Sequence Features and Transcriptional Stalling within Centromere DNA Promote Establishment of CENP-A Chromatin

Citation for published version:
Catania, S, Pidoux, AL & Allshire, RC 2015, ‘Sequence Features and Transcriptional Stalling within Centromere DNA Promote Establishment of CENP-A Chromatin’ PLoS Genetics, vol. 11, no. 3, e1004986. DOI: 10.1371/journal.pgen.1004986

Digital Object Identifier (DOI):
10.1371/journal.pgen.1004986

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
PLoS Genetics

Publisher Rights Statement:
© 2015 Catania et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Abstract

Centromere sequences are not conserved between species, and there is compelling evidence for epigenetic regulation of centromere identity, with location being dictated by the presence of chromatin containing the histone H3 variant CENP-A. Paradoxically, in most organisms CENP-A chromatin generally occurs on particular sequences. To investigate the contribution of primary DNA sequence to establishment of CENP-A chromatin in vivo, we utilised the fission yeast *Schizosaccharomyces pombe*. CENP-A<sup>Cnp1</sup> chromatin is normally assembled on ~10 kb of central domain DNA within these regional centromeres. We demonstrate that overproduction of *S. pombe* CENP-A<sup>Cnp1</sup> bypasses the usual requirement for adjacent heterochromatin in establishing CENP-A<sup>Cnp1</sup> chromatin, and show that central domain DNA is a preferred substrate for de novo establishment of CENP-A<sup>Cnp1</sup> chromatin. When multimerised, a 2 kb sub-region can establish CENP-A<sup>Cnp1</sup> chromatin and form functional centromeres. Randomization of the 2 kb sequence to generate a sequence that maintains AT content and predicted nucleosome positioning is unable to establish CENP-A<sup>Cnp1</sup> chromatin. These analyses indicate that central domain DNA from fission yeast centromeres contains specific information that promotes CENP-A<sup>Cnp1</sup> incorporation into chromatin. Numerous transcriptional start sites were detected on the forward and reverse strands within the functional 2 kb sub-region and active promoters were identified. RNAPII is enriched on central domain DNA in wild-type cells, but only low levels of transcripts are detected, consistent with RNAPII stalling during transcription of centromeric DNA. Cells lacking factors involved in restarting transcription—TFIIS and Ubp3—assemble CENP-A<sup>Cnp1</sup> on central domain DNA when CENP-A<sup>Cnp1</sup> is at wild-type levels, suggesting that persistent stalling of RNAPII on centromere DNA triggers chromatin remodelling events that deposit CENP-A<sup>Cnp1</sup>. Thus, sequence-encoded features of centromeric DNA create an environment of pervasive low quality RNAPII transcription that is an important determinant of CENP-A<sup>Cnp1</sup> assembly. These observations emphasise roles for both genetic and epigenetic processes in centromere establishment.
Author Summary

The kinetochore directs the separation of chromosomes and is assembled at a special region of the chromosome—the centromere. DNA is wrapped around particles called nucleosomes, which contain histone proteins. The nucleosomes at centromeres are specialized, and contain the centromere-specific histone CENP-A. CENP-A nucleosomes form the platform upon which the kinetochore is built. Thus, CENP-A and centromere function go hand-in-hand. How the cell ensures that CENP-A is deposited at centromeres and not elsewhere is not well understood. We investigated the role that DNA sequence plays in defining centromere function in fission yeast. Our observations suggest that it is not the DNA sequence per se that is important for attracting CENP-A, but rather, the particular environment that the sequence creates. During transcription of centromeric DNA, RNA polymerase (RNAPII) appears to get stuck or stalled. Particular proteins—such as TFIIS and Ubp3—are known to help restart RNAPII so it can continue transcribing. We found that when cells lack Ubp3 or TFIIS, CENP-A becomes deposited on centromere sequences. We propose that persistent stalling of RNAPII on centromere DNA attracts factors that help deposit CENP-A. This study highlights the influence of DNA sequence in creating an attractive environment for CENP-A assembly.

Introduction

Centromeres are the chromosomal sites upon which kinetochores are assembled to ensure accurate segregation of sister chromatids into daughter cells. Most kinetochores are built upon a specialized type of chromatin in which canonical histone H3 is replaced by the histone variant CENP-A. Although the centromere-kinetochore complex performs conserved essential functions, and kinetochore proteins are generally conserved [1], centromeric DNA is not conserved, even between related species, and a huge variety of centromere sequences and structures exist [2–5]. The point centromeres of budding yeast consist of 125 bp of DNA and utilize an essential centromere-specific DNA binding protein [6]. At the other extreme, the nematode, Caenorhabditis elegans, has holocentric centromeres, in which kinetochore proteins assemble at multiple loci along each chromosome arm [7,8]. The majority of centromeres studied to date are regional. Centromeres in various plant and animal species are composed of arrays of different types of satellite, repetitive sequences and transposable elements, for instance, human centromeres encompass several megabases of tandem repetitive arrays of alpha-satellite sequence [9–11]. Fission yeast centromeres represent another type of regional centromere, in which a unique central core of 4–7 kb is flanked by inverted repeat elements and blocks of relatively large repeat units, resulting in centromeres of 40–120 kb [12]. Even the centromeres of different chromosomes in individual species are not necessarily homologous; each Candida albicans centromere has a unique central core, whilst chicken and potato each utilize both repeat-rich and unique sequence centromeres [13–15]. Thus, functional centromeres are assembled on diverse types of sequences in different organisms and it remains unknown if there is a universal fundamental property that defines centromeric sequences.

Abundant evidence indicates that centromeres are epigenetically regulated [16]. Although rare, neocentromeres have been observed in many species, forming on DNA sequences that do not normally possess centromere function and share no sequence homology with normal centromeres [17]. The best-characterized example in human is the neocentromere in 1q25 on the long arm of chromosome 10 that arose upon deletion of the centromere and loss of the entire alpha satellite array [18]. In S. pombe, neocentromeres form in close proximity to telomeres.
following the engineered deletion of a centromere [19]. Conversely, centromeres can be inactivated on dicentric human chromosomes despite the continued presence of alpha-satellite sequence at both centromeric loci [20]. In S. pombe one centromere on a dicentric chromosome can be inactivated by mechanisms such as heterochromatinisation or formation of a domain of histone hypoacetylation [21]. These and numerous other examples demonstrate that centromeric sequences are neither necessary nor sufficient for kinetochore assembly.

The histone H3 variant, CENP-A acts as the epigenetic mark that specifies centromere identity [22–24]. CENP-A is found only at active centromeres, including neocentromeres, and is absent at inactivated centromeres. The forced recruitment of CENP-A either by directly tethering CENP-A or its chaperone (HJURP) to a non-centromeric locus leads to the accumulation of CENP-A and kinetochore proteins at that location [24,25]. It is thought that continued deposition of CENP-A at centromere regions through cell and organism generations involves a self-propagation mechanism in which CENP-A chromatin, or features of the kinetochore which is assembled upon it, are recognized and attract additional CENP-A [26,27].

In most organisms there is no obligate coupling of sequence and CENP-A assembly, yet kinetochores are normally assembled upon particular centromeric sequences in any given species [4]. This suggests that centromeric sequences possess underlying properties that promote CENP-A incorporation. Alternatively, the preponderance of particular sequences at centromeres could be driven by properties of CENP-A chromatin or kinetochores themselves [28]. However, centromeric DNA allows the de novo assembly of functional centromeres following its introduction into cells in many organisms. Alpha satellite arrays are able to direct the de novo assembly of centromeres when introduced into certain cell lines as naked DNA [29,30]. De novo assembly of centromeres also occurs when centromeric DNA from S. pombe is introduced into cells. However, de novo establishment does not seem to be a universal property: despite promiscuous neocentromere formation in C. albicans, transformation with bone fide centromeric sequences does not result in kinetochore assembly [13]. At the other extreme, many sequences introduced into the holocentric organism C. elegans appear able to assemble CENP-A chromatin [31,32]. Thus, the relationship between centromeric sequence and the establishment and maintenance of CENP-A chromatin is enigmatic.

Transcription has received a lot of attention as a possible contributor to assembly of CENP-A chromatin. Transcripts emanating from centromeric regions have been detected in many organisms, including maize, human, rice, budding yeast, fission yeast and tammar wallaby [33–38]. Interfering with the chromatin status or transcriptional properties of centromeric repeats affects maintenance of CENP-A chromatin and segregation function on human artificial chromosomes (HACs) [39,40]. RNA Polymerase II (RNAPII) has been detected at mitotic mammalian centromeres where it may influence centromere function [38]. In fission yeast, transient H2B ubiquitylation may loosen centromeric chromatin to promote transcription and CENP-A<sup>Cnp1</sup> incorporation and defective reassembly of H3 chromatin behind elongating RNAPII aids CENP-A<sup>Cnp1</sup> incorporation [41,42]. However, although there are numerous tantalising hints that transcriptional activity contributes to centromere function or identity, much remains to be understood [33,38,43,44].

Here we investigate the contribution of DNA sequence to the establishment of CENP-A chromatin in fission yeast, an organism in which epigenetic mechanisms clearly influence centromere identity. Normally proximal heterochromatin is required to facilitate establishment of CENP-A<sup>Cnp1</sup> chromatin on centromere central domain sequences [45,46]. We show that this requirement can be bypassed by overexpression of CENP-A<sup>Cnp1</sup> and that central domain DNA is a preferred substrate for establishment of CENP-A<sup>Cnp1</sup> chromatin. We find that there is functional redundancy within the central domain but that sub-regions are non-equivalent in their ability to establish CENP-A<sup>Cnp1</sup> chromatin. Analysis of a 2 kb region capable of directing
CENP-ACnp1 assembly indicates that it contains numerous transcriptional start sites, along with promoter elements, and that relatively high levels of RNAPII are recruited, despite low levels of transcripts produced, consistent with the presence of stalled RNAPII. Our observations suggest that redundant sequence features in the centromere central domain create a unique transcriptional environment that is permissive for CENP-ACnp1 establishment. Consistent with this, defective transcriptional elongation where stalled RNAPII is increased promotes the establishment of CENP-ACnp1 chromatin.

**Results**

**Elevated CENP-ACnp1 levels bypass the requirement for heterochromatin in establishing CENP-ACnp1 chromatin**

In wild-type fission yeast cells, de novo CENP-ACnp1 chromatin establishment on circular plasmid-based minichromosomes requires an outer repeat or tethered Clr4 histone H3K9 methyltransferase to form a block of heterochromatin in close proximity to central domain DNA from centromeres [45,46]. CENP-ACnp1 can also be deposited at other non-centromeric locations in the genome when it is overexpressed, however the level incorporated at these ectopic sites is much lower than that detected at natural centromeres [41,47]. To determine whether central domain DNA is a preferential substrate for the establishment of CENP-ACnp1 chromatin, plasmid pMcc2 bearing 8.5 kb of central domain from cen2 (imr2-cc2-imr2) sequence, but no heterochromatic outer repeat sequences, was transformed into cells expressing additional GFP-CENP-ACnp1 (Fig. 1A).

All strains used have 6 kb of cen2 central domain DNA replaced with 5.5 kb of cen1 central domain DNA (cc2Δ::cc1—Fig. 1A, S1 Fig.) so that only 2.5 kb of normal cen2 central domain DNA remains at this modified cen2 (imr2L, regions J, K, R; Fig. 1A). The resulting deletion of fragments L-Q from the cen2 central domain allows detailed and specific analysis of 6 kb of central domain DNA when borne by plasmid-based minichromosomes. Quantitative chromatin immunoprecipitation assays (qChIP) shows that CENP-ACnp1 chromatin does not assemble on regions L, M, N, O or P when a plasmid (pMcc2) containing the 8.5 kb cc2 sequence, but lacking heterochromatin, was transformed into wild-type cells [45]. However, when pMcc2 was transformed into cells over-expressing CENP-ACnp1 (hi-CENP-ACnp1; ~ 15 fold more than wild-type cells [41]), CENP-ACnp1 and the kinetochore proteins CENP-CCnp3 and CENP-KSim4 were easily detected over the central domain of pMcc2 by qChIP (Fig. 1B-E). Importantly, these centromeric proteins were enriched on centromeric DNA but not on the plasmid backbone, indicating that CENP-ACnp1 chromatin assembles specifically on central domain DNA from centromeres (Fig. 1). The relative level of enrichment of CENP-ACnp1 and the other kinetochore proteins on different parts of pMcc2 suggests all proteins are distributed uniformly across this plasmid-borne central domain (Fig. 1B-D). Furthermore, the levels of histone H3 associated with the L-P regions of pMcc2 were reduced in cells expressing additional CENP-ACnp1 compared to control cells (Fig. 1E). We conclude that H3 chromatin is normally assembled on central domain DNA on pMcc2 in wild-type cells but CENP-ACnp1 chromatin assembles instead when pMcc2 is placed in hi-CENP-ACnp1 cells.

**CENP-ACnp1 can assemble on pre-chromatinised substrates and is trans-generationally inherited**

To determine whether CENP-ACnp1 can become established on plasmids that are already assembled in chromatin, the pMcc2 plasmid was transformed into cells expressing wt-CENP-ACnp1 levels and subsequently crossed with hi-CENP-ACnp1 cells. qChIP analyses indicate that
CENP-A<sup>Cnp1</sup> is initially absent from pMcc2 in the wt-CENP-A<sup>Cnp1</sup> parental strain and becomes assembled in CENP-A<sup>Cnp1</sup> chromatin when transferred into the hi-CENP-A<sup>Cnp1</sup> environment, indicating that plasmid-borne cc2 initially assembled in normal (H3) chromatin can be converted to CENP-A<sup>Cnp1</sup> chromatin (Fig. 2A).

In addition, a copy of cc2 (8.5 kb) was inserted on the arm of the 530 kb Ch16 linear minichromosome which carries a complete cen3 [48] (Ch16-cc2; Fig. 2B). When the expression of additional GFP-CENP-A<sup>Cnp1</sup> was repressed (0h+T), no CENP-A<sup>Cnp1</sup> was detected on cc2. However, when GFP-CENP-A<sup>Cnp1</sup> was induced (48h-T) both CENP-A<sup>Cnp1</sup> and CENP-CCnp3 were detected on cc2 (Fig. 2B). Thus, cc2 borne on a linear minichromosome can be converted from a pre-chromatinised state to a CENP-A<sup>Cnp1</sup> state. Moreover, colony colour assays indicate that hi-CENP-A<sup>Cnp1</sup> expression induces increased loss of Ch16-cc2, which is consistent with a second functional kinetochore being formed at cc2 on Ch16 (Fig. 2C). Thus, Ch16-cc2 behaves as an inducible dicentric chromosome controlled by CENP-A<sup>Cnp1</sup> levels.

It is possible that high levels of CENP-A<sup>Cnp1</sup> are continuously required to maintain CENP-A<sup>Cnp1</sup> on pMcc2, or alternatively, once established, CENP-A<sup>Cnp1</sup> and kinetochore proteins may persist even when CENP-A<sup>Cnp1</sup> is returned to wild-type levels (wt-CENP-A<sup>Cnp1</sup>). To investigate the maintenance of CENP-A<sup>Cnp1</sup> chromatin, pMcc2 was first transformed into hi-CENP-A<sup>Cnp1</sup> cells to allow the assembly of centromeric chromatin and subsequently these pMcc2-containing cells were crossed with wt-CENP-A<sup>Cnp1</sup> cells to transfer the pMcc2 plasmid into cells expressing wild-type levels of CENP-A<sup>Cnp1</sup> chromatin. ChIP analyses show that CENP-A<sup>Cnp1</sup> persisted on the pMcc2 in this wild-type background (Fig. 2D). Western analysis of extracts from Parental, F1 and F2 cells confirmed that GFP-CENP-A<sup>Cnp1</sup> was lost in F1 and F2 cells (Fig. 2D). Thus, CENP-A<sup>Cnp1</sup> chromatin behaves as a true epigenetic entity in that once established it carries its own efficient propagation mechanism, persisting even though the original stimulus has been removed. More remarkably, this CENP-A<sup>Cnp1</sup> chromatin is maintained through 2 rounds of meiosis and at least 50 mitotic divisions. Thus central domain sequences are particularly receptive to the establishment and maintenance of CENP-A<sup>Cnp1</sup> chromatin.

Central domain sequence and length affect de novo CENP-A<sup>Cnp1</sup> deposition

To determine if specific regions from the central domain of cen2 are required to establish CENP-A<sup>Cnp1</sup> chromatin, plasmids bearing different sub-fragments from cc2 were transformed into wt-CENP-A<sup>Cnp1</sup> or hi-CENP-A<sup>Cnp1</sup> cells (Fig. 3). We used an unbiased approach to divide the 8.5 kb cc2 into 1 kb regions (J-R). Deletion of 1 kb from the centre of cc2 (N) does not affect CENP-A<sup>Cnp1</sup> establishment (pΔN; Fig. 3A, compare with pMcc2, Fig. 1B). Notably, CENP-A<sup>Cnp1</sup> incorporation on a plasmid carrying identical centromeric DNA as pAN but with the right half (O-R) inverted relative to the native sequence, was less efficient (pAN-rev; Fig. 3B).
Thus, the relative orientation of central domain sequences within \textit{cc2} influences the degree of CENP-A\textsuperscript{Cnp1} deposition; directionality or the juxtaposition of certain sequences may be important for promoting CENP-A\textsuperscript{Cnp1} incorporation. The central CENP-A\textsuperscript{Cnp1} domain at endogenous fission yeast centromeres is composed of inverted \textit{imr} repeats that flank the central core. ChIP analyses demonstrated that, in a plasmid-based establishment assay, the \textit{imr} repeats are dispensable for \textit{de novo} CENP-A\textsuperscript{Cnp1} incorporation on the remaining central domain sequences (p\textit{Δ}imr; Fig. 3C). Deletion of additional regions (LMN) of \textit{cc2} markedly decreased the efficiency of CENP-A\textsuperscript{Cnp1} incorporation relative to pMcc2 and p\textit{Δ}N (Fig. 3D; compare Fig. 1B, Fig. 3A), suggesting that either the LM region is critical for promoting CENP-A\textsuperscript{Cnp1} incorporation or that the overall reduced centromeric DNA length diminishes CENP-A\textsuperscript{Cnp1} deposition (Fig. 3D). However, further investigation using plasmids bearing smaller \textit{cc2} fragments suggests that the specific sequences present have a more significant influence on CENP-A\textsuperscript{Cnp1} deposition than the overall length of \textit{cc2} DNA present (Fig. 3E,F). For example, p\textit{Δ}JM and p\textit{Δ}NR differ by only 500 bp, however, p\textit{Δ}JM incorporated substantially more CENP-A\textsuperscript{Cnp1} than p\textit{Δ}NR (Fig. 3F). We conclude that specific sequences from the central domain of fission yeast centromeres, combined with their overall length, promote the efficient \textit{de novo} assembly of CENP-A\textsuperscript{Cnp1} chromatin.

A 2 kb region of centromeric DNA is sufficient to direct \textit{de novo} CENP-A\textsuperscript{Cnp1} chromatin assembly

It is possible that shorter fragments of centromere DNA from within the central domain can actively promote CENP-A\textsuperscript{Cnp1} assembly but that because longer total lengths are required to stabilise incorporated CENP-A\textsuperscript{Cnp1} the activity of shorter fragments cannot be detected. To address this possibility we selected two distinct sequences from the central domain of \textit{cen2} for analyses. The 2 kb OP region was present on all the pMcc2 derivatives with which we detected significant CENP-A\textsuperscript{Cnp1} incorporation following transformation into hi-CENP-A\textsuperscript{Cnp1} cells (Fig. 3). In addition, ChIP-seq analysis indicates particularly high CENP-A\textsuperscript{Cnp1} nucleosome occupancy within OP at endogenous \textit{cen2} [49 and Fig. 4A]. In contrast, the 2 kb LM region appears to be dispensable for \textit{de novo} CENP-A\textsuperscript{Cnp1} assembly on pMcc2 derived plasmids and exhibits low CENP-A\textsuperscript{Cnp1} nucleosome occupancy (Fig. 3, Fig. 4A).

Initial tests showed that neither OP (p1xOP) nor LM (p1xLM) sequences alone were capable of inducing significant \textit{de novo} CENP-A\textsuperscript{Cnp1} incorporation when introduced into hi-CENP-A\textsuperscript{Cnp1} cells (Fig. 4B). This finding is consistent with a minimal length of central domain DNA being required for stable CENP-A\textsuperscript{Cnp1} chromatin assembly and retention. To satisfy this apparent length requirement, the OP and the LM fragments were multimerised as tandem repeats to create 3xOP and 3xLM (p3xOP, p3xLM; Fig. 4C). Remarkably, when transformed into hi-CENP-A\textsuperscript{Cnp1} cells no CENP-A\textsuperscript{Cnp1} was detectable on p3xOP whereas p3xLM allowed a reasonable level of CENP-A\textsuperscript{Cnp1} incorporation (Fig. 4C). This suggests that in isolation the OP
Fig 3. Centromeric DNA sequence affects CENP-A<sup>Cnp1</sup> chromatin establishment. Plasmids containing fragments of cc2 DNA were transformed into wild-type (wt-CENP-A<sup>Cnp1</sup>) or cells expressing high levels of CENP-A<sup>Cnp1</sup> (hi-CENP-A<sup>Cnp1</sup>). Plasmids contained: (A) Deletion of central 1 kb (N) from cc2, (B) Left half of cc2 (J to M: 4 kb) plus region O to R (3.5 kb) in inverted orientation (region N absent), (C) imr2 regions (J and R) deleted along with 1 kb of cc2 (K), leaving LMNOPQ, (D) Region L to N deleted, leaving JKOPQR. (E) Deletion of J to M, leaving N to R, (F) Deletion from N to R, leaving J to M. The enrichment
region is unable to promote CENP-A<sup>Cnp1</sup> deposition even though in the context of an entire central domain it normally accepts CENP-A<sup>Cnp1</sup> and ends up with high CENP-A<sup>Cnp1</sup> nucleosome occupancy (Fig. 4A). We note that the removal of the LM region from the central domain of pMcc2 derived plasmids greatly reduced the level of CENP-A<sup>Cnp1</sup> incorporated on OP (compare pΔN Fig. 3A with pΔLMN Fig. 3D). Thus, in contrast to OP, the LM region appears to have the ability to induce CENP-A<sup>Cnp1</sup> deposition. To directly test this possibility, a single copy of LM was placed adjacent to two tandem copies of OP (pLM-2xOP) and transformed into hi-CENP-A<sup>Cnp1</sup> cells. High levels of CENP-A<sup>Cnp1</sup> were detected on both the LM and OP regions of pLM-2xOP (Fig. 4D), thus the LM region has an innate ability to stimulate CENP-A<sup>Cnp1</sup>.

Fig 4. A 2 kb region of cc2 is sufficient to establish CENP-A<sup>Cnp1</sup> chromatin. (A) Profile of ChIP-SEQ for CENP-A<sup>Cnp1</sup> [49] and corresponding position on endogenous cc2. (B) ChIP analysis of CENP-A<sup>Cnp1</sup> levels on plasmids containing one copy (p1xLM) or one copy of OP (p1xOP) transformed into wt-CENP-A<sup>Cnp1</sup> or hi-CENP-A<sup>Cnp1</sup>. (C) CENP-A<sup>Cnp1</sup> levels on plasmids containing three copies of a 2 kb fragment, either OP or LM, cloned in tandem repeats (p3xOP and p3xLM) and transformed into wild-type (wt-CENP-A<sup>Cnp1</sup>) or into cells expressing high levels of CENP-A<sup>Cnp1</sup> (hi-CENP-A<sup>Cnp1</sup>). (D) A plasmid containing one copy of the LM fragment adjacent to two copies of OP (direct repeats) was transformed into wt-CENP-A<sup>Cnp1</sup> and hi-CENP-A<sup>Cnp1</sup> and CENP-A<sup>Cnp1</sup> levels analysed by ChIP (n = 3).

doi:10.1371/journal.pgen.1004986.g004
A\textsuperscript{Cnp1} deposition on the OP region. A different arrangement of the same sequences (pOPL-MOP) also attracted CENP-A\textsuperscript{Cnp1} in hi-CENP-A\textsuperscript{Cnp1} cells (S2 Fig.). These analyses indicate that the 2 kb LM sequence contains all the features that are required to promote and accept CENP-A\textsuperscript{Cnp1} assembly, and thus LM defines a 2 kb region of \textit{S. pombe} centromeric sequence that allows the \textit{de novo} assembly of CENP-A\textsuperscript{Cnp1} chromatin.

**The 2 kb LM element is sufficient to form functional centromeres**

Plasmids bearing an entire central core domain flanked by outer heterochromatin repeats assemble functional centromeres when transformed into wild-type cells [45,50]. To determine if the 2 kb LM region imparts centromere function, a plasmid carrying the 3xLM tandem repeat adjacent to a 5 kb outer repeat heterochromatin forming element (pH-3xLM) was transformed into wild-type cells expressing CENP-A\textsuperscript{Cnp1} at normal levels (Fig. 5A). The establishment of functional centromeres in the resulting transformants was monitored by an \textit{ade6}-based colony colour sectoring assay [51]. Minichromosomes carrying full-length \textit{cc2} and 5 kb of outer repeat heterochromatin were able to establish functional centromeres upon transformation (Fig. 5B and S2 Fig.). pH-3xLM and pH-LM-2xOP transformants also established functional centromeres, but at lower frequencies than pH-cc2 (Fig. 5B and S2 Fig.). Differences in the ability of various constructs to form functional centromeres may reflect the particular configuration of sequences in individual minichromosomes. In contrast, pH-3xOP (3xOP flanked by heterochromatin) was unable to establish functional centromeres. Thus the LM sequence in a 3x tandem array, flanked by heterochromatin, is sufficient to form functional centromeres. ChIP analyses confirmed that kinetochores were assembled on pH-3xLM since CENP-A\textsuperscript{Cnp1} and the kinetochore proteins CENP-C\textsuperscript{Cnp3} and CENP-K\textsuperscript{Sim4} were enriched over the LM sequences at levels comparable to endogenous centromeres (Fig. 5C). We conclude that the LM sequence within pH-3xLM not only promotes incorporation of CENP-A\textsuperscript{Cnp1} into chromatin but also supports the assembly of a functional centromere.

**Features within centromeric central domain DNA are required to promote CENP-A\textsuperscript{Cnp1} chromatin establishment**

Nucleosome occupancy is known to be influenced by a combination of DNA sequence and the action of chromatin remodelers [52]. Primary DNA sequence itself influences nucleosome occupancy since DNA sequences with a high GC content and periodic dinucleotide patterns, that are devoid of poly(dA:dT) sequences, are strongly favored for nucleosome occupancy because of biophysical constraints that allow such sequences to wrap more easily around nucleosomes. These constraints have led to the development of algorithms that predict the probability of nucleosome occupancy [53,54]. In common with centromeres of many organisms, fission yeast centromeric DNA is AT-rich with a higher frequency of poly(dA:dT) tracts. It is therefore possible that H3 nucleosomes have a lower affinity for such sequences whereas CENP-A nucleosomes may be unperturbed by such AT rich DNA. To examine the underlying sequence specificity within centromeric DNA that favours the deposition of CENP-A\textsuperscript{Cnp1} nucleosomes, the sequence of LM DNA was altered by randomisation using a 5 bp sliding window throughout the entire 2 kb element. This generated a synthetic LM sequence (SynR-LM) that is 62.6% identical to the wild-type LM sequence, retaining the same AT content and dinucleotide periodicity, and thus the same predicted nucleosome occupancy as the wild-type LM element (Fig. 6A) [55].

Synthesised SynR-LM assembled as a 3xSynR-LM tandem array was placed in the same plasmid backbone as p3xLM to generate p3xSynR-LM. p3xSynR-LM was transformed into wt-CENP-A\textsuperscript{Cnp1} and hi-CENP-A\textsuperscript{Cnp1} cells. In contrast to p3xLM, CENP-A\textsuperscript{Cnp1} was not detectable
Fig 5. Tandem copies of the 2 kb LM region adjacent to heterochromatin/outer repeat are sufficient to establish a functional centromere in wild-type cells. (A) Schematic representation of plasmids containing 5.6 kb of heterochromatin-forming outer repeat element flanking a full length cc2 (pH-cc2) or three tandem repeats (pH-3xLM). (B) Minichromosome frequency of establishment.

<table>
<thead>
<tr>
<th>Minichromosome</th>
<th>Frequency of establishment</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH-cc2</td>
<td>11.2% (61/542)</td>
</tr>
<tr>
<td>pH-3xLM</td>
<td>5.8% (25/259)</td>
</tr>
<tr>
<td>pH-3xOP</td>
<td>0% (0/280)</td>
</tr>
</tbody>
</table>

C1: ChIP: CENP-A$^{Cnp1}$

C2: ChIP: CENP-A$^{Cnp3}$

C3: ChIP: CENP-A$^{Cnp3}$

C4: ChIP: CENP-A$^{Cnp3}$

Fig 5. Tandem copies of the 2 kb LM region adjacent to heterochromatin/outer repeat are sufficient to establish a functional centromere in wild-type cells. (A) Schematic representation of plasmids containing 5.6 kb of heterochromatin-forming outer repeat element flanking a full length cc2 (pH-cc2) or three tandem repeats (pH-3xLM).
on the shuffled LM sequence of pSynR-LM (Fig. 6B, compare with Fig. 4B). These analyses demonstrate that preservation of nucleotide composition (AT-content, dinucleotide periodicity) and predicted nucleosome occupancy within an altered centromeric DNA is not sufficient to allow CENP-A<sup>Cnp1</sup> deposition. The fact that the natural 2 kb LM sequence is active whereas the artificial SynR-LM is inactive reveals that the primary sequence of wild-type centromeric LM DNA encodes properties that somehow allow its recognition in vivo and consequent de novo assembly of CENP-A<sup>Cnp1</sup> chromatin.

**Centromeric DNA produces an unusual transcriptional environment**

Upon transformation into cells innate features within 3xLM sequence must allow it to be either immediately assembled in CENP-A<sup>Cnp1</sup> chromatin, or, initially assembled in H3 chromatin with subsequent remodelling that exchanges canonical H3 for CENP-A<sup>Cnp1</sup>. The process of transcription is obviously accompanied by chromatin remodelling and non-coding transcripts

---

Fig 6. Genetic information contained within DNA sequence is required for centromeric chromatin establishment. (A) Comparison between ChiP-seq data for CENP-A<sup>Cnp1</sup> ([49]; blue) and nucleosome positioning prediction (algorithm developed by [55]; in grey) for the LM region of cc2 or a mutagenised synthetic version of LM (SynRLM). The algorithm predicts nucleosome peaks within the central domain that match those mapped in vivo by ChiP-seq. SynRLM was designed by randomising the LM sequence in a 5 bp window and taking into consideration both periodicity of nucleosome positions and AT/GC content distribution of the original LM sequence. (B) ChiP analysis of CENP-A<sup>Cnp1</sup> levels of wild-type (wt-CENP-A<sup>Cnp1</sup>) and cells overexpressing CENP-A<sup>Cnp1</sup> (hi-CENP-A<sup>Cnp1</sup>) transformed with a plasmid containing three tandem repeats of SynRLM (p3xSynRLM). CENP-A<sup>Cnp1</sup> levels are shown relative to cc1/3 (n = 3).

doi:10.1371/journal.pgen.1004986.g006
synthesised from within the central \( \text{CENP-A}^{\text{Cnp1}} \) domains of fission yeast centromeres have been detected \[37,42\]. The transcription of central domain DNA might influence the assembly of \( \text{CENP-A}^{\text{Cnp1}} \) chromatin. In cells expressing \( \text{CENP-A}^{\text{Cnp1}} \) at wild-type levels, plasmid-borne central domain sequences are assembled in H3 rather than \( \text{CENP-A}^{\text{Cnp1}} \) chromatin (Fig. 1E). Higher levels of RNAPII are detected on plasmid-borne central domain sequences (pMcc2) introduced into wild-type cells than when \( cc2 \) is assembled in \( \text{CENP-A}^{\text{Cnp1}} \) chromatin on pMcc2 or at endogenous centromeres (Fig. 7A, S3 Fig.). Although relatively high levels of RNAPII associate with the pMcc2 central domain when assembled as H3 chromatin in wild-type cells (10–30% of levels at \( act1^+ \)) (Fig. 1E, Fig. 7A), the level of transcripts emanating from the central domain is very low (<0.1% of \( act1^+ \)), even when analysed in exosome defective cells (\( dis3–54 \); Fig. 7B). Thus, although ample RNAPII is recruited to the central domain of pMcc2 few transcripts are generated, suggesting that transcriptional stalling occurs.

To map transcriptional start sites (TSSs) within the LM and OP regions, 5' RACE was performed on RNA extracted from \( dis3–54 \) exosome mutant cells harbouring p3xLM or p3xOP (Fig. 7C, S4 Fig.). Many TSSs were identified within LM and OP, suggesting that these regions contain several promoters (Fig. 7C). 200 bp regions from both LM and OP were tested for their ability to drive production of \( \beta \)-galactosidase when placed upstream of a \( lacZ \) reporter in fission yeast and as shown in Fig. 7D, the regions displayed promoter activity. Mutated or inverted versions of promoter region M2 did not promote transcription of \( LacZ \) (S4 Fig.). Whilst most regions of LM and OP exhibit promoter activity that is lower than that of \( nmt81 \) control promoter, it is notable that region-O1 and region-P2 from OP have equivalent and 10-fold higher activity, respectively (Fig. 7D). It is possible that the higher promoter activity possessed by some regions of OP may affect its ability to establish \( \text{CENP-A}^{\text{Cnp1}} \). We surmise that the central domain from \( cen2 \) is peppered with promoters that can drive the production of transcripts on both strands. Their relative arrangement along with the strength and pattern of transcription may affect \( \text{CENP-A}^{\text{Cnp1}} \) incorporation.

**\( \text{CENP-A}^{\text{Cnp1}} \) establishment is enhanced in mutants that increase RNAPII stalling**

The progression of RNAPII is impeded by obstacles such as nucleosomes, DNA damage, bound proteins and by sequences that are intrinsically difficult to transcribe, causing transcriptional pausing, stalling or arrest [56]. RNAPII-associated proteins ease the passage of RNAPII through such impediments, contributing to the processivity of the polymerase [57]. TFIIIS facilitates transcriptional elongation of stalled/backtracked RNAPII by stimulating cleavage of nascent transcripts [58–60]. Upon stalling an elongating RNAPII becomes mono- then poly-ubiquitylated on the largest Rpb1 subunit. A rescue pathway involving de-ubiquitylation by the ubiquitin hydrolase Ubp3 is deployed to restart stalled RNAPII [56,61].

Our analyses suggest that the central domain chromatin landscape contains numerous promoters on both strands and multiple TSSs. In addition, long poly(dA:dT) tracts are likely to be an intrinsically problematic sequence for RNAPII transcription and present a barrier to RNAPII elongation [62,63]. We reasoned that mutants that are defective in the response to transcriptional stalling might influence the ability of the central domain to become assembled in \( \text{CENP-A}^{\text{Cnp1}} \) chromatin. To test this possibility, wild-type and TFIIIS (\( tfs1 \)Δ) mutant cells expressing hi-\( \text{CENP-A}^{\text{Cnp1}} \) were transformed with pMcc2. Surprisingly, slightly increased levels of \( \text{CENP-A}^{\text{Cnp1}} \) were detected on pMcc2 in the \( tfs1 \)Δ mutant compared to wild-type cells, suggesting that loss of TFIIIS promotes \( \text{CENP-A}^{\text{Cnp1}} \) deposition (S5 Fig.). Consistent with this, even when pMcc2 was transformed into \( tfs1 \)Δ cells expressing wt-\( \text{CENP-A}^{\text{Cnp1}} \) levels, \( \text{CENP-A}^{\text{Cnp1}} \) was detected on the pMcc2 central domain (Fig. 8A). In order to determine whether the
effect on CENP-A<sup>Cnp1</sup> establishment was specific to <i>tfs1</i>Δ or a general consequence of increased RNAPII stalling, we also investigated if loss of the ubiquitin hydrolase <i>Ubp3</i>, which normally rescues arrested RNAPII, affects CENP-A<sup>Cnp1</sup> deposition. Strikingly, CENP-A<sup>Cnp1</sup> was detected at high levels on central domain sequences in <i>ubp3</i>Δ cells transformed with pMcc2. CENP-A<sup>Cnp1</sup> was also detected on p3xLM, but not p3xOP in <i>ubp3</i>Δ (Fig. 8B, S6 Fig.). CENP-CC<sup>Cnp3</sup> and CENP-K Sim<sup>4</sup> centromere proteins were also significantly enriched on pMcc2 in <i>ubp3</i>Δ cells (S8 Fig.). These effects were not due to increased abundance of CENP-A<sup>Cnp1</sup> in <i>tfs1</i>Δ or <i>ubp3</i>Δ cells as protein levels were similar to wild-type cells (S7 Fig.). In fact, a reduction in CENP-A<sup>Cnp1</sup> and CENP-CC<sup>Cnp3</sup> levels was detected at endogenous centromeres in <i>ubp3</i>Δ, but not <i>tfs1</i>Δ cells (S8 Fig.). Tfs1 and Ubp3 were previously reported to modulate RNAi-
Fig 8. Mutants that affect RNAPII elongation allow de novo establishment of CENP-A Cnp1 chromatin. (A) ChIP for CENP-A Cnp1 on a pMcc2 plasmid transformed in wild-type (wt) or a strain deleted for TFIIS (tfs1Δ) (n = 3). (B) ChIP for CENP-A Cnp1 on a pMcc2 plasmid transformed in wild-type (wt) or a strain deleted for Ubp3 (ubp3Δ) (n = 3). (C) ChIP for RNAPII on pMcc2 plasmid in wt, tfs1Δ and ubp3Δ. p<0.05 is indicated with an asterisk. (D) Enrichment of RNAPII phosphorylated at Ser2 of its CTD (RNAPII-Ser2P) at the pMcc2 plasmid in wt, tfs1Δ and ubp3Δ. p<0.05 is indicated with an asterisk (n = 3). (E)
independent heterochromatin assembly [64]. To test whether the effect on CENP-ACnp1 establishment in tfs1Δ or ubp3Δ cells could be due to spurious assembly of heterochromatin on pMcc2, H3K9me2 ChIP was performed. The level of H3K9me2 on pMcc2 in tfs1Δ and ubp3Δ was similar to that on a negative control locus, act1+, and assembly of CENP-ACnp1 on pMcc2 in these mutants was not dependent on the H3K9-methyltransferase Clr4 (S9 Fig.). Thus, CENP-ACnp1 assembly on pMcc2 in the absence of TFIIS or Ubp3 does not result from induction by ectopic heterochromatin.

If lack of TFIIS or Ubp3 hinders transcriptional elongation, an increased level of RNAPII would be expected on affected chromatin templates. Elevated levels of Rpb1/RNAPII were detected on the central domain of pMcc2 in tfs1Δ (TFIIS) and ubp3Δ cells (Fig. 8C). In addition, increased levels of the elongation-specific Phospho-Ser2 form of RNAPII were observed on the central domain of pMcc2 in ubp3Δ cells, suggestive of failure to efficiently clear stalled RNAPII (Fig. 8D). Thus, two mutants, which perturb the progress of RNAPII elongation complexes in different ways, lead to deposition of CENP-ACnp1. These observations suggest that altering the transcriptional properties of the central domain chromatin through increased RNAPII stalling creates an environment that is permissive for establishment of CENP-ACnp1 in place of H3 chromatin.

Discussion

It is thought that once established, CENP-A chromatin has the ability to be ‘self-propagating’, and through the recruitment of factors that are themselves involved in deposition of CENP-A, it ensures its own maintenance [16,17,23,24,26,65]. Epigenetic inheritance can be defined as the propagation of a state in the absence of the initial inducer of that state. In this study, the inducer—overexpression of CENP-ACnp1—causes an event that would not normally occur, the assembly of CENP-ACnp1 chromatin on episomal centromeric DNA (pMcc2). When CENP-ACnp1-assembled pMcc2 is crossed from hi-CENP-ACnp1 cells into wt-CENP-ACnp1 cells, CENP-ACnp1 is propagated in the absence of the initial inducer through many generations and through meiosis. These observations further strengthen the evidence that CENP-A behaves as a bona fide epigenetic entity [24].

It is clear that both epigenetic and genetic factors influence CENP-A assembly. We have investigated the role of DNA sequence in establishment of CENP-A chromatin in fission yeast, an organism where analysis is not confounded by repetitive arrays of short satellite sequences. CENP-ACnp1 is normally restricted to the central domain of centromeres where it forms the basis for the kinetochore. Central domain DNA is a preferred substrate for establishment of CENP-ACnp1 chromatin upon overexpression, whilst other genomic loci do not support accumulation of high levels of CENP-ACnp1 [47], and even vector DNA adjacent to the central domain is not a good substrate. Conditions and mechanisms that influence assembly of CENP-ACnp1 on naïve plasmid DNA are also able to convert pre-chromatinised cc2 present on episomal plasmids or linear minichromosomes. What makes central domain DNA a preferred site for CENP-ACnp1 assembly? The lack of homology between cc2 and cc1/cc3 sequences suggests that it is not a simple case of specific sequence that is critical [66–68]. Our analyses
indicate that there is functional redundancy within the central domain and no one particular sequence is either necessary or sufficient for CENP-A<sup>Cnp1</sup> establishment, consistent with previous findings [50]. Despite this redundancy it appears that there are inherent distinctions between different regions of cc2. The 2 kb sub-regions, LM and OP, are functionally non-equivalent and consistently behaved differently when challenged to assemble CENP-A<sup>Cnp1</sup> chromatin. LM is competent to establish centromeric chromatin upon CENP-A<sup>Cnp1</sup> overexpression, contains sufficient information to make a functional centromere when placed next to heterochromatin (pH-3xLM), and assembles CENP-A<sup>Cnp1</sup> chromatin in cells lacking Ubp3. On the other hand, the OP region fails to become assembled in CENP-A<sup>Cnp1</sup> chromatin in all these situations, yet can accept CENP-A<sup>Cnp1</sup> when adjacent to one copy of LM, which apparently acts as an initiator. The ability of LM, but not OP, to substitute for full-length cc2 sequence indicates that not all sequences are equivalent and LM must contain all information necessary to make this region permissive for CENP-A<sup>Cnp1</sup> establishment. It is possible that the observed higher promoter activity observed in the OP region (Fig. 7D) prevents stabilisation of CENP-A<sup>Cnp1</sup> nucleosomes on this sequence.

In common with many organisms, the central domain of <i>S. pombe</i> centromeres is AT rich and this property might contribute to the propensity of centromeric DNA to attract CENP-A [5,68]. <i>S. pombe</i> central domain DNA has an AT content of 72% (genome average of 64%), as does the establishment competent LM sequence. However, other regions that alone fail to support CENP-A<sup>Cnp1</sup> establishment have a similar AT content, such as OP (71% AT) and intergenic regions (72% AT). Moreover, randomisation of the LM sequence resulted in SynR-LM that, even with identical nucleotide composition (72% AT), was incompetent for CENP-A<sup>Cnp1</sup> establishment. Thus, high AT content alone, even when it mimics natural nucleosome positioning predictions, is not a defining factor in CENP-A<sup>Cnp1</sup> assembly. Together our observations indicate that rather than there being a specific critical sequence, central domain sequences encode unique properties capable of triggering or promoting the establishment of CENP-A<sup>Cnp1</sup> nucleosomes.

Transcription-coupled remodelling is associated with the deposition of histone variants and could potentially contribute to the assembly of CENP-A chromatin [69,70]. However, the simple act of transcription cannot be sufficient to provide specificity to the deposition of CENP-A. Our observations suggest that the transcriptional landscape of the centromeric central domain is unusual: scattered promoters of various strengths resulting in pervasive low quality transcription and numerous TSSs on both strands, in conjunction with poly(dA:dT) tracts that are inherently difficult to transcribe are likely to cause collision between convergently transcribing RNAPIIs and pile-ups at difficult sequences [63,71]. The relatively high density of RNAPII on pMcc2 contrasts with very low levels of transcripts (Fig. 7), consistent with inefficient progress of transcription by RNAPII on cc2, and many stalled elongation complexes. In addition, long tracts of poly(dA:dT) are known to disfavour nucleosome assembly, consistent with the apparently wide spacing of nucleosomes at endogenous centromeres [49,72]. These regions may be de facto nucleosome free regions, similar to those at promoters, allowing cryptic initiation of transcription to occur [72,73]. The randomized synthetic sequence SynR-LM that is a poor substrate for CENP-A<sup>Cnp1</sup> deposition has similar long A tracts, but transcription-related sequence-sensitive elements—such as promoters and transcription factor binding sites—would be destroyed. Thus, the central domain, due to its sequence-encoded properties, may produce a distinctive chromatin and transcriptional environment.

CENP-A<sup>Cnp1</sup> chromatin does not assemble de novo on cc2 sequence alone in wild-type cells expressing normal CENP-A<sup>Cnp1</sup> levels [45]. Instead, we envisage that the unique transcriptional chromatin environment created by the cc2 sequence renders it permissive for CENP-A<sup>Cnp1</sup> establishment, but that establishment occurs only if other favourable conditions exist. CENP-
A\(^{Cnp1}\) is preferentially incorporated on these central domain sequences upon overexpression, when adjacent to heterochromatin, and in the absence of factors that usually enhance transcriptional elongation. Any explanation of CENP-A\(^{Cnp1}\) chromatin establishment on central domain DNA must also account for how CENP-A\(^{Cnp1}\) is incorporated instead of H3. Serine 2 in the CTD heptad repeat of Rpb1 is phosphorylated in elongating RNAPII, and this Ser2P-Rbp1/RNAPII becomes ubiquitylated upon stalling [74–76]. The ubiquitin hydrolase Ubp3 normally acts as a proof-reading activity to prevent degradation of stalled but rescuable RNA-P II [56,61]. Absence of Ubp3 compromises the processing of stalled RNAPII, resulting in the accumulation of ubiquitylated Ser2P-Rbp1/RNAPII complexes. We propose that such modifications contribute to the distinctive status of central domain chromatin, leading to recruitment of factors that promote CENP-A\(^{Cnp1}\) deposition (Fig. 8E). Alternatively, it may create an environment in which H3 nucleosomes are efficiently turned over/evicted, whereas CENP-A\(^{Cnp1}\) nucleosomes are poorly evicted specifically in the context of stalled RNAPII. In cells lacking Ubp3, severe or prolonged stalling, even with normal levels of CENP-A\(^{Cnp1}\), would provide extended opportunities for CENP-A\(^{Cnp1}\) recruitment, or poor eviction of CENP-A\(^{Cnp1}\) during prolonged stalling. TFIIS promotes transcriptional elongation by cleaving nascent transcripts in the context of stalled/backtracked RNAPII [57,58,77]. Although the effects of TFIIS deletion are more subtle than lack of Ubp3, the accumulation of RNAPII correlates with assembly CENP-A\(^{Cnp1}\) chromatin, supporting a mechanism where persistent RNAPII stalling within central domain triggers remodelling that results in CENP-A\(^{Cnp1}\) deposition.

In this model, when naïve central domain DNA (pMcc2) is introduced into wild-type cells, transient stalling occurs but it is efficiently cleared with the aid of factors such as TFIIS and Ubp3 (Fig. 8E). Because in wild-type cells CENP-A\(^{Cnp1}\) levels are extremely low compared to histone H3 there would be little opportunity for CENP-A\(^{Cnp1}\) to gain access to cc2, and with efficient clearing of stalled RNAPII, CENP-A\(^{Cnp1}\) would fail to accumulate in cc2 [78]. CENP-A\(^{Cnp1}\) overexpression would increase the probability of interaction with the transiently stalled RNAPII in central domain chromatin, increasing the likelihood of recruitment. Alternatively, increased access coupled with poor eviction would lead to CENP-A\(^{Cnp1}\) accumulation. In addition, CENP-A\(^{Cnp1}\) nucleosomes themselves, which have distinct N-terminal tails that lack the conserved lysine residues of H3 whose modification aids transcription, are likely to present a greater barrier to transcription than H3 nucleosomes [79]. Thus, once incorporated, CENP-A\(^{Cnp1}\) nucleosomes might exacerbate the poor transcriptional elongation, creating conditions permissive for recruitment of more CENP-A\(^{Cnp1}\) in a self-perpetuating system. Longer regions of central domain DNA would have greater probability of triggering stalling events and thus be more likely to initiate the incorporation of CENP-A\(^{Cnp1}\). In the context of this model, heterochromatin could promote establishment of CENP-A\(^{Cnp1}\) chromatin on adjacent cc2 sequence by drawing plasmids to sites of endogenous heterochromatin such as the spindle pole body where they would encounter a higher concentration of CENP-A\(^{Cnp1}\) than non-heterochromatinized plasmids located in the nuclear interior [80]. Alternatively, heterochromatin-associated chromatin modifying activities may influence transcriptional elongation by RNAPII within cc2, causing enhanced stalling and deposition of CENP-A\(^{Cnp1}\)[41].

Following establishment of CENP-A chromatin and kinetochore assembly, transcription could play a proof-reading role that evicts H3 deposited at centromeres during S phase [81]. Indeed, transcription and RNAPII have been detected at centromeres in mammalian cells and transcription/RNAPII may play a role in centromere integrity [33,34,38]. Transcription of human \(\alpha\)-satellite arrays introduced as HACs is known to occur. Although CENP-A assembly is compatible with targeting of mild transcriptional activators, targeting of a strong transcriptional activator is deleterious [30,38,82]. Thus transcription and/or the transcription-coupled
histone modifications detected at centromeres may promote CENP-A deposition at mammalian centromeres.

In conclusion, we show that the sequence of fission yeast centromere central domain DNA is important only in so far as it encodes for certain properties that contribute to the region’s unusual chromatin and transcriptional landscape. Establishment of CENP-A\textsuperscript{Cnp1} chromatin is driven by these sequence-encoded properties that when combined with the presence of nearby heterochromatin, overexpressed CENP-A\textsuperscript{Cnp1} or increased RNAPII stalling, tips the balance in favour CENP-A\textsuperscript{Cnp1} chromatin assembly. It seems likely that a similar combination of factors, which together favour CENP-A incorporation, must also contribute to the formation of new centromeres at novel chromosomal locations.

**Materials and Methods**

**Cell growth and manipulation**

Standard genetic and molecular techniques were followed. Fission yeast methods were as described [83]. Fission yeast strains are listed in Table 1. Minichromosomes used in this study were transformed by electroporation. Transformants were selected by growth on PMG—ura—ade at 32°C. As circular minichromosomes lack heterochromatin and therefore centromeric cohesion, plasmids were maintained in cells by selection in medium lacking adenine and uracil. 3 independent colonies from each transformation were analysed for the presence of kinetochore proteins by chromatin immunoprecipitation (ChIP).

**Centromere plasmids and minichromosomes**

Plasmids bearing centromere fragments contained a minimal \textit{ars1} element to ensure efficient replication in \textit{S. pombe}, in addition to selectable markers \textit{sup3-5} (complements \textit{ade6-704}), \textit{ura4}\textsuperscript{+} and KAN\textsuperscript{R}. 8.5 kb of central domain DNA (\textit{cc2} plus inner part of \textit{imr2L} and \textit{imr2R}) was cloned into the multiple cloning site as a SalI–NcoI fragment to create pMcc2. Various sub-fragments of \textit{cc2} (J–Q) were amplified by PCR and cloned into the multiple cloning site as BamHI/BglII fragments. 5.6 kb of heterochromatin-forming outer repeat sequence was inserted adjacent to central domain sequences to test ability to form functional centromeres.

A plasmid, pMC28, bearing \textit{cc2}, a KAN resistance marker and an inverted \textit{ura4} sequence was constructed from pMcc2. Linearisation of the plasmid at \textit{NotI} within the inverted \textit{ura4} sequence allowed integration at \textit{ura4}\textsuperscript{+} located on the arm of Ch16-m23: \textit{ura4}\textsuperscript{+}. Ch16-m23: \textit{ura4}\textsuperscript{+} is a derivative of Ch16, a 530 kb minichromosome, itself derived from Chromosome III [48]. It

<table>
<thead>
<tr>
<th>Table 1. Schizosaccharomyces pombe strains.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A7373 cc2Δ::cc1 ade6-704-HYGMX6 ura4-</td>
</tr>
<tr>
<td>A7408 cc2Δ::cc1 ars1:nmt41-GFP-cnp1-NAT ade6-704-HYGMX6 ura4-</td>
</tr>
<tr>
<td>A7528 cc2Δ::cc1 dis3–54 ade6-704-HYGMX6 ura4-</td>
</tr>
<tr>
<td>A9235 cc2Δ::cc1 tfs1Δ::leu2 ade6-704-HYGMX6 ura4-</td>
</tr>
<tr>
<td>A9391 cc2Δ::cc1 ubp3Δ::KAN ade6-704-HYGMX6 ura4-</td>
</tr>
<tr>
<td>A9491 cc2Δ::cc1 tfs1Δ::leu2 clr4Δ::NAT ade6-704-HYGMX6 ura4-</td>
</tr>
<tr>
<td>A9477 cc2Δ::cc1 ubp3Δ::KAN clr4Δ::NAT ade6-704-HYGMX6 ura4-</td>
</tr>
<tr>
<td>A9048 cc2Δ::cc1 [Ch16 m23 ura4::cc2-KAN ade6-216] ade6-210 leu1-32 ura4-</td>
</tr>
<tr>
<td>A9054 cc2Δ::cc1 [Ch16 m23 ura4::cc2-KAN ade6-216] ars1:nmt41-GFP-Cnp1-NAT ade6-210 ura4-</td>
</tr>
<tr>
<td>A9059 cc2Δ::cc1 [Ch16-ade6-216] ade6-210 leu1-32 ura4-</td>
</tr>
<tr>
<td>A9066 cc2Δ::cc1 [Ch16-ade6-216] ars1:nmt41-GFP-Cnp1-NAT ade6-210 ura4-</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pgen.1004986.t001
also bears the ade6-216 allele which complements the ade6-210 allele present on endogenous Chromosome III by interallelic complementation. Integration of linearised pMC28 on Ch16-m23:ura4 allowed selection on the counter-selective drug 5-fluoro-orotic acid and G418 (KAN). Cells that lost the Ch16-m23:ura4::cc2-KAN (abbreviated as Ch16-cc2) became red on limiting adenine and were sensitive to G418. For growth in liquid, cells containing Ch16-cc2 were grown in media lacking adenine.

**ChIP**

ChIP was performed as previously described [84] using anti-CENP-A<sup>cnp1</sup> antibody, anti CENP-C<sup>Cmp3</sup> antibody, anti-CENP-K<sup>Sim4</sup> antibody, anti-H3 antibody (ab1791; Abcam.), anti-H3K9me2 antibody (T. Urano) and anti-total RNA polymerase II (4F8; 61081, Active Motif), anti-Rpb1-Ser2P (3E10; 61083, Active Motif) and analysed by qPCR. Primers are listed in Table 2. P-values were calculated by standard t-test on 3 replicates between wild-type and mutant; p<0.05 was considered significant.

**Establishment assay**

For the establishment assay, cells were transformed with minichromosomes (containing 5.6 kb of outer repeat sequence in addition to cc2 sequences), by electroporation with ~ 200 ng of DNA and plated on selective medium. Resultant colonies were replicated onto rich medium containing limiting adenine. The presence of pale pink/white colonies indicates establishment of a functional centromere on the minichromosome. Establishment efficiency is calculated as percentage of these colonies divided by the total number of transformants. Colonies were streaked on limiting adenine plates to confirm the presence of sectoring that is indicative of centromere function.

**Real-time PCR (qPCR)**

Quantitative PCR reactions were carried out in 10 μl volume, with 5μl Light Cycler 480 Sybr-Green Master Mix (Roche), 0.5μl each primer (10 μM) and 3μl ChIP or total template. The data were analysed using Light Cycler 480 Software 1.5 (Roche).

5′RACE-PCR and RT-PCR

5′RACE-PCR was performed as previously described [37]. In brief, RNA was isolated with RNaseasy mini/midi kit (Qiagen) according to the manufacturer’s protocol. Poly(A) containing RNA was purified from 500 μg of total RNA by affinity purification with biotinylated oligo-dT using PolyATtract mRNA Isolation Systems (Promega). 5′RACE PCR was performed using SMARTer 5′/3′ RACE (Clontech) according to the manufacturer’s protocol. PCR products were then run on 1% agarose gel, purified and cloned into pGEM-T Easy vector (Promega) and subsequently sequenced. Reverse transcription reaction for 5′RACE and qRT-PCR was performed using Superscript III Reverse Transcriptase (Invitrogen) using RNA extracted from 3 independent colonies. For qRT-PCR, transcript levels were normalized over gDNA to take into account differences in copy number between plasmids and normalized relative to act1<sup>+</sup>.

**LacZ assay**

LacZ assay was performed as described [85]. pREP81X-LacZ was digested with XhoI and PstI and the nmt81 promoter upstream of LacZ replaced with sequences from centromere 2. Plasmids were transformed into wild-type and grown on minimal medium (n = 3).
Table 2. Primers used in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>qact1 Fw</td>
<td>CCCAAATCCAACCGTGAGAAGATG</td>
</tr>
<tr>
<td>qact1 rev</td>
<td>CCAGAGTCCAACAGATACCAGT</td>
</tr>
<tr>
<td>qcnt1 Fw</td>
<td>CAGACAATCCTGATGTTACTATC</td>
</tr>
<tr>
<td>qcnt1 rev</td>
<td>AGGTGAACCTGTAAGTGTG</td>
</tr>
<tr>
<td>qpDF16-vect2 Fw</td>
<td>AATATACCACGGCCACACGT</td>
</tr>
<tr>
<td>qpDF16-vect2 rev</td>
<td>CCAAGAAGCTGGTGAAGAT</td>
</tr>
<tr>
<td>qcnt2 L Fw</td>
<td>GAAGATGAGATAGGACTG</td>
</tr>
<tr>
<td>qcnt2 L rev</td>
<td>GTTATCTACATTCTGCTC</td>
</tr>
<tr>
<td>qcnt2 M Fw</td>
<td>ATGACATGGGCTGAAAGTC</td>
</tr>
<tr>
<td>qcnt2 M rev</td>
<td>CATTAACAAAACACCGGACAC</td>
</tr>
<tr>
<td>qcnt2 N Fw</td>
<td>TAAGCCAGCAAATCTGGAG</td>
</tr>
<tr>
<td>qcnt2 N rev</td>
<td>GAATTAATGACACACTGCA</td>
</tr>
<tr>
<td>qcnt2 O Fw</td>
<td>CTGATGATTACTCAAGAAAGC</td>
</tr>
<tr>
<td>qcnt2 O rev</td>
<td>CTGATATTCCTGACTTTAC</td>
</tr>
<tr>
<td>qcnt2 P Fw</td>
<td>AGGCGTGTCCATGCGAAAGG</td>
</tr>
<tr>
<td>qcnt2 P rev</td>
<td>TCTAGGATCCGACCTCGAGATGTTAGTGACTCTAC</td>
</tr>
<tr>
<td>CC2-F1 (J)</td>
<td>TACTACGGATCCCCCATGGAATAT</td>
</tr>
<tr>
<td>CC2-R1(J)</td>
<td>TACTACGTCGACCTCGAGATCTCGAGCTCTGCTCTGTGAA</td>
</tr>
<tr>
<td>CC2-F2 (K)</td>
<td>TACTACGGAATCCTGAGACACTGTTTACATCTAA</td>
</tr>
<tr>
<td>CC2-R2(K)</td>
<td>TACTACGTAACCTGAGATCTCGTAGTACTACGGGTCTCTGCTGTTT</td>
</tr>
<tr>
<td>CC2-F3(L)</td>
<td>TACTACGGAATCCTGAGATCTCGAGAATTGCTTACAGG</td>
</tr>
<tr>
<td>CC2-R3(L)</td>
<td>TACTACGTCGACCTGAGATCTCGAGACTACCTGGCAACAGC</td>
</tr>
<tr>
<td>CC2-F4(M)</td>
<td>TACTACGGAATCCTGAGATCTCGAGACTACCTGGCAACAGC</td>
</tr>
<tr>
<td>CC2-R4(M)</td>
<td>TACTACGTCGACCTGAGATCTCGAGACTACCTGGCAACAGC</td>
</tr>
<tr>
<td>CC2-F5(N)</td>
<td>TACTACGGAATCCTGAGATCTCGAGACTACCTGGCAACAGC</td>
</tr>
<tr>
<td>CC2-R5(N)</td>
<td>TACTACGGAATCCTGAGATCTCGAGACTACCTGGCAACAGC</td>
</tr>
<tr>
<td>RegionL1-Lacz Fw</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionL1-Lacz rev</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionL2-Lacz Fw</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionL2-Lacz rev</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionM1-Lacz Fw</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionM1-Lacz rev</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionM2-Lacz Fw</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionM2-Lacz rev</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionO1-Lacz Fw</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionO1-Lacz rev</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionO2-Lacz Fw</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionO2-Lacz rev</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionP1-Lacz Fw</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionP1-Lacz rev</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionP2-Lacz Fw</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>RegionP2-Lacz rev</td>
<td>TACTACCGACTGACCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>probe cc2 F</td>
<td>CGTGCACATTTGTGAAAAAGG</td>
</tr>
<tr>
<td>probe cc2 rev</td>
<td>TCTGGAGATCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
<tr>
<td>probe cc1 F</td>
<td>CCATTTGTAGATCCACTGCTG</td>
</tr>
<tr>
<td>probe cc1 rev</td>
<td>TCTGGAGATCTGGGAGATTAGATGATTACGTTTCCACC</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pgen.1004986.t002
Southern analysis

DNA was extracted as previously described [83]. The DNA was digested with BglII/SpeI or SphI/SpeI, run on a 1% agarose gel, blotted on nylon membrane (Hybond N, Amersham) and UV-crosslinked. The membrane was hybridized with DNA probes specific for central domain 1 or central domain 2. To make the probes, PCR products were used as template in the labelling reaction using High Prime (Roche). Primers sequences are listed in Table 2.

Western analysis

Western analysis was performed as described previously using anti-GFP antibody (Roche) and anti-H3 antibody (ab1794-abcam) [86]. The intensities of GFP and H3 signals were acquired using LICOR Odyssey Infrared Imaging System software (Li-COR Bioscience).

Supporting Information

S1 Fig. Replacement of central core 2 sequence from endogenous centromere 2 with central core 1 DNA. (A) Schematic representation of endogenous centromere 1 (cnt1), centromere 2 (cnt2) and central core 2 deletion (cc2Δ). The probes on cc2 (grey) and cc1 (black) used for Southern analysis and the expected size of DNA fragments following restriction enzyme digestion are indicated. (B) Genomic DNA was extracted from wild-type cells and cc2Δ strain. DNA was digested with BglII and SpeI and the Southern hybridised with the 2.8 kb cc1 probe. (C) Genomic DNA was extracted from wild-type cells and the cc2Δ strain. The DNA was digested with SphI and SpeI and the Southern hybridised with the 2.8 kb cc2 probe. The absence of the band for cc2 indicates its deletion from the endogenous cen2 locus. (D) Spotting assay of wild-type or clr4Δ cells containing centromere 2 sequence at the endogenous centromere (cc2+) or wild-type cells with cc2 sequence replaced by cc1 (cc2Δ). YES: complete medium. Phloxin B stains dead cells red. Cells with impaired centromere function (e.g. clr4Δ) show sensitivity to the microtubule destabilizing drug thiabendazole (TBZ), while cc2+ and cc2Δ are able to grow in the presence of the drug TBZ suggesting that cc2Δ does not affect centromere function.

S2 Fig. Plasmids containing a combination of the LM and OP fragments are able to CENP-A\(^{\text{Cnp1}}\). (A) ChIP analysis of CENP-A\(^{\text{Cnp1}}\) levels of wild-type (wt-CENP-A\(^{\text{Cnp1}}\)) and cells over-expressing CENP-A\(^{\text{Cnp1}}\) (hi-CENP-A\(^{\text{Cnp1}}\)) transformed with a plasmid containing a combination of LM and OP fragment (OP-LM-OP). (B) Establishment assay using plasmids containing 5.6 kb of heterochromatin-forming outer repeat element flanking a full length cc2 (pH-cc2), three tandem repeats of the LM region (pH-3xLM), three tandem repeats of the OP region (pH-3xOP) or one copy of the LM fragment adjacent to two copies of OP (pH-LM-2xOP). pH-cc2, pH-3xLM, pH-3xOP and pH-LM-2xOP were transformed in wild-type cells and the percentage of transformants that contained plasmids with centromere function was assessed by replica plating onto limiting adenine medium. White colonies indicate formation of functional centromeres.

S3 Fig. RNA polymerase II is present on central core sequences when CENP-A\(^{\text{Cnp1}}\) is not established. ChIP of RNA polymerase II (RNAPII) on the p3xLM (A) and p3xOP (B) plasmids transformed into wild-type cells. Positions of the PCR products quantified for each plasmid are indicated. (C) ChIP of RNA polymerase II (RNAPII) on the pMcc2 plasmid transformed into wild-type (wt-CENP-A\(^{\text{Cnp1}}\)) and cells overexpressing CENP-A\(^{\text{Cnp1}}\) (hi-CENP-A\(^{\text{Cnp1}}\)). The enrichment is shown as relative to act1\(^{+}\) (n = 3).
S4 Fig. Identification of Transcriptional Start Sites (TSS) within the 2 kb LM and OP regions. (A) Example of RT-PCR products obtained from the 5’RACE reaction. Generally, multiple products were identified for the centromeric regions, while one single product was identified from the act1 gene. (B) Sequences of TSSs detected by 5’RACE in LM and OP regions. (C) Analysis of promoter activity of the ~200 bp regions from cc2. Region-M1 and region-M2 were placed upstream of a LacZ reporter gene. The levels of LacZ expression were assessed by measuring absorbance at 420 nm of cell lysates incubated with 2-Nitrophenyl-β-D-galactopyranoside (ONPG). region-M2 inv: region cloned inverted; region-M2 mut: point mutations inserted in the M2 sequence. nmt81: positive control with nmt81 promoter (n = 3).

(S5) S5 Fig. Lack of S. pombe TFII S increases CENP-A<sup>Cnp1</sup> deposition at pMcc2 in hi-CENP-A<sup>Cnp1</sup> cells. Enrichment levels of CENP-A<sup>Cnp1</sup> at the endogenous centromeres (cc1/cc3) and at the pMcc2 plasmid transformed in wild-type cells (wt), in cells expressing high levels of CENP-A<sup>Cnp1</sup> (hi-CENP-A<sup>Cnp1</sup>) and in hi-CENP-A<sup>Cnp1</sup> in which the transcription factor TFII S (Tfs1) is deleted (tfs1Δ+hi-CENP-A<sup>Cnp1</sup>) (n = 3).

(S6) S6 Fig. Lack of Ubp3 allows establishment of CENP-A<sup>Cnp1</sup> chromatin on p3xLM but not p3xOP. ChIP for CENP-A<sup>Cnp1</sup> on a p3xLM and p3xOP plasmids transformed into wild-type (wt) or a strain deleted for Ubp3 (ubp3Δ). CENP-A<sup>Cnp1</sup> levels are shown relative to cc1/3.

(S7) S7 Fig. Deletion of ubp3 and tfs1 does not affect CENP-A<sup>Cnp1</sup> protein levels. Western analysis and protein levels comparison between wild-type (wt), ubp3Δ and tfs1Δ of endogenous N-terminal tagged GFP-CENP-A<sup>Cnp1</sup> and H3. Ponceau staining as internal loading control.

(S8) S8 Fig. Cells lacking Ubp3 show reduced levels of CENP-A<sup>Cnp1</sup> at endogenous centromeres. (A) ChIP for CENP-A<sup>Cnp1</sup>, CENP-C<sup>Cnp3</sup> and CENP-K<sup>Sim4</sup> levels at endogenous centromere cc1/3 in wild-type (wt), tfs1Δ and ubp3Δ. p<0.05 is indicated with an asterisk (n = 3). (B) ChIP for CENP-C<sup>Cnp3</sup> and CENP-K<sup>Sim4</sup> levels at pMcc2 plasmid transformed into wt, tfs1Δ and ubp3Δ. p<0.05 is indicated with an asterisk (n = 3).

(S9) S9 Fig. CENP-A<sup>Cnp1</sup> deposition on pcc2 plasmid is not dependent on H3K9me2 in the tfs1Δ and ubp3Δ mutants. (A) ChIP for H3K9me2 (A) and CENP-A<sup>Cnp1</sup> (B) at act1<sup>+</sup> and at pMcc2 plasmid transformed in wt, clr4Δ, ubp3Δ and ubp3Δclr4Δ. CENP-A<sup>Cnp1</sup> is deposited on pcc2 in ubp3 regardless of whether Clr4 is present or absent. (C) ChIP for H3K9me2 at act1<sup>+</sup> and on the pcc2 plasmid transformed into wt, tfs1Δ and tfs1Δclr4Δ cells. The absence of H3K9me2 on the plasmid suggests that heterochromatinisation is not responsible for CENP-A<sup>Cnp1</sup> deposition. (D) CENP-A<sup>Cnp1</sup> levels on act1<sup>+</sup> and at the pcc2 plasmid in wt, tfs1Δ and tfs1Δclr4Δ. The absence of the methyltransferase Clr4 does not affect CENP-A<sup>Cnp1</sup> deposition on pMcc2.

Acknowledgments

We thank all members of the Allshire lab for helpful advice and discussion. In particular, P. Tong for bioinformatics support, and R. Ard, P. Audergon, and M. Shukla for comments on the manuscript. We also thank S. Webb for designing the program for randomization of DNA
sequence. We are grateful to S. Forsburg for providing the pRep81X-LacZ plasmid and T. Urano for providing the H3K9me2 antibody (mAb5.1.1).

**Author Contributions**

Conceived and designed the experiments: SC ALP RCA. Performed the experiments: SC. Analyzed the data: SC ALP RCA. Wrote the paper: SC ALP RCA.

**References**


