of HJURP/CENP-A complexes to centromeres. We were curious whether the loss of centromeric transcripts directly abrogated CENP-A and HJURP localization at centromere. Because loss of centromeric transcripts resulted in reduced cell density and mitotic defects (Figure 6), there were insufficient cells for biochemical experiments. Thus, we turned to chromatin fibers to investigate this issue. After synchronization at eG1, chromatin fibers were isolated from shRNA\textsuperscript{scram} cells, or the few remaining of shRNA\textsuperscript{sat} transfected cells, and stained for RNAPII and centromeric proteins. In both, shRNA\textsuperscript{scram} or shRNA\textsuperscript{sat} transfected cells, RNAPII\textsuperscript{S2P} remains associated with centromeric chromatin fibers (Figure 7, first top panel), confirming that RNAPII localization
and transcription of centromeres are independent of centromeric transcripts.

Additionally, CENP-B and CENP-C localization was also unaffected at centromeres (Figure 7, second and third panels). In contrast, on centromeric fibers from shRNA\textsuperscript{sat} transfected cells, CENP-A and HJURP were barely detectable (Figure 7, fourth and fifth panels). These data suggest that at eG1, CENP-A targeting through its chaperone HJURP is dependent not just on active transcription itself, nor on processes that facilitate centromeric transcription [23,24], but specifically requires the presence of centromeric transcripts. The long-term loss of these centromeric transcripts leads to mitotic defects (Figure 6), which are physically underpinned by the loss of HJURP recruitment and CENP-A loading (Figure 7) to native centromeres.

**DISCUSSION**

Active transcription is thought to be essential for centromere structure and function [8-11]. In this study, we investigated the mechanistic contribution of transcription, and centromeric transcripts, to centromeric integrity. We show that RNAPII and TBP are loaded onto and transcribe human centromeric chromatin at eG1 (Figure 1). This cell cycle regulated centromeric transcription is required for the synthesis of centromeric RNAs (Figure 2). Biochemical purification and analysis reveal a 1.3 kb transcript which is physically associated with CENP-A and HJURP in the soluble pre-assembly state (Figure 4). Targeted destruction of this centromeric RNA leads to the loss of centromere integrity and subsequent mitotic and cellular defects (Figure 6), which are mechanistically underpinned by
the loss of HJURP and CENP-A recruitment to centromeres at eG1 (Figure 7).

Altogether, these data reveal a hitherto unsuspected function for IncRNAs in RNA-dependent chaperone targeting to centromeres in human cells (Figure 7 – figure supplement 1).

Several specific questions arise from our observations. First, no active genes have ever been described in human centromeres, making the identification of RNAPII (Figure 1 and [11]) and TBP at the centromere surprising. Because the repetitive nature of centromeric α-satellite DNA has thus far disallowed complete sequencing, successful annotation of transcriptional motifs that may exist in human centromeres remains to be accomplished. Our data show that centromeric transcription is an event integral to the epigenetic maintenance of centromere integrity, and discovering precisely where such motifs lie within active centromeres is an exciting avenue of research.

Second, our functional characterization of a 1.3 kb centromeric IncRNA deriving from centromeric transcription reveals its interaction with CENP-A and HJURP at eG1. Specific depletion of centromeric α-satellite transcripts affects the recruitment of both CENP-A and HJURP proteins, directly implicating centromeric RNA in CENP-A and HJURP targeting onto centromeres at eG1. We note that previous data have shown that HJURP loading is also dependent on its interaction with the Mis18 complex, in a CDK1-dependent manner [23-26]. However, mutations of CDK1-phosphorylated sites in HJURP only partially abrogate its recruitment *in vivo*, highlighting the existence of other hitherto
unknown CENP-A loading factors [23]. We speculate centromeric IncRNAs are, in fact, the missing factor. Thus, of immediate interest is the elucidation of the structure of the 1.3 kb centromeric RNA with its cognate binding domains in CENP-A and HJURP.

Third, the exact molecular process involved in targeting IncRNA-nucleoprotein complexes to centromeres is an unexpected and novel avenue to pursue. For example, whereas it is well known that Xist RNA binds its cognate DNA locus only in cis [27], it is unknown if centromeric transcripts can bind in cis solely to the centromere of origin, or in trans, across all centromeres. An attractive possibility is centromeric RNAs originate from multiple centromeres, and serve a redundant function to ensure accurate targeting of CENP-A/HJURP to homologous centromeres.

Fourth, our study has potential evolutionary implications. Prior studies have described RNA originating from centromeres in multiple species [28,29]. In mouse cells, a 120 nucleotides minor satellite RNA is associated with centromeres [29], and, in tammar wallaby, centromeric transcription results in the production of ~40 nucleotides crasiRNAs (centromere repeats-associated short interacting RNAs) [22,28,30]. A logical explanation for the difference in size of ncRNAs generated in different organisms may be the divergent nature of the centromeric DNA sequences across species, which in turn may lead to divergence in the type of centromeric RNAs produced. However, despite this difference, over-expression or down-regulation of mouse minor satellite RNA, or crasiRNAs in tammar wallaby, or the 1.3 kb centromeric human RNA identified in
our study, leads to similar cellular and mitotic defects. Our data reveals that such RNAs generated from human centromeric transcription bind HJURP and CENP-A in the soluble form, and that mitotic loss seen in cells depleted of these IncRNAs is specifically linked to abrogation of HJURP-mediated targeting of CENP-A. Thus, our data suggest an evolutionarily conserved basis for the phenomena of centromeric transcription seen in other organisms. We speculate that accurate CENP-A targeting onto active centromeres probably requires a “dual-lock” system, coupling chromatin-bound centromeric factors (such as Mis18), which facilitate cell-cycle regulated centromeric transcription, which in turn results in the production of a lncRNA/CENP-A/chaperone complex that can effectively target CENP-A back to pre-existing active centromeric sites (Figure 7 – figure supplement 1).

It is noteworthy that transcription-coupled, chaperone-mediated histone variant assembly governs much of chromatin biology. Our report potentially reveals an RNA-based mechanism by which specialized histone-variant driven chromatin structures might be maintained \textit{in vivo}. 
Materials and Methods

Antibodies

Antibodies are commercially available, except the custom CENP-A antibody (available upon request) used for CENP-A detection on Western blot. Supplemental File 2 lists all antibodies used for each experiment.

Cell culture and RNA Polymerase inhibition

HeLa cells were grown at 37°C in a humidified atmosphere containing 5% CO₂, in Dulbecco’s modified Eagle’s medium high glucose and L-glutamine (#11965, Gibco, Grand Island, NY) supplemented with 10% Fetal Bovine Serum (#26140-079, Gibco), and 1X Pen/Strep solution (#10378-016, Gibco). All synchronizations were done by double thymidine block (0.5 mM, #T9250, Sigma-Aldrich, Saint Louis, MO). After a first block of 19 hrs, cells were released for 9 hrs, followed by a second thymidine block of 16 hrs. Cells then were released for the appropriate time (9 hrs for G2, 10 hrs for mitosis and 11 hrs for eG1, Figure 1 - Figure supplement 1). Synchronization was assessed by flow cytometry. Cells were stained with propidium iodide and analyzed on a FACScalibur (Becton Dickinson, San Jose, CA). Synchronized cells were treated with either 0.2 µg/mL of actinomycin D (#A2263, Sigma-Aldrich) or 2 µg/mL of α-amanitin (#A1410, Sigma-Aldrich) to analyze the effect of RNAPI and RNAPII inhibition, respectively, on centromere transcription.

RNA extraction, retro-transcription and Polymerase Chain Reaction (PCR)
RNAs were extracted by Trizol reagent (#15596-026, Ambion, Grand Island, NY) according to manufacturer protocol. Briefly, cells were resuspended in Trizol, and following 5 min incubation at room temperature (RT), 200 µL of chloroform was added. After centrifugation at 12000 rpm for 15 min at 4°C, the clear phase was mixed with 500 µL of isopropanol, and centrifuged. The pellet was washed with 75% ethanol, and resuspended in water complemented with DNase I buffer, DNase I (#M0303, New England Biolabs - NEB, Ipswich, MA) and RNase inhibitor (#M0314, NEB) to avoid genomic DNA contamination. After incubation for 30 min at 37°C, the DNase I activity was inhibited by addition of 5 mM EDTA and incubation at 65°C for 10 min. RNAs were purified a second time by phenol:chloroform:isoamyl alcohol (25:24:1, #AC327115000, Acros Organics, Pittsburgh, PA) method and ethanol precipitated. RNAs were conserved at -80°C until further analysis.

After quantification by UV-spectroscopy (230, 260 and 280 nm) and verification of RNA quality on 1.5% agarose gel, equivalent concentrations of RNA were subjected to retro-transcription, using the SuperScript III First-Strand Synthesis System with random hexamer primers (#18080-051, Invitrogen, Grand Island, NY), and amplified with Takara PCR kit (#RR001B, Clontech Laboratories Inc., Mountain View, CA). Control reactions without the reverse transcriptase or complementary DNA were performed to rule out DNA contamination and non-specific amplification, respectively. Primer sequences are included in Supplemental File 3. PCR conditions were [3 min 94°C; (10 s 98°C, 30 s 57°C, 30s 72°C) 30 cycles; 5 min 72°C] for GAPDH and centromeric α-satellite, [3 min
94°C; (10 s 98°C, 30 s 52°C, 30s 72°C) 30 cycles; 5 min 72°C] for 18S rRNA and
[3 min 94°C; (10 s 98°C, 30 s 57°C, 30s 72°C) 45 cycles; 5 min 72°C] for
cenRNA#1.

**Immuno-fluorescence (IF)**

Cells were grown on poly-D-Lysine-treated coverslips in 6-well plate and
synchronized by double thymidine block. After 2 washes with cold 1X PBS, they
were prefixed for 30 s with cold 4% paraformaldehyde (PFA, #15714-S, Electron
Microscopy Sciences, Hatfield, PA) in PEM (80 mM K-PIPES pH6.8, 5 mM EGTA
pH7.0, 2 mM MgCl₂). Following 3 washes with cold PEM, soluble proteins were
extracted for 5 min on ice with 0.5% Triton X-100 in CSK (10 mM PIPES pH6.8,
100 mM NaCl, 300 mM sucrose, 1 mM EGTA, 3 mM MgCl₂). Few drops of 4%
PFA in PEM were added for 5 min. Slides were then incubated with fresh 4%
PFA in PEM for 40 min on ice. After 3 washes with PEM, cells were
permeabilized with 0.5% Triton X-100 in PEM for 30 min at RT, washed again 3
times, and blocked in 1X TBS, 3% Bovine Serum Albumine, 5% normal goat
serum (#005-000-121, Jackson ImmunoResearch, West Grove, PA) for 1 hr at
RT. Finally cells were incubated with the primary antibody diluted in 1X TBS, 1%
Bovine Serum Albumine, 5% normal goat serum over-night (O/N) at 4°C in a
humidified chamber. Slides were washed 3 times for 5 min at RT with 1X TBS,
0.1% Tween 20, and incubated with secondary antibody for 1 hr at RT. After
washing, the same protocol was repeated for co-IF, and then cells were stained
with DAPI (4',6-diamidino-2-phenylindole, #D9542, Sigma-Aldrich) in 1X TBS and mounted with mowiol solution [31].

For β-actin IF, a classic protocol was used. Briefly, cells were fixed with 2% PFA, 1X PBS for 10 min on ice. After 3 washes with 1X PBS, 0.1% Triton X-100, 1% Bovine Serum Albumine, cells were incubated with the primary antibody diluted in 1X PBS, 0.1% Triton X-100, 1% Bovine Serum Albumine, 5% normal goat serum O/N at 4°C in a humidified chamber. Slides were washed 3 times for 5 min at RT, and incubated with secondary antibody for 1 hr at RT. After washing, cells were stained with DAPI in 1X PBS and mounted with mowiol solution.

Chromatin fiber protocol was adapted from [32]. Trypsinized HeLa cells were incubated in hypotonic buffer (75 mM KCl) for 10 min at RT, before cytopspining for 10 min at 400 rpm. Slides were immersed in freshly prepared fiber lysis buffer (2.5 mM Tris-HCl pH7.5, 0.5 M NaCl, 1% Triton X-100, 0.4 M urea) for 15 min at RT, then in fixation buffer (4% formaldehyde (#F8775, Sigma-Aldrich), 1X PBS, final pH7.4) for 10 min at RT, and finally in permeabilization buffer (1X PBS; 0.1% Triton X-100) for 7 min at RT. After blocking (1X PBS, 0.5% Bovine Serum Albumin, 0.01% Triton X-100), chromatin fibers were stained O/N at 4°C in a humidified chamber with primary antibody diluted in blocking solution complemented with 1% normal goat serum. Slides were washed 3 times for 5 min at RT with 1X PBS, 0.05% Tween 20, before incubation with secondary antibody for 1 hr at RT. After washing, the protocol was repeated for co-IF, and the fibers were then stained with DAPI in 1X PBS and mounted with mowiol solution.
When FISH was performed, antibody–protein complexes were crosslinked (8% formaldehyde diluted in distilled water) for 10 min at RT, denaturated in 70% formamide (#F47670, Sigma-Aldrich), 2X SSC buffer for 8 min at 78°C, and then incubated with denaturated probed (tagged with biotinylated nucleotides or Cy5-nucleotides by nick translation method) O/N at 37°C in a humidified chamber. Slides were washed 5 min at 45°C 3 times with 50% formamide, 2X SSC solution, and 4 times with 2X SSC, 0.05% Tween 20 solution. Slides were blocked in 4X SSC, 0.1% Tween 20, 3% BSA for 30 min at RT. Following the incubation with the secondary antibody at 37°C for 1 hr, slides were washed 4 times for 5 min each at 45 °C with 4X SSC, 0.1% Tween 20, stained with DAPI in 2X SSC and mounted coverslip with mowiol solution.

All samples were observed with a DeltaVision RT system (Applied Precision, Issaquah, WA) controlling an interline charge-coupled device camera (Coolsnap, Roper) mounted on an inverted microscope (IX-70; Olympus, Center Valley, PA). Images were captured by using a 60 X at 0.2 μm z sections for cell and 100X objective at 0.1 μm z sections for chromatin fiber, deconvolved, and projected by using softWoRx (Applied Precision). Three independent experiments were performed and in each 5-10 chromatin fibers or 30-50 cells were analyzed per slide.

**IF analysis**

To quantify IF signals, the acquisition of pictures for all samples of an experiment was performed with the same time of exposure during the same day to avoid
variability from the instrument. Using ImageJ (ImageJ 1.43U), signal intensity of each CENP-A or CENP-B spot inside of the nucleus (as defined by the DAPI staining) was extracted. The background level of the nucleus was subtracted from the average value of the spot intensity per cell. For each experiment, the average value of average value of the spot intensity per cell and the ratio – signal intensity in treated condition versus signal intensity in non-treated condition – were measured. The means and standard deviations of three experiments are presented in the Supplemental File 1.

Chromatin extraction and immuno-precipitation (IP)

Five F175 flasks of HeLa cells (70-80% of confluence) were used for IP. Cells were trypsinized (#25300, Gibco) and washed 3 times with cold 1X PBS, 0.1% Tween 20 coupled with 5 min centrifugation at 800 rpm at 4°C. Nuclei were isolated in TM2 buffer (20 mM Tris-HCl pH8, 2 mM MgCl₂, 0.5 mM PMSF) complemented with 0.5% NP40, and washed once with TM2 buffer. Chromatin was digested 2 min at 37°C with 0.2 unit/mL of MNase (#N3755, Sigma-Aldrich) in 0.1 M TE buffer (0.1 M NaCl, 10 mM Tris-HCl pH8, 0.2 mM EGTA) complemented with 2 mM CaCl₂. The reaction was stopped by addition of 10 mM EGTA and transferred to ice. After centrifugation for 5 min at 800 rpm at 4°C, the nuclear pellet was resuspended in 1 mL low-salt buffer (0.5X PBS, 5 mM EGTA, 0.5 mM PMSF, protease inhibitor cocktail (#05892953001, Roche, Indianapolis, IN)) and the chromatin was extracted O/N at 4°C in an end-over-end rotator. An aliquot of the supernatant obtained after centrifugation for 5 min at 8000 rpm at
4°C was saved as “input” (1.5%). At 4°C, sample was precleared with protein A/G Plus agarose beads (#sc-2003, Santa Cruz Biotechnology, Dallas, TX) for 30 min at 4°C in an end-over-end rotator, incubated with the primary antibody for 4 hrs, followed by IP with protein A/G Plus agarose beads for 2 hrs. After centrifugation, an aliquot was saved as “unbound” (UB, 1.5%), and the bead-associated IP was washed 3 times with low-salt buffer, and stored at -20°C in Laemmli buffer (30 µL) for Western blot analysis.

RNA-chromatin IP and RNA purification

A general scheme is presented on Figure 4 – figure supplement 3. Five F175 flasks of eG1 synchronized HeLa cells (70-80% of confluency) were used for IP. After trypsinization, cells were washed 2 times with cold 1X PBS, 0.1% Tween 20, fixed 10 min at RT with 1% formaldehyde, quenched by addition of 125 mM glycine, and washed twice with cold 1X PBS, 0.1% Tween 20. Samples were treated as described above (i.e., “Chromatin extraction and immunoprecipitation”) in presence of 10 mM Ribonucleoside Vanadyl Complex (RVC, #1402, NEB). After centrifugation 5 min at 800 rpm at 4°C following the MNase digestion, the supernatant was saved and named “soluble fraction”, whereas the nuclear pellet was suspended in 1 mL low-salt buffer (0.5X PBS, 5 mM EGTA, 0.5 mM PMSF, protease inhibitor cocktail, 10 mM RVC) and chromatin was extracted O/N at 4°C. The supernatant obtained after centrifugation 5 min at 8000 rpm at 4°C was named “chromatin fraction”. IPs were performed as described previously with both fractions (i.e., “Chromatin extraction and immuno-
precipitation”). After washes, bead-associated IPs were divided in 2 equal
samples for protein analysis by Western blot and for RNA study.

For RNA study, RNA-protein complexes were eluted from protein A/G Plus
agarose beads by incubation in an end-over-end rotator for 15 min at RT with
250 µL elution buffer (1% SDS; 0.1 M NaHCO₃). The supernatant was saved and
the elution step was repeated once more. All samples (input, unbound and IP)
were denatured at 65°C for 2 hrs with 200 mM NaCl. Proteins were digested with
20 µg of proteinase K (#AM2548, Ambion) in presence of 40 mM Tris-HCl pH6.5,
10 mM EDTA at 42°C for 45 min. To avoid genomic DNA contamination, samples
were treated with DNase I for 30 min at 37°C. The reaction was stopped by
addition of 5 mM EDTA and RNAs were purified by phenol:chloroform:isoamyl
alcohol method and ethanol precipitation. Samples were stored at -80°C until
further use in retro-transcription PCR and Northern blot.

Western blot
Samples in Laemmli buffer were denatured 5 min at 95°C, plunged on ice for 2
min, loaded into a 4-20% SDS-PAGE (#456-1093, Biorad, Hercules, CA) for
separation in 1X Tris Glycine SDS Running Buffer (#161-0732, Biorad), and
transferred to Whatman® nitrocellulose membrane (#10439396, Sigma-Aldrich)
in Tris Glycine transfer buffer (#351-087-131, Quality Biological, Gaithersburg,
MD) diluted in 20% ethanol. Membrane was blocked in Odyssey® blocking buffer
(#)927-40000, Li-Cor, Lincoln, NE) diluted in 1X PBS (1:1) at RT for 1 hr, and
incubated with primary antibody diluted in blocking buffer complemented with
0.1% Tween 20 O/N at 4°C. After 3 washes in 1X PBS, 0.1% Tween 20, the membrane was incubated with the secondary antibody conjugated to IRDye680 (#926-68072 and #926-68073, Li-Cor) diluted in blocking buffer complemented with 0.1% Tween 20 and 0.05% SDS for 1 hr at RT. The membrane was washed in the same conditions than previously and proteins were detected by scan on Odyssey® CLx scanner (Li-Cor).

**Protein quantification**

eG1-synchronized HeLa cells treated with α-amanitin (as described in “Cell culture and RNA Polymerase inhibition”) were resuspended in lysis buffer (20 mM Tris-HCl pH7.5, 400 mM NaCl, 2 mM dithiothreitol, 1% Nonidet P40 substitute (#74385, Sigma-Aldrich), 0.5 mM PMSF, protease inhibitor cocktail). After 3 cycles of freeze and thaw, extract was centrifuged for 20 min at 12000 rpm at 4°C. The cleared suspension was quantified by UV spectroscopy and 50 µg of proteins were resuspended in Laemmli buffer. After separation into a 4-20% SDS-PAGE, transfer to nitrocellulose membrane and incubation with primary and secondary antibodies (as described in “Western Blot”), proteins were detected by scan on Odyssey® CLx scanner and quantified with ImageStudioLite software (Li-Cor). For each experiment, the ratio of signal intensity in treated condition versus signal intensity in non-treated condition was measured. The means and standard deviations of three experiments are presented in the figure.

**Northern blot**
Northern blotting protocol was adapted from [33] and (http://archive.bio.ed.ac.uk/ribosys/protocols/website_Northern_blotting.pdf). Five μg of Trizol extracted RNAs or 1 μg of immuno-precipitated RNAs were separated on 4% urea page against 0.5X TBE buffer at 25 W for 90 min. RNAs were transferred to Amersham Hybond-NX membrane (#RPN203T, GE Healthcare, Pittsburgh, PA) for 2 hrs at 65 V, UV-cross-linked, blocked for 1 hr in SES buffer (0.5 M Na₃PO₄ pH7.2, 7% (w/v) SDS, 1 mM EDTA), and hybridized O/N at 37°C with radiolabelled α-satellite probes (end labeling method using primer extension-system AMV reverse transcriptase kit, #E3030 Promega, Madison, WI) diluted in SES buffer. Membrane was washed in 6X SSPE (1.08 M NaCl, 0.06 M NaH₂PO₄, 20 mM EDTA, pH adjusted to 7.4) 2 times for 15 min at 37°C and 2 times for 30 min at 42°C. Blot was exposed to P32-sensitive film (Hyblot CL film, #E3012, Denville, Saint Laurent, Canada) at -80°C to reveal the potential interaction for a short (<24 hrs) or long (>24 hrs) period of time. The sequences of the radiolabeled probes are indicated in Supplemental File 3.

Computational analysis

Computational prediction of RNA binding residues was performed with BindN⁺ program (http://bioinfo.ggc.org/bindn+) with a specificity equal to 85% [15]. We used human CENP-A (P49450), H3.1 (P68431) and HJURP (Q8NCD3), and Scm3 (Q12334) protein sequences from uniprot database.

Transfection with shRNA
pGFP-V-RS plasmids expressing shRNA sequence were purchased from Origine (Rockville, MD). Two controls were used for each experiment: the empty vector (#TR30007) and the vector expressing scrambled sequence cassette (#TR30013; shRNA$^{\text{scram}}$ 5'-GCACTACCAGAGCTAACTCAGATAGTACT-3'). Two shRNA sequence cassettes were designed from the centromeric α-satellite consensus sequence [3]: shRNA$^{\text{sat1}}$ 5'-TGTGTGCATTCAACTCACAGAGTTG-3' and shRNA$^{\text{sat2}}$ 5'-CAACTCACAGAGTTGAACCTTCCTT-3' (Figure 6 – figure supplement 1); and from the cenRNA#1 sequence (Figure 5 – figure supplement 1): shRNA$^{\text{cenRNA#1}}$ 5'-TGCTAGACAGCCAATGCAATTCCTTATT-3'.

Cells were transfected with Escort II Transfection Reagent (#L6037, Sigma-Aldrich) following manufacturer instruction. Forty height hours after transfection, the medium was replaced every 2 days with fresh medium complemented with 0.5 μg/ml puromycin (#A1113802, Gibco) to select transfected cells. At day six, cells were either treated for IF or RNA extraction.

**Quantitative PCR (qtPCR)**

To detect α-satellite expression level in shRNA transfected cells, RNAs were extracted, quantified by UV-spectroscopy, and equal quantities were retro-transcribed using Superscript III First- Strand Synthesis kit as described above (i.e., RNA extraction, retro-transcription and Polymerase Chain Reaction). To perform qtPCR, complementary DNAs (cDNAs) samples were prepared using the iQ SYBR Green supermix (#170-8880, Biorad) following manufacturer protocol. Control reactions without the cDNA were performed to rule out non-
specific amplification. The qPCR was run on Step one plus Real time PCR system (Applied Biosystem, Grand Island, NY). Primer sequences are available on Supplemental File 3.

The comparative cycle threshold (C\textsubscript{T}) method was used to analyze the level expression of a-satellite transcripts. C\textsubscript{T} values were normalized against the average C\textsubscript{T} value of the housekeeping gene β-actin. The ∆∆C\textsubscript{T} values were determined from the scrambled shRNA samples. Relative fold differences \((2^{-\Delta\Delta C_T}})\) are indicated on figure.

β-galactosidase assay

Senescent cells were detected using the protocol developed by Itahana et al. [34]. Briefly, six days post transfection with shRNA, HeLa cells grown in 6-well plate were washed 2 times in 1X PBS, fixed 5 min at RT with 3.7% formaldehyde in 1X PBS, and wash twice with 1X PBS. Cells were stained with the X-gal staining solution (1 mg/mL of X-gal (#B9146, Sigma-Aldrich), 40 mM citric acid/sodium phosphate buffer pH 6.0, 5 mM potassium ferricyanide (#702587, Sigma-Aldrich), 5 mM potassium ferrocyanide (#P3289, Sigma-Aldrich), 150 mM NaCl, 2 mM MgCl\textsubscript{2}) O/N at 37°C. After rinsing, cells were observed under a light microscope for blue color, indicator of senescence.

Statistical Analysis

Standard deviation was determined for all quantification measures. To test the significance of these measures, a two-tailed, paired Student's \(t\)-test was
performed. For all tests $\alpha$ was assumed to be 0.05. The $p$-value is indicated on the figures or tables each time it was evaluated.
Conflict of interest statement: The authors have declared that no competing interests exist.

Acknowledgments: We thank Shiv Grewal (LBMB, NCI), Tom Misteli (LRBGE, NCI), Rachel O'Neill, and members of our lab for thoughtful discussion and critical feedback; Trent Bowen and Rajbir Gill for technical assistance; David Sturgill for assistance with computational analysis; and James McNally (LRBGE, NCI) and Kathy McKinnon (Vaccine Branch, NCI) for the access to microscopy facility and FACS core facility, respectively. YD and DQ are supported by the Intramural Research Program of the Center for Cancer Research at the National Cancer Institute/NIH.

Author contributions: Conceived and designed the experiment: DQ, YD. Performed the experiments: DQ. Analyzed the data: DQ. Wrote the paper: DQ, YD.
REFERENCES

10.1038/nrm3107


10.1038/nsmb845

10.1038/ng1966

10.1038/nrg2905


10.1101/gr.6022807

10.1073/pnas.1108705109

10.1038/emboj.2010.329


2154-1264-2-3-2 [pii]


FIGURE LEGENDS

Figure 1. Active RNA Polymerase II (RNAPII<sup>S2P</sup>) and TATA-box binding protein (TBP) are associated with centromeric chromatin at eG1. (A) Chromatin fibers prepared from synchronized HeLa cells at eG1, G1/S and G2 phases were co-stained for RNAPII<sup>S2P</sup> and TBP (green) with centromeric proteins CENP-B and CENP-A (red), respectively. The DAPI raw image is shown for a representative chromatin fiber (cyan). Three independent experiments were performed and in each a minimum of 5 fibers were analyzed per slide (co-localization on the same chromatin fiber at eG1: between CENP-B and RNAPII<sup>S2P</sup> = 10/15; and between CENP-A and TBP = 9/16). Scale bar: 1 µm. (B) CENP-A, CENP-C or RNAPII<sup>S2P</sup> were immuno-precipitated, and co-purifying partners were detected on Western blots (1.5% of input and UB, 75% of IP). Co-IF, co-immuno-fluorescence; IP, immuno-precipitate; UB, unbound.

Figure 2. Transcription of centromeres is dependent on RNA Polymerase II (RNAPII) and occurs at eG1. G2 (green), mitotic (M; blue) and eG1 (black) synchronized cells were treated 2 hrs with either actinomycin D or α-amanitin, to block RNA Polymerase I or RNAPII, respectively. After RNA purification and retro-transcription, expression levels of control target genes (18S rRNA and GAPDH) and centromeric α-satellite transcripts were assessed by semi-quantitative PCR. The graph represents the average of three biological replicates, the y-axis plots the ratio (± SD) of gene expression after treatment (actinomycin D or α-amanitin) compared to the non-treated (NT) condition.
values indicating statistical significance are presented where appropriate above the histograms. *P>0.1, **0.001<P<0.05 and ***P<0.001.

**Figure 3.** RNA Polymerase II (RNAPII)-dependent transcription is required for the recruitment of CENP-A and its chaperone HJURP onto the centromere at eG1. (A) eG1 synchronized HeLa cells were treated or not (NT) with α-amanitin (α-ama) for 2 hrs before staining for centromeric proteins CENP-A or CENP-B (green). The DAPI raw image is shown for a representative cell (cyan). Three independent experiments were performed and in each a minimum of 30 cells were analyzed per slide. Scale bar: 5 µm. (B) Signal intensity of CENP-A and CENP-B spots from (A) were quantified using ImageJ, and relative ratios of α-amanitin versus non-treated conditions were determined. Means ± SD from three independent experiments is represented on the graph. Quantification values are tabulated in Supplemental File 1. (C) Centromeric proteins CENP-A and CENP-B (red) were co-stained with HJURP (green) and RNAPII phosphorylated on serine 2 (RNAPII\textsuperscript{S2P}, green), respectively, on centromeric chromatin fibers prepared from eG1 synchronized cells treated or not with α-amanitin for 2 hrs (cyan, DAPI). Three independent experiments were performed and in each a minimum of 5 chromatin fibers were analyzed per slide (co-localization on the same chromatin fiber after α-amanitin compared to non-treated: between CENP-B and RNAPII\textsuperscript{S2P} = 9/15 versus 10/15; and between CENP-A and HJURP = 2/15 versus 9/15). Scale bar: 1 µm. Co-IF, co-immuno-fluorescence.
Figure 4. The pre-assembly soluble HJURP/CENP-A complex binds a 1.3 kb centromeric transcript at eG1. (A) Total RNAs from HeLa cells were separated on denaturing gel, and visualized by Northern blot with radiolabeled centromeric α-satellite probes. (B) Co-immuno-precipitated RNAs by CENP-A or HJURP were analyzed by Northern blot as in (A). CF: chromatin fraction; IP, immuno-precipitate; MW, molecular weight; SF: soluble fraction; UB, unbound.

Figure 5. The identified cenRNA#1 transcript associated with CENP-A has a centromeric origin. (A). CENP-A (green) was co-stained with Xist DNA FISH probe (FISH XistDNA, red) on chromatin fibers (cyan, DAPI). Two independent experiments were performed and in each, a minimum of 8 chromatin fibers were analyzed per slide (co-localization on the same chromatin fiber between CENP-A and FISH XistDNA = 0/18). Scale bar: 1 µm. (B) Chromatin fibers were either co-stained by IF/FISH with CENP-A (green) and cenRNA#1DNA (red), or IF/double-FISH with CENP-A (green), cenRNA#1DNA (red) and centromeric α-satellite probes (grey). The DAPI raw image is shown for a representative chromatin fiber (cyan). Two independent experiments were performed and in each, a minimum of 8 chromatin fibers were analyzed per slide (co-localization on the same chromatin fiber between CENP-A and cenRNA#1DNA and FISH α-satellite compared to no co-localization = 9/20 versus 11/20). Scale bar: 1 µm. FISH, fluorescence in situ hybridization; IF: immuno-fluorescence.
Figure 6. Targeted down-regulation of centromeric α-satellite transcript results in mitotic defect. (A) Six days post-transfection of empty vector, control scrambled (shRNA\textsuperscript{scram}) or α-satellite shRNA (shRNA\textsuperscript{sat1}, shRNA\textsuperscript{sat2}), cell morphology was observed by phase contrast microscopy. Scale bar: 100 pixels. (B) Cells were treated as in (A) and stained for β-actin (green) and DAPI (cyan) to reveal alterations in cell morphology. Scale bar: 5 μm. (C) Cells were treated as in (A), synchronized in mitosis and stained for α-tubulin (green), CENP-B (red) and DAPI (cyan/blue). Three independent experiments were performed and in each a minimum of 30 cells were analyzed per slide. Scale bar: 5 μm. IF: immuno-fluorescence.

Figure 7. Targeted down-regulation of centromeric α-satellite abrogates HJURP/CENP-A targeting to centromeric chromatin at eG1. Chromatin fibers were prepared from shRNA\textsuperscript{scram}, shRNA\textsuperscript{sat1}, or shRNA\textsuperscript{sat2} transfected cells synchronized at eG1. To visualize centromeric domains, co-IF was performed for CENP-B (red), CENP-C (green) and RNA Polymerase II (phosphorylated on serine 2, RNAPII\textsuperscript{S2P}, green), whereas CENP-A (red) and HJURP (red) antibodies were co-stained with a DNA FISH probe against centromeric α-satellite DNA repeats (green). The DAPI raw image is shown for a representative chromatin fiber (cyan). Three independent experiments were performed and in each 5 chromatin fibers were analyzed per slide (co-localization on the same chromatin fiber in shRNA\textsuperscript{sat1} or shRNA\textsuperscript{sat2} transfected cells: between CENP-B and RNAPII\textsuperscript{S2P} = 9/15 and 8/15; CENP-B and CENP-C = 14/15 and 13/15; between...
CENP-A and centromeric $\alpha$-satellite DNA = 3/15 and 1/15; between HJURP and centromeric $\alpha$-satellite DNA = 1/15 and 0/15, respectively). Scale bar: 1 $\mu$m.

FISH, fluorescence *in situ* hybridization; IF, immuno-fluorescence.
Figure 1- figure supplement 1. Scheme presenting the strategy for HeLa cell synchronization and treatment with drugs (actinomycin D or α-amanitin) inhibiting either RNA Polymerase I or RNA Polymerase II activity. HeLa cells were synchronized in G1/S phase by two consecutive treatments of thymidine, and then released until G2 (9 hrs), mitosis (M; 10 hrs) and early G1 (eG1; 11 hrs) cell cycle phases. The synchronization and the potential effect of actinomycin D or α-amanitin treatment (2 hrs) were assessed by cell cycle analysis. NS, non-synchronized cells; PR, post-release.

Figure 1 - figure supplement 2. Active RNA Polymerase II (RNAPII S2P) is associated with centromeric α-satellite sequences on chromatin fibers at eG1. (A) RNAPII S2P (green) was co-stained with centromeric α-satellite DNA FISH probe (red) on chromatin fibers from eG1-synchronized cells. (cyan, DAPI). Three independent experiments were performed, and a minimum of 5 fibers were counted per slide (co-localization on the same chromatin fiber: between CENP-B and centromeric α-satellite DNA = 11/15. (B) The progressive accumulation of mono-nucleosomes after chromatin digestion (0.5-8 min) with MNase was analyzed on agarose gel. The arrow indicates the chosen MNase time used for experiments in the manuscript. (C) Chromatin extract was incubated in presence of agarose beads and co-precipitated proteins were analyzed on Western blot (1.5% of input and UB, 75% of IP). FISH, fluorescence in situ hybridization; co-IF, co-immuno-fluorescence; IP, immuno-precipitate; UB, unbound.
Figure 3 - figure supplement 1. Inhibition of RNA Polymerase II-dependent transcription at eG1 does not affect CENP-A and HJURP protein levels. (A)
eG1-synchronized HeLa cells were treated with α-amanitin for 2 hrs before staining for centromeric protein CENP-B (green) and DAPI (cyan). The
heterogeneity of the CENP-B signal on a picture is shown with magnification of one cell with high signal (1 and 2) and low signal (3 and 4). Scale bar: 5 µm. (B)
eG1-synchronized cells were treated, or not, with α-amanitin (2 hrs), before total protein extraction, and the expression levels of CENP-A and HJURP were quantified by Western blot. The graph represents the average of three biological replicates, the y-axis plots the ratio (± SD) of protein level after α-amanitin compared to the non-treated condition for CENP-A and HJURP Western blot, respectively. IF, immuno-fluorescence.

Figure 4 - figure supplement 1. Centromeric transcripts are 1.3 kb in length (A) To determine the size of the centromeric α-satellite transcripts, the graph of the distance (in y) between the border of the Northern blot and each band of the molecular weight as a function of the number of bases was made. The distance of the centromeric α-satellite transcript band was analyzed using the standard curve from this graph to deduce its size. (B) Total RNAs treated with RNase A were separated on a denaturing gel, and revealed by Northern blot with radiolabeled centromeric α-satellite probes. (C) eG1-synchronized cells were treated, or not, with α-amanitin (2 hrs). RNAs were processed and analyzed on
Northern blot as in (B) to examine whether trace DNA contamination could yield
the same band as in (A).

**Figure 4 – figure supplement 2.** HJURP and CENP-A display potential RNA
binding residues. RNA binding residues in CENP-A, H3, Scm3 and HJURP
were predicted by computational analysis using the BindN+ program. Binding
residues are labeled with '+' and in red, whereas non-binding residues are
labeled with '-' and in green. The confidence of binding potential ranges from low
(0) to high (9).

**Figure 4 - figure supplement 3.** Scheme presenting the strategy for RNA-
chromatin immuno-precipitation, RNA purification, and sub-cloning for
sequencing.

**Figure 4 - figure supplement 4.** Mock IPs demonstrate specificity of the
1.3kb RNA binding to CENP-A and HJURP (A) CENP-A and HJURP were
immuno-precipitated from eG1 cells, separated on SDS-PAGE and revealed on
Western blot (1.5% of input and UB, 100% of IP). To exclude non-specific
binding of CENP-A or HJURP to the secondary beads, immuno-precipitation was
performed with beads in absence of antibody (mock IP). (B) RNAs immuno-
precipitated from the mock IP in (A) were analyzed on Northern blot using
radiolabeled centromeric α-satellite probes. CF: chromatin fraction; IP, immuno-
precipitate; MW, molecular weight; SF: soluble fraction; UB, unbound.
Figure 5 – figure supplement 1. Sequence of cenRNA#1. RNAs immuno-precipitated with CENP-A were retro-transcribed and subcloned. One 675 nucleotide long sequence called cenRNA#1 was obtained by sequencing, and is reported here. The sequence of shRNA$^{cenRNA#1}$ is underlined.

Figure 5 – figure supplement 2. shRNA$^{cenRNA#1}$-transfected cells have a cell survival defect. (A) Six days post-transfection of shRNA$^{scram}$ or shRNA$^{cenRNA#1}$, cell morphology was observed by phase contrast microscopy. Scale bar: 100 pixels (B) Cells were treated as in (A) and stained for β-actin (green) and DAPI (cyan) to reveal alterations in cell morphology. Scale bar: 5 µm.

cDNA, complementary DNA; IF, immuno-fluorescence; NT, non-treated.

Figure 6 – figure supplement 1. Down-regulation of centromeric RNAs by a targeted shRNA approach. (A) The centromeric α-satellite consensus sequence was used to design α-satellite shRNA (shRNA$^{sat1}$ and shRNA$^{sat2}$) compared to control shRNA containing a scrambled sequence (shRNA$^{scram}$). Residues of the CENP-B box are marked by a black box, and the sequences of shRNA$^{sat1/2}$ are underlined. (B) Centromeric α-satellite transcript level after transfection of shRNA$^{scram}$, shRNA$^{sat1}$ or shRNA$^{sat2}$ was measured by qtPCR. The graph represents the average of three biological replicates, with relative expression (± SD) of α-satellite transcript in shRNA$^{sat}$ transfected cells compared to shRNA$^{scram}$ transfected cells.
Figure 6 – figure supplement 2. Down-regulation of centromeric transcripts does not induce senescence. (A) Six days post-transfection of empty vector, control scrambled (shRNA_{scram}) or α-satellite shRNA (shRNA_{sat1} or shRNA_{sat2}), the potential senescent status of transfected cells was tested by a β-galactosidase assay. Senescent BJ cells were used as positive control (blue).

Figure 7 – figure supplement 1. A speculative model proposing a mechanism by which centromeric long non-coding RNAs target soluble HJURP/CENP-A complexes to centromeres at eG1. At mitosis/eG1, Mis18 primes centromeres for active transcription by RNA Polymerase II (RNAPII) (step 1), leading to the synthesis of long non-coding RNAs, which are released into the nucleoplasm at eG1. These centromeric transcripts bind pre-assembly HJURP/CENP-A/H4 in the nucleoplasm (step 2), and specifically target the resulting nucleoprotein complex back to active native centromeres. Thus, a “dual lock” system for active centromeric loading is provided by the Mis 18 complex facilitating eG1 centromeric transcription, and the IncRNA/HJURP/CENP-A complex binding centromeres (step 3), potentially via the formation of RNA-DNA hybrids with centromeric DNA.
Supplemental File 1: RNA Polymerase II inhibition results in CENP-A loss at centromere at early G1. eG1 synchronized cells were treated 2 hrs with or without α-amanitin, and stained for centromeric protein CENP-A or CENP-B. After image acquisition, immuno-fluorescent signals were quantified using ImageJ.

Supplemental File 2: List of antibodies used in this study.

Supplemental File 3: List of primer sequences used in this study.
Figure 2.
Figure 5.

A

No colocalization

Chromatin fiber IF/FISH

DAPI
CENP-A
DNA FISH Xist DNA
Colocalization

B

No colocalization

Chromatin fiber IF/FISH

DAPI
CENP-A
DNA FISH cenRNA#1 DNA
Colocalization

54.8%

Colocalization

45.2%