

Pericentric and centromeric transcription: a perfect balance required

Laura E. Hall · Sarah E. Mitchell ·
Rachel J. O'Neill

Published online: 4 July 2012
© Springer Science+Business Media B.V. 2012

Abstract The pericentromere and centromere regions of the genome have previously been considered tightly compacted and transcriptionally inert. However, there is mounting evidence that these regions not only actively produce transcripts but that these pericentric and centromeric transcripts are also vital to maintaining genome stability and proper cell division. In this review, we define the pericentromere and centromere of eukaryotic chromosomes in terms of their histone modifications and their nascent transcripts. In addition, we present the currently known roles these transcripts play in heterochromatin formation, development, and differentiation, as well as their interaction with centromeric proteins, and ultimately centromere function. Recent work has added considerable complexity to the theoretical framework defining the innate requirement for pericentric and centromeric transcription. It is clear that maintaining a fine balance of transcriptional output is critical, as deviations from this balance result in centromere dysfunction and genomic instability.

Keywords Centromere transcripts · Pericentromere transcripts · CENP-A · Genome instability

Abbreviations

CT	Centromere transcripts
PCT	Pericentromere transcripts
H3	Histone H3
H3K9me2/3	Histone H3 di- or trimethylated at lysine residue 9
H3K27me2/3	Histone H3 di- or trimethylated at lysine residue 27
H3K4me1/2	Histone H3 mono- or dimethylated at lysine residue 4
H3K36me2/3	Histone H3 di- or trimethylated at lysine residue 36
CENP	Centromere protein
RDRP	RNA-dependent RNA polymerase
tRNA	Transfer RNA
RNAi	RNA interference
RITS	RNA-induced initiation of transcriptional silencing
CLRC	Histone methyltransferase containing Clr4 complex
WDHD1	WD repeat and HMG-box DNA binding protein 1
RNAPII/III	RNA polymerase II or III
TFIIIC	Transcription factor IIIC
RpL	Ribosomal protein
HP1	Heterochromatin protein 1
CRM	Centromeric retroelements of maize
HAC	Human artificial chromosomes

Responsible Editor: Rachel O'Neill and Beth Sullivan

L. E. Hall and S. E. Mitchell has contributed equally to this work.

L. E. Hall · S. E. Mitchell · R. J. O'Neill (✉)
Department of Molecular and Cell Biology,
University of Connecticut,
Storrs, CT 06269, USA
e-mail: rachel.oneill@uconn.edu

FACT	Facilitates chromatin transcription
CHD1	Chromodomain-helicase-DNA-binding protein 1
nt	Nucleotide
LINE	Long interspersed nuclear element
HSF1	Heat shock factor 1 protein
nSBs	Nuclear stress bodies
tonEBP	Tonicity-responsive enhancer binding protein
KDM2A	Lysine-specific demethylase 2A
PDAC	Pancreatic ductal adenocarcinomas
DNMT3B	DNA methyltransferase 3B
ICF	Immunodeficiency centromere instability, facial anomalies syndrome
BRCA1	Breast cancer susceptibility gene 1
JMJD2B	Jumonji domain containing protein

Introduction

As active participants in mitosis, centromeres are the location of the assembled kinetochore, a proteinaceous structure that binds microtubules allowing for proper chromosome congression. Most complex eukaryotic centromeres have not been traversed and fully assembled by modern sequencing technology, but are known to be composed of highly repetitive sequences, mainly satellites and retroelements (Schueler et al. 2001; Jiang et al. 2003; Dawe and Henikoff 2006; Birchler et al. 2011). While the function of the centromere is evolutionarily conserved in all forms of life, the DNA sequences and several of the corresponding DNA binding proteins found at centromeres are rapidly evolving. Henikoff et al. (2001) have termed this conundrum the “centromere paradox.” The lack of a conserved DNA satellite sequence across species that demarcates the position of the centromere has led to the formation of the hypothesis that the location of the centromere is determined epigenetically by the presence of a histone H3 variant [CENP-A (mammals), cenH3 (plants), or CID (*Drosophila*)], found only at active centromeres (Henikoff et al. 2001; Sullivan et al. 2001; Allshire and Karpen 2008). While we are gaining a better understanding of the pathway that results in the assembly of CENP-A nucleosomes at centromeres, the precise mechanism for determining the genomic location of CENP-A deposition is unknown. Recent work implicates centromeric transcripts as active participants in CENP-A deposition

and centromere function, adding complexity to a pathway previously thought to be restricted to a large protein network (Topp et al. 2004; Chueh et al. 2009; Ferri et al. 2009; Bergmann et al. 2011, 2012).

Historically, centromeres were considered simply heterochromatin-rich and thus transcriptionally silent. Over the past decade, pericentromere and centromere regions have been characterized in more detail by differences in chromatin compaction and histone modifications (Fig. 1a). Densely packed heterochromatin is found in most eukaryotic pericentromeres, most commonly marked by di- and trimethylation of lysine residues 9 and 27 of histone H3 (H3K9me2, H3K9me3, H3K27me2, and H3K27me3), histone modifications typically associated with transcriptional silencing (Gopalakrishnan et al. 2009). The chromatin encompassing the centromere core, also referred to as “centrochromatin,” is distinct from that of pericentromeres and contains the histone H3 variant CENP-A interspersed with histone H3 methylation and dimethylation of lysine 4 and di- and trimethylation of lysine 36 of histone H3 (H3K4me1, H3K4me2, H3K36me2, and H3K36me3), histone modifications associated with transcriptionally active chromatin (Sullivan and Karpen 2004; Gopalakrishnan et al. 2009; Bergmann et al. 2011, 2012). Interestingly, transcripts emanating from both the pericentromere and centromere core have been identified in a multitude of organisms (Eymery et al. 2009a; Stimpson and Sullivan 2010). Moreover, recent studies have identified proteins that interact with these transcripts (Topp et al. 2004; Wong et al. 2007; Chueh et al. 2009; Ferri et al. 2009; Du et al. 2010; Hsieh et al. 2011), analyzed the effects of increased and decreased transcription on genome stability (Bergmann et al. 2011; Ohkuni and Kitagawa 2011; Bergmann et al. 2012), or identified changes in transcription in stressed or diseased cells (Eymery et al. 2009b; Gopalakrishnan et al. 2009; Ting et al. 2011; Zhu et al. 2011). In this review, we will present pericentric and centromeric noncoding RNA transcription and discuss its diverse roles in modified histone recruitment, centromere function and stability, and insulator activity as well as newly discovered correlations between centromere and pericentromere transcription and human disease. What is emerging from these studies is the synthesis of a molecular model wherein both the process of active transcription and the noncoding RNA species themselves are involved in the complex system required for proper centromere formation and function.

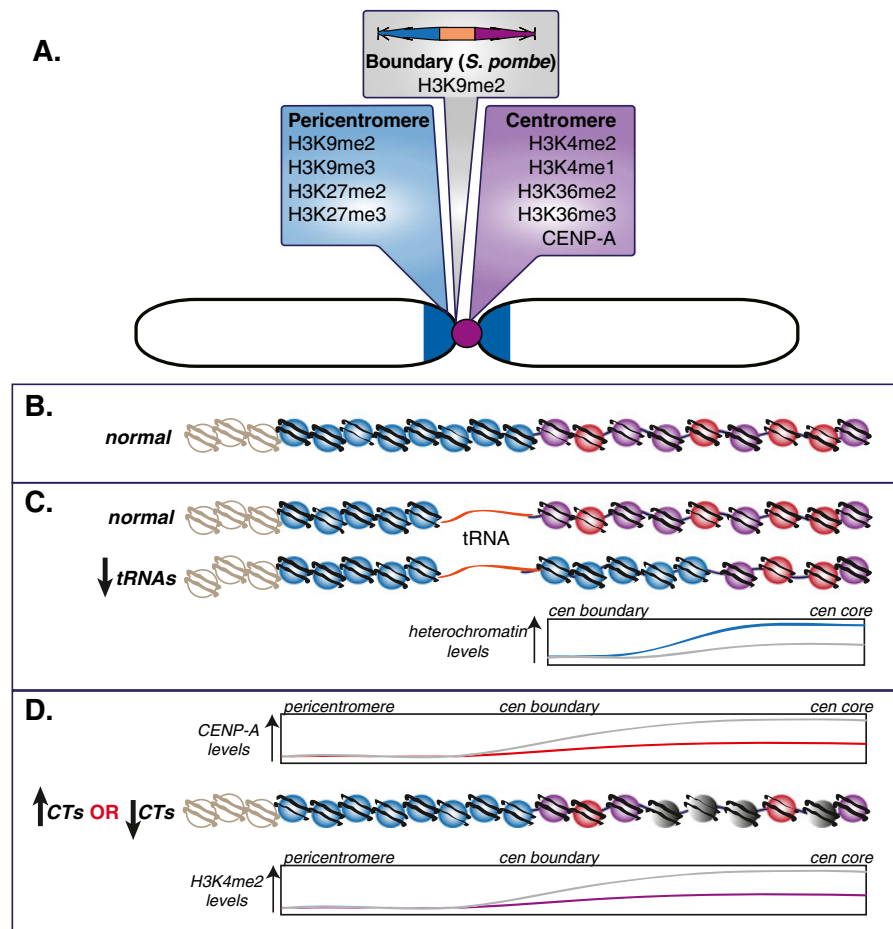


Fig. 1 **a** Modified histone marks found at the pericentromere (blue) and centromere (purple). The boundary between these regions within *S. pombe* (gray) is marked by the presence of tRNA (orange in inset). **b** Nucleosomes found in normal eukaryotic chromatin coded by the color of the region: blue for pericentromere (H3K9me2, H3K9me3, H3K27me2, and H3K27me3), purple for centrochromatin (H3K4me1, H3K4me2, H3K36me2, and H3K36me3), red for CENP-A containing nucleosomes, and light brown for unannotated nucleosomes with respect to histone modifications. **c**. In *S. pombe*, a

decrease in tRNA transcription leads to a spread of pericentric chromatin into the centromere region. *Graph inset* levels of heterochromatin under normal (gray) and tRNA misregulated (blue) conditions. **d** Loss of centrochromatin associated with either an increase or decrease in CTs from human artificial chromosomes. *Graph inset* levels of CENP-A (red) and H3K4me2 (purple) within centrochromatin under normal (gray) and CT misregulated conditions. Gray nucleosomes represent unknown histone replacements

Pericentric transcription

The pericentromere is a distinct chromatin structure found on both sides of the centromere core region of monocentric chromosomes (Fig. 1a, b) and performs a variety of functions such as maintaining the boundary that separates the euchromatin from the centromere core (Chen et al. 2008), providing sites for sister chromatid cohesion during mitosis (Lippman and Martienssen 2004), and repressing meiotic recombination around the centromere (Ellermeier et al. 2010). Transcripts

emanating from this region, known as pericentric transcripts, or PCTs, recruit heterochromatin factors that maintain the heterochromatic histone modifications, specifically H3K9me2, H3K9me3, H3K27me2, and H3K27me3 (Lippman and Martienssen 2004; Chen et al. 2008; Djupedal et al. 2009; Reyes-Turcu et al. 2011).

The fission yeast has been an instrumental model system in determining the mechanism that facilitates heterochromatin formation. To date, three different mechanisms of heterochromatin formation in the fission yeast have been identified, all involving transcription

and processing of the pericentric sequences into shorter fragments (Lippman and Martienssen 2004; Djupedal et al. 2009; Reyes-Turcu et al. 2011). The first mechanism of heterochromatin formation involves the RNA interference (RNAi) pathway through the action of RNA-dependent RNA polymerase (RDRP) (Lippman and Martienssen 2004). This enzyme produces double-stranded RNA from single-stranded PCTs that can be cleaved by the RNase III cleavage enzyme Dicer to form short interfering RNAs (siRNAs) (Lippman and Martienssen 2004). The siRNA then associates with the RNA-induced initiation of transcriptional silencing (RITS) complex, which in turn binds another nascent pericentric transcript forming double-stranded RNA (Djupedal et al. 2009). The process then cycles again, starting with Dicer cleaving the newly formed double-stranded RNA. The association of RITS with PCTs leads to recruitment of the histone methyltransferase containing Ctr4 complex (CLRC) that methylates lysine 9 of histone H3, thus maintaining regional heterochromatin (Lippman and Martienssen 2004; Djupedal et al. 2009).

Recent work in yeast has shown that an alternate RNAi pathway exists, in which certain PCTs form secondary stem loop structures that are recognized and cleaved by Dicer, thus bypassing the need for RDRP (Djupedal et al. 2009). The resulting siRNAs associate with the argonaute containing complex RITS to assist in heterochromatin formation in a manner analogous to that produced from the RDRP pathway. The third mechanism of heterochromatin formation in yeast is an RNAi-independent mechanism that acts in parallel to the RNAi pathway (Reyes-Turcu et al. 2011). Heterochromatin was shown to form in yeast cells carrying a deletion of dicer or argonaute and mlo3, an RNA binding protein that exports messenger RNA from the nucleus (Reyes-Turcu et al. 2011). In these mutants, the exosome degrades aberrant PCTs into shorter RNA fragments that are then capable of forming de novo heterochromatin through an as yet unknown mechanism.

The RNAi pathway is not a feature unique to yeast as similar pathways for establishing heterochromatin at pericentromeres have also been identified in other organisms, including rice, *Arabidopsis*, and *Drosophila* (Lippman and Martienssen 2004; Neumann et al. 2007). Moreover, in mouse, major and minor satellite transcripts specific to the pericentromere and centromere, respectively, may also be involved in heterochromatin maintenance (Hsieh et al. 2011). Transcripts

from both types of mouse satellites associate with WDHD1 (WD repeat and HMG-box DNA binding protein 1), an acidic nucleoplasmic DNA-binding protein whose activity is coupled to RNA polymerase II (RNAPII) transcription and may play a role in RNA processing (Hsieh et al. 2011). WDHD1 knock-down results in an increase in major and minor satellite transcription and a decrease in the compaction of heterochromatin, ultimately leading to a cell cycle progression deficiency (Hsieh et al. 2011). These data implicate WDHD1 as a member of a heterochromatin maintenance pathway analogous to the RNAi pathway in yeast.

From studies across major eukaryotic lineages, it is clear that PCTs have other roles in addition to heterochromatin formation and that some of these functions may be strand specific (Chen et al. 2008; Eymery et al. 2009a; Probst et al. 2010). Recent research suggests that PCTs may play a role in the formation of chromocenters (Eymery et al. 2009a; Probst et al. 2010), nuclear structures formed from the aggregation of heterochromatin from multiple chromosomes. The number of chromocenters present within a nucleus can be tissue specific and can change during cell differentiation (Ceccarelli et al. 1998). PCTs and centromeric transcripts or CTs are shown to localize to chromocenters as mouse cells differentiate into muscle cells (Eymery et al. 2009a). Another study examining developing mouse embryos illustrates that major satellite PCTs are required for the formation of chromocenters at the two-cell stage of development (Probst et al. 2010). Interestingly, the sense and antisense strand of major satellite PCTs in mouse are expressed at different developmental times and levels and localize to different places within the cell. Moreover, the sense strand is expressed in a parent of origin manner, emanating only from the paternal chromosome, beginning at the two-cell stage of development. Towards the end of the two-cell stage, when chromocenters have formed just prior to the second mitotic division, there is a burst in expression from the antisense strand of both maternal and paternal chromosomes. While the sense strand PCTs localize to the nucleus and the cytoplasm, the antisense strand PCTs are confined within the nucleus, demonstrating that nuclear retention is also strand-specific (Probst et al. 2010).

Differences in strand expression from the pericentromere are also observed in adult mouse tissues and human cells [reviewed in Eymery et al. (2009a)]. For

example, expression of the antisense strand of PCTs in mouse testis was found only to be present within seminiferous tubules lacking mature sperm (Rudert et al. 1995). The sense strand was not limited to a specific developmental time point within the mouse testis and was also present in the liver (Eymery et al. 2009a). Within normal and stressed HeLa cells, PCT levels of satIII are more abundant in the sense orientation [Valgardsdottir et al. (2008) and reviewed in Eymery et al. (2009a)]. The same phenomenon of differences in sense and antisense strand transcription of pericentric sequences has been observed in yeast. Chen et al. (2008) showed that the sense strand is not transcribed in the presence of heterochromatin, but the antisense strand is actively transcribed. During the S phase of the cell cycle, there is an increase in sense strand transcription coincident with high RNAPII occupancy; the authors propose that this is due to less heterochromatin compaction during this time (Chen et al. 2008). Why transcription of one strand over another would occur is unclear, especially if heterochromatin limits RNAPII binding altogether. However, it is possible that there are strand-specific DNA or histone marks, analogous to those found at imprinted loci where there are parent of origin marks, which define strand specific expression control (Ferguson-Smith 2011).

Boundary elements

While the pericentromere itself is important for providing a boundary around the centromere core, another critical region in maintaining chromosome integrity is the chromatin barrier between the pericentromere and the centromere (Saffery et al. 2003; Scott et al. 2007). In fission yeast, nucleosome-free tRNA act as a barrier preventing the spreading of heterochromatin into the centromere (Scott et al. 2007) (Fig. 1c). Transcription factor IIIC binds the A and B box sequences of specific tRNAs, thereby recruiting RNA polymerase III (RNAPIII). Whether transcription of the tRNA is necessary for a functional barrier is unclear, but recruitment of RNAPIII is required (Scott et al. 2007). Interestingly, RNAPIII genes (tRNA and 5SRNA) found throughout the genome are known to cluster at the centromere and associate with condensin in the presence of a specific threshold level of transcription (Iwasaki et al. 2010). Ribosomal proteins (RpLs), specifically RpL7, RpL11, and RpL25

associate with centromeric tRNA clusters; unlike RpL association at euchromatic loci, the centromeric association is sensitive to RNaseA and RNaseT1 treatments, indicating the RNA involved is single stranded (De et al. 2011). While the purpose of ribosomal protein recruitment to these areas is unknown, De et al. (2011) speculate a role for RpLs in centromere function, possibly through an association with nascent centromere transcripts (Choi et al. 2011). tRNAs may not be the only type of barrier sequence between pericentric and centromeric regions. In the human Mardel 10 neocentromere, an active gene was identified between the heterochromatin protein 1 (HP1) domain and the CENP-A domain. Thus, it is possible that, analogous to tRNAs in yeast, this expressed gene acts as a barrier to protect centromeric chromatin in this newly formed centromere (Saffery et al. 2003). However, species that lack genes within or near their highly repetitive centromeres may have other means for regulating the various centromere domains.

While there are commonalities among instances of PCT, it is clear that diverse forms of PCTs produce an equally diverse array of functions (from RNAi and heterochromatin formation to chromocenters). The next challenge is to understand how these different sizes and forms of PCTs are transcriptionally regulated and to what extent PCTs affect the function of CTs.

Centromeric transcription

Centromeres are composed of unique chromatin, centromeric chromatin, which is marked by the modified histone 3 CENP-A, H3K4me1, H3K4me2, H3K36me2, and H3K36me3 in most species (Sullivan and Karpen 2004; Bergmann et al. 2011) (Fig. 1a, b). However, maize and rice appear unique as the centromeres of these species do not contain enrichment for H3K4me2 (Wu et al. 2011; Gent et al. 2012). Instead, the specific type of centromere sequence present dictates the class of corresponding histone modifications. For instance, genes present at rice centromeres have the same histone modifications as genes present in euchromatin (Yan et al. 2005; Wu et al. 2011). Thus, the lack of enrichment for H3K4me2 could be due to the presence of actively transcribed genes embedded in between clusters of centC centromeric satellite and centromeric retroelements of maize (CRM) retrotransposon sequences (Gent et al. 2012). In support of this theory,

actively transcribed genes are also found in human neocentromeres, and their expression does not affect the presence of centrochromatin (Saffery et al. 2003).

Centromeric satellite and retroelement transcripts have been identified in a multitude of different organisms, including rice, maize, beetle, tammar wallaby, mouse, and human, and are of various sizes, from 35 to 5,000 nt in length (Topp et al. 2004; Bouzinba-Segard et al. 2006; Lee et al. 2006; Yan et al. 2006; Pezer and Ugarković 2008; Carone et al. 2009; Chueh et al. 2009; Ferri et al. 2009; Du et al. 2010). CT levels, like PCT levels, can change depending on developmental stage and tissue type. For instance, in the beetle, more CTs are observed in the pupae stage than adult (Pezer and Ugarković 2008). In both the beetle and tammar wallaby, transcription of centromeric sequences can be seen from both strands (Pezer and Ugarković 2008; Carone et al. 2009), implicating nearby bidirectional promoters within or adjacent to the sequences being transcribed (Lee et al. 2006; Pezer and Ugarković 2008; Carone et al. 2009). Since retroelement sequences contain their own promoters and are found at most centromeres, it is plausible that these promoters are utilized to transcribe retroelement and adjacent satellite sequences (Carone et al. 2009). However, beetle satellite sequences (PRAT) contain bidirectional promoters that may also facilitate nascent centromeric transcription at least in this species (Pezer and Ugarković 2008).

Studies have shown that the overall level of transcription of centromeric sequences is lower than that of pericentric sequences (Ohkuni and Kitagawa 2011). In some cases, CTs are almost undetectable due to the rapid turnover of the RNA (Choi et al. 2011; Ohkuni and Kitagawa 2011; Chan et al. 2012). However, maintaining the correct level of centromeric transcription in a cell is crucial for centromeres to assemble and function properly during cell division. For example, when centromeric transcription is substantially decreased or increased in budding yeast, there is a marked increase in chromosome missegregation during cell division (Ohkuni and Kitagawa 2011). Chan et al. (2012) showed a similar effect in HeLa cells, wherein inhibition of RNAPII transcription resulted in a decrease in CT levels and CENP-C deposition, concomitant with an increase in lagging chromosomes during cell division.

Epigenetic modifications of engineered human artificial chromosomes (HACs) confirmed the need for a critical balance in transcription levels for properly

functioning centromeres (Bergmann et al. 2011, 2012 and reviewed in this issue). These studies also showed that active transcription, not histone modifications, is a key element to maintaining centromere function. When the levels of H3K4me2 at an engineered HAC centromere were decreased, a significant decrease in transcription was coincident with a decrease in the loading of newly synthesized CENP-A and decrease in CENP-C at the kinetochore (Bergmann et al. 2011) (Fig. 1d). Alteration of the HAC centrochromatin to an open chromatin state likewise resulted in disrupted CENP-A loading; the altered chromatin state did not directly affect the HAC's centromere activity, rather the resulting increase in transcription lead to the observed detrimental effects (Bergmann et al. 2012) (Fig. 1d). Thus, a tight regulation of transcription is essential for proper centromere assembly (O'Neill and Carone 2009). Okada et al. (2009) identified protein complexes, including facilitates chromatin transcription (FACT), which contains CENP-H and CHD1 (chromodomain-helicase-DNA-binding protein 1) that are required for CENP-A deposition. While active transcription and CENP-A loading was not assayed in this study, the localization of FACT at the centromere support the model that active transcription is part of the process required for CENP-A deposition.

There is increasing evidence of a direct RNA–protein interaction between CTs and CENP-A bound chromatin in maize, mouse, and human. The specific function of each type of CT may be dependent on their size, with observed CTs associating with CENP-A ranging from 35 to >900 nt in length (Topp et al. 2004; Chueh et al. 2009; Ferri et al. 2009; Du et al. 2010). For example, the knockdown of a specific sized long interspersed nucleotide element (LINE), which is shown to interact with CENP-A at a human neocentromere, causes a decrease in CENP-A loading (Chueh et al. 2009). CenH3 in maize was shown to interact with long CentC satellite transcripts (Du et al. 2010). Interestingly, in the tammar wallaby, small centromeric transcripts align to the same centromeric sequences found specifically within CENP-A nucleosomes (Renfree et al. 2011). Conceivably, the small and long RNAs interact with CENP-A at different times during the cell cycle, thereby temporally separating their respective functions. CTs of various sizes are also part of a non-nucleosomal protein complex that includes CENP-A. CTs in mouse are needed for the proper activity of Aurora B kinase and the proper association

of CENP-A with Aurora B kinase and Survivin (Ferri et al. 2009). CTs also associate with other centromeric proteins including CENP-C (Wong et al. 2007; Du et al. 2010) and INCENP (Wong et al. 2007). The localization of CENP-C and INCENP to both the nucleolus and centromere is dependent upon the presence of centromeric RNA (Wong et al. 2007).

Due to the variety of interactions CTs have with various proteins, culminating in the proper loading of newly synthesized CENP-A required for centromere assembly, it is plausible that, like PCTs, there are specific size classes of CTs, each possessing a unique function. The type of RNAs that cooperate with specific proteins in centromere function may also vary in their structure, dependent on whether they are single-stranded, double-stranded, or DNA–RNA heteroduplexes. While no RNA-binding domain has been defined for CENP-A, several studies have shown CENP-A can associate directly with RNA (e.g. Topp et al. 2004). Moreover, since CENP-C is known to bind single-stranded RNA in the same domain shown to bind DNA (Du et al. 2010), it is possible that the same holds true for CENP-A. Remarkably, the nucleotide sequence of CENP-C associated, single-stranded RNA does not affect its ability to bind CENP-C; instead, the size of the transcript affects binding capability (Du et al. 2010), supporting the idea that there are different size classes of transcripts with potentially different functions. The function of the RNA may also be dependent on the type of repeat element being transcribed. For example, there are different subtypes of centromeric retroelements in rice and wallaby, and some of these subtypes are processed into small RNAs while others remain as long transcripts (Neumann et al. 2007; Carone et al. 2009; Ferreri et al. 2011). Interestingly, some of these retroelement transcripts are also alternatively spliced resulting in the production of slightly different transcripts (Neumann et al. 2007).

Centromeric transcription in cellular stress and disease

The conservation of centromeric and pericentric transcription across major eukaryotic lineages indicates that these transcripts may play a critical role in the cell. Unsurprisingly, there has been mounting evidence in the past few years that if the level of PCTs and CTs is not kept in a perfect balance, there are potentially dire consequences for the organism.

The first indication of the functional importance of satellite transcripts was the discovery that transcription of certain pericentric satellites was induced under cellular stress in human cells (Jolly et al. 2004; Rizzi et al. 2004). Stress can be induced in cells by exposing them to any condition outside of their optimal growth range; this includes subjecting cells to high temperature (heat shock), heavy metals, hazardous chemicals, ultraviolet radiation, and hyperosmotic or oxidative conditions (Valgardsdottir et al. 2008). After human cells are exposed to high temperatures, heat shock factor 1 protein (HSF1) is upregulated and associates with distinct nuclear structures, termed nuclear stress bodies (nSBs), that accumulate on the pericentric regions of chromosomes (Jolly et al. 2004). The nSBs recruit RNAPII and the pericentric satellite III is then highly transcribed. While this process is best characterized in heat-shocked cells, satellite III PCTs are also induced during many other kinds of cellular stresses, albeit under control of different transcription factors dependent on the type of stress the cell is subjected to, i.e., the tonicity-responsive enhancer binding protein, tonEBP, during hyperosmotic stress (Valgardsdottir et al. 2008). This finding indicates that PCTs are a commonality to multiple stress response and recovery pathways.

The cellular response of satellite transcript accumulation during stress is, like the production of the PCTs and CTs themselves, highly conserved. In mouse cells stressed by chemical exposure, minor satellite transcription increases, resulting in an accumulation of 120nt transcripts (Bouzinba-Segard et al. 2006). While unstressed cells were found to contain a basal level of these CTs, the forced accumulation of transcripts impaired centromere function, leading to decondensed centromeres and mitotic defects, such as multiple spindle attachments, loss of sister chromatid cohesion, and aneuploidy (Bouzinba-Segard et al. 2006). Notably, CenpB and CenpC were retained normally in these cells, implying that inner kinetochore function was perhaps unaffected. The accumulation of satellite transcripts during stress conditions also occurs in insects (Pezer et al. 2011; Pezer and Ugarkovic 2012) and *Arabidopsis* (Pecinka et al. 2010; Tittel-Elmer et al. 2010), although the precise role of the transcripts in these other organisms remains to be elucidated.

Eymery et al. (Eymery et al. 2009b) compared satellite transcription during cellular stress using a microarray strategy based on specific satellite variants

from centromere regions (delineated by enrichment for CENP-A) and pericentromeres (delimited by association with H3K9me3 and HP-1) in stressed and diseased human cells. These cell stress experiments supported the well-documented upregulation of PCTs during recovery from heat shock; however, it was also discovered that while PCTs were globally upregulated, CTs were not. This finding supports the hypothesis that PCTs and CTs are under different transcriptional controls, at least during stress recovery. Perhaps more intriguing, however, are the observed transcript differences among the various normal and tumorigenic tissues assayed in this study. One of the most striking differences was found in testis, in which antisense PCTs were highly expressed in normal tissue but were silenced in testicular cancer tissue from the same patient (Eymery et al. 2009b). This result contrasted with that of lung samples, in which pericentric transcription was repressed in normal tissue and upregulated in the tumor. CTs were also present in normal ovary, placenta, fetal liver, and fetal kidney samples, which raises interesting questions about the role these CTs may be playing in development and differentiation. Since HSF1 protein was not upregulated in any of the normal or tumor samples, the dysregulated PCTs and CTs were not considered linked to the heat shock pathway (Eymery et al. 2009b). Thus, this work poses the questions of how and why satellite transcription changes in tumorigenesis and what functions, if any, these satellite transcripts perform specific to tumor progression.

A few recent studies have addressed how and why centromeric transcription changes during oncogenesis (Frescas et al. 2008; Iotti et al. 2011; Slee et al. 2012; Ting et al. 2011; Zhu et al. 2011). The tumor suppressor lysine-specific demethylase 2A (KDM2A) was identified as a heterochromatin associated protein that is downregulated in prostate cancer (Frescas et al. 2008). Through its Jumonji domain, KDM2A demethylates pericentric H3K36me₂, thereby maintaining a closed chromatin state and silencing nascent transcription. KDM2A knockdown resulted in a loss of HP1 from the pericentromere and a large increase in major or alpha satellite transcripts in both mouse and human cells, respectively (Frescas et al. 2008) (Fig. 2a). This loss of the heterochromatic state leads overall to genomic instability, including the misalignment of centromeres along the mitotic plate and segregation defects such as chromosome breaks and bridges.

Interestingly, the lower the level of KDM2A expression in prostate cancer, the more severe the tumor grade, linking an increase in PCT-mediated instability to cancer prognosis.

Satellite derepression was linked to DNA methylation when next generation digital gene expression analysis was used to measure the transcriptional output of pancreatic ductal adenocarcinomas (PDACs) and a variety of other epithelial primary tumors in mice and human (Ting et al. 2011). In mice, pericentric (major) satellite expression in tumor tissue was greatly increased over normal tissue and in human tumor samples, both alpha satellite and satellite II had significantly higher expression levels compared to normal tissues. Satellite derepression was also shown to be specific to *in vivo* cancer conditions, since PDAC tumor cells immortalized *in vitro* no longer expressed the same satellite repeats (Ting et al. 2011). A defect in normal heterochromatin DNA methylation was implicated as the cause of the satellite derepression; once the immortalized PDAC cells were treated with the demethylase 5-aza-2'-deoxycytidine, the cells began to re-express the satellite repeats at levels similar to that observed within *in vivo* conditions. To address what may be promoting transcription in the tumor tissues, a linear regression analysis was used to find transcripts that were co-regulated with major satellite (mouse) or alpha satellite (human) (Ting et al. 2011). LINE1 transposable elements were found to be upregulated in tumor tissues and as a consequence, genes with LINE1 insertions were differentially regulated compared to normal tissue. Interestingly, many of these differentially regulated genes function in neural cell fate and stem cell pathways, implicating the misregulation of PCT expression in the neuronal differentiation pathway leading to these cancers (Ting et al. 2011).

DMNT3B is a DNA methyltransferase responsible for maintaining the proper methylation levels at the pericentromere and centromere (Gopalakrishnan et al. 2009). DMNT3B is recruited to satellite repeats through an interaction with the centromere protein, CENP-C. Any impairment of that interaction prevents methylation in the pericentric and centromeric regions and results in overexpression of PCTs and CTs (Gopalakrishnan et al. 2009). Mutations in DMNT3B are known to lead to immunodeficiency, centromere instability, and facial anomalies syndrome (ICF) (Hansen et al. 1999). ICF patients have hypomethylated

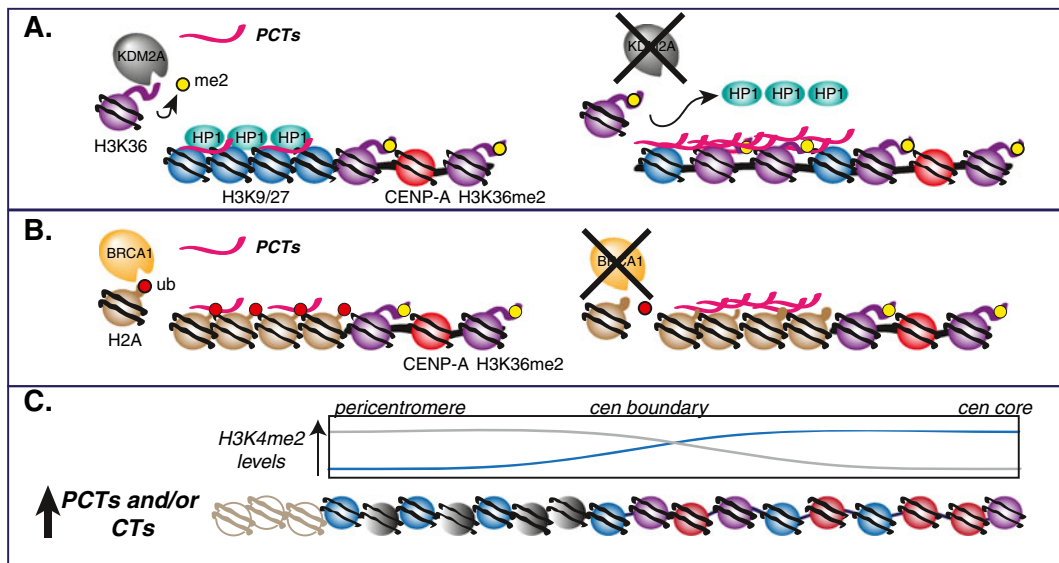


Fig. 2 **a** KDM2A (gray) is a demethylase that targets H3K36me2. Under normal conditions (*left*), HP1 is associated with PCTs (*red*) and pericentric nucleosomes (*blue*), adjacent to centrochromatin containing H3K36me2 (*purple*) and CENP-A (*red*) nucleosomes. Loss of KDM2A (*right*) results in an increase of H3K36me2 within the pericentromere concomitant with a loss of HP1 and dramatic increase in PCTs. **b** BRCA1 (*orange*) targets H2A within the pericentromere (*brown*) for

ubiquitination (*small red circle*) under normal conditions (*left*). Loss of BRCA1 (*right*) leads to loss of ubiquitination of H2A within the pericentromere and a dramatic increase in PCTs. **c** An increase in PCTs or CTs leads to a shift of H3K9me3 (*blue*) from mostly pericentric (indicated in *gray* on the graph inset) to spreading into centrochromatin (indicated in *blue* on the graph inset). Nucleosome key as per Fig. 1a and b

heterochromatic DNA and, consequently, derepressed heterochromatic genes and satellite repeats, while euchromatic gene methylation remains at normal levels (Brun et al. 2011). DNMT3B then is yet another protein specifically responsible for maintaining normal PCT and CT levels.

The hereditary ovarian and breast cancer susceptibility gene (BRCA1) has been extensively studied, with conflicting results regarding its normal function and how it behaves as a tumor suppressor. Since BRCA1 mutation leads to genomic instability, it has been predicted to function in DNA replication, DNA damage repair, cell cycle control, and a host of other mitotic and regulatory functions. Recently, BRCA1 was discovered to be involved in maintaining specific epigenetic states within centromeric and pericentric regions (Zhu et al. 2011). BRCA1 protein, through an E3 ligase activity in its RING finger domain, is responsible for monoubiquitinating the histone H2A (Chen et al. 2002). When BRCA1 is knocked out, there is a global increase in both major and minor satellite transcription in mouse cells, and of alpha satellite in human cells, concomitant with a loss of

H2A ubiquitination (Fig. 2b); other effects included a reduction in heterochromatin centers as well as numerous mitotic defects and an increase in DNA double-strand breaks (Zhu et al. 2011). In contrast to the work of Ting et al. (2011), there was no observable increase in LINE or other retrotransposable element activity associated with PCT and CT increases (Zhu et al. 2011). Thus, while the loss of H2A ubiquitination, via loss of BRCA1 function, may be responsible for converting heterochromatin into an open chromatin state and allowing transcription, it remains unknown what factors or sequences promote the observed satellite transcription in BRCA1-deficient cells. Through ectopic expression of satellite RNA in cells lacking BRCA1 mutations, Zhu et al. (2011) were able to support the previous prediction that increased satellite expression facilitated the genomic instability, as they observed many of the same defects in these cells as in BRCA1-deficient cells.

Other work suggests that PCT and CT derepression may be a common feature to certain breast cancers, regardless of the protein involved. Jumonji domain containing protein, JMJD2B, is a demethylase and an

oncogene in certain breast tumors (Slee et al. 2012). When overexpressed, JMJD2B causes a large decrease in centromeric H3K9me3, which leads to chromosome instability. Although the levels of PCTs and CTs in these JMJD2B tumors were not measured, it is plausible that the transcripts are derepressed, as they are in BRCA1-null tumors.

Taken together, all of these studies support the theory that if heterochromatic epigenetic marks are altered to a transcriptionally active state, the resulting overexpression of satellite sequences can lead to genomic instability and oncogenesis. In contrast, the action of the tumor suppressor Prep1 indicates that complete silencing of satellite expression may be just as detrimental as derepression. Prep1 (also known as PKNOX1 or PBX/knotted 1 homeobox1 in human) is implicated in controlling DNA damage and regulating histone methylation levels (Iotti et al. 2011). When Prep1 levels are downregulated in mouse or human cells, DNA damage increases, which leads, through an unknown mechanism, to a widespread increase in the repressive histone mark H3K9me3 (Iotti et al. 2011) (Fig. 2c). Consequently, major satellite transcription in mouse, and alpha satellite transcription in human decreased 62 and 45 %, respectively, when compared to controls with normal Prep1 levels. Remarkably, this decrease in PCTs and CTs leads to precisely the same cellular phenotypes observed in cells with an increase in PCTs and CTs, such as aneuploidy, miniature chromosomes, Robertsonian translocations, and centromere duplications.

As observed in the engineered HAC studies described above, a tight control of PCT and CT transcription levels are required to maintain genomic stability at native centromeres. While the work described herein demonstrates potential mechanisms for how PCT and CT levels change during oncogenesis, the critical question of how exactly the overabundance or lack of transcripts leads to genomic instability remains less clear. Genomic instability is a broad term that is characterized by a host of replicative, mitotic, and chromosomal defects. While it will likely take years to determine all the mechanisms leading to genomic stability in cancers with PCT and CT involvement, the answers could lead to broader treatment options. As work in the field continues, it will be interesting to determine how satellite transcription might be involved in other known oncogenic consequences. If satellite transcripts do play a role in the recruitment of centromere and kinetochore proteins,

this could explain the upregulation and mislocalization of CENP-A observed in colon cancer (Tomonaga et al. 2003). In like manner, if PCTs and CTs aide in defining or maintaining heterochromatin boundaries, their oncogenic dysregulation may explain why there is an observed expansion of CENP-A nucleosomes on alpha satellite arrays in cancer cells (Sullivan et al. 2011).

Concluding remarks

Despite the fact that the biology of centromeric/pericentric transcription is seemingly complex, the distinction between PCTs and CTs is important to make as these transcripts emanate from separate chromatin environments, each with a specific function with respect to the centromere. While many general conclusions have been made as to the contribution of PCTs and CTs in a variety of cellular processes, precise mechanisms remain elusive. The protein, DNA, and other noncoding RNA interactions that PCTs and CTs undertake in both normal and abnormal conditions are not fully characterized, nor are the diversity of size classes and types of RNAs that are derived from PCTs and CTs. Once the normal interactions and their functional consequences are revealed, work can begin on addressing therapies to correct the abnormal situations that lead to genome instability and oncogenesis. Understanding the transcriptional framework that controls CT and PCT production may also provide insight into long unanswered mysteries, such as the biological foundation of the centromere paradox and the complete pathway that determines CENP-A positioning (both established and novel) as well as centromere inactivation in the genome. While there are many questions left to answer, it is clear that proper transcription of pericentric and centromeric sequences is crucial to proper centromere function, genome stability, and accurate cell division.

Acknowledgments We thank Judy Brown for editorial comments. LH, ST, and RO were supported by a grant from the NSF. We apologize to all whose work could not be included herein due to space constraints.

References

- Allshire RC, Karpen GH (2008) Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat Rev Genet* 9 (12):923–937

- Bergmann JH, Rodriguez MG, Martins NMC et al (2011) Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *EMBO J* 30(2):328–340
- Bergmann JH, Jakubsche JN, Martins NM et al (2012) Epigenetic engineering: histone H3K9 acetylation is compatible with kinetochore structure and function. *J Cell Sci* 125(Pt 2):411–421
- Birchler JA, Gao Z, Sharma A, Presting GG and Han F (2011) Epigenetic aspects of centromere function in plants. *Current Opinion in Plant Biology* 14:217–222
- Bouzinba-Segard H, Guais A, Francastel C (2006) Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. *Proc Natl Acad Sci U S A* 103(23):8709–8714
- Brun ME, Lana E, Rivals I et al (2011) Heterochromatic genes undergo epigenetic changes and escape silencing in immunodeficiency, centromeric instability, facial anomalies (ICF) syndrome. *PLoS One* 6(4):e19464
- Carone DM, Longo MS, Ferreri GC et al (2009) A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. *Chromosoma* 118(1):113–125
- Ceccarelli M, Morosi L, Cionini PG (1998) Chromocenter association in plant cell nuclei: determinants, functional significance, and evolutionary implications. *Genome* 41:96–103
- Chan FL, Marshall OJ, Saffery R et al (2012) Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proc Natl Acad Sci* 109(6):1979–1984
- Chen A, Kleiman FE, Manley JL, Ouchi T, Pan ZQ (2002) Autoubiquitination of the BRCA1-BARD1 RING ubiquitin ligase. *J Biol Chem* 277:22085–22092
- Chen ES, Zhang K, Nicolas E et al (2008) Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature* 451(7179):734–737
- Choi ES, Stralfors A, Castillo AG et al (2011) Identification of noncoding transcripts from within CENP-A chromatin at fission yeast centromeres. *J Biol Chem* 286(26):23600–23607
- Chueh AC, Northrop EL, Brettingham-Moore KH, Choo KHA, Wong LH (2009) LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin. *PLoS Genetics* 5(1):e1000354
- Dawe RK, Henikoff S (2006) Centromeres put epigenetics in the driver's seat. *Trends Biochem Sci* 31(12):662–669
- De S, Varsally W, Falciani F, Brogna S (2011) Ribosomal proteins' association with transcription sites peaks at tRNA genes in *Schizosaccharomyces pombe*. *RNA* 17(9):1713–1726
- Djupedal I, Kos-Braun IC, Mosher RA et al (2009) Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. *EMBO J* 28(24):3832–3844
- Du Y, Topp CN, Dawe RK (2010) DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet* 6(2):e1000835
- Ellermeier C, Higuchi, EC, Phadnis, N et al (2010) RNAi and heterochromatin repress centromeric meiotic recombination. *Proc Natl Acad Sci USA* 107:8701–8705
- Eymery A, Callanan M, Vourc'h C (2009a) The secret message of heterochromatin: new insights into the mechanisms and function of centromeric and pericentric repeat sequence transcription. *Int J Dev Biol* 53(2–3):259–268
- Eymery A, Horard B, El Atifi-Borel M et al (2009b) A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. *Nucleic Acids Res* 37(19):6340–6354
- Ferguson-Smith AC (2011) Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet* 12(8):565–575
- Ferreri GC, Brown JD, Oberfell C et al (2011) Recent amplification of the kangaroo endogenous retrovirus, KERV, limited to the centromere. *J Virol* 85(10):4761–4771
- Ferri F, Bouzinba-Segard H, Velasco G, Hube F, Francastel C (2009) Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. *Nucleic Acids Res* 37(15):5071–5080
- Frescas D, Guardavaccaro D, Kuchay SM et al (2008) KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. *Cell Cycle* 7(22):3539–3547
- Gent JI, Dong Y, Jiang J, Dawe RK (2012) Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. *Nucleic Acids Res* 40(4):1550–1560
- Gopalakrishnan S, Sullivan BA, Trazzi S, Della Valle G, Robertson KD (2009) DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. *Hum Mol Genet* 18(17):3178–3193
- Hansen RS, Wijmenga C, Luo P et al (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc Natl Acad Sci U S A* 96(25):14412–14417
- Henikoff S, Ahmad K, Malik H (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293(5532):1098–1102
- Hsieh CL, Lin CL, Liu H et al (2011) WDHD1 modulates the post-transcriptional step of the centromeric silencing pathway. *Nucleic Acids Res* 39(10):4048–4062
- Iotti G, Longobardi E, Masella S et al (2011) Homeodomain transcription factor and tumor suppressor Prep1 is required to maintain genomic stability. *Proc Natl Acad Sci U S A* 108(29):E314–E322
- Iwasaki O, Tanaka A, Tanizawa H, Grewal SIS, Noma K-I (2010) Centromeric localization of dispersed Pol III genes in fission yeast. *Mol Biol Cell* 21(2):254–265
- Jiang J, Birchler JA, Parrott WA, Dawe RK (2003) A molecular view of plant centromeres. *Trends Plant Sci* 8(12):570–575
- Jolly C, Metz A, Govin J et al (2004) Stress-induced transcription of satellite III repeats. *J Cell Biol* 164(1):25–33
- Lee HR, Neumann P, Macas J, Jiang J (2006) Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. *Mol Biol Evol* 23(12):2505–2520
- Lippman Z, Martienssen R (2004) The role of RNA interference in heterochromatic silencing. *Nature* 431(7006):364–370
- Neumann P, Yan H, Jiang J (2007) The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. *Genetics* 176(2):749–761

- Ohkuni, K and Kitagawa, K (2011) Endogenous transcription at the centromere facilitates centromere activity in budding yeast. *Curr Biol* 21:1695–1703
- Okada M, Okawa K, Isobe T, Fukagawa T (2009) CENP-H-containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1. *Mol Biol Cell* 20(18):3986–3995
- O'Neill RJ, Carone DM (2009) The role of ncRNA in centromeres: a lesson from marsupials. *Prog Mol Subcell Biol* 48:77–101
- Pecinka A, Dinh HQ, Baubec T et al (2010) Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in *Arabidopsis*. *Plant Cell* 22(9):3118–3129
- Pezer Ž, Ugarković Đ (2008) RNA Pol II promotes transcription of centromeric satellite DNA in beetles. *PLoS One* 3(2):e1594
- Pezer Z, Ugarkovic D (2012) Satellite DNA-associated siRNAs as mediators of heat shock response in insects. *RNA Biol* 9(5):1
- Pezer Z, Brajkovic J, Feliciello I, Ugarkovic D (2011) Transcription of satellite DNAs in insects. *Prog Mol Subcell Biol* 51:161–178
- Probst AV, Okamoto I, Casanova M et al (2010) A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. *Dev Cell* 19(4):625–638
- Renfree MB, Papenfuss AT, Deakin JE et al (2011) Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol* 12(8):R81
- Reyes-Turcu, FE, Zhang, K, Zofall, M, Chen, E and Grewal, SIS (2011) Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. *Nat Struct Mol Biol* 18:1132–1138
- Rizzi N, Denegri M, Chiodi I et al (2004) Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock. *Mol Biol Cell* 15(2):543–551
- Rudert F, Bronner S, Garnier JM, Dolle P (1995) Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. *Mamm Genome* 6(2):76–83
- Saffery R, Sumer H, Hassan S et al (2003) Transcription within a functional human centromere. *Mol Cell* 12(2):509–516
- Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. *Science* 294(5540):109–115
- Scott KC, White CV, Willard HF (2007) An RNA polymerase III-dependent heterochromatin barrier at fission yeast centromere 1. *PLoS One* 2(10):e1099
- Slee, RB, Steiner, CM, Herbert, BS et al (2012) Cancer-associated alteration of pericentromeric heterochromatin may contribute to chromosome instability. *Oncogene* (in press)
- Stimpson KM, Sullivan BA (2010) Epigenomics of centromere assembly and function. *Curr Opin Cell Biol* 22(6):772–780
- Sullivan BA, Karpen GH (2004) Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol* 11(11):1076–1083
- Sullivan BA, Blower MD, Karpen GH (2001) Determining centromere identity: cyclical stories and forking paths. *Nat Rev Genet* 2(8):584–596
- Sullivan LL, Boivin CD, Mravinac B, Song IY, Sullivan BA (2011) Genomic size of CENP-A domain is proportional to total alpha satellite array size at human centromeres and expands in cancer cells. *Chromosome Res* 19(4):457–470
- Ting DT, Lipson D, Paul S et al (2011) Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science (New York, NY)* 331(6017):593–596
- Tittel-Elmer M, Bucher E, Broger L et al (2010) Stress-induced activation of heterochromatic transcription. *PLoS Genet* 6(10):e1001175
- Tomonaga T, Matsushita K, Yamaguchi S et al (2003) Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. *Cancer Res* 63(13):3511–3516
- Topp CN, Zhong CX, Dawe RK (2004) Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc Natl Acad Sci U S A* 101(45):15986–15991
- Valgardsdottir R, Chiodi I, Giordano M et al (2008) Transcription of Satellite III non-coding RNAs is a general stress response in human cells. *Nucleic Acids Res* 36(2):423–434
- Wong LH, Brettingham-Moore KH, Chan L et al (2007) Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res* 17(8):1146–1160
- Wu Y, Kikuchi S, Yan H et al (2011) Euchromatic subdomains in rice centromeres are associated with genes and transcription. *Plant Cell* 23(11):4054–4064
- Yan H, Weiwei J, Kiyotaka N et al (2005) Transcription and histone modifications in the recombination-free region spanning a rice centromere. *Plant Cell* 17(12):3227–3238
- Yan H, Ito H, Nobuta K et al (2006) Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere. *Plant Cell* 18(9):2123–2133
- Zhu Q, Pao GM, Huynh AM et al (2011) BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* 477(7363):179–184