

# Histone H3 variants specify modes of chromatin assembly

Kami Ahmad and Steven Henikoff\*

Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, A1-162, Seattle, WA 98109

Histone variants have been known for 30 years, but their functions and the mechanism of their deposition are still largely unknown. *Drosophila* has three versions of histone H3. H3 packages the bulk genome, H3.3 marks active chromatin and may be essential for gene regulation, and Cid is the characteristic structural component of centromeric chromatin. We have characterized the properties of these histones by using a *Drosophila* cell-line system that allows precise analysis of both DNA replication and histone deposition. The deposition of H3 is restricted to replicating DNA. In striking contrast, H3.3 and Cid deposit throughout the cell cycle. Deposition of H3.3 occurs without any corresponding DNA replication. To confirm that the deposition of Cid is also replication-independent (RI), we examined centromere replication in cultured cells and neuroblasts. We found that centromeres replicate out of phase with heterochromatin and display replication patterns that may limit H3 deposition. This confirms that both variants undergo RI deposition, but at different locations in the nucleus. How variant histones accomplish RI deposition is unknown, and raises basic questions about the stability of nucleosomes, the machinery that accomplishes nucleosome assembly, and the functional organization of the nucleus. The different *in vivo* properties of H3, H3.3, and Cid set the stage for identifying the mechanisms by which they are differentially targeted. Here we suggest that local effects of "open" chromatin and broader effects of nuclear organization help to guide the two different H3 variants to their target sites.

Nucleosomes are the fundamental units of chromatin, consisting of 146 bp of DNA wrapped around an octamer of four core histones. Histone deposition occurs primarily as DNA replicates to complete chromatin doubling (1). During S phase of the cell cycle, new histones are produced in abundance for immediate replication-coupled deposition. In most metazoans, this abundant S-phase synthesis results from the tight regulation of tens to hundreds of intronless histone genes that have special 3' untranslated regions instead of poly(A) tails (2). However, some histones are produced from orphan genes outside of S phase. In *Drosophila*, orphan genes encode two H3 variants: one encodes Cid, the centromeric histone (3), and two encode H3.3, the replacement variant (Fig. 1; refs. 4 and 5). These variants have equivalents in many other eukaryotes (6, 7). The H3.3 histone is nearly identical to H3, differing at only four amino acid positions. Cid differs profoundly from H3 in sequence, showing some significant identity only within the histone fold domain. Surprisingly, these three histones have different deposition properties. H3 and H3.3 are deposited as DNA replicates, but both H3.3 and Cid can be deposited at sites that are not undergoing DNA replication (Fig. 2; refs. 8 and 9). Whereas only a minor fraction of the bulk genome is packaged into Cid- and H3.3-containing nucleosomes, each variant is targeted to different specialized sites, with Cid localizing to centromeres and H3.3 to transcriptionally active genes. Specific localization of centromeric H3-like histones (CenH3s) has been observed in various animals (10), fungi (11, 12), and plants (13). Also, an H3.3-like histone targets the transcriptionally active macronucleus in

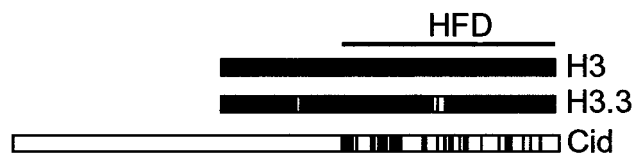


Fig. 1. *Drosophila* produces three versions of H3 histones. H3 and H3.3 are nearly identical throughout their N-terminal tail and histone fold domains (HFD), with only four amino acid residue differences. Cid is much more diverged, and can only be aligned to H3 in the HFD (49). Black boxes indicate identities to H3.

ciliates (14). Thus, the targeting of H3 variants is likely a feature of every eukaryotic cell, where centromeres and transcribed regions are the major loci of activity in metaphase and interphase, respectively. Both kinds of loci use a distinct pathway for nucleosome assembly (8, 9), and here we explore the properties of this process.

## H3 Variants Determine the Nucleosome Assembly Pathway

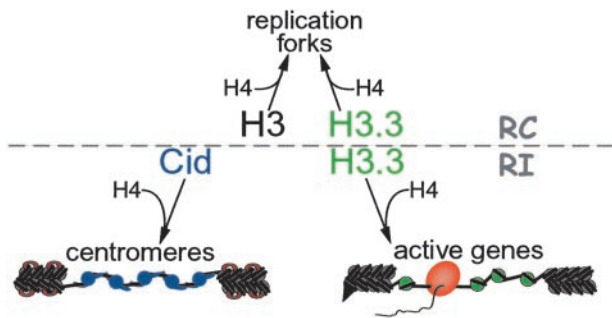
Studies of histone deposition have generally been done using crude extracts, purified components or pools of cells from which bulk chromatin is extracted (15). These methods reveal the average properties of chromatin, and have shown that the bulk of chromatin doubles as DNA replicates. Extensive *in vitro* work has demonstrated that the assembly of nucleosomes is a stepwise process in which deposition of an (H3·H4)<sub>2</sub> tetramer is followed by addition of two H2A·H2B dimers (1). The new histones are brought to the replication fork in a complex with chromatin assembly factor 1 (CAF1). CAF1 appears to be recruited to the replication fork by binding to the ring-shaped proliferating cell nuclear antigen (PCNA) that encircles the DNA template at each replication fork (16). Histones from the parent DNA are distributively segregated to the two sister chromatids behind the replication fork, and the gaps in their nucleosomal arrays are rapidly filled by step-wise assembly of new nucleosomes. These nucleosomes are then matured by addition of linker histones and covalent modification of histone tails to complete chromatin.

Nucleosomes containing H3 variants comprise only a small proportion of bulk chromatin, and thus their properties have been generally undetectable. However, replacement H3 variants can become enriched in the chromatin of nonreplicating cells (17–19). This means that other ways of depositing histones must exist; but because such variant enrichment was only detectable in unusual cell types (such as long-lived neurons or spermatogonia),

This paper results from Arthur M. Sackler Colloquium of the National Academy of Sciences, "Self-Perpetuating Structural States in Biology, Disease, and Genetics," held March 22–24, 2002, at the National Academy of Sciences in Washington, DC.

Abbreviations: RI, replication-independent; CenH3, centromeric H3-like histone; HP1, heterochromatin protein 1; H3K9me, H3-methylated lysine-9; PCNA, proliferating cell nuclear antigen.

\*To whom reprint requests should be addressed. E-mail: steveh@fhrc.org.



**Fig. 2.** The three versions of histone H3 determine the mode of nucleosome assembly. The deposition of H3 is strictly replication-coupled (RC), and H3 is recruited to replication forks for chromatin doubling. Deposition of Cid (blue) is exclusively replication-independent (RI), and normally occurs only at centromeres. H3.3 (green) undergoes RC deposition, and RI deposition at active loci. Open chromatin at centromeres and at active genes may promote histone replacement. Transcriptional activity, chromatin remodeling factors, and RNA polymerases (orange) will unfold the chromatin fiber and disrupt nucleosome (gray) structure (open chromatin). Transcriptionally inactive regions are not subjected to these forces and remain in a closed configuration. Flanking heterochromatin and H3-containing blocks within the centromeric domain presents the H3K9me epitope, thereby binding HP1 (red) and resulting in a compacted, closed chromatin structure. Cid-containing nucleosomes cannot be methylated in this way, and thus remain comparatively open. The specialized N-terminal tail of Cid may alter the linker DNA between nucleosomes, also contributing to the open chromatin configuration. RI deposition of H3.3 is limited to the open chromatin at active genes and RI deposition of Cid is limited to the open chromatin in centromeric domains.

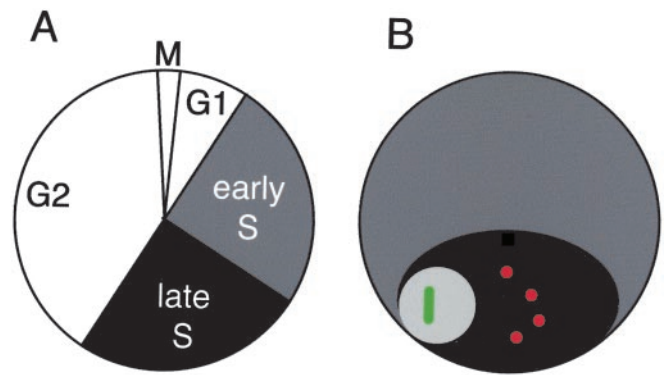
cytes), studies of the phenomenon have been limited. The ability to tag histones and examine their deposition properties in single cells has allowed us to gain insight into chromatin assembly processes.

We developed a cytological assay system for studying replication and chromatin assembly by using *Drosophila Kc* cells, a cell line that displays a regular cell division schedule (Fig. 3A) and a consistent tetraploid karyotype. Organization of the *Drosophila* nucleus is visually simple, because the late-replicating heterochromatin typically coalesces into a compartment in the nucleus, termed the chromocenter (Fig. 3B). This provides both a temporal and spatial distinction between the early replicating, gene-rich euchromatin, and the late-replicating heterochromatin. DNA replication can be tracked either by pulse-labeling with nucleotide analogs or by using anti-PCNA antibody. Furthermore, by introducing histone-GFP fusion constructs and producing a pulse of the tagged protein, we can track histone deposition during the cell cycle. Using this system, we have been able to quantitatively examine DNA replication and histone deposition in unsynchronized populations of cells (3, 8, 9).

GFP-tagged H3 shows exclusively replication-coupled deposition, displaying co-localization with replication markers and showing no detectable deposition in cells in which replication has been blocked (9). The N-terminal tail of H3 is required, suggesting that the H3 tails of tetramer particles interact with accessory factors at some early step in nucleosome assembly *in vivo*.

In contrast to the properties of GFP-tagged H3 in cells, tagged H3.3 deposits in a replication-independent manner at actively transcribing loci (9). Deposition can occur in any stage of the cell cycle, and we demonstrated that it is not accompanied by unscheduled DNA synthesis. Incorporation of H4 also occurs at these target sites, as expected for deposition of (H3.3-H4)<sub>2</sub> tetramers; but how replication-independent (RI) histone deposition occurs is virtually unknown.

Tagged Cid can also deposit throughout the cell cycle (8), suggesting that its deposition is also replication-independent.



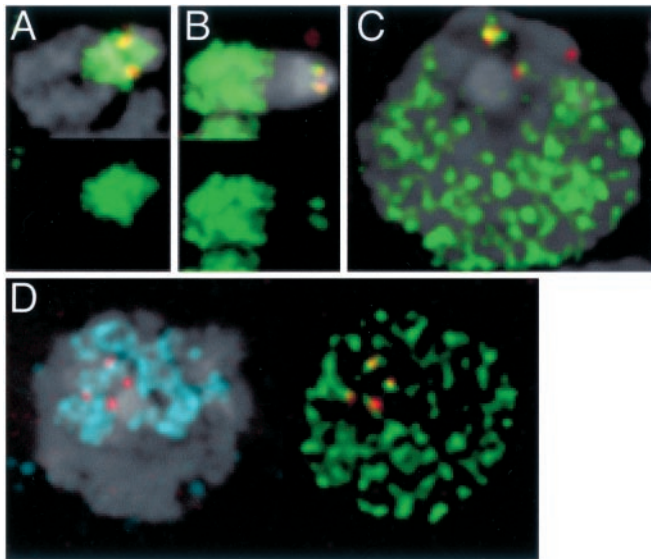
**Fig. 3.** *Drosophila Kc* cells. (A) The cell cycle is  $\approx 20$  h long, and S phase has two distinct periods: early S phase, when all euchromatin (gray) replicates, and late S phase, when all heterochromatin (black) replicates. Eighty percent of cells show  $\approx 15$  chromosomes, and this karyotype has been stable for  $>2$  years. (B) The morphology of a *Drosophila* interphase nucleus. All heterochromatin typically associates into a chromocenter (black). Centromeres (red) are enclosed within the chromocenter, with the nucleolus (light gray) next to it. The active rDNA genes (green) are located within the nucleolus.

However, this conclusion depends on knowing the timing of centromere replication. We have shown that centromeres replicate within a defined portion of S phase (8). The evidence for this conclusion has been challenged (20), and so here we examine the available data on centromere replication timing. We confirm that *Drosophila* centromeres replicate as isolated domains within later-replicating heterochromatin.

### Centromeres Replicate Before Their Surrounding Heterochromatin

Historically, centromeres have been thought to replicate very late in the cell cycle. This is because they are embedded within pericentric heterochromatin, which replicates late. Analysis has usually relied on visualization at mitosis; but mitotic chromosomes have inherently low resolution because they are highly condensed. Indeed, a recent study showed that *Drosophila* centromeres cannot be resolved from heterochromatin in 44% of spread mitotic chromosomes (21). Despite this limitation, Sullivan and Karpen (20) concluded from the analysis of normal mitotic chromosomes that Cid-containing chromatin replicates on the same late schedule as pericentric heterochromatin. However, this could be late replication in pericentric heterochromatin that was mis-scored as replication of centromeres.

We have addressed this uncertainty by analyzing mitotic chromosome replication patterns, providing brief 15-min pulses to *Kc* cells and examining mitotic figures after a chase. This provides a “snapshot” of replication at single points in the cell cycle. We observed examples of heterochromatin replication patterns similar to those previously reported (20), where labeling overlaps Cid spots (Fig. 4A). However, we also observed unambiguous examples of chromosomes that were intensely labeled throughout the euchromatic arms, with foci directly coinciding with centromeres (Fig. 4B). These centromeric foci are surrounded by heterochromatin that did not replicate during the labeling pulse. We attribute the inability of Sullivan and Karpen to observe early centromeric replication foci to the continuous labeling protocol they used, where all surrounding heterochromatin will always be labeled when earlier-replicating sites acquire label. When late-replicating heterochromatin is labeled, this will often overlap Cid-containing chromatin. If overlapping heterochromatin accounts for the apparent late replication of centromeres when analyzed on mitotic chromosomes, then its removal should improve the visualization of earlier-replicating



**Fig. 4.** Centromeres replicate with euchromatin in tetraploid *Kc* cells and in larval diploid neuroblasts. (A) Mitotic X chromosomes from cells pulsed with dig-dUTP nucleotide analog (green) and then chased for 4 h show heavy labeling in the heterochromatin surrounding centromeres (Cid, red), as expected for incorporation during late S phase. (B) Mitotic X chromosomes from cells pulsed with dig-dUTP and then chased for 10 h were in early S phase at the time of the pulse, because they show heavy labeling in the euchromatic arms. There are also foci of incorporation corresponding to both sister centromeres. (C) Pulse-labeling and imaging of interphase *Kc* cells shows that centromeres replicate in the early S-phase period when euchromatin is also replicating. We tracked cell survival and S-phase progression over a 5-h period, and mitotic index over a 25-h period in all labeling experiments. These parameters were indistinguishable from control, untreated cultures. In labeled cultures after 7 h we observed  $\approx 98\%$  labeling of mitotic figures, indicating that virtually all cells in S phase at the time of the pulse received the nucleotide analog. (D) Neuroblast centromeres are contained within one to three heterochromatic chromocenters (H3K9me, blue). Pulse-labeling with dig-dUTP reveals foci of DNA replication in two centromeric spots and in euchromatin. Cultured cells and dissected larval brains were labeled and prepared as described (8).

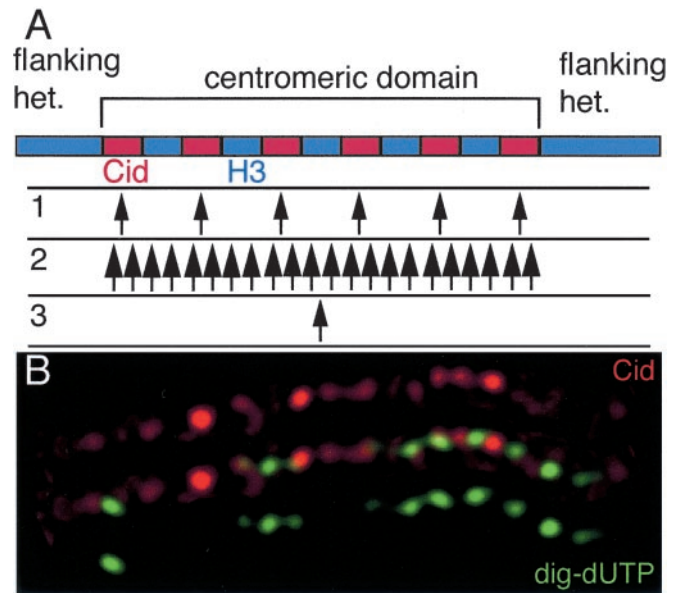
centromeres. Indeed, this has been observed for minichromosomes deficient in flanking heterochromatin (20).

Our previous experiments using interphase *Kc* cells revealed that  $\approx 90\%$  of centromere replication occurs when euchromatin is replicating. The remaining 10% may have been late replication in centromeric regions, but is more likely the result of nearby heterochromatic replication foci that could not be resolved from sites with Cid. Such early replication of centromeres is not limited to tetraploid *Kc* cells (Fig. 4C)—we have observed similar replication patterns in diploid larval neuroblasts (Fig. 4D)—although the much shorter cell cycle time and the more irregular chromocenter limits quantitative analysis. Therefore, this early timing of centromere replication appears to be general for *Drosophila* cells.

This feature of centromeres extends to other eukaryotes. It has long been known that budding yeast centromeres replicate early in the cell cycle (22). In mammals and plants, centromeres appear to replicate conspicuously earlier than similarly repetitive heterochromatic DNA (23–25). Thus, whereas the absolute timing of centromere replication in the cell cycle appears variable (26), the relative timing of replication in euchromatin, centromeres, and heterochromatin is consistent.

#### High-Resolution Mapping of Centromere Replicons

A series of progressively more direct experiments have provided insight into the fine structure in the centromere region. A model



**Fig. 5.** Replication within centromeric domains. (A) Cid and H3 appear to be interspersed as alternating blocks in the centromeric domain. Replication within the domain occurs before replication of the flanking heterochromatin. We consider three possible arrangements of replication origins within this domain. 1: Multiple origins coincide precisely with each block of Cid-containing chromatin. Firing of these origins would replicate every Cid block first, after which replication forks proceed into H3-containing chromatin. 2: Multiple origins are distributed throughout the domain, without regard to the kind of chromatin. At any one time, some Cid-containing and some H3-containing chromatin would be replicating. 3: A single origin lies in the centromeric domain, and bidirectional replication duplicates the entire region. Labeling from short nucleotide analog pulses can distinguish these arrangements. (B) A pulse-labeled centromeric domain fiber. *Kc* cells were labeled with dig-dUTP (8) and fibers prepared according to (30), except that a high salt buffer described in (60) was used. The stretched fiber shows an array of Cid spots (red). In this case, the nucleotide analog (green) has incorporated in the intervening gaps between Cid chromatin. These replication tracts are scattered throughout the centromeric domain, and this pattern demonstrates that multiple origins are scattered throughout the domain. The statistically significant association of replication tracts with non-Cid chromatin across the domain suggests that origins have a fixed relationship to the chromatin blocks, and that H3-containing blocks replicate out of phase from Cid-containing blocks.

for the centromeric constriction has suggested that loops of DNA coil through the constriction, with centromeric nucleosomes lying in the outward parts of these coils, and conventional nucleosomes in the interior portions (21, 27, 28). This would account for the polar structure of the entire centromere if centromeric nucleosomes nucleate kinetochore formation (and thus microtubule capture) and conventional nucleosomes recruit cohesins (and thus centromeric cohesion). The linear arrangement of nucleosomes along centromeric DNA would then be alternating blocks of centromeric and conventional nucleosomes within the centromeric domain (Fig. 5A). A recent study using stretched chromatin fibers has demonstrated that Cid and H3 are interspersed in *Drosophila*, although these are not included in the same nucleosome (21). Apparently, blocks packaged in one kind of nucleosome alternate with blocks packaged in the other. How could the duplication of such regular but discontinuous arrays of nucleosomes occur?

The alternating pattern of nucleosomes on stretched chromatin fibers is reminiscent of replication patterns on fibers from normal chromatin (29). Replication origins within a chromatin domain often appear to be regularly spaced with an interval of 50–100 kb, and these origins fire synchronously. Perhaps the



nucleosome blocks in the centromeric regions correspond to an underlying regular arrangement of replication origins throughout the entire centromeric domain. If Cid-containing blocks include the origins for these domains, and if replication initiates at a time when H3 is not available, ultimately only the RI deposition of Cid will package these blocks. The later replicating stretches would incorporate H3 as it becomes available. In this way, the fine pattern of replication would maintain the discontinuous Cid arrays over an extended region.

Our model for maintaining the higher-order chromatin structure of the entire centromere has precise requirements for replication patterns in this region: a discontinuously spaced arrangement of origins must correspond to the blocks of Cid-containing chromatin (Fig. 5A, pattern 1). At least two other patterns of replication in this region can be imagined. Firstly, all Cid- and H3-containing blocks might replicate simultaneously (pattern 2). Secondly, a single origin might replicate the entire domain (pattern 3).

In a study of stretched centromeric fibers (21), Blower *et al.* showed that centromeric domains replicate out of phase with surrounding heterochromatin. This result is consistent with our previous demonstration that centromere replication precedes that of surrounding heterochromatin (8). Within centromeric domains, it was reported that H3- and CenH3-containing chromatin replicate concurrently (21). However, any fine structure to the replication patterns in these domains might have been obscured by the 2–2.5-h labeling period this study used, because this is enough time for even just two bidirectional replication forks from a single replicon to transit the entire domain. Despite this qualification, a data trace in this work (figure 5C in ref. 21) was intriguing. This trace appears to show that edges of CenP A signal significantly coincided with edges of nucleotide incorporation, with gaps between the CenP A blocks. If confirmed, this might be an example of discontinuous replication tracks corresponding to blocks of centromeric chromatin.

We investigated this possibility by pulse-labeling cells for only 15 min. To prepare stretched chromatin fibers, we disrupted nuclei spread on a glass slide in a high-salt buffer. As the buffer runs off the slide, it pulls chromatin fibers behind it. We identified stretched centromeres and examined those fibers in which nucleotide incorporation was unambiguous. In each of these cases it was clear that replication was occurring in discrete patches scattered throughout the centromeric domain (Fig. 5B). These replication tracks must arise from multiple origins, and thus we can rule out the two possibilities that the entire domain replicates from a single origin, or that the whole domain replicates simultaneously.

These patches corresponded significantly with the segments *between* Cid-containing chromatin. Thus, from published experiments (21) and the experiments described here it appears that replication occurs in two discrete phases: all CenH3-containing chromatin within a domain replicates, and at a different time all H3-containing chromatin replicates. Therefore, replication within this domain is discontinuous and initiates from multiple origins.

Identifying the precise location of origins within the centromeric domain is problematic, because of three difficulties inherent to interpreting stretched chromatin fibers. Firstly, although these fibers are stretched to about 50–100 times their interphase size (21), the radius of H3 or CenH3 spots is substantially less than the resolution of light microscopes. Secondly, the intensity of spots along a fiber is variable, implying that the fiber is unevenly stretched. Thirdly, fibers are inherently spotted—even DNA *in situ* hybridization with probes that should uniformly label an extended region always appear spotty (30). These effects mean that a lack of overlap between different signals could result from artifacts in fiber preparation, and conversely that overlap can result from unstretched segments. It

is apparent from fiber preparations that artifacts are occurring, because some sites on fibers do not appear to be packaged with any histones (21). These concerns can only be addressed with improvements in fiber technology.

Identifying the position of origins within centromeric domains is a critical issue to address, because the deposition of H3 must occur as its DNA substrate replicates. The maintenance of this interspersed arrangement of chromatin must involve the differential regulation of the replication-coupled deposition of H3 and the RI deposition of Cid.

### Inferring the RI Machinery

Given that deposition of any H3 must occur in the form of (H3·H4)<sub>2</sub> tetramers, there must be discrimination of H3-containing tetramers from tetramers containing variants. Our analysis of RI assembly initiated the mapping of discriminating sites within the histone variants (9). We found that one type of discrimination is a cluster of three residues within the histone fold domain (HFD) of H3 that limits it to replication-coupled deposition. Furthermore, because both Cid and H3.3 undergo RI deposition but have mutually exclusive targets, there must be additional discrimination between these variants.

Replication-coupled nucleosome assembly is aided by accessory factors that are recruited to the replication fork by binding to PCNA. However, the process of RI deposition must be different, because RI deposition of H3.3 does not require portions of the histone that are required for replication-coupled deposition (9). Furthermore, the lack of PCNA during gap phase deposition raises the question of what is recruiting histones to the sites. The phenomenon of CenH3 targeting has raised expectations that a specific, localized chromatin assembly factor or histone modification will be involved in the targeting of CenH3s (26, 31, 32). Indeed, a chromatin remodeler of the RSC family, P/BAF, localizes to kinetochores during mitosis of mammalian cells (33). Furthermore, RSC mutations in budding yeast have been previously described to have altered chromatin structure specifically around centromeres (34), and perhaps RSC activity is involved in assembly of centromeric nucleosomes. Mutations in CAF and Hir genes also give centromere defects, and it has been suggested that these factors are involved in loading the yeast CenH3 Cse4p (35). However, a role for any of these factors does little to explain the specific targeting of CenH3s, because these factors are all widely distributed in the nucleus (33, 36). The best candidate for a uniquely centromere-localized chromatin assembly factor is the Mis6 protein in fission yeast (12). This protein is required for centromeric localization of the CenH3 SpCENP-A, but Mis6 homologs in budding yeast (Ctf3; ref. 37) and in mammals (CENP-I; ref. 38) localize to centromeres but are not required for targeting CenH3s. Thus, Mis6 proteins appear to be structural components of centromeres, not histone assembly factors.

An alternative model is that some feature of centromeric chromatin facilitates the targeting of its specialized histones. An obvious candidate for this feature is that centromeric nucleosomes themselves bind to and thereby recruit new CenH3 tetramers for future deposition. Such an interaction is a possible molecular mechanism for direct templating of centromere duplication (39). Regardless of whether CenH3 targeting involves specialized co-factors, templating, or both, the question remains as to why it should use an RI pathway.

The targeted deposition of H3.3 to active genes is likewise replication-independent, although transcription-coupled assembly may facilitate (H3.3·H4)<sub>2</sub> deposition. Perhaps H3.3 targeting is mediated by a component of RNA polymerase complexes. Because RNA polymerases move processively along the DNA during transcription, a contiguous transcribed segment of DNA might incorporate the H3.3 variant. Alternatively, RI deposition of H3.3 may be facilitated by any of a number of ATP-dependent

chromatin remodeling complexes to target specific sites near transcription units. Any candidate factor might be expected to preferentially use H3.3 instead of H3, but whether there is any such discriminating factor is unknown, because all *in vitro* studies of higher eukaryotic chromatin assembly have been performed with H3. We anticipate that this will soon be addressed. However, the prospects for identifying a unique remodeler that is required for RI deposition are uncertain, because budding yeast mutants that eliminate any known chromatin assembly factors do not eliminate chromatin assembly (1). Thus, we need to consider the possibility that RI deposition at active genes and at centromeres uses generic remodeling activities, and that components or structural aspects common to *both* centromeres and actively transcribed genes may result in RI histone deposition at both kinds of sites.

### Opening a Space for Histone Replacement

The deposition of histones throughout the cell cycle by a replication-independent process implies that previously existing nucleosomes are unraveled, and their histones released. It is known that the process of transcription results in a local unfolding of the chromatin fiber and an "open" chromatin configuration (Fig. 2; ref. 15). Although transcription of nucleosomal templates with bacterial polymerases can occur *in vitro* without displacing histone octamers from DNA (40), *in vivo* assays demonstrated that a measurable amount of transcription-dependent histone displacement does occur in eukaryotic nuclei (41). In fact, recent experiments revealed that, even *in vitro*, RNA polymerase II is virtually unable to transcribe nucleosomal DNA under physiological conditions (42). Transcription requires that histone-DNA contacts be broken for polymerase to transit the nucleosomal DNA. Although transcription can occur without histone displacement if the histone octamer releases some contacts with DNA and maintains others (40), at some frequency all contacts might be released. The histone octamer would then simply fall off. Additionally, localized remodeling factors will disrupt nucleosome structure as they act. The *in vitro* and *in vivo* observations can be reconciled if histone displacement occurs occasionally as nucleosomes are disrupted.

Constraints on nucleosomes in a compacted chromatin fiber (i.e., "closed" chromatin) would limit histone displacement. Although internucleosome forces within inactive chromatin are uncharacterized, they have been inferred from numerous experiments, including the tendency of nucleosomes within heterochromatin to form extremely regular and fixed arrays (43). A likely constraint in heterochromatin arises from the multimeric associations that occur between heterochromatin-specific non-histone chromatin proteins. Attention has focused on the heterochromatin protein-1 (HP1). HP1 is recruited to heterochromatic DNA by binding, through its chromodomain, to the H3 tail when it is methylated at lysine-9 (H3-K9me; ref. 44). The chromo shadow domain of HP1 mediates associations between HP1 molecules, and multimers of HP1 bound to methylated histone tails provides one basis for constraining arrays of nucleosomes.

Although the state of chromatin in heterochromatin and in actively transcribed regions is well known, less is known about the chromatin fiber packaged by centromeric nucleosomes. However, these regions appear to be open. Centromeric DNA is sensitive to micrococcal nuclease digestion both in budding yeast (45) and in the central core region of fission yeast centromeres where SpCCENP-A-containing nucleosomes reside (12), and plant meiotic centromeres appear decondensed (46). In addition, early replication is a feature of open chromatin, and centromeric chromatin replicates before surrounding heterochromatin. An open configuration may arise from at least three sources. First, all CenH3s lack a canonical H3 tail (3). Because methyl-modification of lysine-9 appears to be the key epitope to maintain heterochromatin, the lack of this site in centromeric

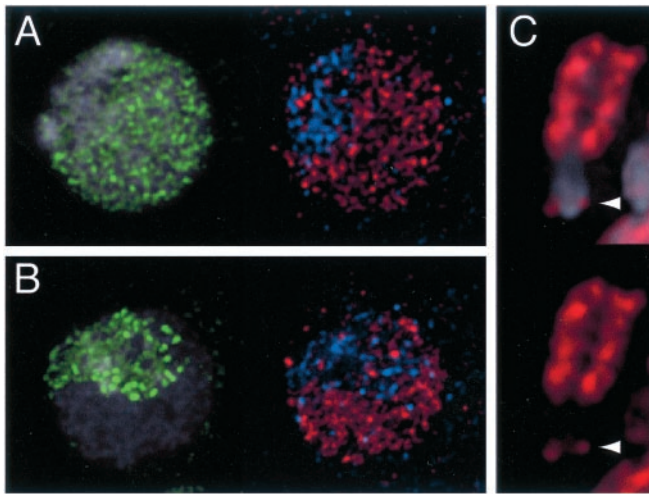
nucleosomes means that such regions cannot become heterochromatic. Indeed, the heterochromatin protein HP1 is not associated with chromatin packaged by CenH3s (47, 48). Second, our recent study of Cid homologs in *Drosophila* has uncovered DNA minor-groove binding motifs in the Cid tail outside of the nucleosome core (49). Extension of the Cid tail along linker DNA between nucleosomes may inhibit compaction of the nucleosome strand, thus maintaining these regions in an open configuration (Fig. 2). Third, chromatin remodeling factors that destabilize nucleosomes are found both at active genes and centromeres, and their activity will promote histone replacement. We suggest that an open chromatin configuration is the common basis for RI deposition at centromeres and at actively transcribed genes.

### The RI Target Sites for H3 Variants Are in Distinct Nuclear Compartments

If open chromatin were the sole basis for RI deposition, then we would expect that active genes and centromeres would incorporate both H3.3 and CenH3s. However, their deposition is mutually exclusive. This exclusivity is likely to rely on multiple mechanisms that act on all steps in nucleosome assembly. Factors that discriminate between H3.3 and Cid would be the best candidates for directing these variants to their targets. However, the organization of the nucleus provides a clue as to another way in which exclusive targeting may be accomplished. Centromeric DNA in *Drosophila* is flanked by repeated sequences that are packaged into heterochromatin, and this forms a compartment at interphase in which centromeres are embedded in heterochromatin (Fig. 3B). The active rDNA genes are the primary sites of H3.3 deposition and they are also found in a distinct nuclear compartment, the nucleolus, next to the chromocenter. This functional nuclear organization is very simple to see in *Drosophila*, where all heterochromatin typically associates into one large chromocenter, and the active rDNA arrays also often associate to present one large nucleolus. In fact, this general compartmentalization is almost invariant in eukaryotes, and has led to the idea that heterochromatin somehow protects centromeres and NORs (50). Although both Cid and H3.3 undergo RI deposition, their exclusive targeting could in part be accomplished by restricting one or both variants within the nucleus. For example, unincorporated (Cid-H4)<sub>2</sub> tetramers might be sequestered within the heterochromatic chromocenter. Cid deposition would then appear targeted to the centromere, because this is the only site within the chromocenter with open chromatin.

Whether (Cid-H4)<sub>2</sub> tetramers are actually sequestered in this way is unknown. Indeed, whether sequestering substrates can have *any* effect on reactions within the nucleus has become a pressing issue (51). Many nuclear components remain mobile, but functional experiments argue that certain effects in the nucleus actually only occur when components are sequestered (52). It is likely that some reactions in the nucleus are relatively independent of localization because they associate efficiently with their partners and their reactions proceed quickly. Conversely, reactions that involve weak interactions or multiple steps may require raising the effective concentration of their substrates by nuclear sequestration.

We have previously suggested that the heterochromatic compartment is involved in histone traffic within the nucleus (8). The basis of this hypothesis was our realization that Cid-containing chromatin behaves unusually during S phase. Generally, the deposition of H3 quickly follows DNA replication. However, the replication of Cid-containing centromeric DNA occurs without H3 deposition (8), implying that the normal coupling between replication components and nucleosome assembly components must be broken. Because this coupling is thought to result from an interaction between chromatin assembly factor 1 (CAF1)-



**Fig. 6.** Overexpression of Cid mis-localizes to euchromatin by RI deposition. *Kc* cells were transfected with a *HS-CidGFP* construct (3) with a modified translational start sequence (61). Cells were induced to produce high levels of Cid-GFP (red), and then immediately pulse-labeled with nucleotide analog (green) to identify cell cycle stages. The heterochromatic compartment is labeled with an anti-HP1 antibody (blue; ref. 61). (A and B) Overexpressed Cid incorporates at centromeres but also mis-incorporates throughout euchromatin. Mis-incorporation in euchromatin occurs by a replication-independent process, because it occurs both in early S phase (A) and in late S phase (B) cells. (C) After a chase of 6 h, mitotic chromosomes that were induced during S phase show labeling at both sister centromeric foci (arrowhead), and throughout the euchromatic arms.

histone complexes and PCNA, the simplest explanation for uncoupling the two processes would be to sequester replicative nucleosome assembly factors away from centromeres. We imagined that unincorporated H3-containing tetramers might be sequestered in euchromatin in the first half of S phase, and would thus never (productively) see the replication forks at centromeres within the heterochromatic compartment. This uncoupling might be necessary to prevent dilution of centromeric nucleosomes by conventional nucleosomes that would assemble after replication-coupled deposition. Genetic experiments in budding yeast and *Drosophila* suggest that CenH3s and H3 do compete for assembly (21, 53).

One way that a competition between CenH3 and H3 histones can be probed is to change their relative concentration. We have previously reported that a tagged Cid protein exclusively deposits at centromeres when it was ectopically expressed at low levels from a heat-shock-inducible promoter (3). However, it was apparent that expression from this construct remained low. Re-engineering the transcriptional start region of the construct to include a translational initiation consensus site now allows overproduction of Cid in cells.

To analyze the behavior of excess quantities of Cid protein, we introduced an overexpression construct into *Drosophila Kc* cells (Fig. 6). Cells receive varying amounts of transfected DNA, and thus express Cid over a wide range of levels. In cells that express low amounts of the ectopic protein, Cid localizes to centromeres, as expected. However, a new localization pattern for Cid is seen at high expression levels: the tagged protein localizes to centromeres and throughout euchromatin. The incorporation pattern of ectopic Cid is especially clear on mitotic chromosomes from these transfections, where the tagged protein is incorporated throughout the euchromatic arms as well as at centromeres (Fig. 6C). We conclude from this result that excess Cid can be deposited at sites other than centromeres. Normal cells must have mechanisms to prevent euchromatic deposition, but overexpression is sufficient, by itself, to overcome this restriction.

The mis-incorporation pattern of Cid shows an interesting specificity: Cid can deposit at centromeres and euchromatin but not in heterochromatin (Fig. 6). Therefore, heterochromatin must lack the feature that tolerates mis-incorporation, or must actively exclude Cid. As we have argued above, centromeres and euchromatin share the feature of open chromatin, which we have proposed is the first prerequisite for RI deposition of histone variants. Indeed, the mis-incorporation of Cid into euchromatin is replication-independent, because it occurs both when euchromatin is replicating in early S phase (Fig. 6A), and in late S phase when euchromatic replication is complete (Fig. 6B). We suggest that Cid is contaminating open chromatin in the euchromatic compartment when it is overexpressed.

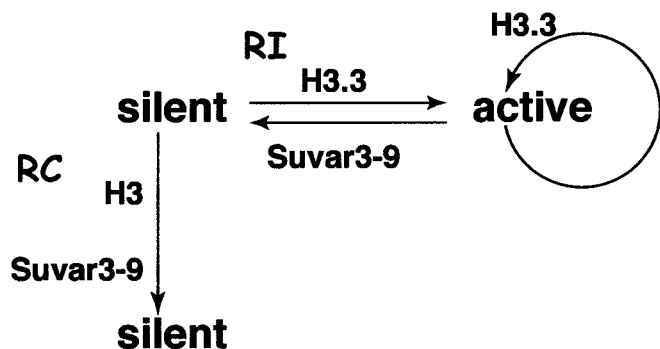
What normally prevents the deposition of Cid into euchromatin? Endogenous Cid is present only at low levels, and mis-incorporation could be avoided if Cid were sequestered away from euchromatin in the nucleus. If unincorporated Cid were sequestered in the heterochromatic chromocenter, it would be unable to deposit in the closed chromatin of this compartment. Thus, sequestration might serve two purposes: deposition in euchromatin would be prevented and deposition at centromeres would be promoted. Overexpression of CenpA in mammalian cells also mis-incorporates into euchromatin (54). Although it has not been examined whether CenpA mis-incorporation is replication-independent, we expect this to be the case, because this is how CenpA deposits at centromeres (24).

The idea that histone variants may respect nuclear compartments was first raised by our experiments expressing heterologous CenH3s in *Drosophila Kc* and human HeLa cells (3). These extremely diverged heterologous histones did not localize to centromeres in these cells, implying that there is some kind of specificity for depositing the correct CenH3 at centromeres. Surprisingly, heterologous histones were preferentially enriched in the heterochromatic blocks. We suggested that it is a default ability of cells to enrich diverged H3 variants in the heterochromatic compartment. Perhaps heterochromatic enrichment is a normal first step in the deposition of the endogenous CenH3s. Those experiments and our overexpression results encourage the view that nuclear compartments may guide histone variants to the correct subset of their potential deposition sites. Compartment effects may also affect the RI deposition of H3.3 in an inverse way to Cid: i.e., sequestering to promote H3.3 deposition at active genes, and preventing its deposition at centromeres. Because H3.3 is largely identical to H3, the hypothetical element that is recognized in H3 and results in its exclusion from chromocenters during centromere replication may also be present in H3.3. Perhaps this discrimination against canonical H3 histones also serves to prevent the RI deposition of H3.3 at centromeres.

### Why RI Assembly?

RI assembly permits immediate chromatin repair. The unfolding of chromatin during transcription may be damaging, in that the forces RNA polymerases apply to their template DNA should at least occasionally displace histone octamers from DNA (55). Additionally, histone octamers may sometimes be displaced by chromatin remodeling factors associated with transcriptional activity. In either case, these regions must be repackaged into nucleosomes. Similarly, replacement of CenH3s may be required to maintain the nucleosomal configuration of centromeres after mitosis. Bundles of microtubules drag a chromosome to the pole during anaphase, and the forces they apply (56) may be sufficient to occasionally pull off histone octamers. Chromatin would then be stripped of some CenH3 histone octamers. RI deposition allows repair of this damage. In fact, the RI deposition of CenpA in mammalian cells seems to occur around the time of mitosis (24). The deposition of Cid in *Drosophila* cells occurs throughout the cell cycle, but may only be required at two points: as





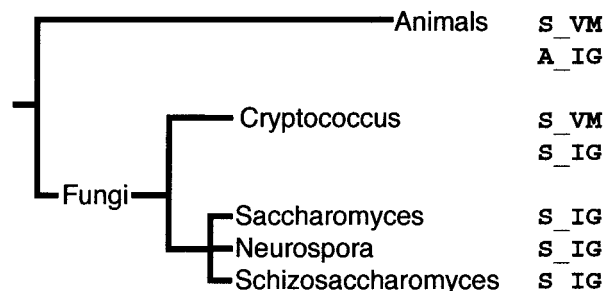
**Fig. 7.** RI deposition allows switching of heritable chromatin states. Nucleosomes in silent heterochromatin are distinctively modified by methylation, and thereby recruit the HMT Suvar3-9. The silencing epitope can be perpetuated by Suvar3-9 through the cell cycle by the methylation of H3 after replication-coupled deposition (vertical arrow). A gene can be activated (rightward arrow) at any time in the cell cycle, and the unraveling of methylated nucleosomes and RI deposition of H3.3 will remove the silencing epitope. This abolishes Suvar3-9 recruitment and allows stable activation. RI deposition of H3.3 will continue as long as the gene is transcribed. Switching from an active to a silent state (leftward arrow) can occur by repressing transcription and methylating the N-terminal tail of H3.3 at Lys-9, once again recruiting the Suvar3-9 complex.

centromeric DNA replicates to double its chromatin, and after mitosis to repair stripped chromatin.

The process of RI assembly at active genes provides a novel level of control over histone modifications. Replacement of nucleosomes in one modification state by new histones could switch chromatin to an active state. Initiation of transcription would start this process, and successive transits of RNA polymerases would promote RI assembly. The replacement H3 histone in alfalfa is hyperacetylated (57), and RI assembly with acetylated histones could enrich such modifications in active chromatin. However, histone modification by methylation has appeared more problematic (43). A number of histone methyltransferases (HMTs) have been characterized (44), but no histone demethylase is known. Methylated lysine-9 in the H3 tail (H3K9me) is a critical epitope for recruiting heterochromatic chromatin proteins, because this is the binding site for HP1. HP1 recruits additional heterochromatic proteins including the Suvar3-9 HMT. Therefore, it is straightforward to imagine how these recruited proteins could perpetuate a heterochromatic state through replication-coupled nucleosome assembly and cell division (Fig. 7).

Because an irreversible methyl modification appears to specify the heterochromatic state, it has been unknown how a heterochromatic site could switch to an active state. One route for switching might be to prevent the methylation of nucleosomes assembled during replication. Successive cell cycles could then dilute methylated nucleosomes, allowing eventual activation. However, more rapid mechanisms for activating silenced chromatin must exist. Induction of silenced genes can occur within a single cell cycle; for example, X chromosomes become reactivated and lose H3K9me during diplotene in the *Caenorhabditis* ovary (58). In addition, our work using a reporter for heterochromatic gene silencing suggests that switching to an active state can occur in somatic cells without cell division (59). Thus, H3K9me can be removed without replication-coupled nucleosome assembly.

RI deposition implies that the entire heterochromatic nucleosome may be unraveled and replaced (9). The process of transcriptional activation may force the disassembly of H3K9me-containing nucleosomes, followed by RI assembly of an unmarked nucleosome. Although we do not know the fate of the



**Fig. 8.** The relationship of the canonical H3 histones in animals and fungi. Three residues in the HFD differ between H3 and the replacement H3.3 histone in metazoans. Basidiomycete fungi (*Cryptococcus*) also have two canonical H3 histones that can be classed as an H3 and a replacement histone. Ascomycetes (*Saccharomyces*, *Neurospora*, and *Schizosaccharomyces*) have only one type of canonical H3, resembling the replacement histone in basidiomycetes. This phylogeny identifies that ascomycetes have lost their H3 histone, and retain only a replacement variant.

displaced methylated H3, we do know that RI deposition can occur at any time in the cell cycle, and thus should be able to rapidly derepress silencing (Fig. 7). Conversely, an active gene could be silenced by methylating the tail of H3.3, which presents the same lysine-9 epitope. The stability of histone methylation gives it a distinct advantage over other histone modifications for heritable effects on chromatin. The possibility of RI deposition circumvents the irreversible nature of methylation, thus retaining the potential to switch the heritable chromatin state at a later time.

### Conclusions

H3 variants are used to package functionally specialized chromatin, where they play vital functional roles. Localizing these variants to centromeres and to transcriptionally active regions utilizes an RI process that is distinct from the nonspecific, replication-coupled method of packaging the bulk genome. We have argued that RI deposition is the consequence of the activities that impinge on these sites in the genome and create an open chromatin structure. This flexibility in histone deposition may be necessary to maintain the nucleosomal structure of these regions. In higher eukaryotes, the RI deposition process allows specialized chromatin to be distinguished at the most basic level, where histone variants are incorporated into chromatin. The differences between the generic H3—which packages the bulk of the genome—and the H3 variants may contribute to the physical properties of specialized regions and recruit particular non-histone chromatin proteins. Because histones remain associated with DNA through mitosis, these variants establish heritable distinctions in chromatin.

Centromeres are a defining feature of eukaryotes, and all are likely to have a CenH3. However, the utilization of two conserved versions like H3 and H3.3 is not universal. For example, budding yeast has only one canonical H3 histone, which undergoes both replication-coupled and RI deposition (1). Surprisingly, this is H3.3: phylogenetic analysis reveals that ascomycetes have lost H3, whereas their sister clade basidiomycetes have both H3 and H3.3, as do animals (Fig. 8; ref. 9). Therefore, an H3.3 gene performs all general functions in some organisms. The extraordinary conservation of H3.3, which is identical from mollusks to mammals, speaks to its fundamental role in the eukaryotic nucleus.

We thank Hillary Hayden for help with stretched fiber imaging, Jim Smothers for plasmid constructs, and Harmit Malik, Paul Talbert, and Danielle Vermaak for helpful discussions. This work was supported by the Howard Hughes Medical Institute.

1. Verreault, A. (2000) *Genes Dev.* **14**, 1430–1438.
2. Osley, M. A. (1991) *Annu. Rev. Biochem.* **60**, 827–861.
3. Henikoff, S., Ahmad, K., Platero, J. S. & van Steensel, B. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 716–721.
4. Akhmanova, A. S., Bindels, P. C. T., Xu, J., Miedema, K., Kremer, H. & Hennig, W. (1995) *Genome* **38**, 586–600.
5. Fretzin, S., Allan, B. D., van Daal, A. & Elgin, S. C. (1991) *Gene* **15**, 341–342.
6. Henikoff, S., Ahmad, K. & Malik, H. S. (2001) *Science* **293**, 1098–1102.
7. Waterborg, J. H. & Robertson, A. J. (1996) *J. Mol. Evol.* **43**, 194–206.
8. Ahmad, K. & Henikoff, S. (2001) *J. Cell Biol.* **153**, 101–110.
9. Ahmad, K. & Henikoff, S. (2002) *Mol. Cell.* **9**, 1191–1200.
10. Warburton, P. E., Cooke, C. A., Bourassa, S., Vafa, O., Sullivan, B. A., Stetten, G., Gimelli, G., Warburton, D., Tyler-Smith, C., Sullivan, K. F., *et al.* (1997) *Curr. Biol.* **7**, 901–904.
11. Stoler, S., Keith, K. C., Curnick, K. E. & Fitzgerald-Hayes, M. (1995) *Genes Dev.* **9**, 573–586.
12. Takahashi, K., Chen, E. S. & Yanagida, M. (2000) *Science* **288**, 2215–2219.
13. Talbert, P. B., Masuelli, R., Tyagi, A. P., Comai, L. & Henikoff, S. (2002) *Plant Cell.* **14**, 1053–1066.
14. Allis, C. D. & Wiggins, J. C. (1984) *Dev. Biol.* **101**, 282–294.
15. Wolffe, A. P. (1998) *Chromatin: Structure and Function* (Academic, San Diego), 3rd Ed.
16. Shibahara, K. & Stillman, B. (1999) *Cell* **96**, 575–585.
17. Akhmanova, A., Verkerk, T., Langeveld, A., Grosveld, F. & Galjart, N. (2000) *J. Cell Sci.* **113**, 4463–4474.
18. Akhmanova, A. S., Miedema, K., Wang, Y., van Bruggen, M., Berden, J. H. M., Moudrianakis, E. N. & Hennig, W. (1997) *Chromosoma* **106**, 335–347.
19. Pina, B. & Suau, P. (1987) *Dev. Biol.* **123**, 51–58.
20. Sullivan, B. A. & Karpen, G. H. (2001) *J. Cell Biol.* **154**, 683–690.
21. Blower, M. D., Sullivan, B. A. & Karpen, G. H. (2002) *Dev. Cell* **2**, 319–330.
22. McCarroll, R. M. & Fangman, W. L. (1988) *Cell* **54**, 505–513.
23. O’Keefe, R. T., Henderson, S. C. & Spector, D. L. (1992) *J. Cell Biol.* **116**, 1095–1110.
24. Shelby, R. D., Monier, K. & Sullivan, K. F. (2000) *J. Cell Biol.* **151**, 1113–1118.
25. Jasencakova, Z., Meister, A. & Schubert, I. (2001) *Chromosoma* **110**, 83–92.
26. Choo, K. H. A. (2001) *Dev. Cell* **1**, 165–177.
27. Zinkowski, R. P., Meyne, J. & Brinkley, B. R. (1991) *J. Cell Biol.* **113**, 1091–1110.
28. Buchwitz, B. J., Ahmad, K., Moore, L. L., Roth, M. B. & Henikoff, S. (1999) *Nature (London)* **401**, 547–548.
29. Berezney, R., Dubey, D. D. & Huberman, J. A. (2000) *Chromosoma* **108**, 471–484.
30. Haaf, T. & Ward, D. C. (1994) *Hum. Mol. Genet.* **3**, 697–709.
31. Sullivan, K. F. (2001) *Curr. Opin. Genet. Dev.* **11**, 182–188.
32. Sullivan, B. A., Blower, M. D. & Karpen, G. H. (2001) *Nat. Rev. Genet.* **2**, 584–596.
33. Xue, Y., Canman, J. C., Lee, C. S., Nie, Z., Yang, D., Moreno, G. T., Young, M. K., Salmon, E. D. & Wang, W. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 13015–13020.
34. Tsuchiya, E., Hosotani, T. & Miyakawa, T. (1998) *Nucleic Acids Res.* **26**, 3286–3292.
35. Sharp, J. A., Franco, A. A., Osley, M. A. & Kaufman, P. D. (2002) *Genes Dev.* **16**, 85–100.
36. Ng, H. H., Robert, F., Young, R. A. & Struhl, K. (2002) *Genes Dev.* **16**, 806–819.
37. Measday, V., Hailey, D. W., Pot, I., Givan, S. A., Hyland, K. M., Cagney, G., Fields, S., Davis, T. N. & Hieter, P. (2002) *Genes Dev.* **16**, 101–113.
38. Nishihashi, A., Haraguchi, T., Hiraoka, Y., Ikemura, T., Regnier, V., Dodson, H., Earnshaw, W. C. & Fukagawa, T. (2002) *Dev. Cell* **2**, 463–476.
39. Willard, H. F. (1998) *Curr. Opin. Genet. Dev.* **8**, 219–225.
40. Studitsky, V. M., Clark, D. J. & Felsenfeld, G. (1994) *Cell* **76**, 371–382.
41. Jackson, V. (1990) *Biochemistry* **29**, 719–731.
42. Kireeva, M. L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M. & Studitsky, V. M. (2002) *Mol. Cell* **9**, 541–552.
43. Grewal, S. I. S. & Elgin, S. C. R. (2002) *Curr. Opin. Genet. Dev.* **12**, 178–187.
44. Jenuwein, T. & Allis, C. D. (2001) *Science* **293**, 1074–1080.
45. Bloom, K. S. & Carbon, J. (1982) *Cell* **29**, 305–307.
46. Franz, P. F., Armstrong, S., de Jong, J. H., Parnell, L. D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T. & Jones, G. H. (2000) *Cell* **100**, 367–376.
47. Partridge, J. F., Borgstrom, B. & Allshire, R. C. (2000) *Genes Dev.* **14**, 783–791.
48. Blower, B. D. & Karpen, G. H. (2001) *Nat. Cell Biol.* **3**, 730–739.
49. Malik, H. S., Vermaak, D. & Henikoff, S. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 1449–1454.
50. Manuelidis, L. (1990) *Science* **250**, 1533–1540.
51. Carmo-Fonseca, M. (2002) *Cell* **108**, 513–521.
52. Feuerbach, F., Galy, V., Trelles-Sticken, E., Fromont-Racine, M., Jacquier, A., Gilson, E., Olivo-Marin, J.-C., Scherthan, H. & Nehrbass, U. (2002) *Nat. Cell Biol.* **4**, 214–221.
53. Glowczewski, L., Yang, P., Kalashnikova, T., Santisteban, M. S. & Smith, M. M. (2000) *Mol. Cell. Biol.* **20**, 5700–5711.
54. Van Hooser, A. A., Ouspenski, I. I., Gregson, H. C., Starr, D. A., Yen, T. J., Goldberg, M. L., Yokomori, K., Earnshaw, W. C., Sullivan, K. F. & Brinkley, B. R. (2001) *J. Cell Sci.* **114**, 3529–3542.
55. Bennink, M. L., Leuba, S. H., Leno, G. H., Zlatanova, J., de Grooth, B. G. & Greve, J. (2001) *Nat. Struct. Biol.* **8**, 606–610.
56. King, J. M. & Nicklas, R. B. (2000) *J. Cell Sci.* **113**, 3815–3823.
57. Waterborg, J. H. (1993) *J. Biol. Chem.* **268**, 4912–4917.
58. Kelly, W. G., Schaner, C. E., Dernberg, A. F., Lee, M.-H., Kim, S. K., Villeneuve, A. M. & Reinke, V. (2002) *Development (Cambridge, U.K.)* **129**, 479–492.
59. Ahmad, K. & Henikoff, S. (2001) *Cell* **104**, 839–847.
60. Balajee, A. S. & Geard, C. R. (2001) *Nucleic Acids Res.* **29**, 1341–1351.
61. Smothers, J. F. & Henikoff, S. (2001) *Mol. Cell. Biol.* **21**, 2555–2569.