DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES

DER FAKULTÄT FÜR BIOLOGIE

DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Expression and functional analysis of minor satellite RNA during mitosis in mouse embryonic stem cells

Yung-Li Chen

Diese Dissertation wurde angefertigt

unter der Leitung von Prof. Dr. Maria-Elena Torres-Padilla

im Institute of Epigenetics and Stem Cells

Helmholtz Center Munich

Erstgutachter: Prof. Dr. Maria - Elena Torres - Padilla

Zweitgutachter: PD. Dr. Tamara Mikeladze-Dvali

Tag der Abgabe: 22.03.2024

Tag der mündlichen Prüfung: 11.09.2024

EIDESSTATTLICHE ERKLÄRUNG

Ich versichere hiermit an Eides statt, dass ich meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt habe.

Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt.

Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 22.03.2024

Yung-Li Chen

Copyright authorization

Some of my results in this thesis have been published and are available in peer-reviewed article.

1. **Chen, Y. L.**, Jones, A. N., Crawford, A., Sattler, M., Ettinger, A., & Torres-Padilla, M. E. (2024). Determinants of minor satellite RNA function in chromosome segregation in mouse embryonic stem cells. The Journal of cell biology, 223(7), e202309027. https://doi.org/10.1083/jcb.202309027

For Figure 1, the scheme is modified and reproduced from published journal article. According to MDPI Open Access Information and Policy (https://www.mdpi.com/openaccess), all parts of the article can be re-used with proper article citation.

1. Balzano, E., & Giunta, S. (2020). Centromeres under Pressure: Evolutionary Innovation in Conflict with Conserved Function. Genes, 11(8), 912. https://doi.org/10.3390/genes11080912

Table of contents

Table of contents **2**

Abstract **5**

Introduction **6**

Chromosome segregation 6

Centromere formation and kinetochore assembly 6

Evolution and types of centromeres 8

Transcription of centromeres 10

The function of centromeric RNAs 10

Function and evolutionary conservation of non-coding RNAs (ncRNAs) 12

Scope and aim of the thesis 14

Results **16**

Mouse minor satellite RNAs are expressed from both strands in mouse embryonic stem cells 16

The endogenous MinSat RNA level reveals a cell cycle-related pattern 17

Overexpression of MinSat RNA leads to chromosome missegregation and reduction in cell proliferation 19

Specificity validation of CENPC antibody ab193666 22

Overexpression of reverse MinSat transcripts reduces total protein level of CENPA and disrupts the binding of CENPC to chromatin 24

Strand-specific-knockdown of MinSat RNAs with antisense oligos reveals a mutual feedback loop regulating MinSat RNA homeostasis 26

Knock-down of MinSat RNAs results in chromosome missegregation and genome instability 27

MinSat RNA knock-down interferes with the binding of CENPC and CENPA to chromatin during mitosis 29

CENPC binds to MinSat forward transcripts via its RNA binding region 31

MinSat RNA forward transcripts adopt a specific secondary structure 33

Human and mouse centromeric RNAs both possess similar stem-loop structures with apical loops containing the CENPB box sequence. 36

Overexpression of human α-satellite RNA leads to an increase in chromosome missegregation events in mouse cells 38

The apical loop structure rather than CENPB box is critical for MinSat forward RNA function 40

Binding affinity to CENPC decreases in RNA with CENPB box mutants 42

Enrichment of MinSat transcripts from total RNA for Oxford Nanopore sequencing 43

Mapping of MinSat RNA reads 45

Centromeric RNAs are transcribed in *Xenopus laevis* and *Bos taurus* 49

Discussion 51

An optimal expression level of centromeric RNA is required to maintain chromosome segregation fidelity 51

Imbalanced levels of forward or reverse MinSat RNA lead to chromosome missegregation 52

CENPC binds RNA potentially via a stem-loop motif 53

Secondary structure of centromeric RNA and evolution 55

Aberrant expression or oncogenic lncRNAs can lead to mislocalization of centromere formation 55

The apical loop may contribute to the function of centromeric RNAs 56

While MinSat Swap mutant leads to chromosome missegregation, the missegregation may not be mediated by MinSat RNA- CENPC protein interaction 57

Formation of a functional centromere across species 58

Material and Methods 60

Cell culture 60

Plasmid transfection 60

Total RNA extraction and Real-Time qPCR 60

Northern blot 61

Knockdown of MinSat RNAs 62

Immunofluorescence 62

Chromatin extraction and Western blot analysis 62
Cell proliferation assay 63
Flow cytometry analysis of cell cycle distribution 63
3'end biotinylated RNA pulldown assay 64
RNA transcription and purification 64
Selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) 65
Nuclear magnetic resonance 65
CatRAPID prediction of RNA binding ability 65
Protein expression and purification 65
Binding gel shift assays 66
Sequence-specific enrichment using Biotinylated DNA oligos and Streptavidin magnetic beads 66
Library preparation for sequence-specific cDNA nanopore sequencing 67
Statistical analyses 68
References 69
Appendix 86
Contributions 91
Curriculum Vitae 92
Acknowledgments 95

Abstract

The centromere is a specialized higher-order region on the chromosome that is necessary for chromosome segregation fidelity during cytokinesis. Thus, faithful formation of centromere structure is crucial for the progress of cell division. Transcripts derived from the centromeric region are involved in the regulation of centromere function. The mechanism whereby centromeric RNAs (cenRNAs) interact with centromeric proteins appears conserved in multiple species. However, in contrast, centromeric DNA is one of the fastest-evolving regions within the genome. The low similarity of centromeric sequences between species creates a paradox for the model in which conserved centromeric proteins can interact with diverse DNA or RNA sequences. One possibility is that the secondary structure, as opposed to nucleotide sequence, is responsible for the function of cenRNAs. By manipulating minor satellite (MinSat) RNA levels, I first discovered that disrupting the balance of endogenous MinSat RNA levels in mouse embryonic stem cells (ESCs) induces chromosome missegregation. I showed that CENPC, an inner kinetochore protein interacting with CENPA nucleosomes, has its chromatin binding ability decrease after manipulation of MinSat RNA levels. I further identified a stem-loop secondary structure on MinSat RNA, which is also conserved in human α-satellite transcripts. In contrast, mutants that disrupt the identified secondary structure did not lead to missegregation phenotypes. This suggests that this structural component indeed triggers chromosome missegregation events. In this study, I provide a potential puzzle piece for unraveling the function of cenRNAs suggesting that the conserved function of cenRNAs is due to their RNA secondary structure. Based on the idea that the conservation of RNA on a structural level conveys cenRNA function, cenRNAs and proteins may co-evolve without losing their interaction.

Introduction

Chromosome segregation

Multiple cellular mechanisms ensure chromosome segregation fidelity, equal division of genetic material, and chromosome stability during mitosis and meiosis (Marston, 2014, Tanaka & Hirota, 2016). In eukaryotic cells, aberrant sister chromatid segregation can not only disrupt cell cycle progression but also lead to chromosomal rearrangements. These are known as potential factors for neoplastic transformation and tumor formation (Ly, Brunner et al., 2019). During the formation of gametes, abnormal segregation of chromosomes is one of the causes of genetic disorders like trisomies or can lead to developmental arrest and miscarriage (Hassold & Hunt, 2001, Tsuiko, Jatsenko et al., 2019, Zickler & Kleckner, 1999). During gametogenesis, accurate chromosome segregation is a key indicator of successful meiosis. Failure of chromosome distribution in meiosis I or meiosis II can contribute to aneuploidy of offspring or infertility of adults (Webster & Schuh, 2017).

On a molecular level, chromosome segregation includes strictly controlled steps such as centromere formation and kinetochore assembly. Mature centromeres and kinetochores serve as a platform for mitotic spindle attachment (Fukagawa & Earnshaw, 2014, Santaguida & Musacchio, 2009). To achieve precise segregation of chromosomes, the attachment of chromosomes to the mitotic spindle must be highly regulated. Failures in spindle attachment can cause lagging chromosomes and micronuclei, which are hallmarks of defective DNA replication, chromosome fragmentation and chromosome rearrangements (Kapoor, 2004, Thompson & Compton, 2011).

Centromere formation and kinetochore assembly

In 1882, Walter Flemming first described a specialized central structure on mitotic chromosomes (Flemming, 1882). The centromere was then characterized as an element on a chromosome that contributes to mitotic stability of chromosomes in budding yeast (Clarke & Carbon, 1980, Clarke & Carbon, 1985) a hundred years afterward. Centromeric DNA is one of the critical components of centromere formation. Centromeric DNA consists of repetitive sequences and this feature is evolutionarily conserved (Biscotti, Canapa et al., 2015). However, the similarity of centromeric DNA sequences is low across species (Henikoff, Ahmad et al., 2001, Rosic, Kohler et al., 2014). In mammals, a consensus sequence named CENPB-box has been characterized as a binding motif for centromeric protein CENPB (Iwahara, Kigawa et al., 1998,

Masumoto, Masukata et al., 1989, Suntronpong, Kugou et al., 2016). Other than this motif, no other features at DNA sequence level or general rules that govern centromeric sequence evolution have been found. So far, the heterochromatic region containing repetitive sequences is the only conserved feature recognized at centromere loci across species.

A potential function of the repetitive sequences that form the centromere is to maintain (peri)centromeric heterochromatin, thus isolating centromeres from the rest of chromosomal structures. Compared to peri-centromeric regions, which are abundant in H3K9me3, the current definition describes centromeres as epigenetically defined regions enriched in CENPA nucleosomes. The CENPA nucleosomes are scattered among canonical H3 nucleosomes modified with H3K4me1, H3K4me2, H3K36me2 and H3K36me3. These modifications provide a euchromatic environment amenable to PolII binding and CENPA deposition (Allshire & Karpen, 2008, Corless, Hocker et al., 2020, Hall, Mitchell et al., 2012, Smurova & De Wulf, 2018). Remarkably, increasing evidence has shown that the formation of peri-centromeric heterochromatin plays also an important role during de novo centromere formation. Due to the physical feature of heterochromatin, peri-centromeric region can insulate between centromeric repeats to avoid recombination in between repeats in the genome (Bernstein & Allis, 2005). Importantly, SUV4-20H2 methyltransferase can compact peri-centromeric region and further recruit cohesin to the heterochromatic region (Hahn, Dambacher et al., 2013). Cohesin recruitment is critical for maintaining the tension on sister chromatid during cytokinesis (Bernard, Maure et al., 2001, Sakuno, Tada et al., 2009, Yamagishi, Honda et al., 2010, Yi, Chen et al., 2018).

After centromere formation, a functional kinetochore is important as an attaching point for microtubules on the chromosomes before moving on cytokinesis. The kinetochore complex can be distinguished into four layers (Musacchio & Salmon, 2007). The first layer is the centromeric chromatin which is composed of CENPA-specific nucleosomes. Initiated by specific binding between CENPC and CENPA, the second layer named inner kinetochore is formed by the constitutive centromere associated network (CCAN) complex. The CCAN complex contains multiple centromeric proteins such as CENPC, CENP-T and CENP-N. The main function of the CCAN complex is to re-organize or compact CENPA nucleosomes to form a rigid platform for a stronger connection to the outer kinetochore and mitotic spindle (McAinsh & Meraldi, 2011, Pesenti, Raisch et al., 2022). Within the outer kinetochore, the NDC80/HEC1 complex acts as a key component to connect mitotic spindle microtubules to the entire kinetochore complex. In the corona layer – the fourth layer -

of the kinetochore, the ROD-Zwilch-ZW10 (RZZ) protein complex and the protein Spindly can stabilize the interaction between the mitotic spindle and the kinetochore (Raisch, Ciossani et al., 2022). While most of the components and functional interactions between these four layers are thought to be conserved, it is the human kinetochore that has been studied the most.

Evolution and types of centromeres

From the first described centromere of budding yeast *Saccharomyces cerevisiae* to metazoans such as human and mouse, there is a great heterogeneity of centromeres between species. Based on the region size, distribution and structure, we can categorize most centromeres into several types: point centromeres, regional centromeres, holocentric centromeres and meta-polycentric centromeres (Balzano & Giunta, 2020). First, point centromeres (~125 bp), which are only in fungi such as *Saccharomyces cerevisiae* (Furuyama & Biggins, 2007). Second, regional centromeres which span certain genomic regions on the chromosomes are the most common type of centromere across species. Of note, regional centromeres can also be separated into short and long regional centromeres using the size of 40kb as a cutoff (Mandal, 2017). Third, holocentric centromeres are centromeres spreading throughout the entire chromosome; they have been discovered in the plant *Luzula nivea*, silk moth *Bombyx mori* and nematode *C. elegans* (Barlow & Nevin, 1976, Consortium, 1998, International Silkworm Genome, 2008, Kawamoto, Jouraku et al., 2019). A recent study shows that indeed holocentric centromeres consist of multiple budding yeast-like point centromeres simultaneously (Steiner & Henikoff, 2014). Lastly, meta-polycentric centromeres are centromeres that appear within a genomic region but alternate at multiple positions. This newly characterized type of centromeres is so far only described in the pea *Pisum sativum* (Neumann, Navratilova et al., 2012) **(Figure 1)**.

At the DNA sequence level, centromeres show a variety of features in different species. For point centromeres, taking budding yeast *S. cerevisiae* as an example, the centromeres consist of ~125 bp AT-rich sequences (Furuyama & Biggins, 2007, Henikoff & Henikoff, 2012, Meluh, Yang et al., 1998). They contain mainly three centromere DNA elements (CDEs): CDEI (~8 bp), CDEII (~78-86 bp) and CDEIII (~25 bp) (Brogaard, Xi et al., 2012, Hegemann & Fleig, 1993, Henikoff, Ramachandran et al., 2014, Krassovsky, Henikoff et al., 2012, Tatchell & Van Holde, 1979). Importantly, Cse4, the centromeric H3 (cenH3) variant in budding yeast, occupies specifically the CDEII region suggesting that CDEII is the core centromere (Meluh et al., 1998). In fission yeast *S. pombe*, the representative of short regional centromeres (~35-110 kb),

the centromere is composed of a central sequence (cnt) flanked by two inverted repeat sequences (ImrL and ImrR) (Nakaseko, Adachi et al., 1986, WoodGwilliam et al., 2002). The long regional centromeres, for instance, the centromeres of *Drosophila melanogaster,* contain satellite DNA spanning 420kb. Within this region, it has been discovered that there are transposable elements (TEs) in between (Sun, Wahlstrom et al., 1997). Like *D. melanogaster*, plants such as *Arabidopsis thaliana*, *Oryza sativa* and *Zea mays* also contain centromeres with repetitive sequences but interrupted by interspersed TEs (Ananiev, Phillips et al., 1998, Dong, Miller et al., 1998, Round, Flowers et al., 1997). In mice, the centromeric region consists mainly of two parts: major and minor satellites. The major satellite which forms pericentromeric heterochromatin is made of repeats of 234 bp units. The core centromeric region is organized by repeats of 120 bp units next to major satellite sequences (Biscotti et al., 2015). In human, the core centromeres called α-satellite regions are repeats composed of 171 bp monomers. Various numbers of these 171 bp monomers can organize as blocks named higher order repeat (HOR). These HORs can span 2-5 Mb on chromosomes (Aldrup-Macdonald & Sullivan, 2014, Lo, Craig et al., 2001). Interestingly, the numbers of HORs are different between chromosomes and of note also across individuals within species **(Figure 1)**.

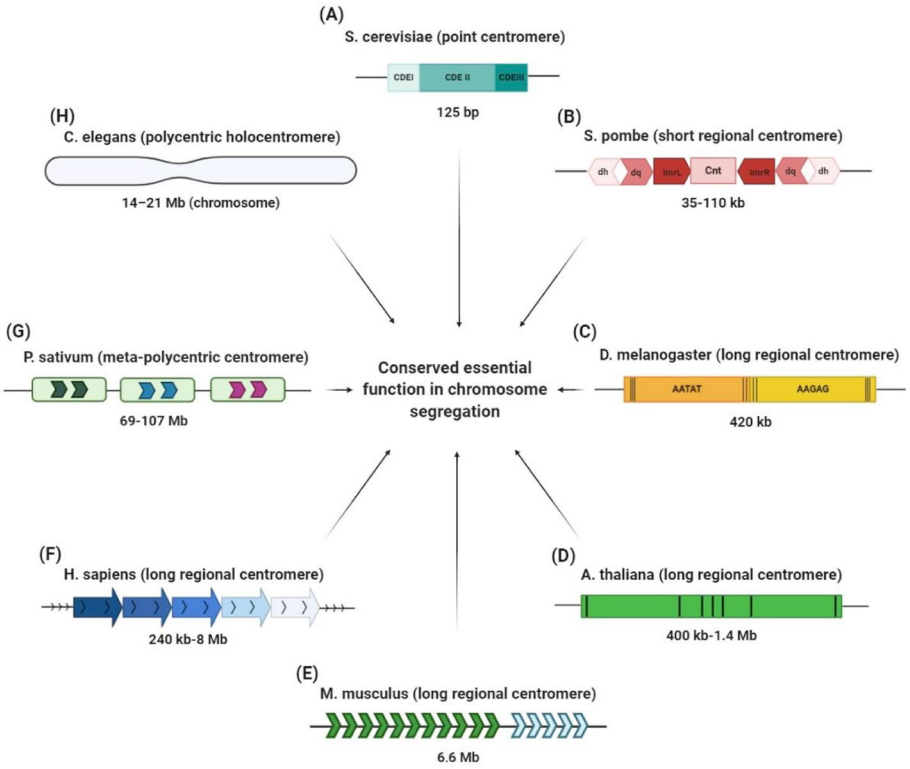


Figure 1. Schemes of various structures of centromeric DNA sequences (Balzano & Giunta, 2020)

Transcription of centromeres

Centromeric repeats have been characterized as transcriptionally active regions. The transcription is operated bidirectionally by RNA polymerase II (PolII) (Corless et al., 2020, Hall et al., 2012, Smurova & De Wulf, 2018). While the transcriptional regulation of centromeric repeats is not fully understood, it has been shown that CENPB and transcription factors are important for driving cenRNA transcription. In yeast, the transcription factor Centromere-binding protein 1 (CBF1) and its upstream regulators Pheromone alpha factor receptor (STE2) and Down-regulator of invasive growth 1 (DIG1) can induce centromeric transcription. Loss of function in either of these factors disrupts the function of the centromere (Ohkuni & Kitagawa, 2012). In *Drosophila*, cenRNA transcription is driven by the CENPA chaperone Chromosome alignment defect 1 (CAL1), which recruits the chromatin remodeler factor FACT (facilitates chromatin transcription) to promote PolII transcription (Chen, Bowers et al., 2015). In mice, CENPB also has been proposed to have the ability to facilitate the occupancy of zinc-finger transcriptional regulator ZFAT and recruit PolII to centromeric regions (Ishikura, Yoshida et al., 2021). For human centromeres, TATA box motifs and SV40 enhancer-core sequence have been characterized and they have been shown to facilitate PolII transcription (Vissel, Nagy et al., 1992). Moreover, the FCP1 phosphatase also acts as a stimulator to promote the elongation of PolII transcription within centromeric regions (Mandal, Cho et al., 2002). Due to the diverse centromeric DNA sequences across species, although there is no conserved transcription factor involved, the machinery of centromeric transcription is critical for down-stream centromere formation remains conserved. Therefore, further studies of centromeric transcription remain indispensable.

The function of centromeric RNAs

An increasing amount of evidence suggests that, as transcription happens, cenRNAs are involved in centromere formation. For example, cenRNAs can form R-loop structure to activate ATR pathway (Kabeche, Nguyen et al., 2018). The most recognized function of MinSat RNA in the field is that cenRNAs may guide centromeric proteins such as CENPA, CENPC and components of the CPC complex. Based on multiple studies, cenRNAs share conserved functions in different species (Ideue & Tani, 2020). For instance, in budding yeast, a certain level of cenRNA must be maintained, since up-regulation of cenRNA decreases the chromatin binding of centromeric proteins. Interestingly, knockdown of cenRNA can cause the loss of minichromosomes suggesting the import role of cenRNA in chromosome replication and maintenance

(Ling & Yuen, 2019). In *Drosophila*, RNA transcribed from SATIII repeats is an essential element for CENPA and CENPC incorporation into the centromeric region. Knockdown of SATIII RNA induces severe chromosome missegregation defects (Bobkov, Gilbert et al., 2018, Rosic et al., 2014). In *Xenopus*, the kinase Aurora B is proposed to localize to the centromere using cenRNA as a localization signal followed by mitotic centromere-associated kinesin (MCAK) phosphorylation (Blower, 2016, Jambhekar, Emerman et al., 2014). In *Xenopus*, it has been shown that cenRNA is processed by the RNA splicing machinery to promote kinetochore assembly. Treatment with the spliceosome inhibitor Isogingketin blocked cenRNA processing and was sufficient to abolish mitotic spindle formation (Grenfell, Heald et al., 2016). For mouse minor satellite (MinSat) RNA, pull-down assays suggest that MinSat RNA interacts with Aurora B, Survivin, and Inner centromere protein (INCENP) (Ferri, Bouzinba-Segard et al., 2009). These proteins are known to be the key players in centromere formation, they serve as check point proteins to ensure correct chromosome segregation during Metaphase (Vader, Cruijsen et al., 2007). Knockdown of MinSat RNAs can cause the disruption of nucleus structure suggesting the role of MinSat RNAs in prevention of DNA damage (Ideue, Cho et al., 2014). In humans, α-satellite RNA has been suggested to be involved in safeguarding the fidelity of chromosome division. Using immunofluorescence, it has been shown that knockdown of α-satellite RNA prevents localization of CENPA to centromeres. Moreover, Actinomycin D and RNase A treatments both cause mislocalization of CENPC, suggesting an essential role of centromeric transcription and cenRNAs in human centromere establishment (Wong, Brettingham-Moore et al., 2007). Thus, cenRNAs are structural components of the centromere, with known interactions with centromeric proteins and further contributes to centromere formation.

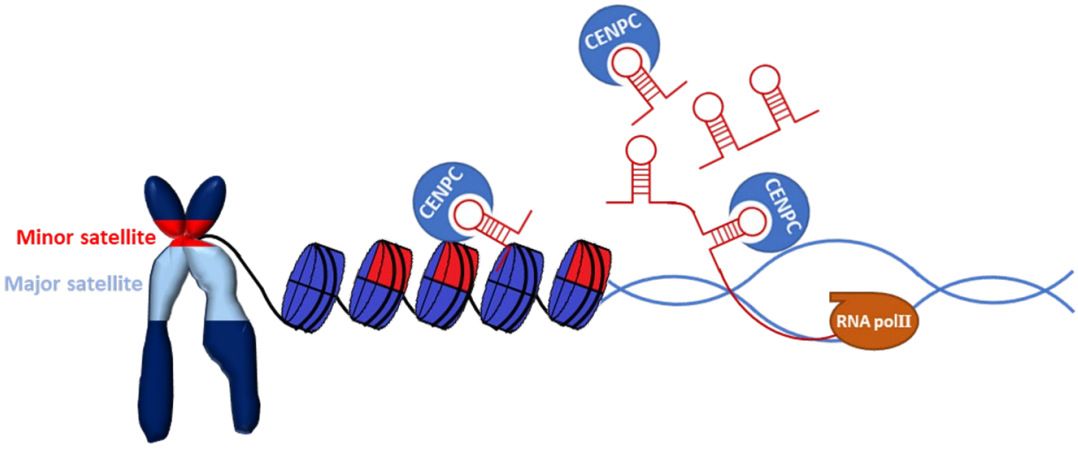


Figure 2. Potential model of centromeric transcription or transcript recruiting centromeric protein to chromatin.
In the minor satellite region, the centromeric chromatin is composed of a mixture of histone H3 (all blue) and CENPA (with partially red) nucleosomes. In mammals, cenRNAs are transcribed at this centromeric region by RNA polII during G2/M phase (Bouzinba-Segard, Guais et al., 2006, Bury, Moodie et al., 2020). The cenRNAs can interact with centromeric proteins such as CENPC (light blue) to localize to CENPA nucleosomes (Red) and bind directly to CENPA protein. This process can be considered the first step for kinetochore formation.

Function and evolutionary conservation of non-coding RNAs (ncRNAs)

CenRNAs are considered as non-coding RNAs within transcriptome. Depending on the centromeric DNA sequences in different species, some of them can be transcribed as long non-coding RNAs (lncRNAs) whereas some of them are shorter than 200 nts which are defined as small RNAs or non-coding RNAs (ncRNAs) (Arunkumar & Melters, 2020). However, as cenRNAs are initially transcribed into RNA containing multiple repeats and subsequently processed into smaller fragments, they exist within cells in a mixture of sizes. (Bouzinba-Segard et al., 2006). It is challenging to clearly define the categories of cenRNAs based on size, I hereby would like to introduce the evolution of cenRNAs using the idea applied to lncRNAs due to their incompetent translational potential.

Among the entire transcriptome, lncRNAs or ncRNAs are more abundant than mRNAs of protein-coding genes (Mattick, Amaral et al., 2023). The current definition of lncRNAs is that they are transcripts longer than 200 nt and have low protein-coding potential such as a lack of canonical open reading frames (ORFs) or sequence components for translational signaling (Mattick et al., 2023, Ulitsky & Bartel, 2013). Due to the technical limitations and incomplete annotation of each gene, the molecular features such as function, expression and splicing of lncRNAs are still poorly studied. Hence, an important way to study the biology of lncRNAs is through the conservation between species (Diederichs, 2014, Johnsson, Lipovich et al., 2014, Szczesniak, Kubiak et al., 2021). The evolutionary conservation of lncRNAs is mainly investigated from four aspects (Diederichs, 2014, Szczesniak et al., 2021): First, the most conventional angle to study RNA conservation is according to the similarity of primary sequences (Diederichs, 2014, Szczesniak et al., 2021). In recent years, many sequence alignment tools have been developed to search for sequence similarity in RNAs

(Hezroni, Koppstein et al., 2015, Noviello, Di Liddo et al., 2018, Pegueroles, Iraola-Guzman et al., 2019). Also, several lncRNAs have been reported to show sequence conservation in different species, but in most of the cases, only small stretches of conserved sequences were found. Compared to protein-coding genes, lncRNAs show weaker sequence conservation in general (Guttman, Amit et al., 2009, Ponjavic, Ponting et al., 2007). The second aspect is based on the conservation of gene position. Syntenic lncRNAs are discovered in the same relative genomic region or syntenic coding genes between compared species. This may also indicate a likelihood of lncRNA functions. For example, some lncRNAs can act *in cis* to neighboring genomic loci or, through reverse complementarity to protein-coding genes, influence the expression of their targets (Katayama, Tomaru et al., 2005, Lehner, Williams et al., 2002). Third, RNA splicing signals may also provide a hint of lncRNA conservation. Part of lncRNAs results from non-coding exons after alternative splicing. By analyzing splice sites throughout the genome, scientists developed an algorithm to map the conservation of lncRNA transcripts. In a previous study, 87% of alternative splice sites in human have been found to exist in other vertebrates such as mice, rats, cows and dogs (Yeo & Burge, 2004). Fourth, the aspects above are not sufficient to explain lncRNA function on the molecular level. The action of a majority of lncRNAs greatly relies on the interaction with their protein partners (Huang, Li et al., 2021, Sweta, Dudnakova et al., 2019). To achieve this, lncRNAs form certain secondary or tertiary structures for the recognition of specific proteins (Sanchez de Groot, Armaos et al., 2019). For example, lncRNA XIST can form complicated secondary structures based on tandem repeat sequences, especially the well-characterized A-repeat region. This special secondary structure enables RNA to interact with its protein partner such as the silencing factor SPEN (Lu, Zhang et al., 2016). Throughout evolution, lncRNAs and their interacting proteins can co-evolve and undergo a purifying selection process (He, Valkov et al., 2023, Weinreb, Riesselman et al., 2016). Thus, investigating the conservation via the structure of lncRNAs may give a better understanding of the exact modes of action of lncRNA molecules. For example, a recent study indicates that lncRNA MALAT1 in human reveals only 28%, 25% and 25% sequence identity across primates, mammals, and vertebrates, respectively. However, based on secondary structure analysis, 153 out of 194 helices within human MALAT1 appear beyond 90% structural similarity to 43 mammalian MALAT1 (McCown, Wang et al., 2019). This suggests that along with lncRNA evolution, structure conservation reflects more to a functional consensus than what might appear from RNA sequence alone.

Scope and aim of the thesis

Self-renewal is a crucial biological process to maintain the number of stem cells which strongly depends on cell division (Balboa, Iworima et al., 2021, Bogliotti, Wu et al., 2018, Hanna, Saha et al., 2010). To study cell division mechanism during self-renewal process, mESCs were used in this project. Furthermore, mESCs have been shown to have high tolerance to aneuploidy (Mantel, Guo et al., 2007). This brings me the advantage of capturing abnormal cell division events without severe cell death. Intending to understand chromosome segregation in mESCs, I set to investigate the aspect of centromere formation and function.

CenRNAs were discovered several decades ago (Rieder, 1979). Since then, researchers have described their potential functions and biogenesis in multiple systems across species (Ideue & Tani, 2020). Previously considered as a by-product of centromeric transcription, the role of cenRNAs has only emerged and been characterized during recent years. However, due to the complicated nature of cenRNAs such as repetitive sequence, low expression level and transcription from both DNA strands, it remains challenging to unravel the detailed molecular roles and features inside the cells (Biscotti et al., 2015). Particularly, the function and the regulatory roles of mouse minor satellite RNA are poorly understood.

The objectives of my Ph.D were to characterize the impact of MinSat RNAs expression on mESC cell division and the determinants that contribute to MinSat RNA function. In order to determine the crosstalk between forward and reverse MinSat transcripts, I first developed a strand-specific quantitative PCR (qPCR) approach to characterize the forward and reverse transcripts separately. Next, to examine the effect of both MinSat transcripts, I performed overexpression and knock-down experiments in mouse embryonic stem cells (mESCs). I observed that both forward and reverse transcripts can play a role in chromosome segregation fidelity. To further look for the decisive protein factors that cause chromosome missegregation, I measured the binding of centromeric proteins to chromatin. I then narrowed down to CENPC as a potential key player mediating the phenotypes. Furthermore, to see if CENPC can directly interact with MinSat RNA, I showed that MinSat forward transcripts can pull-down CENPC from mESC lysate which is in line with other publications tested in other species. Next, to explore how MinSat RNA interacts with its protein partners, I decided to address this from the angle of the RNA secondary structure. I found that MinSat RNA obtains a stable stem-loop structure which may contribute to the RNA function. I also identified similar secondary structure on human cenRNA indicating an evolutionarily conserved stem-loop structure that appears on both human and mouse cenRNAs. Of note, I could demonstrate that this stem-loop contributes to the function of cenRNA

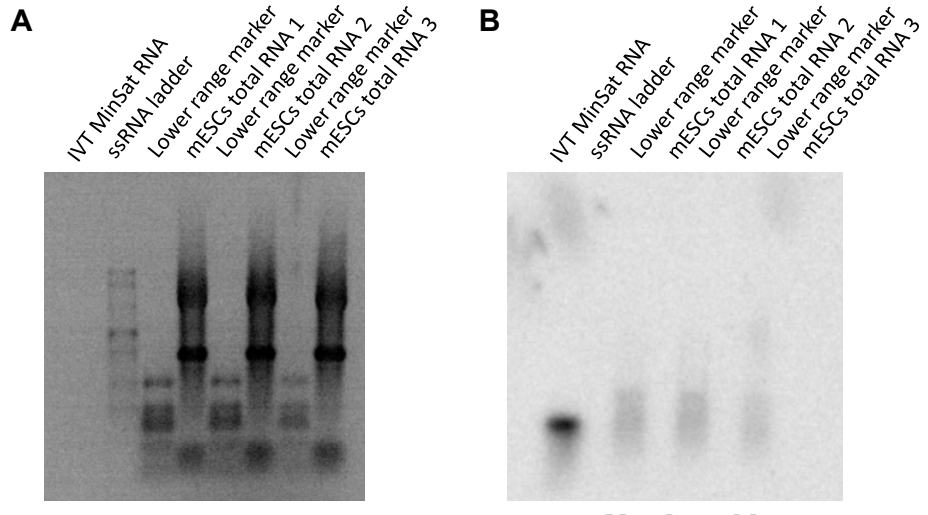
across species.

Centromeric transcription and RNA have been shown to be a conserved mechanism responsible for recruiting centromeric proteins to the centromere region across species (Ideue & Tani, 2020). Nevertheless, the absence of consensus and similarity in centromeric sequences contradicts the interaction concerning conserved centromeric proteins. By examining the function and structural similarity of human and mouse cenRNA molecules, I found a structural motif may contribute to the function of cenRNA. This work will provide a potential explanation to the field of why centromeric proteins can recognize cenRNAs based on the secondary structure formed instead of sequences.

Results

Mouse minor satellite RNAs are expressed from both strands in mouse embryonic stem cells

From previous literature, minor satellite regions have been shown to generate centromeric RNAs (cenRNAs) from both strands (Blower, 2016, Bury et al., 2020, Rosic et al., 2014). We therefore asked whether forward, reverse or both minor satellite (MinSat) transcripts have functional roles in mitosis. To first investigate the endogenous expression profile of MinSat RNAs in mESCs, I performed Northern blot analysis to detect MinSat RNA in total RNA from mESCs. I processed RNA for northern blot, firstly by performing gel electrophoresis under denatured conditions. I verified RNA integrity using ethidium bromide staining **(Figure 3A)**. The clear bands of ribosomal RNAs (18S, 5.8S, and 28S) are the indicators of RNA quality. Then, after transferring the RNAs to the membrane, I hybridized the membrane with probes against MinSat sequences generated from DNA template converted from MinSat RNA. As a result, the radioactive signal was not detected in 20 µg total RNA input samples whereas the *in vitro* transcribed MinSat RNA positive control was detected suggesting the Northern blot protocol is working properly **(Figure 3B)**. This result can be due to the insufficient amount of total RNA loaded. This also reflects the fact that the expression of MinSat RNA is low. As an alternative, a more sensitive detection method was needed, I turned to real-time quantitative PCR assay (RT-qPCR). For this, I designed primers for measuring either total (forward+ reverse) transcripts or strand-specific MinSat transcripts **(Appendix Table 1)**. I performed non-strand-specific **(Figure 3C)** and strand-specific **(Figure 3D)** qPCR on wild-type mESCs. Based on raw threshold cycle (Ct)-values, the expression of MinSat RNAs is from both strands of DNA templates and the expression level is consistent throughout biological replicates.



EtBr staining **Northern blot**

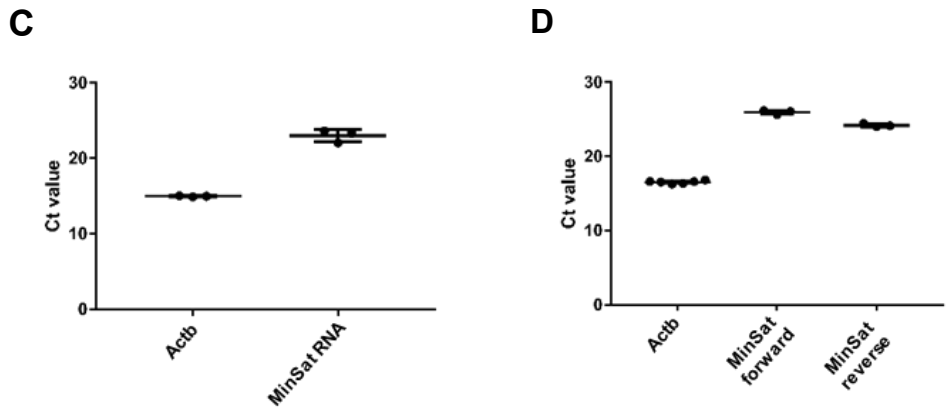


Figure 3. Endogenous MinSat RNA expression profile in mESCs.

(A) Image of an agarose gel after EtBr staining. The RNA samples from three biological replicates are labeled on top.

(B) Northern blot image after hybridization with P^{32} radioactive probes against MinSat.

(C) Raw Ct values of PCR amplification for beta-actin and MinSat after reverse transcription with oligo dT and random hexamers. Solid line, mean of N=3 independent experiments; error bars, standard deviation.

(D) Raw Ct values of PCR amplification after reverse transcription with primers specific for beta-actin, MinSat forward, and MinSat reverse RNA, respectively. Solid line, mean of N=3 (MinSat) or N=6 (beta-actin) experiments; error bars, standard deviation.

The endogenous MinSat RNA level reveals a cell cycle-related pattern

To test if MinSat RNA expression levels fluctuate in different cell cycle stages, I utilized mESCs harboring the Fucci (Fluorescence Ubiquitin Cell Cycle Indicator) system (Nakatani, Lin et al., 2022, Sakaue-Sawano, Kurokawa et al., 2008). The Fucci system allows to sort cells in G1, S and G2/M phases, respectively, by flow cytometry **(Figure 4A)** and determine expression levels in the respective cell cycle stages by RT-qPCR. I observed that in total, MinSat RNA increases at G2/M phase and decreases during S phase **(Figure 4B)**. Not surprisingly, both forward and reverse transcripts follow a similar pattern as total **(Figure 4C)**. These results are in line with previous publications indicating cenRNAs are transcribed during mitosis. However, in other cell types, the expression level is reduced at G1 phase. One possible explanation for this difference is the very short G1 phase of mESCs (Coronado, Godet et al., 2013) which may lead to the peak of RNA levels last until S-phase.

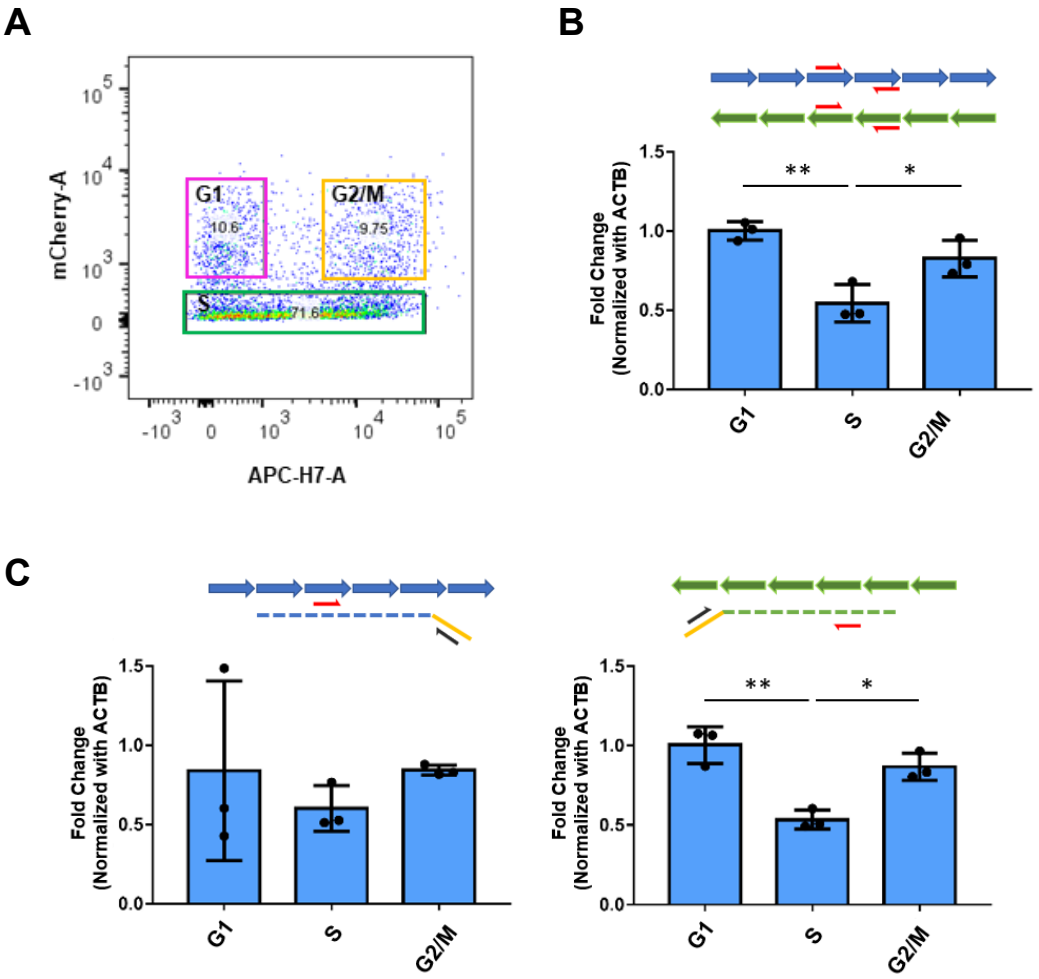


Figure 4. Expression level of MinSat RNA along with cell cycle stages of mESCs

(A) Representative gating strategy for Fluorescence-Activated Cell Sorting (FACS) of ESCs harboring the Fucci reporter. (red, G1 phase; green, S phase; yellow, G2/M phase).

(B-C) RT-qPCR analysis of total (F), forward or reverse (G) MinSat transcripts in G1-, S-, and G2/M-phase populations of ESCs harboring the Fucci cell cycle reporter system. The schematics on top indicate the position of the PCR primers (Red) used and the strategy for the analysis of strand-specific transcripts. The yellow bars represent adaptor sequences incorporated into cDNA and the black arrows represent primer against adaptor sequence. Bars, mean values of N=3 independent experiments (individual dots), normalized to beta-actin mRNA and control sample; error bars, standard deviation; p-values * ≤ 0.05, ** ≤ 0.01

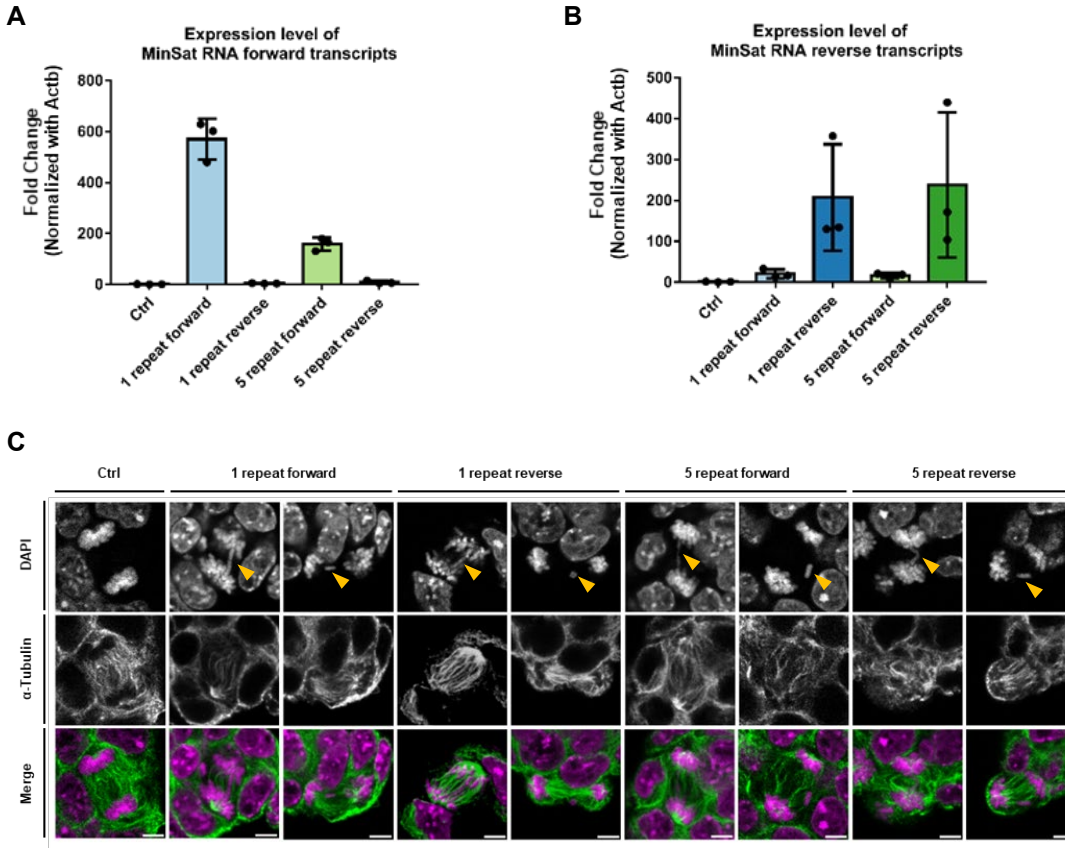
Overexpression of MinSat RNA leads to chromosome missegregation and reduction in cell proliferation

Previous studies have shown that cenRNA levels must be maintained at a certain level and both up- and down-regulation of cenRNAs can contribute to aberrant chromosome segregation (Chan, Moralli et al., 2017, Ling & Yuen, 2019). To understand if the expression level of MinSat RNA influences chromosome segregation in mESCs, I constructed mammalian expression vectors using the pCAG plasmid, which contains the pCAG promoter **(Appendix Figure 1)**, driving the expression of 1 or 5 tandem repeats of MinSat sequences to generate either forward or reverse transcripts. After transfecting each of these plasmids into mESCs, I applied the strand-specific RT-qPCR strategy as mentioned above to assess the overexpression level and the strand-specific transcripts derived from the expression vectors. As expected, the expression levels of either strand of MinSat transcripts only increase when the corresponding strand-specific primers are used **(Figure 5A, B)**.

Next, I quantified chromosome segregation effects. For this, I performed immunofluorescence against α-Tubulin to help distinguish stages of cell division based on microtubule distribution and DAPI staining for DNA to see chromosome morphology. By this approach, I was able to distinguish anaphase and telophase mitotic figures and count chromosome missegregation events such as lagging chromosomes and inter-chromosomal bridges. Interestingly, overexpression of MinSat RNAs shows 10-20% more chromosome missegregation events relative to the control group, which I transfected with an empty vector alone. Of note, within all the experimental groups, forward transcripts induced more missegregation events than reverse transcripts (26% ± 14% for 1 repeat forward and 13% ± 3% for 1 repeat reverse transcript) **(Figure 5C, D)**. Chromosome instability often comes with a delay in cell cycle or even cell death (Bakhoum & Landau, 2017, Ippolito, Martis et al., 2021, Lukow, Sausville et al., 2021). To address whether the chromosome segregation defects elicited by MinSat overexpression lead to defective cell proliferation, I also conducted a cell proliferation assay. The results show that cell numbers increased more slowly only when MinSat reverse transcripts were expressed from the pCAG promoter **(Figure 5E)**. However, I did not observe a change in cell cycle distribution in the overall mESCs population **(Figure 5F)**. To test if the observed phenotype is also present in different cell types, I expressed 1 repeat of forward transcript in mouse embryonic fibroblasts (MEFs) **(Figure 5G, H)**. Indeed, I observed that expression of 1 forward MinSat repeat transcript can also induce chromosome missegregation in MEFs, suggesting that this

mitotic phenotype is not cell type-specific.

Above all, we confirmed that strand-specific MinSat RNA overexpression can be achieved using pCAG mammalian expression vectors. By using this approach, we can specifically increase MinSat RNA level on either strand of the. Based on the result, I observed that overexpression of MinSat RNA induces chromosome missegregation. Interestingly, overexpression of reverse transcripts has a more severe chromosome segregation defect and reduced cell numbers. In most of the cell types, it is known that DNA damage or chromosome missegregation can lead to delay of cell cycle or even cell cycle arrest, especially during M phase. This is due to the activation of DNA repair pathway and cell cycle check point such as serine/threonine kinase complex CDK1/Cyclin B to safeguard genome stability (Yam, Lim et al., 2022). Based on this, I performed propidium iodide (PI) staining followed by flow cytometry cell cycle analysis. Regardless of all the defects observed above after MinSat RNA overexpression, none of them led to a change in cell cycle distribution in mESCs population. For overexpression of MinSat forward RNA, it can be that the mESCs managed to tolerate the missegregation and move on to the next cell cycle. However, overexpression of reverse transcripts causes more severe damage to the chromosomes, which further triggers cell death leading to reduced cell numbers in the following time points.



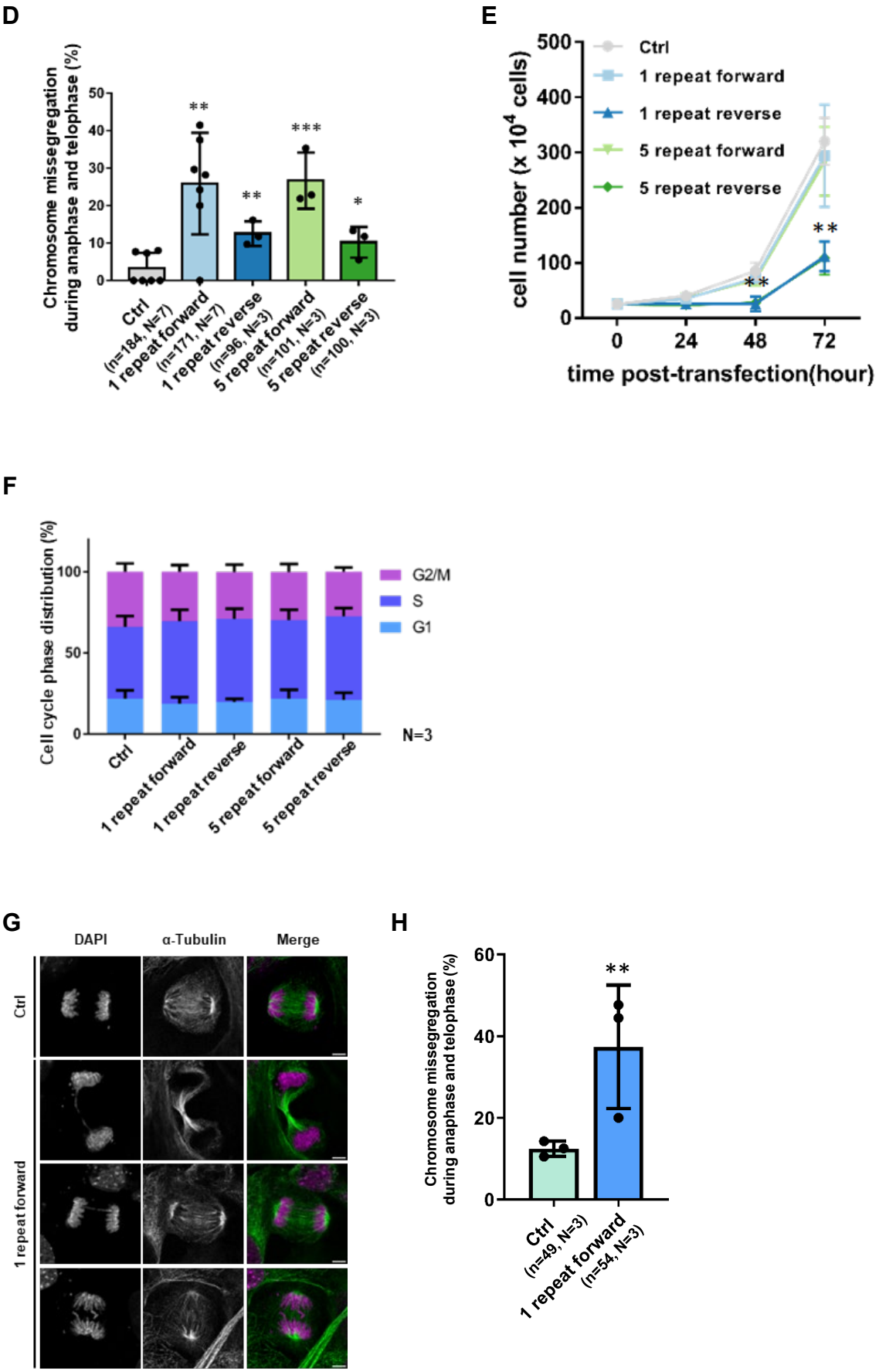

D
Chromosome missegregation during anaphase and telophase (%)
50
40
30
20
10
0
**
**
***
*
Ctrl (n=184, N=7)
1 repeat forward (n=171, N=7)
1 repeat reverse (n=96, N=3)
5 repeat forward (n=101, N=3)
5 repeat reverse (n=100, N=3)
E
cell number (x 10^4 cells)
500
400
300
200
100
0
Ctrl
1 repeat forward
1 repeat reverse
5 repeat forward
5 repeat reverse
**
**
0
24
48
72
time post-transfection(hour)
F
Cell cycle phase distribution (%)
100
50
0
G2/M
S
G1
N=3
Ctrl
1 repeat forward
1 repeat reverse
5 repeat forward
5 repeat reverse
G
DAPI
α-Tubulin
Merge
Ctrl
1 repeat forward
H
Chromosome missegregation during anaphase and telophase (%)
60
40
20
0
**
Ctrl (n=49, N=3)
1 repeat forward (n=54, N=3)

Figure 5. Overexpression of MinSat RNA transcripts causes defects in chromosome segregation.

(A) Strand-specific qPCR analysis after MinSat RNA overexpression. The total RNA was reverse transcribed using primer against MinSat forward transcripts. Bars, mean value； error bars, standard deviation.

(B) Strand-specific qPCR analysis after MinSat RNA overexpression. The total RNA was reverse transcribed using primer against MinSat reverse transcripts. Bars, mean value； error bars, standard deviation.

(C) Representative images of immunofluorescence on mESCs overexpressed MinSat RNAs. Scale bar, 5 μm. The orange arrows point to chromosome missegregation features.

(D) Quantification of chromosome missegregation events in mitotic mESCs. Bars, mean; error bars, standard deviation. Paired-wise student t-test with empty vector control. p-values * ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001. The number of biological replicates (N) and sampled mitotic cells (n) are indicated.

(E) Growth curve of mESCs after MinSat RNA overexpression. The curves show the average cell numbers of three biological replicates. Error bars, standard deviation; p-value ** ≤ 0.01

(F) Flow cytometry analysis of cell-cycle stage distribution using propidium iodide (PI). Bars indicate mean percentages of the respective cell cycle phases from three independent experiments; Error bars, standard deviation.

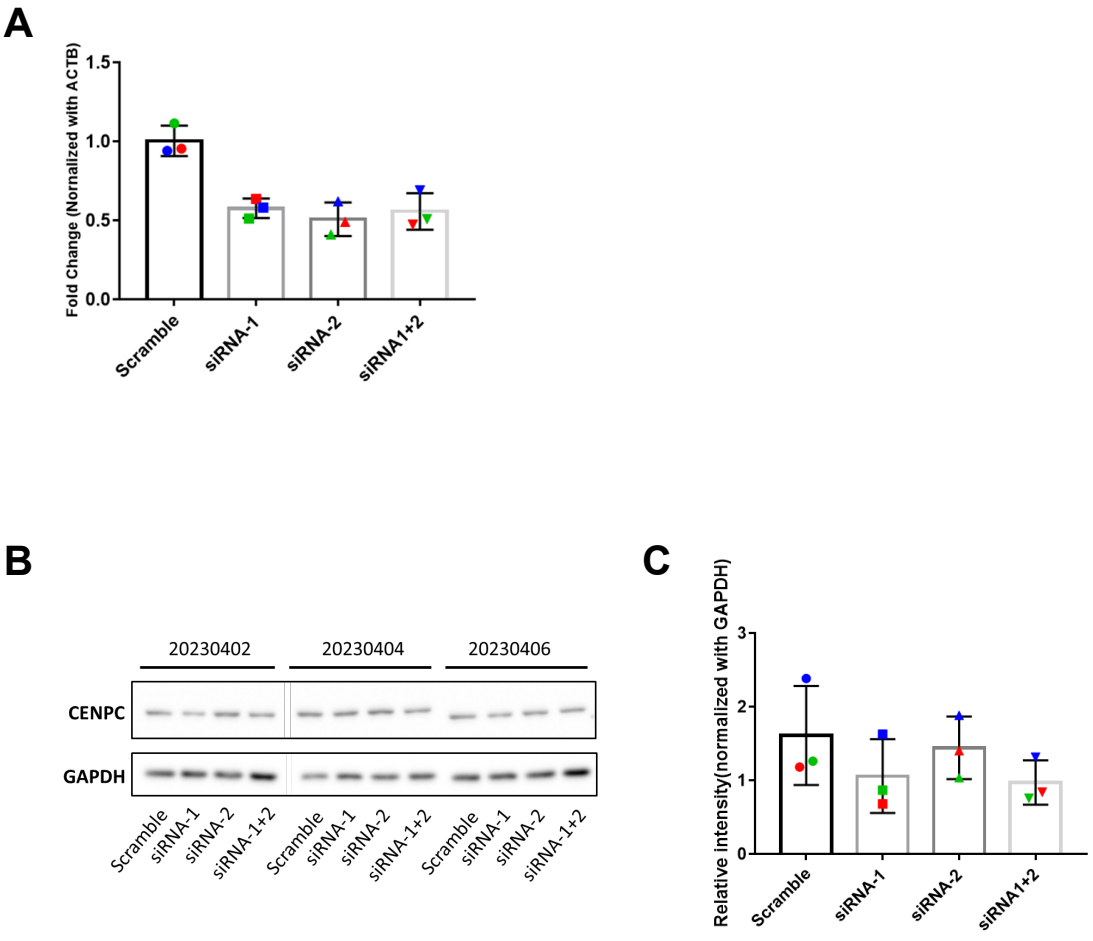
(G) Representative images of immunofluorescent staining of MEFs. Scale bar, 5 μm.

(H) Quantification of chromosome missegregation events in MEFs. Paired-wise student t-test with empty vector control. p-values ** ≤ 0.01. The number of biological replicates (N) and sampled mitotic cells (n) are indicated.

Specificity validation of CENPC antibody ab193666

Most of the CENPC antibodies used in previous publications are either raised in-house by their lab or discontinued by the companies. The CENPC antibodies used in this project have not been utilized frequently in current studies of mouse CENPC protein. Thus, this raises our concern about the specificity of CENPC antibody. It requires further validation to ensure the material used here is reliable. To confirm the specificity of the CENPC antibody used throughout the study, two siRNAs **(Appendix Table 2)** against *Cenpc* mRNA were transfected into mESCs. After 48 hours post-transfection, I performed qPCR to measure endogenous *Cenpc* mRNA expression

(Figure 6A). Compared to scramble control, the two siRNAs achieved ~50% knock-down. After making sure that the siRNA can knock down *Cenpc* on the mRNA level, I performed Western blot to examine CENPC protein level using antibody ab193666 **(Figure 6B)**. The quantification result shows that mESCs transfected with siRNA-1 alone have a minor knock-down effect confirming that antibody ab193666 indeed recognizes mouse CENPC **(Figure 6C)**. To further confirm that the antibody binds to CENPC, I used recombinant CENPC purified from bacteria. I used a prokaryotic expression vector (pET-M11) containing the sequence encoding the CENPC protein, kindly provided by Alisha Jones **(Appendix Figure 2)**. pET-M11 is a bacterial expression plasmid containing lactose operon (lac operon) regulated T7 promoter. The expression can be only induced in the presence of Isopropyl β- d-1-thiogalactopyranoside (IPTG). Using empty vector and non-IPTG induced bacteria as control, the sample from bacteria with IPTG induction showed a strong signal on an immunoblot probed with the CenpC antibody **(Figure 6D)**. Overall, these results indicate that antibody ab193666 recognizes CENPC protein with high specificity. Therefore, this antibody was used for further experiments in this study.



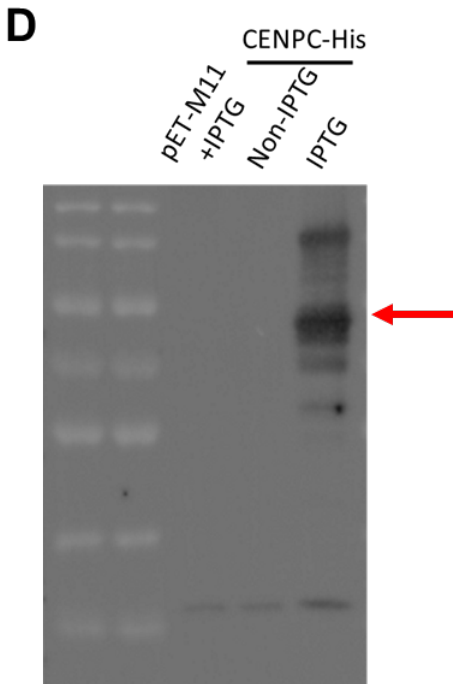


Figure 6. Specificity verification of CENPC antibody ab193666

(A) Measurement of Cenpc mRNA level in mESCs using RT-qPCR. Fold change is calculated by normalizing to beta-actin and the average of control experiments. Bar, mean.

(B) Images of immunoblot of biological replicates. siRNAs transfected were labeled at the bottom. The date of each biological replicates are labeled on the top of the images.

(C) Intensity quantification of immunoblots. The relative intensity (CENPC) is normalized to GAPDH (Dot). Bar, mean.

(D) Immunoblot of bacterial expressed CENPC protein.

Overexpression of reverse MinSat transcripts reduces total protein level of CENPA and disrupts the binding of CENPC to chromatin

To understand the molecular mechanism behind the phenotypes mentioned above, I asked whether overexpression of MinSat RNAs will influence the binding of the key centromeric proteins CENPC and CENPA to chromatin during mitosis. For this experiment, cells were treated with KaryoMAX™ Colcemid™ for 3 hours after 45 hours post-transfection to arrest cells at mitosis. Using acidic chromatin extraction, I then isolated the chromatin-bound protein fraction and performed Western blot with antibodies for CENPC, CENPA, histone H3 and GAPDH. In the chromatin-bound fraction, I observed a significant decrease of CENPC binding only when overexpressing

5 repeats of MinSat reverse transcripts **(Figure 7A, B)**, although 1 repeat reverse, and 5 repeat forward MinSat RNA also resulted in a slight decrease. For CENPA, despite the variability between experiments, I observed a general trend of downregulation after overexpression of reverse transcripts **(Figure 7B)**. To address whether the effect I saw results from an overall decrease in total protein level, I performed Western blot with total cell lysate as well **(Figure 7C, D)** and did not detect consistent differences in total CENPC levels across samples indicating that the overall stability and expression of CENPC protein remains consistent. However, total CENPA also displayed a trend towards lower protein levels after reverse MinSat transcripts overexpression **(Figure 7D)**. Based on this result, I conclude that the MinSat RNA reverse transcripts disrupts the chromatin binding ability of CENPC and this may further cause a down-regulation of CENPA protein stability.

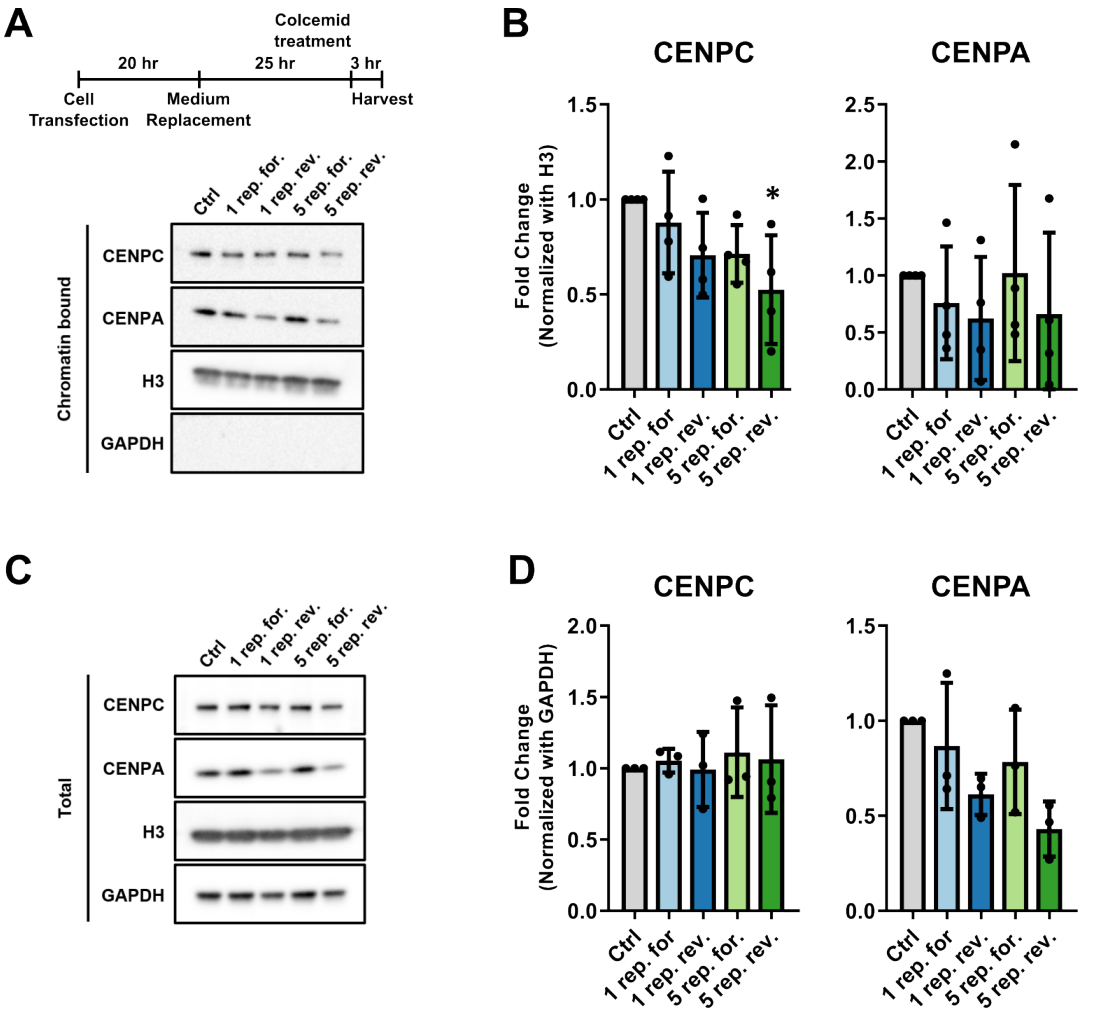


Figure 7. The overall or chromatin-bound CENPC and CENPA level of centromeric proteins after MinSat RNA overexpression.

(A) Timeline scheme of sample treatment and representative images of immunoblot of proteins in chromatin fraction after MinSat RNA overexpression in mESCs. The plasmids transfected are labeled on top.

(B) Intensity quantification of immunoblots. Bar, relative intensity to Histone H3 and normalized to pCAG empty vector control in each biological replicate (dots). p-values * ≤ 0.05.

(C) Representative immunoblot images of overall protein level in mESCs overexpressing MinSat RNAs. The plasmids transfected are labeled on top.

(D) Quantification of immunoblot of overall endogenous proteins. Bar, relative intensity to GAPDH and normalized to pCAG empty vector control in each biological replicate (dots).

Strand-specific-knockdown of MinSat RNAs with antisense oligos reveals a mutual feedback loop regulating MinSat RNA homeostasis

To investigate the function of endogenous MinSat RNAs, I conducted a loss of function experiment using 2'-O-methoxyethyl (2'MOE) antisense oligos (ASOs) with scramble sequence and sequence against MinSat forward or reverse transcripts **(Appendix Table 3)**. After cell transfection of ASOs, I first tested the knock-down efficiency by non-strand-specific **(Figure 8A)** or strand-specific **(Figure 8B, C)** RT-qPCR. In non-strand-specific qPCR which reverse transcribed with oligo dT and random hexamer primers, I only observed a strong knock-down after reverse ASO transfection and when I used a combination of forward and reverse ASOs. Next, to properly characterize the knock-down effect of specific ASOs, I performed strand-specific qPCR. When looking at forward transcripts alone, a 50-60% knock-down efficiency could be achieved **(Figure 8B)**. However, in reverse transcript-specific qPCR, the knock-down efficiency remained comparable to non-strand-specific analysis **(Figure 8C)**, suggesting that the signal in non-strand-specific qPCR is mainly masked or dominated by reverse transcripts. All of these observations point towards that reverse transcripts are more highly transcribed than forward transcripts. Interestingly, strand-specific knockdown of MinSat RNAs results in a slight induction of the respective opposite strand transcripts. Especially in the case of reverse ASO transfection, this led to a 1.32 ± 0.3 -fold increase in the levels of forward MinSat transcripts **(Figure 8B)**. This could imply a feedback loop between MinSat forward and reverse transcripts, potentially due to a reciprocal regulation between both transcripts.

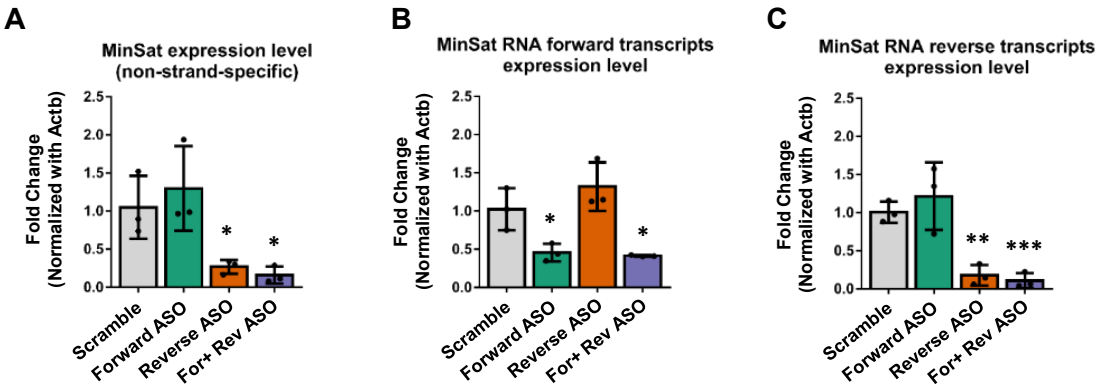


Figure 8. Strand-specific knock-down of MinSat RNAs using 2'MOE ASOs in mESCs.

(A) Non-strand-specific RT-qPCR analysis of endogenous MinSat RNA. Bars, mean values of N=3 independent experiments (dots), normalized to beta-actin mRNA and control samples; Error bars, standard deviation; p-values * ≤ 0.05.

(B) Strand-specific RT-qPCR of MinSat forward transcripts. Bars, mean values of N=3 independent experiments (dots), normalized to beta-actin mRNA and control samples; Error bars, standard deviation; p-values * ≤ 0.05.

(C) Strand-specific RT-qPCR of MinSat reverse transcripts. Bars, mean values of N=3 independent experiments (dots), normalized to beta-actin mRNA and control samples; Error bars, standard deviation; p-values ** ≤ 0.01, *** ≤ 0.001.

Knock-down of MinSat RNAs results in chromosome missegregation and genome instability

After confirming the knock-down efficiency of MinSat RNAs in mESCs, I performed DNA and immunofluorescent staining to assess whether the depletion resulted in chromosome instability or related phenotypes during cell division. Analogously to the previous overexpression experiment, I used DAPI to identify chromosome morphology and microtubule staining to distinguish specific mitotic stages and cell boundaries **(Figure 9A, C)**. I observed an increase in chromosome segregation defects of 23% ± 12% for the forward ASO and 31% ± 16% for the reverse ASO compared to 3% ± 4 in control ASO-treated ESCs (**Figure 9A, B**). At the same time, I noticed an increase in micronuclei formation, which is one of the hallmarks of chromosome instability **(Figure 9C, D)**. In interphase cells, micronuclei appeared in 10% ± 3% for the forward ASO and in 16% ± 2% for reverse ASO compared to 1% ± 1% in control ASO-transfected mESCs, **Figure 9D**. Taken together, these findings substantiate the notion that MinSat transcripts are critical factors for chromosome segregation fidelity and genome stability.

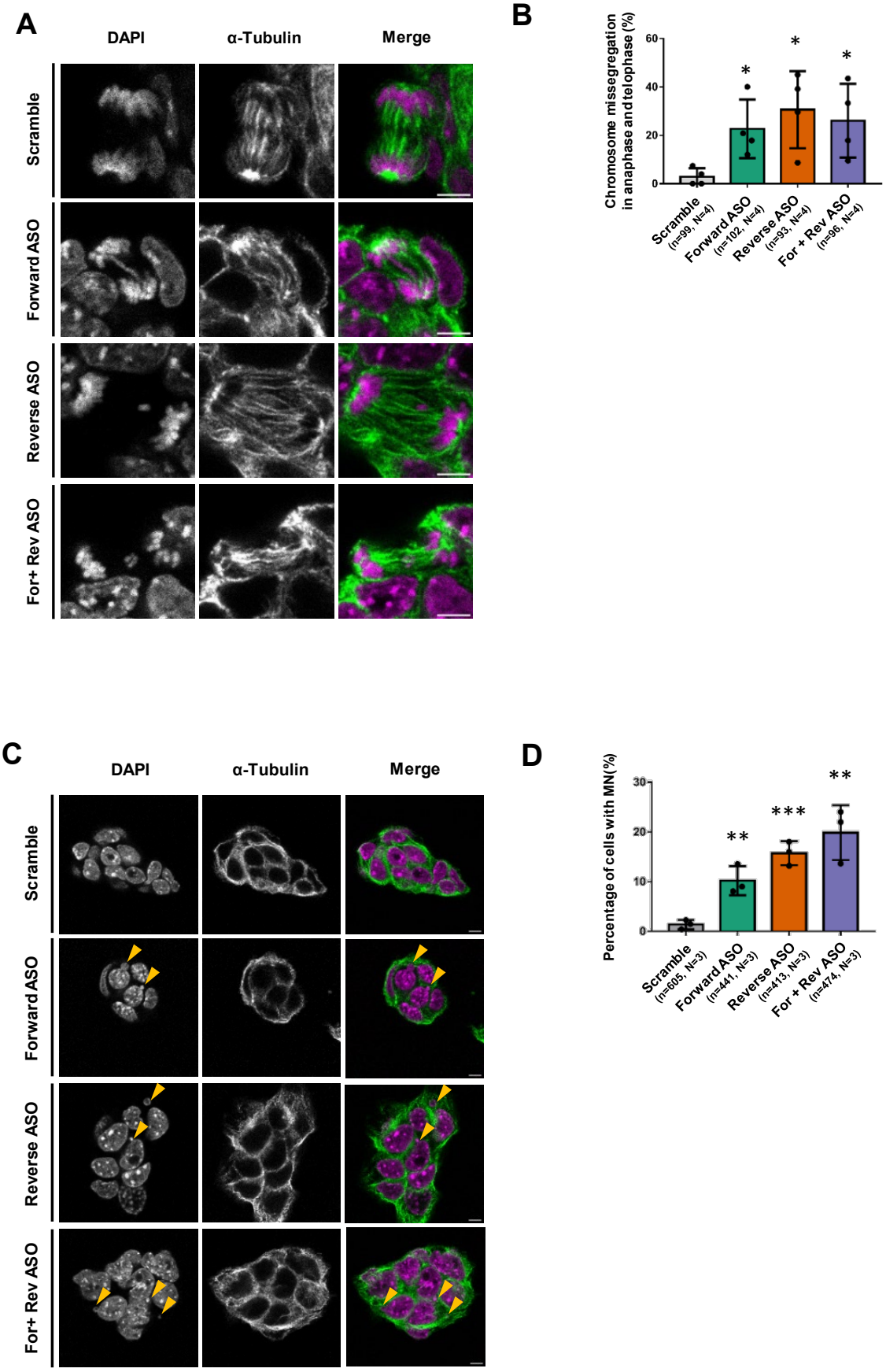

A
DAPI
α-Tubulin
Merge
Scramble
Forward ASO
Reverse ASO
For+ Rev ASO
B
Chromosome missegregation in anaphase and telophase (%)
60
40
20
0
*
*
*
Scramble (n=99, N=4)
Forward ASO (n=102, N=4)
Reverse ASO (n=93, N=4)
For + Rev ASO (n=96, N=4)
C
DAPI
α-Tubulin
Merge
Scramble
Forward ASO
Reverse ASO
For+ Rev ASO
D
Percentage of cells with MN(%)
30
20
10
0
**
***
**
Scramble (n=605, N=3)
Forward ASO (n=441, N=3)
Reverse ASO (n=413, N=3)
For + Rev ASO (n=474, N=3)

Figure 9. Knock-down of MinSat RNAs leads to chromosome instability.

(A) Representative confocal images of mESCs transfected with ASOs against MinSat RNAs. Scale bar, 5 μm.

(B) Quantification of chromosome missegregation events in mESCs during ana- and telo- phase. Bars, mean; error bars, standard deviation. Paired-wise student t-test with empty vector control. p-values * ≤ 0.05. The number of biological replicates (N) and sampled mitotic cells (n) are indicated.

(C) Representative images of mESCs transfected with ASOs against MinSat RNAs. Scale bar, 5 μm. The orange arrows point at the micronuclei structure in cells.

(D) Quantification result of interphase mESCs containing micronuclei. p-values ** ≤ 0.01, *** ≤ 0.001. The number of biological replicates (N) and sampled interphase cells (n) are indicated.

MinSat RNA knock-down interferes with the binding of CENPC and CENPA to chromatin during mitosis

To investigate if the binding of centromeric proteins to the centromere involves centromeric transcripts, I repeated the chromatin extraction experiments on mESCs transfected with 2'MOE ASOs. By Western blot analysis, I observed less binding of CENPC and CENPA to chromatin, especially when both transcripts are knocked down simultaneously, albeit the variability between biological replicates was relatively high. **(Figure 10A, B)**. This suggests a synergistic or complementary action between these two transcripts. To address if the overall protein levels of CENPC and CENPA are not drastically changed, I conducted immunoblot analysis on total mESC lysate after ASO transfection **(Figure 10C)**. Surprisingly, CENPC overall level showed a slight decrease when either or both ASOs were transfected, whereas CENPA protein levels remained unchanged **(Figure 10D)**. This indicates that the loss of MinSat RNA weakens the binding of CENPC and CENPA to chromatin. However, knock-down of either forward or reverse MinSat RNAs may already slightly lower CENPC protein stability and thus global protein levels.

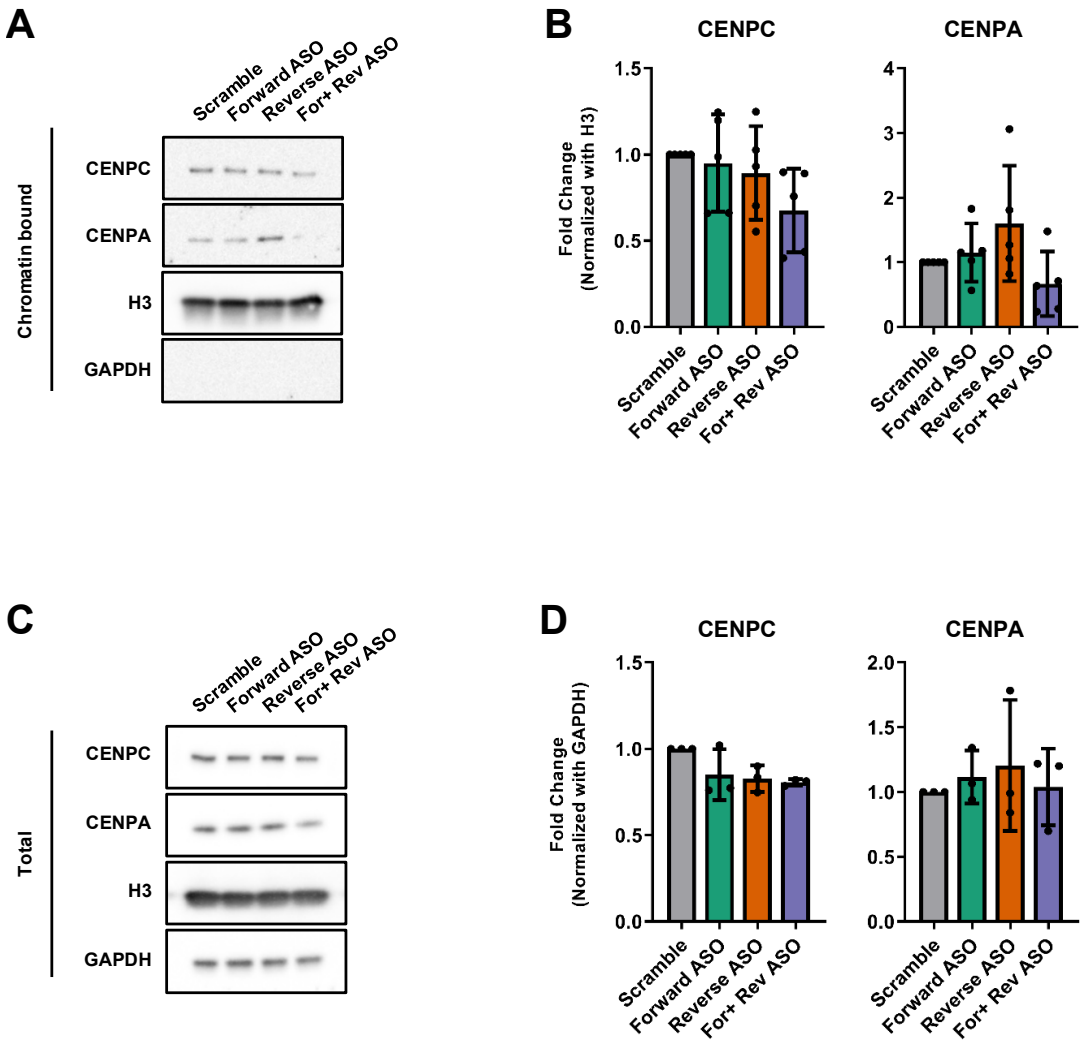


Figure 10. Western blot analysis of chromatin-bound or overall CENPC and CENPA levels in mESCs after strand-specific knock-down by 2'MOE ASOs.

(A) Representative immunoblot images of protein at chromatin-bound fraction after MinSat RNA knock-down in mESCs.

(B) Intensity quantification of immunoblot signals. The intensity of proteins of interest is normalized with Histone H3. Bar, relative intensity to Histone H3 and normalized to scramble ASOs in each biological replicate (dots).

(C) Representative immunoblots of total protein lysates after MinSat RNA knock-down in mESCs.

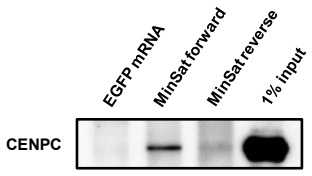
(D) Intensity quantification of immunoblot signals. The intensity of proteins of interest is normalized with GAPDH. Bar, relative intensity to GAPDH and normalized to scramble ASOs in each biological replicate (dots).

CENPC binds to MinSat forward transcripts via its RNA binding region

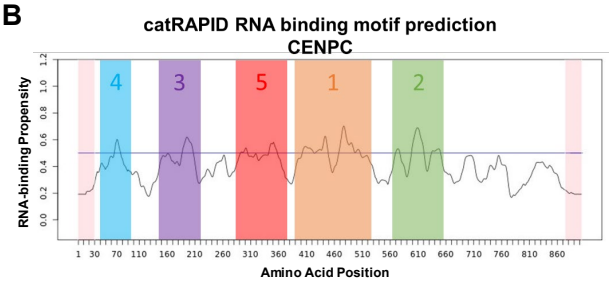
The results from MinSat RNA overexpression as well as its knock-down both indicate that MinSat RNAs potentially influence the binding of CENPC to chromatin. Previous studies in human and maize have indicated that CENPC may localize to centromeres through its RNA binding ability (Du, Topp et al., 2010, Wong et al., 2007). To test whether mouse CENPC can directly bind MinSat RNA and determine if there is a preference for binding to either strand of MinSat RNA, I conducted an RNA pull-down experiment using in vitro transcribed and biotinylated forward or reverse MinSat RNA as bait. Based on the Western blot result, indeed, CENPC has a selective binding to MinSat RNA instead of *Egfp* mRNA which is used as a negative control. Interestingly, using equimolar amounts of one repeat RNA, MinSat forward transcripts were able to pull down more CENPC protein from mESC total cell lysate compared to reverse MinSat RNA **(Figure 11A)**. This suggests that CENPC RNA binding is specific and has a preferential binding to MinSat forward transcripts over reverse transcripts. Furthermore, to predict potential RNA binding regions in CENPC and further investigate if murine CENPC can interact with RNA directly, I applied a web-based prediction algorithm tool named catRAPID (Livi, Klus et al., 2016). By providing amino acid sequences, the catRAPID algorithm can predict the likelihood of a protein as RNA binding protein (RBP) based on potential secondary structure, hydrogen bonding and van der Waals contributions from amino acid sequences. Furthermore, it also gives scores of certain regions that potentially act as RNA binding regions. According to the prediction result, the overall prediction score is 0.58 (a score above 0.5 is considered as RBP). There are five potential RNA binding regions identified. In the following, the different predicted regions are numbered according to their potential RNA binding ability **(Figure 11B, C)**, with region 1 being suggested as the most potent binding region and region 5 with the least binding potential. To address if the RNA binding regions of mouse CENPC can directly bind to 1 repeat of mouse forward MinSat RNA, we performed an *in vitro* gel shift assay by incubating each catRAPID region to 1 repeat MinSat RNA. With the staining of SYBR gold, we saw the shift of the RNA staining signal toward the upper part of the gel suggesting the formation of protein-RNA complex. In line with the prediction result, catRAPID region 1 has the highest binding affinity to (forward) MinSat RNA. Although CENPC catRAPID regions 2 and 3 also interact slightly with the MinSat RNA, the binding is much weaker than region 1. Regions 4 and 5 bind the lowest to MinSat RNA **(Figure 11D)**. To compare these results with previously published human CENPC RNA binding regions,

I aligned both CENPC amino acid sequences, The RNA binding region 1 characterized in this study perfectly lines up with the potential RNA binding region identified in human CENPC, suggesting this RNA binding region may be evolutionarily conserved (Wong et al., 2007) **(Figure 11E)**.

A



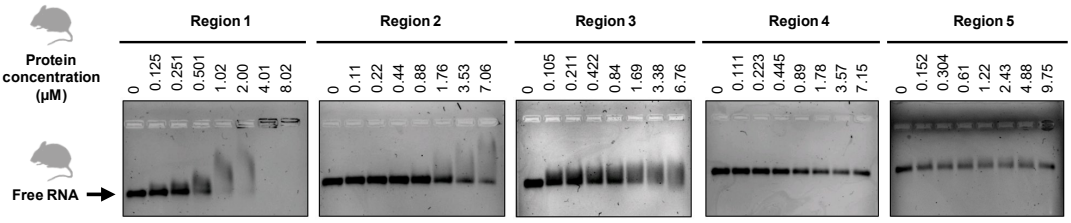
B



C

Amino acid range	Name	Sequence
PUTATIVE_AA_391_528	Catrapid 1	SKPKAAEELRKGQSSWENSNVSNTGQDKLQINSKRNMKDCEEVRNEPNPKKQKPALEN KKKTNSTQTNKEKSGKKFFSGGSKNKFVPKKVTLTSRRSCRISQRPSEWWRVKSDESSVD RNPSKENNSPVVYPNKKKQT
PUTATIVE_AA_566_655	Catrapid 2	GTVSGHDDTSRSQRKPLKIIEADPTQKSLAISRPKRGCKYRNNVMTSPNVHLKSHTEEYTS KTQMESASNSEMSKRSVWEESGPSRFKNY
PUTATIVE_AA_147_220	Catrapid 3	AGQKRVASVSRSPVDRQASNKNISFKTRKRLNFEDKVTLSTAETENSVLQVEDNLSKGQE GTSSEITQKRDDLS
PUTATIVE_AA_40_95	Catrapid 4	CFEDQSKASFLDDFTESLTSSTQKKKANYSQSSSKKCPESHSKPVPVSSRTGEASL
PUTATIVE_AA_284_376	Catrapid 5	RLWVMIPSKDRHLSAHKPSPENTALLQGKKSREKSHSLSAMTFARNTQSDKAHPIEEAQL SVEENPATTCTDELENDCRSPENKMQSETAKTP

D



E

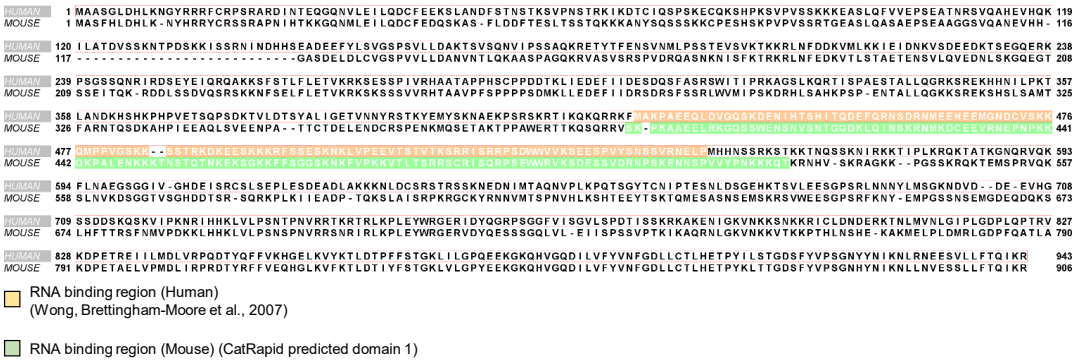


Figure 11. RNA binding of CENPC protein to MinSat RNA

(A) Immunoblot analysis of CENPC protein pulled down by biotinylated RNA. The *in vitro* transcribed RNAs used were labeled on top of the image.

(B) Scoring profile based on the amino acid sequence of CENPC protein predicted by catRAPID webpage. The line represents the 0.5 cut-off of the score.

(C) Prediction result of potential RNA binding regions from catRAPID. The numbering within names corresponds to the ranking of RNA binding ability.

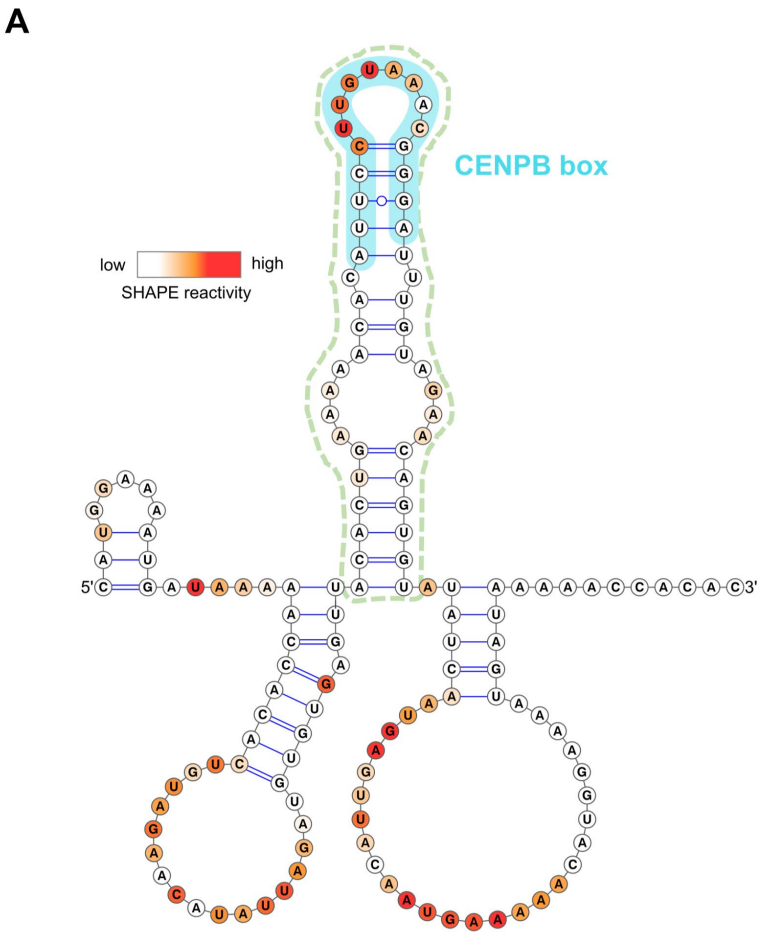
(D) Representative gel images of gel shift experiments. The signal shows the mobility of one repeat of MinSat RNA. The CENPC protein peptides and concentrations incubated were labeled on top.

(E) Amino acid sequence alignment of Human and Mouse CENPC proteins. Orange and green highlighted regions are the potential RNA binding regions from human and mouse (CATRAPID 1), respectively.

MinSat RNA forward transcripts adopt a specific secondary structure

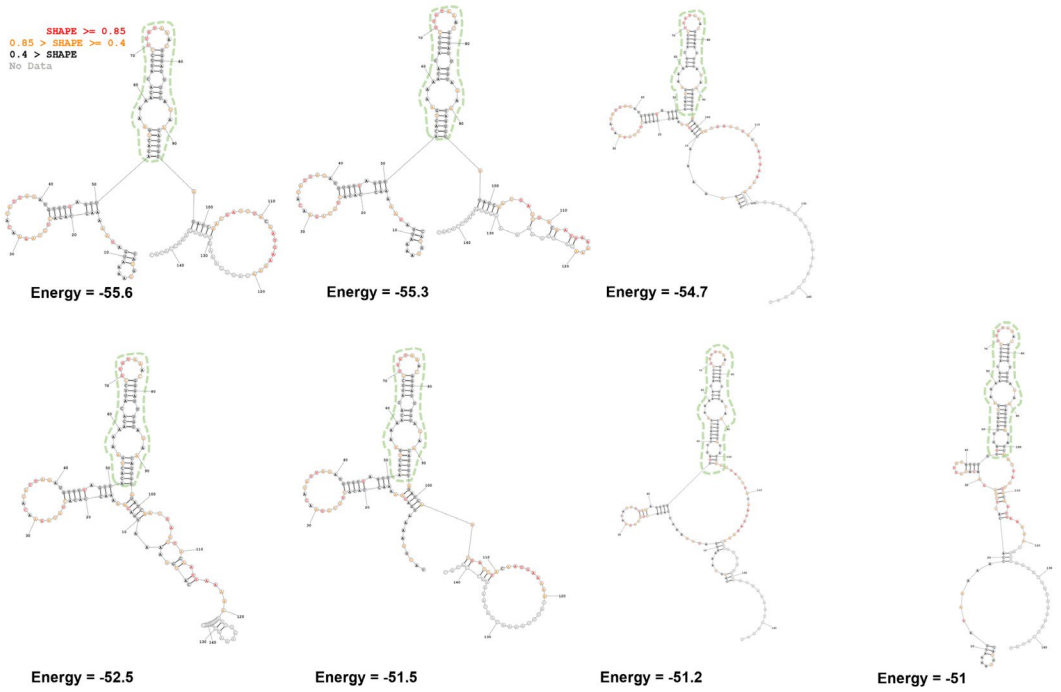
One of the features of an ncRNA is the formation of a secondary structure that in many instances is necessary to interact with protein-binding partners (Sanchez de Groot et al., 2019). Because of this ncRNA property, I wondered if MinSat RNA forward transcripts also possess a specific secondary structure that may mediate the RNA-protein interaction. To address this question, we conducted selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) analysis in collaboration with the Sattler lab (Merino, Wilkinson et al., 2005). SHAPE assay is a chemical probing technique to characterize RNA secondary structure. Taking the advantage that N-methylisatoic anhydride (NMIA) only modifies open loops on the RNA structure, it blocks reverse transcriptase at these unbase-paired regions when performing primer extension. By sequencing these cDNA fragments, we can predict the structure of the RNA of interest. Applying SHAPE assay on MinSat RNA forward transcripts, we find that MinSat RNA can form a secondary structure. The structure with the lowest entropy contains three stem-loop structures. However, among three of them, there is one major stem-loop that contains a longer stem region which seems to be relatively stable **(**green dot-lined in **Figure 12A)**. To test this, we also predicted alternative secondary structures with different thermodynamic parameters. The major stem-loop is present/forms in all the thermodynamic conditions tested, whereas the other two stem-loops may still base-pair differently with the rest of the RNA regions suggesting that the identified stem-loop is a stable structure of MinSat forward transcripts **(Figure 12A, B)**. Surprisingly, we noticed that a known consensus sequence in centromeric DNA across vertebrates, the CENPB box (Masumoto et al., 1989, Suntronpong et al., 2016) is located at the

apical loop of the stem-loop structure (blue outlined) **(Figure 12A)**. To confirm the existence of the stem-loop structure shown by SHAPE result, we conducted a 1H NOESY NMR analysis on the major stem-loop (green dot-lined). The NMR spectra show signals from different types of hydrogen atoms which can reflect the formation of hydrogen bond formation. This revealed that the nucleotides at the stem region are base-paired, supporting the SHAPE result that a stem-loop is formed in this region **(Figure 12C)**. Furthermore, we did the same SHAPE analysis with two consecutive repeats of MinSat forward transcripts. According to the raw SHAPE reactivity spectrum, we see a repetitive pattern of peaks (colored-lined highlighted) which indicates that the stem cell loop structure is repeated along the RNA sequence **(Figure 12D)**.

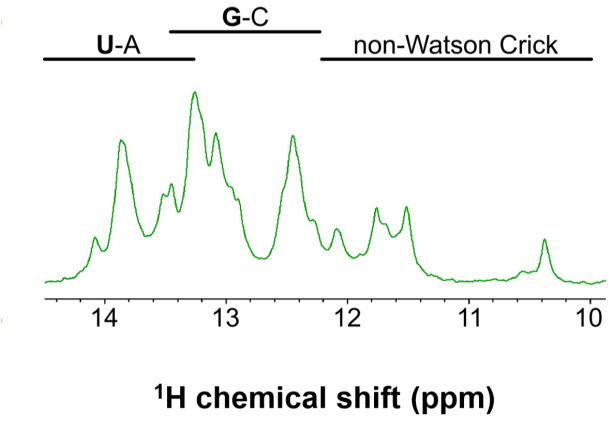


MinSat RNA 1 repeat forward (E=-55.6 kcal/mol)

B


SHAPE >= 0.85
0.85 > SHAPE >= 0.4
0.4 > SHAPE
No Data
Energy = -55.6
Energy = -55.3
Energy = -54.7
Energy = -52.5
Energy = -51.5
Energy = -51.2
Energy = -51

C


G-C
U-A
non-Watson Crick
14
13
12
11
10
¹H chemical shift (ppm)

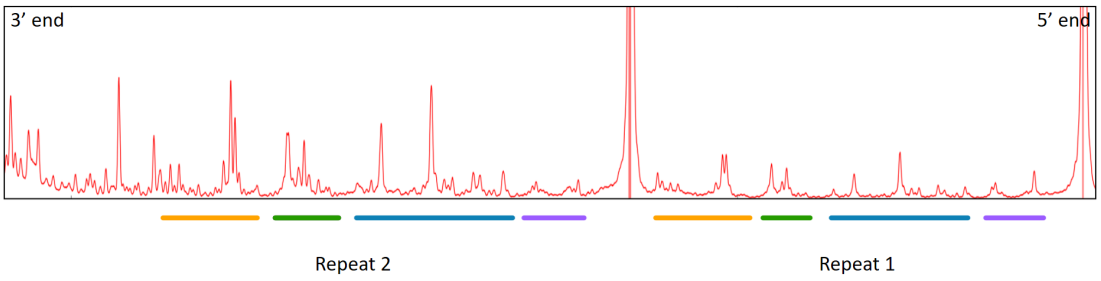

3’ end
5’ end
Repeat 2
Repeat 1

Figure 12. MinSat RNA forward transcript forms a specific secondary structure.

(A) Secondary structure of one repeat MinSat RNA forward by SHAPE analysis. SHAPE reactivity is revealed using a color gradient on the top left. The green outlined stem-loop (Position 51-96) is predicted as a thermodynamically stable region. The light blue outlined region corresponds to the CENPB box motif (Position 65-81). The numbers indicate the nucleotide position relative to the start of the primer extension reaction.

(B) Alternative secondary structures predicted using variable thermodynamic parameters. The green outlined regions represent a constant stem-loop structure across conditions.

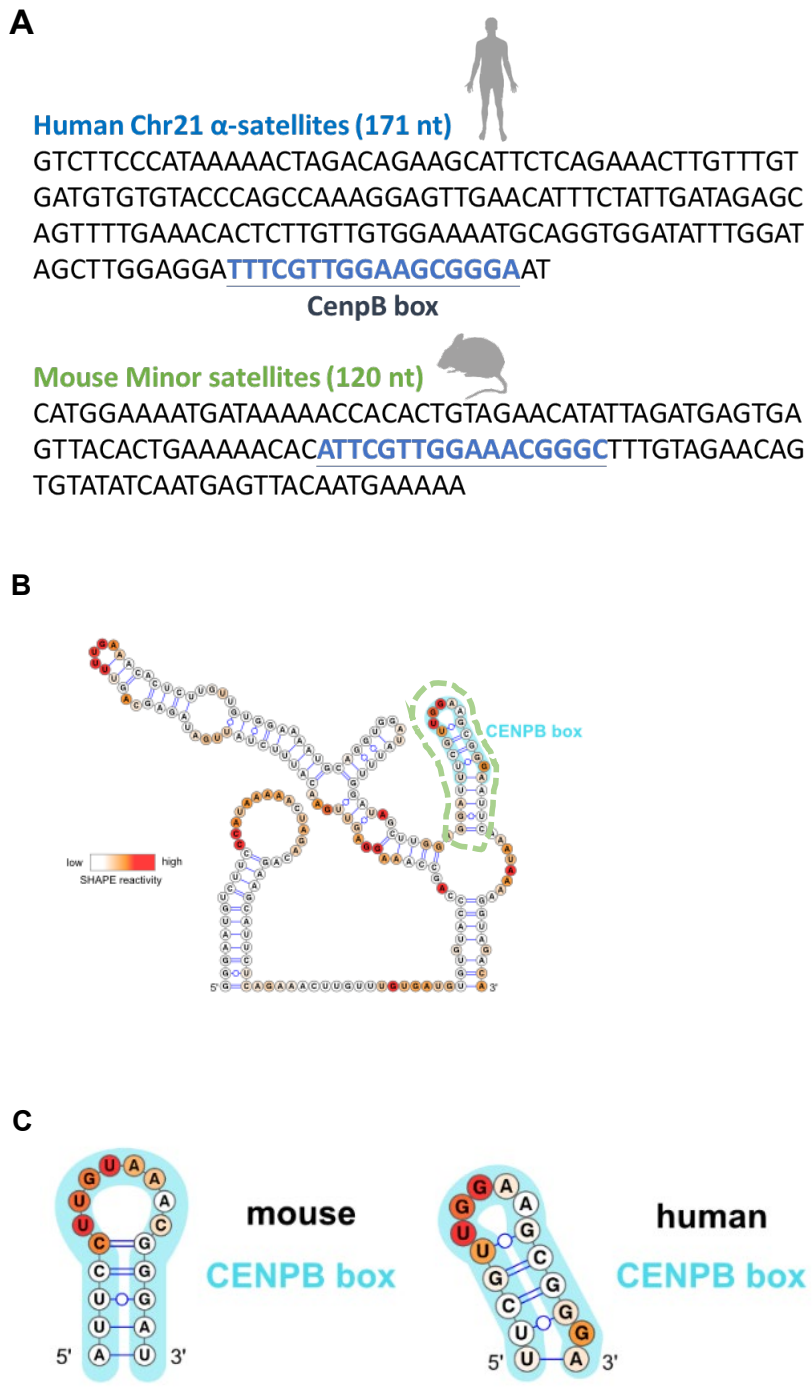
(C) 1D imino NMR spectrum of the MinSat RNA stem-loop (corresponding to the green dotted-lined region in Figure 8A)

(D) Raw SHAPE reactivity spectrum of two repeats of MinSat forward RNA. The colored outline reveals the repetitive patterns of peaks.

Human and mouse centromeric RNAs both possess similar stem-loop structures with apical loops containing the CENPB box sequence.

The lack of similarity in centromeric DNA sequences between species has long been considered a paradox to a shared function of centromeres (Henikoff et al., 2001). As a consequence, also cenRNAs have a high degree of sequence variability. As a model that could explain the divergence in sequence while maintaining similar functions, I hypothesized that instead of sequence similarity, cenRNAs may have similar secondary structures across species. For example, human α-satellite RNA and mouse minor satellite RNA are both AT-rich sequences but the sequence similarity is poor. The CENPB box is the only suggested shared sequence between the two sequences **(Figure 13A)**. To test this hypothesis, we performed SHAPE analysis on one repeat forward of human α-satellite RNA and compared if there is similar secondary structure surrounding CENPB box motif. Interestingly, human α-satellite RNA is structured and, like mouse MinSat RNA, forms a stem-loop with a central CENPB box motif **(Figure 13B)**. The loops on the stem-loop structure from humans and mice follow the following pattern: the beginning of the CENPB box sequence is base pairing with the end of the sequence and the apical loop is formed in the center of the CENPB box **(Figure 13C)**. Next, based on this structural similarity, I suspected that human α-satellite RNA can also bind to mouse CENPC RNA binding regions. To test this, we performed the same gel shift experiment as previously described by incubating 1 repeat

human α-satellite RNA with different catRAPID-predicted RNA binding regions. We indeed found that the human RNA formed a complex with the mouse protein regions depending on its concentration. Of note, catRAPID region 1 showed the best binding ability also in this inter-species binding assay. This suggests that CENPC and cenRNA interaction can happen in a cross-species manner within mammals **(Figure 13D)**. Moreover, this potentially points to a direction that the cenRNA molecules were evolutionarily selected by centromeric proteins at structure level instead of sequence and co-evolve throughout species evolution.



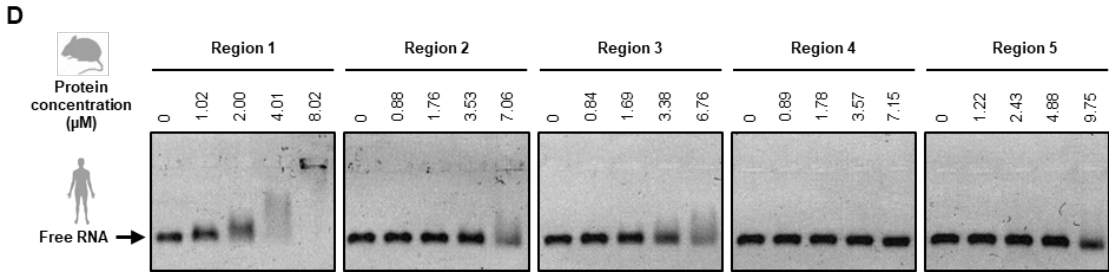


Figure 13. Structural conservation of cenRNA between mouse and human.

(A) Sequences of 1 repeat mouse MinSat RNA and human α-satellite RNA. The blue underlined region corresponds to the CENPB box.

(B) Secondary structure of human one repeats α-satellite forward RNA. SHAPE reactivity is revealed using a color gradient on the left. The green outlined stem-loop shows a similarity with the one in mouse MinSat RNA. The light blue outlined region corresponds to the CENPB box motif.

(C) Apical loop structures of MinSat RNA and human α-satellite RNA based on SHAPE analysis. Note that in both species the CENPB box motif (blue) is embedded in the apical loop structures.

(D) RNA EMSA analysis between human α-satellite RNA and CENPC RNA binding regions. The concentration and purified peptides used are labeled on the top.

Overexpression of human α-satellite RNA leads to an increase in chromosome missegregation events in mouse cells

Considering the observation of similarities on the structural level, I hypothesized that human cenRNA may elicit the same phenotype as mouse MinSat RNAs. To test this, I cloned one repeat of human α-satellite sequence into the same mammalian expression vector used for the mouse MinSat RNAs, the pCAG plasmid **(Appendix Figure 1)** and transfected it into mESCs. To assess the overexpression level, I performed strand-specific qPCR which confirmed that the α-satellite RNAs were overexpressed after 48 hours post-transfection **(Figure 14A)**. Next, to look at chromosome missegregation, DAPI and α-tubulin were stained as described above **(Figure 14B)**. From immunofluorescence micrographs, I counted the percentage of chromosome bridges and lagging events among total anaphase and telophase events. I found that overexpression of either forward or reverse α-satellite RNA can contribute to an increase in chromosome missegregation events to 28% ± 8%, or 33% ± 11%, respectively **(Figure 14C)** whereas only 12% ± 4% of missegregation events were detected in empty vector-transfected mESCs. This data suggests that α-satellite RNA

may function through similar molecular mechanisms as mouse MinSat RNA regardless of the specific centromere sequence of each species.

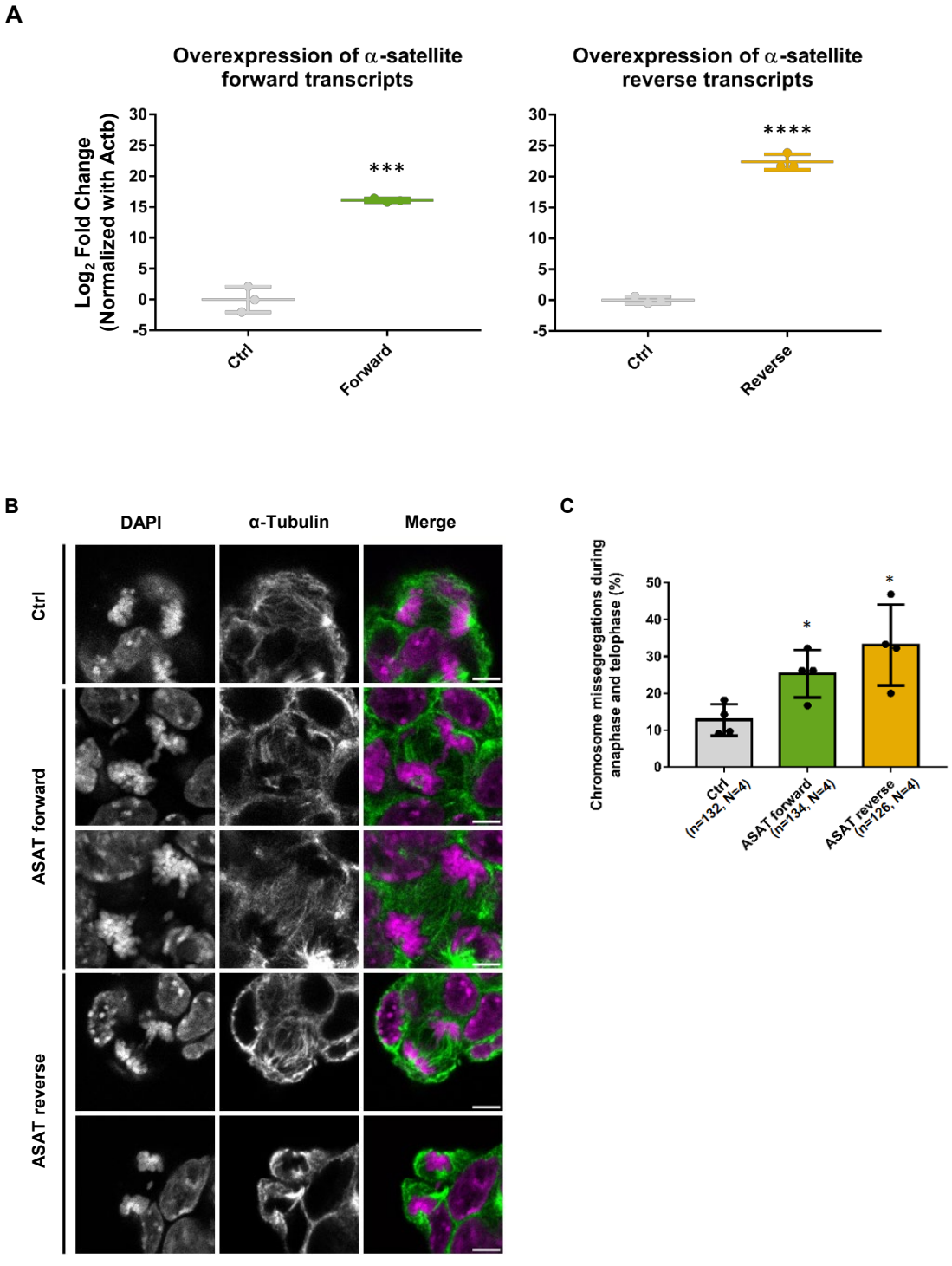


Figure 14. Overexpression of α-satellite RNAs induces chromosome missegregations in mESCs.

(A) RT-qPCR for human α-satellite RNA in ESCs transfected with forward (left panel) or reverse (right panel) α-satellite expressing plasmids using strand-specific primers. Ctrl, empty pCAG vector control; Solid line, mean Ct values of N=3 independent experiments (dots) normalized to beta-actin mRNA and control; error bars, standard deviation, pair-wise comparisons with control (empty pCAG vector) p-values *** ≤ 0.001, **** ≤ 0.0001.

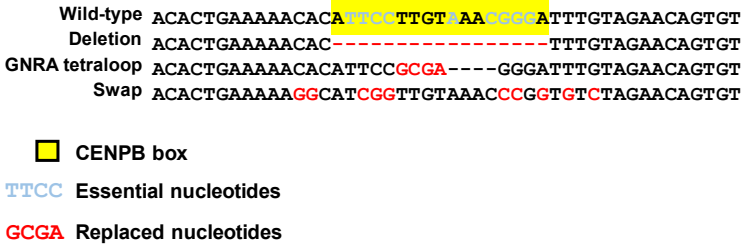
(B) Representative images of immunofluorescent staining of mESCs after transfection of pCAG plasmid expressing α-satellite RNAs. Scale bar, 5 μm.

(C) Quantification of chromosome bridge and lagging events during anaphase and telophase in mESCs under overexpression of α-satellite RNAs. Ctrl, empty vector control. Bars indicate the mean percentages; error bars, standard deviation; *, p-value ≤ 0.05 (pair-wise comparison to control). The number of biological replicates (N) and the total number of mitotic figures analyzed (n) are indicated.

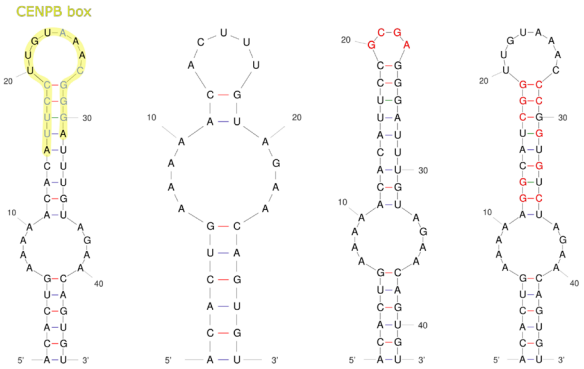
The apical loop structure rather than CENPB box is critical for MinSat forward RNA function

To further characterize the relationship between MinSat RNA structure and its function, I designed mutants focusing on the loop and CENPB box motif. I generated three different RNA mutants, as follows: i) a stem-loop without the CENPB box motif (Deletion); ii) a stem-loop in which we replaced the apical loop with a GNRA tetraloop (GNRA tetraloop) (Heus & Pardi, 1991), and iii) a stem-loop in which we swapped the CG base pairing in the stem region of the stem-loop (Swap) **(Figure 15A)**. The RNA mutants that we generated are thus: i) with both the CENPB box motif and the stem-loop structure disrupted (Deletion); ii) with CENPB box intact but the loop structure was replaced by GNRA tetraloop, and iii) with the stem-loop structure maintained but without the CENPB box motif (Swap) **(Figure 15B)**. I expressed these MinSat RNA mutants in mESCs and confirmed that the expression level is comparable across experimental groups by qPCR **(Figure 15C)**. Using chromosome missegregation as a readout, I observed that only the Swap mutant, which maintains the stem-loop structure, shows a chromosome missegregation phenotype comparable to the wild-type MinSat forward RNA. Moreover, mutants for which the CENPB box sequence was removed entirely or for which the apical loop was replaced by only four nucleotides, lose the ability to cause aberrant chromosome segregation **(Figure 15D, E)**. This result supports my hypothesis that the apical loop plays a critical role in the function of MinSat forward transcripts rather than the RNA sequence.

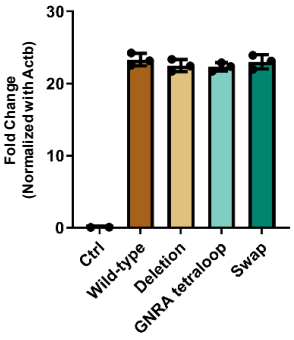
A


Wild-type ACACTGAAAAACACATTCCTTGTAAACGGGCATTTGTAGAACAGTGT
Deletion ACACTGAAAAACAC----------TTTGTAGAACAGTGT
GNRA tetraloop ACACTGAAAAACACATTCCGCGA----GGGATTTGTAGAACAGTGT
Swap ACACTGAAAAAGGCATCGGTTGTAAACCCGGTGTCTAGAACAGTGT
CENPB box
TTCC Essential nucleotides
GCGA Replaced nucleotides

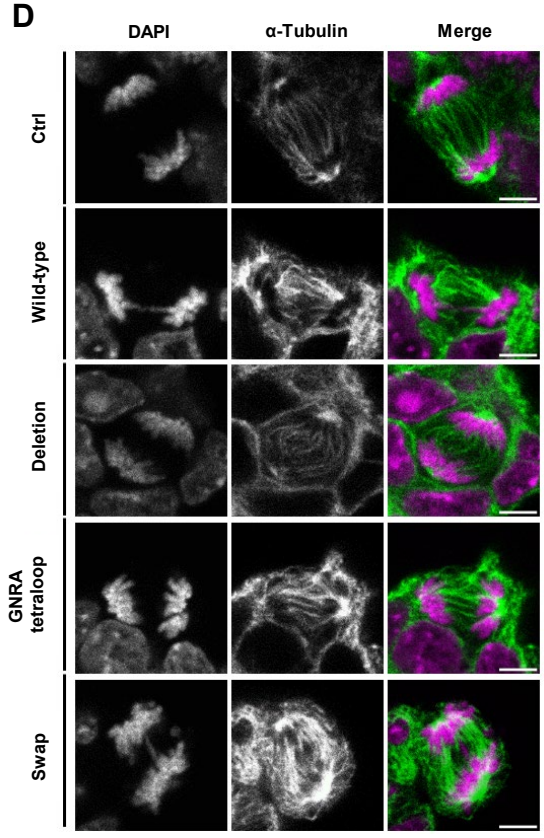
B


Wild-type
Deletion
GNRA tetraloop
Swap
CENPB box

C


Fold Change (Normalized with Actb)
30
20
10
0
Ctrl
Wild-type
Deletion
GNRA tetraloop
Swap

D


DAPI
α-Tubulin
Merge
Ctrl
Wild-type
Deletion
GNRA tetraloop
Swap

E

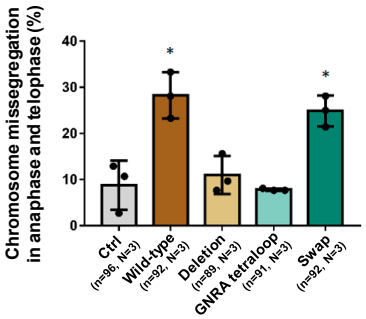

Chromosome missegregation in anaphase and telophase (%)
40
30
20
10
0
*
*
Ctrl (n=96, N=3)
Wild-type (n=92, N=3)
Deletion (n=89, N=3)
GNRA tetraloop (n=91, N=3)
Swap (n=92, N=3)

Figure 15. Overexpression of CENPB box mutants in mESCs.

(A) Sequence comparison showing the position of mutated or deleted (red) nucleotides. The essential nucleotides are characterized by previous publication (Iwahara et al., 1998).

(B) Predicted secondary structure using mFold (Zuker, 2003). Yellow highlighted regions represent CENPB box motif in a wild-type sequence.

(C) Quantification of MinSat RNA levels in ESCs after transfection of the pCAG plasmids containing different mutants of MinSat using RT-qPCR analysis. Bars, mean values of N=3 independent experiments (dots) normalized to beta-actin mRNA and empty pCAG vector control (Ctrl); error bars, standard deviation.

(D) Representative confocal images of mouse ESCs transfected with plasmids expressing empty plasmid, Wild-type, or different mutants. Cells were immunostained with alpha-tubulin and DAPI. Scale bars, 5 µm.

(E) Quantification of chromosome missegregation events in anaphase and telophase mouse ESCs after overexpression of the structural RNA mutants, Wild-type or empty pCAG vector as control (Ctrl), respectively. Bars indicate the mean percentage of N=3 independent experiments (dots); error bars, standard deviation; *, p-value ≤ 0.05 (pair-wise comparison to control).

Binding affinity to CENPC decreases in RNA with CENPB box mutants

Based on these results, I hypothesized that the chromosome missegregation phenotype observed in **Figure 15D and E** may be caused by the differential binding of CENPC to different MinSat RNA mutants. To test this, I performed again an RNA pull-down assay with the different MinSat RNAs versions with mutated CENPB box (Deletion, removal of CENPB box motif; GNRA, replacement of apical loop with GNRA tetraloop; Swap, invert G-C base-pairing at stem region without disrupt stem-loop structure). Under conditions where RNA is present in equimolar amount, I observed that all the RNA mutants decrease in CENPC pull-down efficiency compared to wild-type MinSat RNA control **(Figure 16), which** strongly binds to CENPC. This result indicates that the apical loop in the MinSat RNA forward transcript may indeed play an important role in the interaction between CENPC and MinSat RNAs. However, based on the data shown in **Figure 15**, Swap mutant also leads to chromosome missegregation. This potential difference could be due to the fact that the in vitro transcribed Swap RNA mutant obtains a different secondary structure than the RNAs expressed in cells through plasmids.

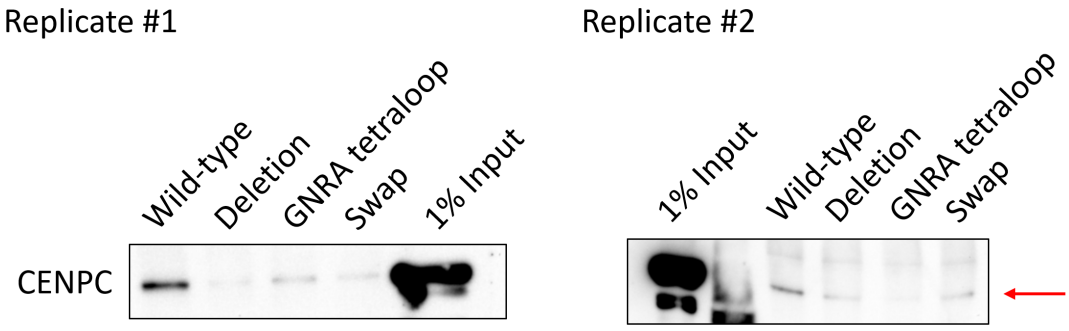


Figure 16. Biotinylated RNA pull-down using different CENPB box mutant
Gel images of biological replicates for Western blot analysis. The red arrow points to CENPC immunoblot signal.

Enrichment of MinSat transcripts from total RNA for Oxford Nanopore sequencing

Although we were able to characterize the structure of a single repeat of MinSat RNA by SHAPE, there remains a gap in our understanding of the MinSat RNA molecules from their 5' to the 3' end. Due to the repetitive nature of MinSat sequences, comprehensive sequence information at the genomic level remains missing.

With conventional short-read sequencing methods, it continues to pose difficulties to assemble a reliable reference genome for this specific region *de novo.* Mapping short fragments with high similarity repeats is also one of the challenging tasks in sequence analysis. Thus, a long-read sequencing technique like Oxford Nanopore technology (ONT) may potentially provide the data to obtain complete sequences for such repetitive regions. To sequence sufficient depth, MinSat RNAs, which are present in low numbers compared to total RNA, must be enriched first. I applied biotinylated DNA oligos **(Appendix Table 4)** coupled with streptavidin beads to pull down both MinSat forward and reverse transcripts **(Figure 17A)** to enrich these RNAs in terms of concentration for Nanopore sequencing and on the other hand to remove the majority of ribosomal RNAs. Given that hybridization relies on a complementary strand approach, with binding occurring between opposite strands, the effectiveness of a probe mixture is greatly influenced by the temperature of hybridization. To optimize the enrichment condition, I tested the hybridization step under different temperatures (40, 50, and 60°C). After conducting MinSat RNA

enrichment protocol, I performed a q-RT-PCR quantification using primers recognizing MinSat RNA and ribosomal RNA large subunit (LSU). Based on the q-RT-PCR result, the hybridization step under 40°C shows the best (600-fold) enrichment when using ribosomal RNA large subunit (LSU) as a comparative standard **(Figure 17B)**. In this case, I reached a hybridization condition that performs the best specificity of the probes which can capture the most MinSat RNA or eliminate the most ribosomal RNAs. Furthermore, biological triplicates were performed and beta-actin was used as a negative control to show that beta-actin mRNA is, in contrast to MinRat transcripts, not enriched and even partially reduced in amount. In summary, I found that the enrichment measured with RT-qPCR can successfully lower the proportion of rRNA and concentrate MinSat RNAs which largely improved the rRNA contamination problem for subsequent sequencing with the Nanopore platform **(Figure 17C)**.

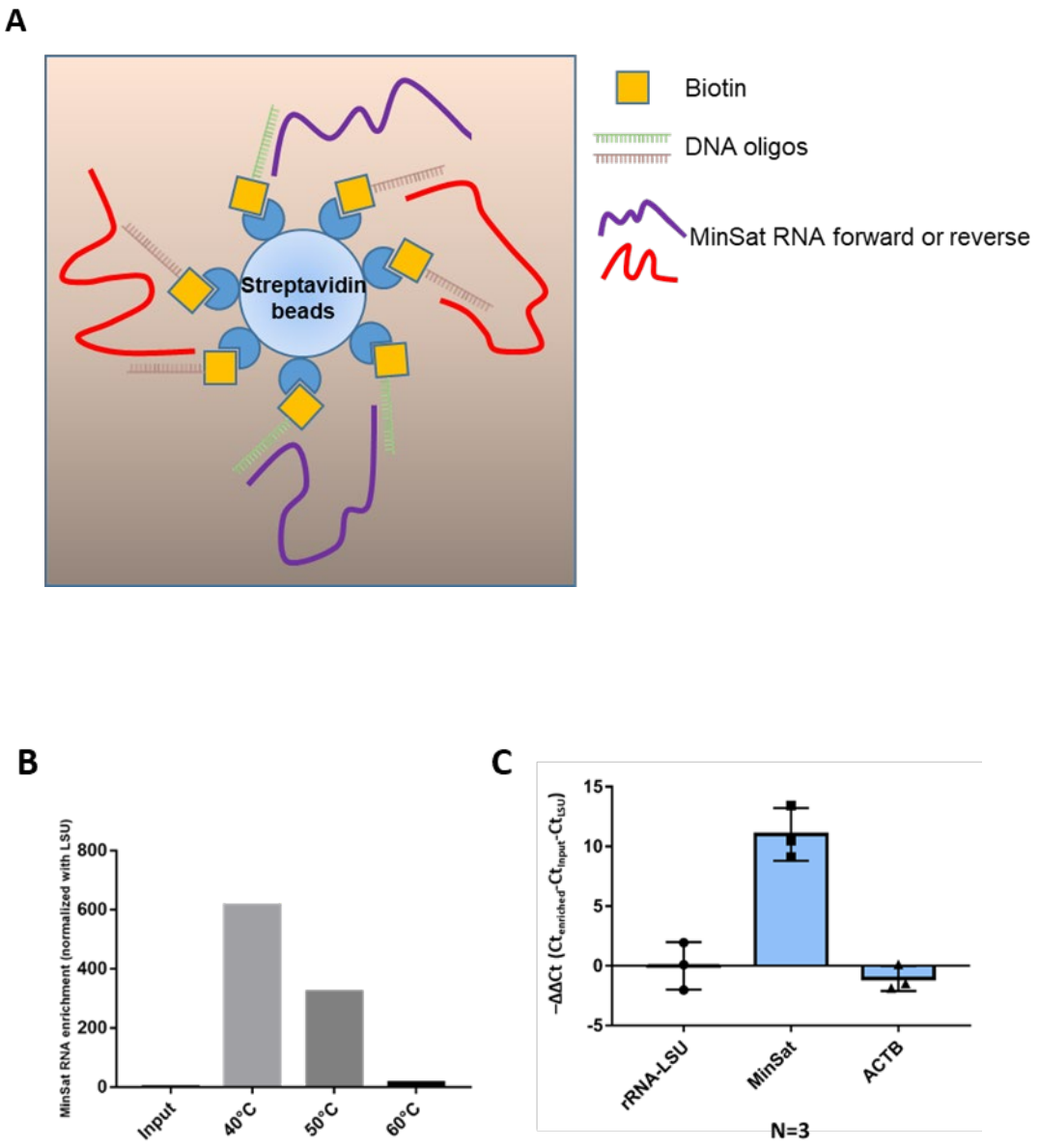


Figure 17. MinSat RNA enrichment using biotin-labeled DNA oligos

(A) Scheme of biotinylated DNA oligos coupled with streptavidin beads to pull down MinSat forward and reverse transcripts.

(B) RT-q-PCR quantification of MinSat RNA under different hybridization temperatures. Bar, fold-change related to Ribosomal RNA long subunit (LSU).

(C) RT-q-PCR quantification of MinSat RNA after enrichment. Bar, fold-change related to Ribosomal RNA large subunit (LSU).

Mapping of MinSat RNA reads

In order to map MinSat transcripts, we performed a computational analysis in collaboration with Anna Danese and Stefan Stricker. To first confirm the quality of Nanopore sequencing reads in three different biological replicates, we plotted the read length distribution and read quality using NanoPlot tool (De Coster & Rademakers, 2023) **(Figure 18A)**. Within three replicates, most of the reads are shorter than 500 bp suggesting that the captured sequencing reads mainly belong to one repeat of MinSat RNAs. This can also be due to the limitation of PCR-based method on repetitive sequences. During the cDNA PCR amplification step, PCR primers can anneal multiple repeats on the cDNA. However, only the primer closest to 5' end will become the final amplicons meaning that this PCR has bias toward amplicons with fewer repeats.

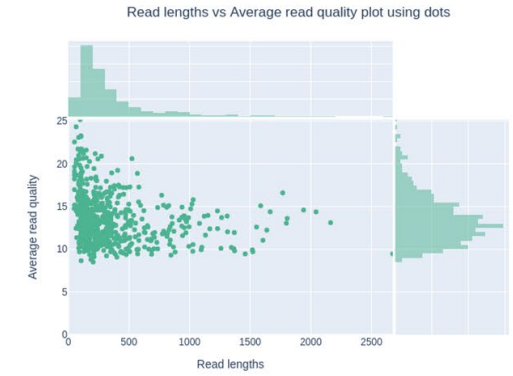
One of the concerns of long-read sequencing is the error rate and sequencing quality. Similarly, to short-read sequencing methods, Nanopore sequencing provides its scoring system as well to reveal the quality of base calling. Using the threshold suggested by Oxford Nanopore, a cut-off score of 7 was established, indicating that reads with an overall quality score above 7 will be classified as passed reads. Thus, we plotted the correlation between read length and overall read quality of the passed reads. We found that indeed the read quality below 7 was completely filtered away. However, we still observed a correlation indicating that longer reads are prone to have lower read quality scores, suggesting that errors may accumulate along with the single molecule sequencing **(Figure 18B)**. After doing the quality check of the sequencing reads based on the score, read length distribution and read counts of sequencing reads using NanoPlot, we first distinguished the direction of the reads using chimeric sequences generated from RT primers **(Appendix Table 5)**, we combined reads from all three datasets and mapped to mouse genome reference including MinSat sequences. The majority of the reads are mapped to MinSat sequences indicating that the enrichment of MinSat RNA and sequence-specific reverse transcription are indeed working **(Figure 18C)**. We further checked the mapping results specific to the reads that map to

MinSat forward or reverse transcripts. Although the alignment method is not optimal due to the repetitive sequences, we decided to examine the mapping results from either primary or secondary alignments. Single nucleotide polymorphisms (SNPs), insertions, or deletions frequently exist among individual repeats and transcripts **(Figure 19A, B)**. We found that centromeric reads are not only heterogeneous in between transcripts but also repeats within one transcript which reflect the complexity of centromeric DNA sequence.

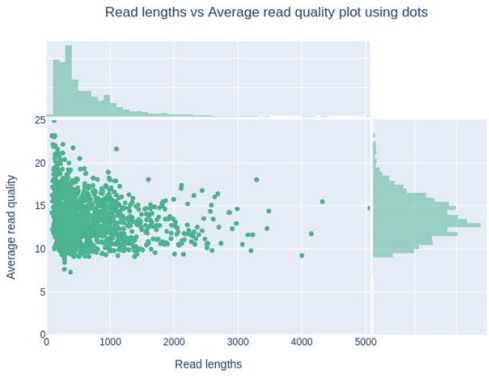
A

B

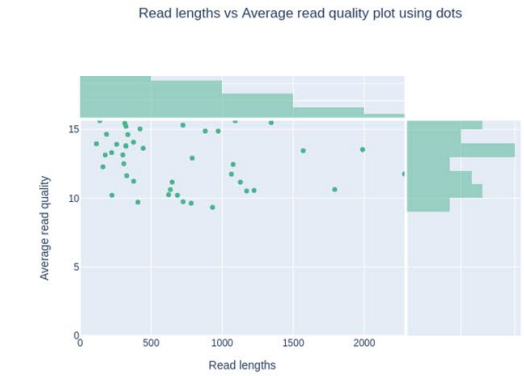
Sample 1



Sample 2



Sample 3



C

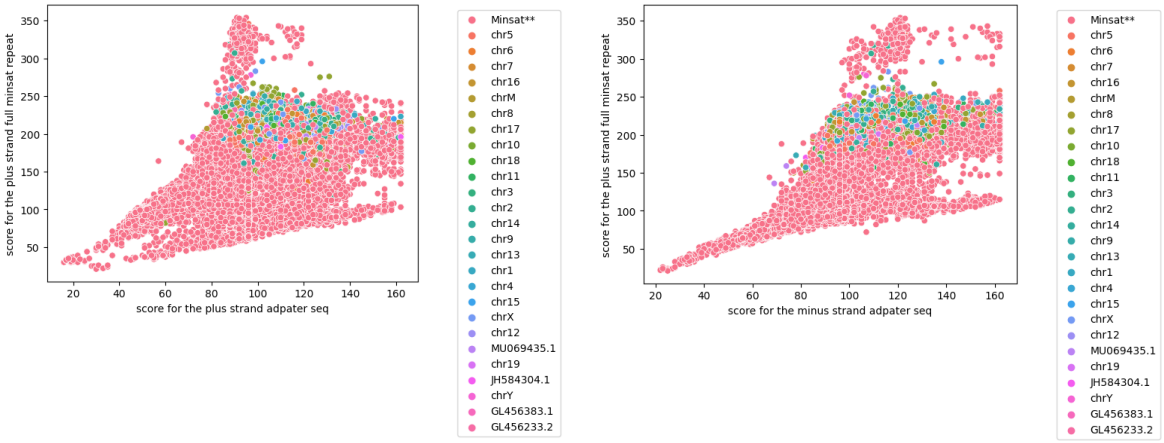


Figure18. Quality check of three Nanopore cDNA sequencing datasets

(A) Total read counts and histograms showing read length distribution and corresponding read counts. Each sample represents different biological replicates.

(B) Dot blots reveal a correlation between read length and base calling quality. The density of dots is shown using histograms aside. Each sample represents different biological replicates.

(C) Scatter plots showing the likelihood of our targeted sequencing reads. Each dot represents a single read. Dot colors imply the chromosomes that the read aligns to.

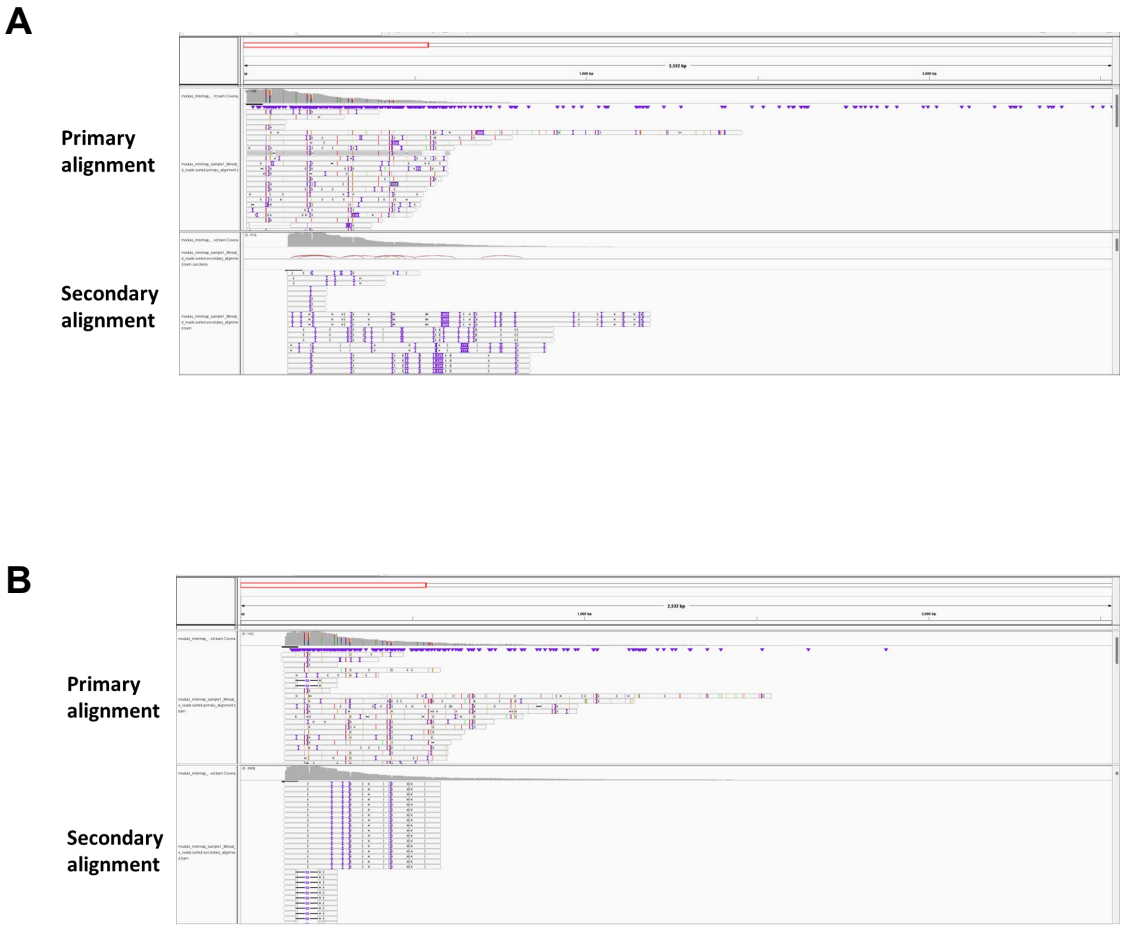


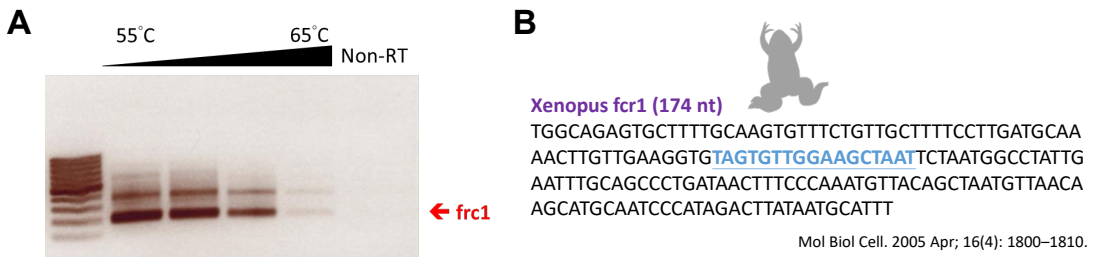
Figure 19. IGV tracks showing the mapping result to customized MinSat references.

(A) Mapping of MinSat forward transcripts to MinSat customized reference.

(B) Mapping of MinSat reverse transcripts to MinSat customized reference.

Centromeric RNAs are transcribed in *Xenopus laevis* and *Bos taurus*

Centromeric transcription is an evolutionarily conserved feature between species (Ideue & Tani, 2020). To characterize centromeric RNAs in other vertebrates, I cloned centromeric RNA sequences from two other species, *Xenopus laevis* and *Bos taurus*. For *Xenopus laevis*, the total RNA was extracted from testis tissue and followed by reverse-transcription (RT)-PCR. The PCR primers used are listed in **Appendix Table 6**. The PCR product showed several bands of different sizes, indicating that the PCR product may contain multiple repeats **(Figure 20A)**. The non-RT control shown did not yield any band confirming that the product indeed comes from cDNA retro-transcribed from RNA. From this RT-PCR result, I could confirm that the frog centromere, fcr1 is indeed actively transcribed. The *Xenopus fcr*1 sequences have been previously published and a putative CENPB box motif has also been described within a single fcr1 repeat **(Figure 20B)** (Edwards & Murray, 2005). Similar to mouse centromeres, *Xenopus* contains a relatively homogenous sequence among different repeats. Thus, using the same pair of primers can amplify multiple numbers of repeats and this is the reason why a pattern of PCR products of several sizes was detected on the gel. For *Bos taurus*, total RNA was purified from bovine epiblast stem cells (courtesy of Marion Genet). Similarly, RT-PCR was performed with specific primers and separated by gel electrophoresis on an agarose gel. The expected sequence size of *Bos taurus* centromeric RNA is 656 bp (Escudeiro, Adega et al., 2019). A PCR product of this size was cloned into pCAG plasmid **(Figure 20C)**. Although a centromeric DNA sequence has been published (Escudeiro et al., 2019), this is the first evidence that *Bos taurus* centromere SAT1.723 is indeed transcriptionally active. Multiple PCR products were detected by gel electrophoresis. However, only the PCR product pointed by red arrow is SAT1.723 whereas the rest of bands are just non-specific PCR product checked by Sanger sequencing **(Figure 20C)**. In the future, solving the secondary structure of cenRNAs from both *Xenopus* and the cow (or bovine cells) can give us a hint if our discovery about CENPB box mediated stem-loop structure is a general rule across vertebrates.



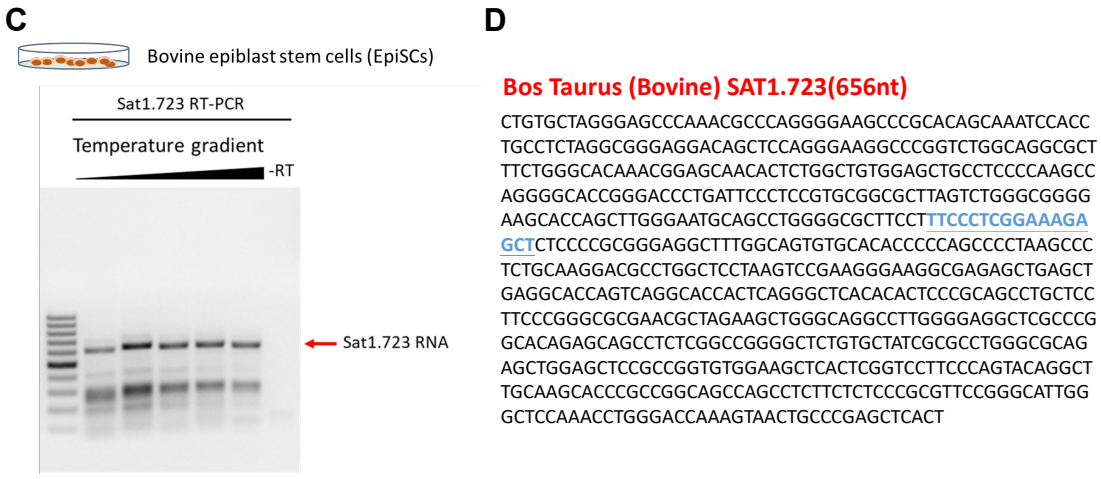


Figure 20. Centromere is transcriptionally active in both *Xenopus laevis* and *Bos Taurus*

(A) Agarose gel image of fcr1 PCR product from Xenopus cDNA. A gradient of primer annealing temperature was used and non-reverse-transcription control are labeled on top.

(B) Reference sequence of *Xenopus laevis* centromere fcr1. The blue highlighted sequence represents CENPB box sequence characterized (Edwards & Murray, 2005).

(C) Agarose gel image of SAT1.723 PCR product. The red arrow indicates the PCR product used for the following experiments.

(D) Reference sequence of *Bos Taurus* centromere SAT1.723. The blue highlighted sequence represents CENPB box sequence characterized (Escudeiro et al., 2019).

Discussion

An optimal expression level of centromeric RNA is required to maintain chromosome segregation fidelity

Centromeric transcription and transcripts have emerged as critical components for centromere formation and kinetochore assembly (Corless et al., 2020, Ideue & Tani, 2020). Interestingly, the transcription timing of centromeric RNAs is different from most other transcription events. To synchronize with mitotic events, centromeric transcription may happen accordingly. In mammals, centromeric RNAs are uniquely transcribed during mitosis (Bury et al., 2020, Ferri et al., 2009, Wu, Lane et al., 2021) and the transcript levels are reduced at the G1 phase in most cell types tested. However, based on my results using Fucci system under mESC background, centromeric RNAs are present at a level similar to G2/M phase across G1 phase. One possible explanation for this observation is that the degradation process of MinSat RNA is comparable in most the cell types, but the very short G1 phase of mESCs allows MinSat RNAs to remain comparably high throughout the entire G1 phase **(Figure 4B, C)**. Hence, a decrease of MinSat RNA was only observed during S phase for either forward or reverse transcripts. In general, centromeric transcription is low (Duda, Trusiak et al., 2017, Perea-Resa & Blower, 2018, Smurova & De Wulf, 2018) but it can vary across cell types or under stress conditions (Bouzinba-Segard et al., 2006, Valgardsdottir, Chiodi et al., 2008). One of the explanations is that due to the fact that centromeric loci are highly compacted regions on the chromosomes, the accessibility of transcription factors or RNA PolII may be low. In the future, for detecting MinSat RNA, I can either increase the amount of total RNA input or remove ribosomal RNAs to enlarge the proportion of the rest of the RNA species in the samples.

Centromeric transcription has been characterized to happen on both strands of DNA template (Ferri et al., 2009, Ling & Yuen, 2019, Topp, Zhong et al., 2004). However, the functional role of two strands of centromeric RNAs has not yet been fully characterized. Thus, it's important to study centromeric RNA in a strand-specific manner. One of the findings of this study was that, at the expression level, the reverse transcript is dominant over the forward transcript in mESCs suggesting that the two RNA molecules do not exist in a 1:1 ratio. However, the detailed molecular mechanisms of centromeric RNA biogenesis and degradation are still an open question. Furthermore, since the two transcripts are generated from identical DNA templates, double-stranded RNA (dsRNA) can potentially form due to sequence reverse-complementarity. Normally, such intracellular dsRNA will be processed by Dicer or other dsRNA sensors as a rapid response (Chen & Hur, 2022). Since MinSat forward

and reverse transcripts are fully reverse-complementary, dsRNA is very likely to form endogenously. Additionally, MinSat RNA is recognized not only for its accumulation but also for its retention in larger sizes when DICER is knocked out in mESCs, suggesting that MinSat RNAs are targeted by the dsRNA pathway for scavenging (Kanellopoulou, Muljo et al., 2005). Now, by knocking down either forward or reverse transcripts, I perturbed the endogenous balance between the two transcripts. Because one strand of the RNA is repressed, a decrease of dsRNA formation and further accumulation of opposite strand of MinSat transcripts may occur. This explains potentially the seesaw effect in **Figure 8B-C** and points toward a mechanism of how centromeric RNA can be maintained at a balance between forward and reverse transcripts.

Imbalanced levels of forward or reverse MinSat RNA lead to chromosome missegregation

As mentioned previously, the function of cenRNA cannot be studied from either strand alone. A reciprocal regulation between forward and reverse transcripts must be considered. In the case of MinSat RNA overexpression experiments, increasing the level of both MinSat transcripts leads to chromosome missegregation. For forward transcripts, exogenous expression has been shown to trigger mislocalization of Aurora B, heterochromatin protein 1 gamma (HP1γ) and alteration in histone mark H3K9me3 (Bouzinba-Segard et al., 2006). These factors are implicated in chromosome segregation and hence, excess amount of forward transcripts may impact directly via these pathways. In contrast, when overexpressing reverse transcripts, the homeostasis of forward and reverse transcripts is highly imbalanced in favor of reverse transcripts. Excessive amounts of reverse transcripts could lead to the sequestration of ribonucleoproteins (RNPs) composed of centromeric proteins and MinSat forward transcripts by RNA-RNA interaction. In the biotinylated RNA pull-down assay **(Figure 11A)**, CENPC protein preferentially binds MinSat forward transcripts. Our experiments would thus be consistent with a model where CENPC binding to chromatin decreases after overexpression of reverse transcripts. In summary, direct and indirect effects will need to be disentangled, although high expression of either forward or reverse transcripts can cause aberrant chromosome segregation, the molecular mechanisms behind these observations may be different.

On the other hand, in MinSat RNA knock-down experiments, down-regulation of cenRNAs also induces chromosome segregation defects **(Figure 9A-D)**. CenRNA has been suggested to function as a molecular guide for centromeric proteins (Mallm &

Rippe, 2015, Quenet & Dalal, 2014, Rosic et al., 2014). A less effective assembly of kinetochore components may underlie the chromosome missegregation phenotypes observed in the MinSat knock-down experiments. This could be due to the loss of mediators for centromeric proteins to the centromeric region. However, a lower amount of CENPC on the chromatin is only observed when both forward and reverse transcripts were knocked down simultaneously. This indicates a potential synergistic or cooperative action of both strands of MinSat transcripts in shuttling CENPC to centromeric chromatin.

Even though levels of CENPC and CENPA in chromatin extraction experiments varied substantially across biological replicates, I observed a general trend: in the overexpression experiments, the binding of CENPA drops similarly as CENPC levels. Interestingly, the overall level of CENPA decreases as well, suggesting that either the translation or the turnover rate of CENPA protein is decreased. One of the explanations for this phenomenon could be feedback between CENPC and CENPA. For instance, it has been shown that CENPC promotes the recruitment of CENPA through Mis18-binding protein 1 (M18BP1) and Holliday Junction Recognition Protein (HJURP) (Dambacher, Deng et al., 2012, Falk, Lee et al., 2016, Moree, Meyer et al., 2011). In mammals, CENPC can recruit M18BP1 to centromeric regions (Dambacher et al., 2012). M18BP1 can then capture HJURP which is a chaperone protein in charge of shuttling and incorporating CENPA into centromeric nucleosomes (Foltz, Jansen et al., 2009, Wang, Liu et al., 2014). This may explain why the decrease of chromatin-bound CENPA is concomitant with that of CENPC. Furthermore, since CENPA is not efficiently incorporated into centromeric nucleosomes **(Figure 7A, B)**, the half-life of the total amount of CENPA may drastically drop due to this (**Figure 7C, D**).

CENPC binds RNA potentially via a stem-loop motif

CENPC is an RNA-binding protein in maize and humans (Du et al., 2010, Wong et al., 2007). However, how CENPC protein interacts with cenRNAs has remained unclear. Especially interactions shown in mammalian systems in previous studies are not based on *in vitro* assay done by one-on-one incubation of protein and RNA, this can create a concern about indirect interaction. For mouse CENPC we narrowed down the potential RNA binding region *in silico* to catRAPID region 1, which overlaps with the RNA binding region of CENPC characterized in humans (Wong et al., 2007). My work presents the first RNA binding region characterized in mouse CENPC, indicating that the RNA binding ability and regions are evolutionarily conserved. CENPC is a protein enriched in Intrinsically disordered regions (IDRs) according to Uniprot database

(UniProt, 2023). Also, based on Alphafold (Jumper, Evans et al., 2021), none of the catRAPID regions was predicted as a structured peptide **(Figure 21)**. Such IDRs can mediate non-canonical RNA binding (Ottoz & Berchowitz, 2020, Ray, Laverty et al., 2023, Zeke, Schad et al., 2022) which may also be the case for CENPC-RNA interaction. Despite the unstructured nature of these regions, the binding of CENPC to RNA displays a high degree of specificity. According to RNA pull-down assay (**Figure 11A)**, endogenous CENPC can bind both forward and reverse MinsSat RNAs instead of *Egfp* mRNA, even if it has a preference for binding to MinSat forward transcripts over MinSat reverse transcripts. Furthermore, besides MinSat and α-satellite RNAs in this study, by combining infrared-crosslinking immunoprecipitation (irCLIP) and SHAPE-MaP a previous study also suggested that CENPC can recognize a stem-loop structure of a lncRNA CCTT in HeLa cells. Altogether, these data suggest that CENPC shows preferential binding to a stem-loop structure (Zhang, Wang et al., 2022). Interestingly, the lingering question pertains to the modes of interaction that contribute to the specificity of this particular RNA secondary structure. In the future, a further structural study applying, for example, CryoEM or NMR on CENPC-RNA RNP 3D structure would provide insight into how CENPC protein and centromeric or non-centromeric RNA complex forms.

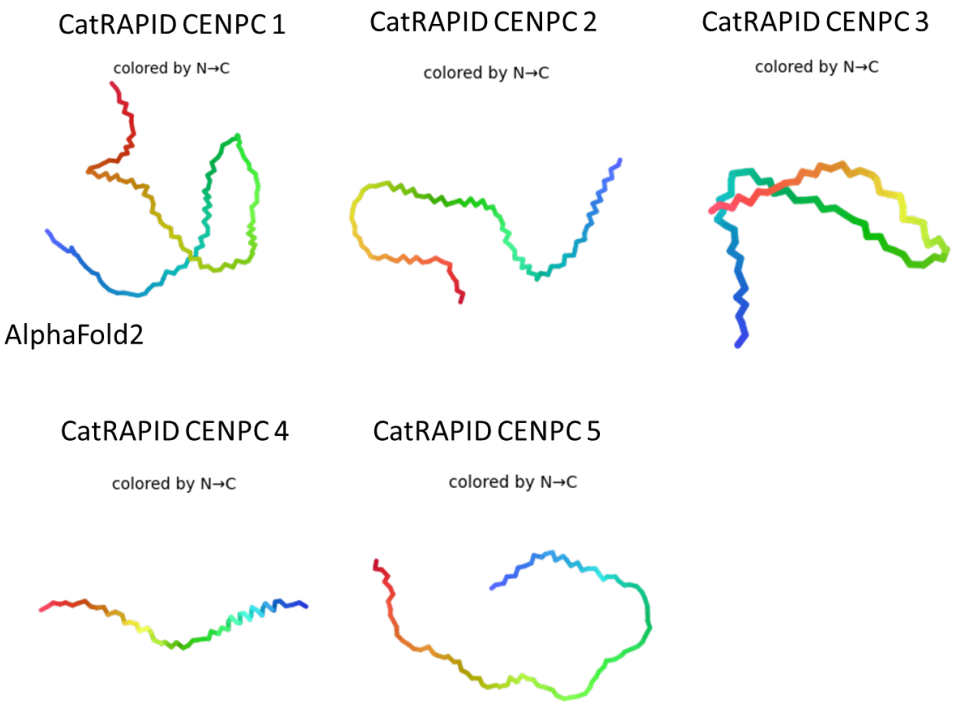


Figure 21. Alphafold prediction of catRAPID regions in mouse CENPC.
In silico prediction of 3D structure of RNA binding regions suggested by catRAPID. Amino acid sequences are used as a query for the prediction.

Secondary structure of centromeric RNA and evolution

As mentioned previously, centromeric sequences have been considered as rapidly evolving elements in the genome. Because the centromere proteins are relatively conserved despite a notable lack of centromeric DNA or RNA sequence similarity, it remains difficult to explain centromere evolution, the so-called centromere paradox (Henikoff et al., 2001). To address this, scientists aimed to understand centromere evolution from multiple angles, one of them being the molecular structure of DNA. At the DNA level, DNA structures such as dyad symmetric sequences, non-B-form DNA and R-loop formation have been proven to be prone to trigger centromere formation (Kabeche et al., 2018, Liu, Yi et al., 2023, Mishra, Chakraborty et al., 2021) and may thus be a distinguishing feature of centromere-forming nucleic acids. Particularly, R-loop formation is frequently coupled with active transcription (Belotserkovskii, Tornaletti et al., 2018). This is in line with the centromeric transcription model where (neo)centromere formation always occurs together with active transcription. Regardless of the sequence, this unique molecular structure may provide a distinguishing molecular signature conferring centromere identity and serve as a platform for kinetochore assembly. On the RNA level, many lncRNAs have been found to be evolutionarily conserved across species without high degree sequence similarity (Diederichs, 2014). I believe that cenRNAs also follow the same strategy throughout evolution. Hence, the working model of cenRNA as a guide for centromeric proteins has been widely recognized in the field (Ideue & Tani, 2020, Liu, Liu et al., 2021, Mallm & Rippe, 2015). Since centromeric proteins are relatively conserved between species, a missing piece of this puzzle is how these proteins recognize RNAs without sequence similarity. Based on our results, we use human and mouse cenRNAs as proof of principle and show a similar secondary structure between species. Along with the concept of structural consensus on lncRNAs, we propose this would be the case for cenRNAs.

Aberrant expression or oncogenic lncRNAs can lead to mislocalization of centromere formation

Besides our study, two research papers on human cancer cells have revealed that two oncogenic lncRNAs can also interact with and recruit CENPC to specific loci. In

colorectal cancer cell line SW480, lncRNA PCAT2 was expressed at locus 8q24 (Arunkumar, Baek et al., 2022). Expression of PCAT2 can recruit CENPC to the 8q24 locus, a non-centromeric region that may cause an ectopic number of centromeres on a single chromosome. Furthermore, according to immunofluorescent staining, mislocalization only happens locally instead of a global disruption of CENPC distribution. This suggests that lncRNA PCAT2 is unlikely to interact with distant loci to cause incorrect occupancy of CENPC on the chromosomes. However, whether lncRNA PCAT2 obtains a specific secondary structure or how PCAT2 binds to CENPC remains unclear. In the other case, in Hela cells, lncRNA CCTT was found to contain two main domains, of which one can interact with CENPC directly and another domain can serve as a DNA binding domain to help locate the RNA molecule to the centromeric region (Zhang et al., 2022). In combination with irCLIP and SHAPE-Map data, the authors characterized a specific CENPC binding region (127-177 nt) on the CCTT RNA and the secondary structure of the region. These pieces of data provide strong evidence that CENPC binding requires a certain secondary structure, most likely the stem-loop that we identified as well. Interestingly, loss of the CENPC binding region 127-177 nt does not affect RNA localization to centromeric regions, but CENPC occupancy on centromeres drops severely (Zhang et al., 2022). This indicates that lncRNA CCTT is playing a role as a shuttle to bring CENPC to the centromere and the CCTT RNA association with the centromere is independent of CENPC protein.

Together with these recent findings, a model emerges in which lncRNAs can interact with CENPC, and promote CENPC localization to specific genome loci. Although the two examples above trigger different localization of CENPC on chromosomes, lncRNA and CENPC signals consistently overlapped suggesting that CENPC functions in an RNP complex. How these different lncRNAs influence CENPC protein localization and function, and whether other lncRNAs influence the dynamic distribution of CENPC during cell division will be intriguing questions for further study.

The apical loop may contribute to the function of centromeric RNAs

In this study, we characterized the involvement of the apical loop of the major stem-loop structure in MinSat RNA function. **(Figure 15)**. Notably, the apical loop fully encompasses the previously described CENPB box sequences. The CENPB box is thought to be an evolutionarily conserved motif across vertebrate centromeres (Alkan, Cardone et al., 2011, Edwards & Murray, 2005, Escudeiro et al., 2019). According to our SHAPE analysis on human and mouse cenRNAs, the high consensus part of the

CENPB box sequence (Iwahara et al., 1998), UUCG at the 5' end and CGGG at the 3' end are constantly base-paired between each other and this potentially contributes to the stem region **(Figure 22)**. On the other hand, the less conserved region of the CENPB box motif based on the previous sequence alignment study is embedded in the open loop region of both the apical loop we identified in human and mouse cenRNAs. This opens the question of whether this arrangement is a common feature between species for all the cenRNAs containing a CENPB box motif. Of note, a 3D cenRNA structure together with its potential protein partners will be even more informative to gain better insight regarding structural conservation. If this is the case, it will provide an explanation for the conservation of centromeric transcription mechanism.

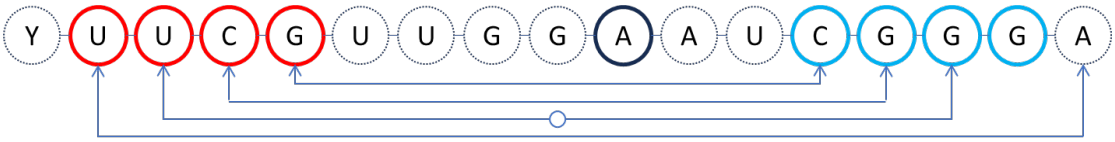


Figure 22. Scheme showing base-paring of essential sites within CENPB box motif. The nucleotides highlighted in bold circles are characterized as essential sites for CENPB box motif. The arrows indicate the base-pairing potential between nucleotides based on SHAPE analysis.

While MinSat Swap mutant leads to chromosome missegregation, the missegregation may not be mediated by MinSat RNA- CENPC protein interaction

CENPC is one of the best-characterized centromeric proteins that binds to RNA molecules. Other than CENPC, there are other centromeric proteins such as the components of CPC complex that could interact with lncRNAs. In biotinylated RNA pull-down assay, CENPC is captured by MinSat forward RNA from mESCs cell lysate. I anticipated that chromosome missegregation phenotype after MinSat forward RNA overexpression mentioned previously was caused by MinSat RNA-CENPC interaction **(Figure 5C, D)**. However, how CENPC protein binds to MinSat RNA remains unclear. Based on **Figure 15**, although the Swap mutant shows a similar chromosome missegregation phenotype as wild-type control, I do not observe a differential binding to CENPC comparing Swap to Deletion or GNRA tetraloop mutants **(Figure 16)**. This can be due to Swap mutant does not have the structure predicted when performing RNA pull-down *in vitro*. The Swap mutant is predicted *in silico*, the RNA secondary structure in real conditions such as working temperature can cause switching of structures and may further lead to differences in protein-RNA affinity.

Although the RNA pull-down result shows an apparent discrepancy to our model that CENPC binds to apical loop of MinSat RNA, it can also be explained if one considers that there are other protein players involved in addition to CENPC alone. Another possibility is that the similar chromosome segregation phenotypes I saw between wild-type and Swap MinSat RNA overexpression were coming from different mechanisms. Other protein partners, such as those belonging to the CPC complex, may also experience disrupted binding following sequence manipulation of CENPB box mutants. To answer this, a proteome study is required to characterize further action of this unique apical loop function by RNA pull-down followed by mass spectrometry analysis. In the future, by conducting comparative analysis of pull-down proteome by all these mutants, a clear model of the entire RNPs may be solved. This will significantly help us understand the molecular function of cenRNAs. Since the apical loop is formed by a conserved CENPB box motif, this can prove of principle to show a shared mechanism of centromere formation across species.

Formation of a functional centromere across species

In 1987, Allshire, R C et al. showed for the first time that fission yeast chromosomes can replicate and function in mouse cells and pointed out that centromere is critical for maintaining foreign chromosomes (Allshire, Cranston et al., 1987). In mammals, human chromosomes can also be maintained in mESCs and even possess the potential to develop into adult mice as Down syndrome models (O'Doherty, Ruf et al., 2005, Shen, Yang et al., 1997, Shinohara, Tomizuka et al., 2001). However, all these approaches were achieved by transplanting (Irradiation microcell-mediated chromosome transfer (XMMCT)) (Dowdy, Scanlon et al., 1990) specific chromosomes alone without coding sequences for centromeric proteins. Interestingly, this suggests that foreign centromeres can function and form regardless of species specificity at the protein level. Along with this, for centromeric protein to recognize centromeric regions, centromeric DNA or RNA is likely to be functional beyond the sequence level. For example, at the DNA level, DNA secondary structures such as i-motif, non-B-form DNA and Dyad symmetric sequences are predicted to form or be enriched in centromeric regions (Thakur, Packiaraj et al., 2021) which are more prone to be recognized by certain centromeric proteins. Since these special structures are common features of centromeric DNA in multiple species, the adoption of these particular structures can offer a potential explanation of how centromeric DNA becomes functional in different systems. In this project, we provide evidence to show that the function and action of centromeric transcripts also utilized similar principles. Especially, the hairpin structure mentioned previously, Dyad symmetric sequence may

play a role in forming this RNA secondary structure. Similar to DNA, RNA can form specific secondary structure to interact conserved centromeric proteins and co-evolve during speciation.

In conclusion, despite the humongous diversity of centromere structures or DNA sequences, the majority of eukaryotic species share a common task of chromosome segregation to maintain genome stability and successful cell division. Understanding the mechanism and evolution of centromeres is crucial to fully explain the conservation of chromosome segregation process. This thesis aims to highlight unique features of mouse cenRNA as well as universal aspects of cenRNA across species and offer a potential explanation for to centromere paradox.

Material and Methods

Cell culture

Mouse E14 ESCs and Fucci cells were grown in Dulbecco's modified Eagle's medium (DMEM) with GlutaMAX (31966-021, Gibco) containing 15% fetal calf serum (Lot P-10397, Cat P30-3302, PAN-Biotech), 2× leukemia inhibitory factor (IGBMC), penicillin-streptomycin (15140122, Gibco), 0.1 mM 2-mercaptoethanol (31350010, Gibco), 3 µM CHIR99021 (13122-25, Cayman Chemical) and 1 µM PD0325901 (13034, Cayman Chemical) on gelatin (P06-20410, PAN-Biotech)-coated plates. MEFs were cultured in DMEM with 10% fetal calf serum and penicillin-streptomycin.

For passaging the cells, cells were first washed two times with DPBS (14040091, Gibco) followed by treatment with 0.25% Trypsin-EDTA (25200-056, Gibco). An equal amount of culture medium was used to neutralize Trypsin-EDTA. After spinning down the cells, 10% of the cells were re-plated for daily maintenance. For mESCs, the medium was replaced daily.

Plasmid transfection

2.5 µg pCAG plasmid DNA containing centromeric sequences and 5 µl Lipofectamine were first incubated separately with 250 µl Opti-MEM (31985062, ThermoFisher) for 5 min. Then, both mixtures were combined at room temperature for 15 min. For each condition, a total of 250,000 mESCs in 500 µl were mixed with the transfection mix. The cells were transfected in suspension in a single well of 6-well plates in a final volume of 2.5 ml. Cells were harvested or fixed 48 hr after transfection.

Total RNA extraction and Real-Time qPCR

Cells were first lysed in 1 ml TRIzol reagent (15596018, Invitrogen). 200 µl of Chloroform (C2432-1L, Sigma Aldrich) was added to the samples and centrifuged at 15000 rpm for 15 min. Afterward, aqueous phases were moved to new tubes and precipitated using an equal amount of isopropanol (149320025, Thermo Scientific). Samples were centrifuged again to pellet down RNAs and washed twice with 80% ethanol. The pellets were then air-dried and dissolved in water. To remove DNA contaminants, each 10 µg of total RNA was treated with 2 units of TURBO™ DNase for 1 hour at 37°C. A second round of TRIzol RNA extraction was performed to remove

DNase and buffer.

1 µg of total RNA was used for cDNA synthesis by applying GoScript™ Reverse Transcriptase (A5001, Promega). For strand-specific reverse transcription, sequence-specific primers were used to discriminate between strands of transcripts. The qPCR primers are listed in Appendix Table 1. The expression level (Ct-value) was determined by using GoTaq qPCR Master Mix (A6002, Promega) on a LightCycler 96 Real-time PCR system (Roche). PCR program was used as follows. The samples were preincubated at 95°C for 5 min, followed by a 2-step amplification at 95°C for 10 second followed by 60°C for 30 second for 45 cycles, then melting at 95°C for 10 second, 65°C for 60 second and 97°C for 1 second and finally cooling at 37°C for 30 sec. The following formula was used to calculate the fold change for relative gene expression.

$$2^{\wedge}(Ct_{\text{gene of interest}} - Ct_{\text{internal control}} - Ct_{\text{experimental control}})$$

Northern blot

20 µg of total RNA was mixed with RNA loading buffer (R1386-5VL, Sigma-Aldrich) and denatured at 65°C for 5 minutes. Then, the samples were run in a single lane in a 2% agarose gel and the RNA integrity was checked by EtBr staining. The gel was then washed twice in 20X SSC for 15 minutes. The transfer system was set up on a Nylon membrane (11MEMP0001, MP Biomedicals) and gel transfer was allowed to proceed overnight. The membrane was next washed in 2X SSC for 5 minutes followed by another wash with buffer containing 20Mm Tris-HCl (pH 8.0) for 25 minutes. Sequentially, the membrane was cross-linked using UV light at 1000 J.

For the preparation of probes, DNA fragments of five copies MinSat were used and radioactivity labeled with ^{32}P using the RadPrime DNA labeling system (18428011, Invitrogen). Afterward, the probes were cleaned up by size selection columns MicroSpin G-25 Columns, cat, no. 27532501, Cytiva.

For hybridization, the membrane was first incubated with ULTRAhyb™ Ultrasensitive Hybridization Buffer (AM8670, Invitrogen) in a hybridization tube to pre-hybridization for 1 hour at 42°C. Then, the prepared radioactive labeled probes were added directly into the hybridization tube while taking care that probes should not be dropped on the membrane. The hybridization step happened in the oven at 42°C overnight. The next day, the membrane was washed twice with 30 ml Wash Buffer 1

(0.1% SDS in 2X SSC) and each wash lasted for 10 min at 42°C. Next, Wash Buffer 2 (0.1% SDS in 0.1X SSC) was used to continue the washes for 20 minutes at 42°C. The membrane was placed in a cassette with radioactive exposure film overnight. Images were finally acquired with a Typhoon™ FLA 7000 (Cytvia).

Knockdown of MinSat RNAs

MinSat RNA knockdown was performed with 2'-O-MOE-modified antisense oligos (Integrated DNA Technologies). The ASOs are listed in Appendix Table 2. 200 µmol of ASOs and Lipofectamine RNAiMAX (13778075, ThermoFisher) were first incubated with 150 µl Opti-MEM in two separate reactions. The two mixtures were combined and incubated at room temperature for 15 min. Then, the transfection mix was incubated with 250,000 mESCs in 500 µl. Cells were plated in a well of a 6-well plate for each condition. Of note, a total amount of 200 µmol (100 µmol each) ASOs were used at a final concentration of 100 µM in the 2 ml final mixture. For the assays, cells were harvested or fixed at 48 hours post-transfection.

Immunofluorescence

The transfected cells were cultured on gelatin-coated coverslips for 48 hr. Cells were fixed with 4% PFA in BRB80 buffer (80 mM K-PIPES, 1 mM $MgCl_2$, 1mM EGTA) for 10 min at room temperature. After fixation, cells were washed twice with DPBS (14190-144, Gibico) and permeabilized with 0.3% Triton-X100 in DPBS for 10 min at room temperature. Blocking was performed with 3% BSA and 0.1% Triton-X100 in DPBS for 20 min. For antibody labeling, primary antibodies were diluted with antibody binding buffer (1% BSA and 0.1 Triton-X100 in DPBS) and incubated overnight at 4°C, washed four times for 5 minutes each with binding buffer and followed by an hour of incubation with secondary antibodies at room temperature. All the antibodies used are listed in **Appendix Table 7**. Then, the samples were washed again four times with a binding buffer and three times with DPBS. Mowiol was used for sample mounting. Images were acquired on a Leica SP8 confocal microscope with a 63x plan-apochromate NA 1.4 oil immersion objective. Pixel size in the final images was typically set to 70-80 nm/pixel and optical sections were acquired every 300 nm.

Chromatin extraction and Western blot analysis

Cells were first synchronized with 100ng/ml KaryoMAX™ Colcemid™ (15210-

040, Gibco) for 3 hours. Next, cells were trypsinized and washed twice with cold DPBS. A million cells were lysed in 100 µl Triton Extraction Buffer (PBS containing 0.5% Triton-X100 and Protease inhibitor mix (80-6501-23, cytiva)) on ice for 10 min. Nuclei were collected by centrifuging at 6,500g for 10 min at 4°C and washed with 50 µl Triton Extraction Buffer. After washing, nuclei were then treated with 20 µl 0.2N HCl overnight at 4°C. The supernatant was transferred and neutralized with 2M NaOH at 1/10 of the volume of the supernatant the next day.

To measure protein concentration, Bradford assay was performed using Protein Assay Dye Reagent (5000006, Bio-Rad). For SDS-PAGE, 1ug of chromatin extract was used for each sample and run with running buffer (0.3% Tris-base, 1.44% Glycine and 0.1% SDS in water). Protein gel was then transferred with transfer buffer (25 mM Tris-HCl (pH 7.6), 192 mM glycine, 0.03% SDS and 20% Methanol) on a PVDF membrane. The membrane was then blocked with 5% skim milk in TBS-Tween 20 (TBST) for an hour. For primary antibody hybridization, the antibodies were diluted in TBST with 1% skim milk and incubated at 4°C overnight. All the antibodies used are listed in **(Appendix Table 6)**. The next day, the membrane was washed four times in TBST followed by secondary antibody hybridization for an hour at room temperature. Afterward, the membrane was washed again four times and developed with SuperSignal™ West Pico PLUS Chemiluminescent Substrate (34580, Thermo Scientific). Finally, the chemiluminescent image was taken with ChemiDoc™ Touch Imaging System (BioRad).

Cell proliferation assay

Cell proliferation was performed by seeding 750,000 cells for transfection. The transfection mix was split into three wells in 6-well plates. Cells were trypsinized and hand-counted using a cell counting chamber (8100104, Hirschmann). The total cell numbers were documented every 24 hours for three days.

Flow cytometry analysis of cell cycle distribution

Cells were first trypsinized and washed with DPBS twice. Cold absolute ethanol was used for cell fixation on ice for 30 min. After fixation, cells were treated with 250 µg/ml (Invitrogen) RNase for 5 min at room temperature. Propidium iodide (PI) was then added to the cell mixture in a final concentration of 50 µg/ml and stained for 30 min on ice. For flow cytometry data analysis, Flowjo 10 software was used. Cell cycle analysis is calculated by fitting the Dean-Jett-Fox model.

3’end biotinylated RNA pulldown assay

RNAs of interest were first *in vitro* transcribed using mMESSAGE mMACHINE™ T7 Transcription Kit (AM1344, Invitrogen). Then, equimolar (20 µmol) of RNAs were heated up at 85°C for 5 mins and labeled with biotin following manufacturer protocol of Pierce™ RNA 3' End Biotinylation Kit (20160, Thermo Scientific). The ligation was performed at 16°C for overnight. After RNA clean-up with RNAClean XP (A63987, Beckman Coulter) beads, we heated the RNA at 70 °C for 10 minutes and cooled it down gradually until it reached room temperature. 10^7 cells were harvested and lysed on ice in Pierce™ IP Lysis Buffer (87787, Thermo Scientific) containing RNasin® Ribonuclease Inhibitor (N2511, Promega) and PMSF (93482-50ML-F, Sigma-Aldrich). Dynabeads™ MyOne™ Streptavidin C1 (65001, Invitrogen) were incubated with biotin-labeled RNA. We next centrifuged cell lysate at 20,000 rcf for 20 minutes and the supernatant was pre-cleared with Dynabeads for 1 hour at 4°C. The RNA-conjugated beads were then added into the cell lysate and incubated at 4°C on a rotor for 3 hours. The RNA-protein complexes were then washed twice with NT2 (50 mM Tris-HCl, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% IGEPAL® CA-630), NT2-middle (50 mM Tris-HCl, 500 mM NaCl, 1 mM $MgCl_2$, 0.05% IGEPAL® CA-630), NT2-high (50 mM Tris-HCl, 1 M NaCl, 1 mM $MgCl_2$, 0.05% IGEPAL® CA-630) and NT2-KSCN (50 mM Tris-HCl, 750 mM KSCN, 1 mM $MgCl_2$, 0.05% IGEPAL® CA-630) buffer. Finally, all samples were boiled in SDS-PAGE sample buffer (#1610747, Biorad) and processed by immunoblotting as described above.

RNA transcription and purification

RNA-encoding DNA templates (possessing the T7 promoter sequence) were either prepared using polymerase chain reaction (PCR; primers ordered from Eurofins Genomics) or ordered as single-stranded DNA templates from Eurofins Genomics. PCR reactions were carried out with the Phusion polymerase (New England Biolabs, NEB) according to manufacturer instructions. RNAs were in vitro transcribed using in-house prepared T7 polymerase. Briefly, 50 microliter transcriptions containing 0.64 µM DNA (supplemented with 0.64 µM T7 top primer for single-stranded DNA templates), 20-80 µM MgCl2, 8 mM of each rNTP, 5% PEG 8000, 1X transcription buffer (5 mM Tris pH 8, 5 mM spermidine, 10 mM DTT), and 0.6 mg of T7 polymerase were incubated at 37°C for 1 hour. Following transcription, the RNAs were purified on urea-denaturing polyacrylamide gels, followed by extraction of the RNA from the gel using the crush and soak method (Morl & Schmelzer, 1993). Extracted RNAs were equilibrated against water and stored at -20°C until further use.

Selective 2’ hydroxyl acylation analyzed by primer extension (SHAPE)

RNAs possessing 5’ (5’-GGAACAACAAGGCCGGAGUACGGCCAAA-3’) and 3’ (5’ AAAAGCAGCGAGUAGCUGCAACAAAAGAAACAACAACAACAAC-3’) SHAPE cassettes and a universal primer binding site were diluted to 1 μM and SHAPE was carried out as previously described (Wilkinson, Merino et al., 2006). Briefly, RNA was snap cooled (95°C for 3 minutes, followed by 5 minutes on ice) in a buffer containing 200 mM NaCl, 100 mM HEPES, pH 8, 0.2 mM EDTA, followed by folding at 37°C (30 min) in a buffer containing 100 mM NaCl, 50 mM HEPES, pH 8, and 16.5 mM MgCl2. RNA was then treated with 1.65, 3.2, or 6.4 mM 1M7 for 5 minutes at 37°C followed by purification using Poly(A) purist magnetic beads (Invitrogen) according to manufacturer instructions. Purified RNAs were reverse transcribed using the Superscript III reverse transcriptase (Invitrogen) and a 5’ Fam labeled primer (Eurofins Genomics) according to manufacturer instructions, followed by purification using magnetic beads. Fam-labeled cDNA fragments were dissolved in HiDi formamide and sequenced using a SeqStudio Fragment Analyzer. Data was analyzed using HiTRACE (Yoon, Kim et al., 2011).

Nuclear magnetic resonance

RNAs were prepared and equilibrated in a buffer containing 15 mM NaCl, 25 mM sodium phosphate pH 6.4. The RNA was snap cooled (95°C for 3 min, 4°C for 5 min) and 1D and 1H NOESY NMR spectra were recorded at 278K on a 1.2 GHz NMR spectrometer equipped with a cryogenic TCI probe.

CatRAPID prediction of RNA binding ability

Amino acid sequence of CenpC protein was submitted to CatRAPID (Livi et al., 2016) web-based prediction algorithm. RNA binding regions were predicted using CatRAPID signature.

Protein expression and purification

DNA encoding for the CENPC protein (amino acids 1 - 906) was ordered as a codon-optimized GBlock from Integrated DNA Technologies (IDT, Leuven, Belgium). Primers (containing NcoI and KpnI enzyme restriction sites) corresponding to the terminal ends of each CatRAPID fragment were ordered from Eurofins Genomics.

PCRs were performed using the Phusion polymerase (NEB) according to manufacturer instructions. CatRAPID PCR fragments were double digested with NcoI (NEB, Frankfurt am Main, Germany) and KpnI (NEB) according to manufacturer instructions, and ligated into a NcoI, KpnI double-digested pETM-11 cloning vector using the NEB Quick Ligase kit. Mini-prepped DNA (obtained from transformations of ligated plasmids into DH10b E. coli competent cells) was sequenced and then transformed into BL21 E. coli competent cells. Colonies were grown up in 1L of lysogeny broth at 37°C and induced with 1M isopropyl beta-D-1-thiogalactopyranoside (IPTG) at 18°C. Cells were lysed using a french press in a buffer containing 500 mM NaCl, 10 mM imidazole, 50 mM Tris, pH 8, 5 mM beta-mercaptoethanol. Protein (possessing an N-terminal His6-tag followed by a tobacco enterovirus (TEV) protease cleavage site) was purified by IMAC against an increasing imidazole concentration gradient. The N-terminal His6-tag was cleaved using in-house-prepared TEV protease, followed by size exclusion chromatography on a Superdex S75 purification column and eluted in buffer containing 300 mM NaCl, 50 mM Tris pH 7.5, and 5 mM DTT. Protein quality was assessed by SDS PAGE.

Binding gel shift assays

RNAs were diluted to 0.250 micromolar and incubated with increasing concentrations of protein as indicated in the figures. Samples were loaded onto a 1% TBE agarose gel supplemented with DNA Stain G (Serva, Heidelberg, Germany) and separated by electrophoresis for 40 minutes at 70V at room temperature. Gels were imaged on a GelDoku imager.

Sequence-specific enrichment using Biotinylated DNA oligos and Streptavidin magnetic beads

Equimolar of biotin-labeled DNA oligos **(Appendix Table 4)** were mixed and dissolved in 500 µl Wash/Binding Buffer (0.5 M NaCl, 20 mM Tris-HCl pH7.5. 1mM EDTA) at a final concentration of 8 pmol/µl. Next, for one sample, 125 µl of Streptavidin Magnetic Beads (S1420S, NEB) were first removed from the buffer applying a magnetic stand and washed once with 100 µl Wash/Binding Buffer. 25 µl of probes mix was incubated with beads at room temperature for 5 minutes with occasional vortex by hand. The beads conjugated with DNA oligos were then washed twice with 100 µl Wash/Binding Buffer.

For hybridization, 100 µg total RNA was extracted and dissolved in 50 µl

Wash/Binding Buffer. To denature the RNA molecules, RNA samples were first heated at 65°C for 5 minutes and snap-cooled on ice for 3 minutes. The conjugated magnetic beads were mixed thoroughly with RNA samples and incubated at 40°C on a thermomixer for an hour with occasional agitation by hand. After hybridization, the supernatant was discarded and the beads were washed twice with 100 µl Wash/Binding Buffer. 100 µl of cold Low Salt Buffer (0.15 M NaCl, 20 mM Tris-HCl pH 7.5, 1 mM EDTA) was added to wash the beads once. At last, after removing the supernatant the RNA was eluted with 70°C 30 µl prewarmed Elution Buffer (10 Mm Tris-HCl pH 7.5, 1 mM EDTA).

Library preparation for sequence-specific cDNA nanopore sequencing

After enrichment of MinSat RNAs, 10 µl RNA sample was first mixed with 1 µl of 2 µM RT primers (specific to MinSat forward and reverse transcripts) and 1 µl of 10 µM dNTP (N0447, NEB) for reverse transcription reaction. The mixture was then incubated 65°C for 5 minutes and put on ice immediately. Next, 4 µl 5X RT Buffer, 1 µl RNaseOUT (10777019, Life Technologies), 2 µl 10 µM Strand-specific primers against MinSat forward or reverse transcripts and 1 µl Maxima H Minus Reverse Transcriptase (EP0751, ThermoFisher). The sample was incubated at 42°C for 90 minutes, 85°C for 5 minutes and cooled at 4°C.

For selecting full-length cDNA by PCR, the original 20 µl cDNA was split into 4 reactions mentioned here: 5 µl final cDNA product was mixed with 25 µl 2x LongAmp Taq Master Mix (M0287, NEB), 1.5 µl cDNA Primer (SQK-PCS109, Oxford Nanopore) and 18.5 µl nuclease-free water. The PCR program was set as follows: the samples were preincubated (95°C for 30 sec), amplified (95°C for 15 sec followed by annealing at 62°C for 15 sec and extension at 65°C for 8 minutes) for 14 cycles, finally extended at 65°C for 8 minutes and held at 4°C. PCR products were treated with 1 µl Exonuclease I (NEB, M0293) at 37°C for 15 minutes and heat inactivated at 80°C for 15 minutes. DNA samples were combined and cleaned up using 2.5 sample volume size of Ampure XP beads (Beckman Coulter, A63882).

For sample loading, 12 µl of DNA libraries were first ligated with 1 µl Rapid Adapter (SQK-PCS109, Oxford Nanopore). The loading steps were performed according to standard instructions from Oxford Nanopore.

Statistical analyses

Prism (GraphPad Software, Boston, USA) and ‘R’ were used for statistical analysis. Statistical tests were performed as indicated in the figure legends. For most experiments, we used a paired t-test for statistical evaluation, unless otherwise stated in the Figure Legends. For multiple treatments in immunoblotting experiments, we used the ‘multcomp’ package in ‘R’ on an ANOVA model, evaluating control to treatment conditions.

References

Aldrup-Macdonald ME, Sullivan BA (2014) The past, present, and future of human centromere genomics. Genes (Basel) 5: 33-50

Alkan C, Cardone MF, Catacchio CR, Antonacci F, O'Brien SJ, Ryder OA, Purgato S, Zoli M, Della Valle G, Eichler EE, Ventura M (2011) Genome-wide characterization of centromeric satellites from multiple mammalian genomes. Genome Res 21: 137-45

Allshire RC, Cranston G, Gosden JR, Maule JC, Hastie ND, Fantes PA (1987) A fission yeast chromosome can replicate autonomously in mouse cells. Cell 50: 391-403

Allshire RC, Karpen GH (2008) Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet 9: 923-37

Ananiev EV, Phillips RL, Rines HW (1998) Chromosome-specific molecular organization of maize (Zea mays L.) centromeric regions. Proc Natl Acad Sci U S A 95: 13073-8

Arunkumar G, Baek S, Sturgill D, Bui M, Dalal Y (2022) Oncogenic lncRNAs alter epigenetic memory at a fragile chromosomal site in human cancer cells. Sci Adv 8: eabl5621

Arunkumar G, Melters DP (2020) Centromeric Transcription: A Conserved Swiss-Army Knife. Genes (Basel) 11

Bakhoum SF, Landau DA (2017) Chromosomal Instability as a Driver of Tumor Heterogeneity and Evolution. Cold Spring Harb Perspect Med 7

Balboa D, Iworima DG, Kieffer TJ (2021) Human Pluripotent Stem Cells to Model Islet Defects in Diabetes. Front Endocrinol (Lausanne) 12: 642152

Balzano E, Giunta S (2020) Centromeres under Pressure: Evolutionary Innovation in

Conflict with Conserved Function. Genes (Basel) 11

Barlow PW, Nevin D (1976) Quantitative karyology of some species ofLuzula. Plant Systematics and Evolution 125: 77-86

Belotserkovskii BP, Tornaletti S, D'Souza AD, Hanawalt PC (2018) R-loop generation during transcription: Formation, processing and cellular outcomes. DNA Repair (Amst) 71: 69-81

Bernard P, Maure JF, Partridge JF, Genier S, Javerzat JP, Allshire RC (2001) Requirement of heterochromatin for cohesion at centromeres. Science 294: 2539-42

Bernstein E, Allis CD (2005) RNA meets chromatin. Genes Dev 19: 1635-55

Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M (2015) Transcription of tandemly repetitive DNA: functional roles. Chromosome Res 23: 463-77

Blower MD (2016) Centromeric Transcription Regulates Aurora-B Localization and Activation. Cell Rep 15: 1624-33

Bobkov GOM, Gilbert N, Heun P (2018) Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. J Cell Biol 217: 1957-1972

Bogliotti YS, Wu J, Vilarino M, Okamura D, Soto DA, Zhong C, Sakurai M, Sampaio RV, Suzuki K, Izpisua Belmonte JC, Ross PJ (2018) Efficient derivation of stable primed pluripotent embryonic stem cells from bovine blastocysts. Proc Natl Acad Sci U S A 115: 2090-2095

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: 8709-14

Brogaard K, Xi L, Wang JP, Widom J (2012) A map of nucleosome positions in yeast at base-pair resolution. Nature 486: 496-501

Bury L, Moodie B, Ly J, McKay LS, Miga KH, Cheeseman IM (2020) Alpha-satellite RNA transcripts are repressed by centromere-nucleolus associations. Elife 9

Chan DYL, Moralli D, Khoja S, Monaco ZL (2017) Noncoding Centromeric RNA Expression Impairs Chromosome Stability in Human and Murine Stem Cells. Dis Markers 2017: 7506976

Chen CC, Bowers S, Lipinszki Z, Palladino J, Trusiak S, Bettini E, Rosin L, Przewloka MR, Glover DM, O'Neill RJ, Mellone BG (2015) Establishment of Centromeric Chromatin by the CENP-A Assembly Factor CAL1 Requires FACT-Mediated Transcription. Dev Cell 34: 73-84

Chen YG, Hur S (2022) Cellular origins of dsRNA, their recognition and consequences. Nat Rev Mol Cell Biol 23: 286-301

Clarke L, Carbon J (1980) Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature 287: 504-9

Clarke L, Carbon J (1985) The structure and function of yeast centromeres. Annu Rev Genet 19: 29-55

Consortium CeS (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282: 2012-8

Corless S, Hocker S, Erhardt S (2020) Centromeric RNA and Its Function at and Beyond Centromeric Chromatin. J Mol Biol 432: 4257-4269

Coronado D, Godet M, Bourillot PY, Tapponnier Y, Bernat A, Petit M, Afanassieff M, Markossian S, Malashicheva A, Iacone R, Anastassiadis K, Savatier P (2013) A short G1 phase is an intrinsic determinant of naive embryonic stem cell pluripotency. Stem Cell Res 10: 118-31

Dambacher S, Deng W, Hahn M, Sadic D, Frohlich J, Nuber A, Hoischen C, Diekmann S, Leonhardt H, Schotta G (2012) CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus 3: 101-10

De Coster W, Rademakers R (2023) NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics 39

Diederichs S (2014) The four dimensions of noncoding RNA conservation. Trends Genet 30: 121-3

Dong F, Miller JT, Jackson SA, Wang GL, Ronald PC, Jiang J (1998) Rice (Oryza sativa) centromeric regions consist of complex DNA. Proc Natl Acad Sci U S A 95: 8135-40

Dowdy SF, Scanlon DJ, Fasching CL, Casey G, Stanbridge EJ (1990) Irradiation microcell-mediated chromosome transfer (XMMCT): the generation of specific chromosomal arm deletions. Genes Chromosomes Cancer 2: 318-27

Du Y, Topp CN, Dawe RK (2010) DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. PLoS Genet 6: e1000835

Duda Z, Trusiak S, O'Neill R (2017) Centromere Transcription: Means and Motive. Prog Mol Subcell Biol 56: 257-281

Edwards NS, Murray AW (2005) Identification of xenopus CENP-A and an associated centromeric DNA repeat. Mol Biol Cell 16: 1800-10

Escudeiro A, Adega F, Robinson TJ, Heslop-Harrison JS, Chaves R (2019) Conservation, Divergence, and Functions of Centromeric Satellite DNA Families in the Bovidae. Genome Biol Evol 11: 1152-1165

Falk SJ, Lee J, Sekulic N, Sennett MA, Lee TH, Black BE (2016) CENP-C directs a structural transition of CENP-A nucleosomes mainly through sliding of DNA gyres. Nat Struct Mol Biol 23: 204-208

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: 5071-80

Flemming W (1882) Zellsubstanz, kern und zelltheilung. Vogel,

Foltz DR, Jansen LE, Bailey AO, Yates JR, 3rd, Bassett EA, Wood S, Black BE, Cleveland DW (2009) Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell 137: 472-84

Fukagawa T, Earnshaw WC (2014) The centromere: chromatin foundation for the kinetochore machinery. Dev Cell 30: 496-508

Furuyama S, Biggins S (2007) Centromere identity is specified by a single centromeric nucleosome in budding yeast. Proc Natl Acad Sci U S A 104: 14706-11

Grenfell AW, Heald R, Strzelecka M (2016) Mitotic noncoding RNA processing promotes kinetochore and spindle assembly in Xenopus. J Cell Biol 214: 133-41

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458: 223-7

Hahn M, Dambacher S, Dulev S, Kuznetsova AY, Eck S, Worz S, Sadic D, Schulte M, Mallm JP, Maiser A, Debs P, von Melchner H, Leonhardt H, Schermelleh L, Rohr K, Rippe K, Storchova Z, Schotta G (2013) Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin. Genes Dev 27: 859-72

Hall LE, Mitchell SE, O'Neill RJ (2012) Pericentric and centromeric transcription: a perfect balance required. Chromosome Res 20: 535-46

Hanna JH, Saha K, Jaenisch R (2010) Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell 143: 508-25

Hassold T, Hunt P (2001) To err (meiotically) is human: the genesis of human aneuploidy. Nat Rev Genet 2: 280-91

He S, Valkov E, Cheloufi S, Murn J (2023) The nexus between RNA-binding proteins and their effectors. Nat Rev Genet 24: 276-294

Hegemann JH, Fleig UN (1993) The centromere of budding yeast. Bioessays 15: 451-60

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293: 1098-102

Henikoff S, Henikoff JG (2012) "Point" centromeres of Saccharomyces harbor single centromere-specific nucleosomes. Genetics 190: 1575-7

Henikoff S, Ramachandran S, Krassovsky K, Bryson TD, Codomo CA, Brogaard K, Widom J, Wang JP, Henikoff JG (2014) The budding yeast Centromere DNA Element II wraps a stable Cse4 hemisome in either orientation in vivo. Elife 3: e01861

Heus HA, Pardi A (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. Science 253: 191-4

Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I (2015) Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep 11: 1110-22

Huang H, Li L, Wen K (2021) Interactions between long non-coding RNAs and RNA-binding proteins in cancer (Review). Oncol Rep 46

Ideue T, Cho Y, Nishimura K, Tani T (2014) Involvement of satellite I noncoding RNA in regulation of chromosome segregation. Genes Cells 19: 528-38

Ideue T, Tani T (2020) Centromeric Non-Coding RNAs: Conservation and Diversity in Function. Noncoding RNA 6

International Silkworm Genome C (2008) The genome of a lepidopteran model insect, the silkworm Bombyx mori. Insect Biochem Mol Biol 38: 1036-45

Ippolito MR, Martis V, Martin S, Tijhuis AE, Hong C, Wardenaar R, Dumont M,

Zerbib J, Spierings DCJ, Fachinetti D, Ben-David U, Foijer F, Santaguida S (2021) Gene copy-number changes and chromosomal instability induced by aneuploidy confer resistance to chemotherapy. Dev Cell 56: 2440-2454 e6

Ishikura S, Yoshida K, Hashimoto S, Nakabayashi K, Tsunoda T, Shirasawa S (2021) CENP-B promotes the centromeric localization of ZFAT to control transcription of noncoding RNA. Journal of Biological Chemistry 297

Iwahara J, Kigawa T, Kitagawa K, Masumoto H, Okazaki T, Yokoyama S (1998) A helix-turn-helix structure unit in human centromere protein B (CENP-B). EMBO J 17: 827-37

Jambhekar A, Emerman AB, Schweidenback CT, Blower MD (2014) RNA stimulates Aurora B kinase activity during mitosis. PLoS One 9: e100748

Johnsson P, Lipovich L, Grander D, Morris KV (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim Biophys Acta 1840: 1063-71

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596: 583-589

Kabeche L, Nguyen HD, Buisson R, Zou L (2018) A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. Science 359: 108-114

Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev 19: 489-501

Kapoor TM (2004) Chromosome segregation: correcting improper attachment. Curr Biol 14: R1011-3

Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J, Suzuki H, Carninci P, Hayashizaki Y, Wells C, Frith M, Ravasi T, Pang KC, Hallinan J, Mattick J, Hume DA et al. (2005) Antisense transcription in the mammalian transcriptome. Science 309: 1564-6

Kawamoto M, Jouraku A, Toyoda A, Yokoi K, Minakuchi Y, Katsuma S, Fujiyama A, Kiuchi T, Yamamoto K, Shimada T (2019) High-quality genome assembly of the silkworm, Bombyx mori. Insect Biochem Mol Biol 107: 53-62

Krassovsky K, Henikoff JG, Henikoff S (2012) Tripartite organization of centromeric chromatin in budding yeast. Proc Natl Acad Sci U S A 109: 243-8

Lehner B, Williams G, Campbell RD, Sanderson CM (2002) Antisense transcripts in the human genome. Trends Genet 18: 63-5

Ling YH, Yuen KWY (2019) Point centromere activity requires an optimal level of centromeric noncoding RNA. Proc Natl Acad Sci U S A 116: 6270-6279

Liu Q, Liu Y, Shi Q, Su H, Wang C, Birchler JA, Han F (2021) Emerging roles of centromeric RNAs in centromere formation and function. Genes Genomics 43: 217-226

Liu Q, Yi C, Zhang Z, Su H, Liu C, Huang Y, Li W, Hu X, Liu C, Birchler JA, Liu Y, Han F (2023) Non-B-form DNA tends to form in centromeric regions and has undergone changes in polyploid oat subgenomes. Proc Natl Acad Sci U S A 120: e2211683120

Livi CM, Klus P, Delli Ponti R, Tartaglia GG (2016) catRAPID signature: identification of ribonucleoproteins and RNA-binding regions. Bioinformatics 32: 773-5

Lo AW, Craig JM, Saffery R, Kalitsis P, Irvine DV, Earle E, Magliano DJ, Choo KH

(2001) A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. EMBO J 20: 2087-96

Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, Davidovich C, Gooding AR, Goodrich KJ, Mattick JS, Mesirov JP, Cech TR, Chang HY (2016) RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell 165: 1267-1279

Lukow DA, Sausville EL, Suri P, Chunduri NK, Wieland A, Leu J, Smith JC, Girish V, Kumar AA, Kendall J, Wang Z, Storchova Z, Sheltzer JM (2021) Chromosomal instability accelerates the evolution of resistance to anti-cancer therapies. Dev Cell 56: 2427-2439 e4

Ly P, Brunner SF, Shoshani O, Kim DH, Lan W, Pyntikova T, Flanagan AM, Behjati S, Page DC, Campbell PJ, Cleveland DW (2019) Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. Nat Genet 51: 705-715

Mallm JP, Rippe K (2015) Aurora Kinase B Regulates Telomerase Activity via a Centromeric RNA in Stem Cells. Cell Rep 11: 1667-78

Mandal SS (2017) Gene Regulation, Epigenetics and Hormone Signaling. John Wiley & Sons,

Mandal SS, Cho H, Kim S, Cabane K, Reinberg D (2002) FCP1, a phosphatase specific for the heptapeptide repeat of the largest subunit of RNA polymerase II, stimulates transcription elongation. Mol Cell Biol 22: 7543-52

Mantel C, Guo Y, Lee MR, Kim MK, Han MK, Shibayama H, Fukuda S, Yoder MC, Pelus LM, Kim KS, Broxmeyer HE (2007) Checkpoint-apoptosis uncoupling in human and mouse embryonic stem cells: a source of karyotpic instability. Blood 109: 4518-27

Marston AL (2014) Chromosome segregation in budding yeast: sister chromatid

cohesion and related mechanisms. Genetics 196: 31-63

Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J Cell Biol 109: 1963-73

Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, Chen R, Dean C, Dinger ME, Fitzgerald KA, Gingeras TR, Guttman M, Hirose T, Huarte M, Johnson R, Kanduri C, Kapranov P, Lawrence JB, Lee JT, Mendell JT et al. (2023) Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol 24: 430-447

McAinsh AD, Meraldi P (2011) The CCAN complex: linking centromere specification to control of kinetochore-microtubule dynamics. Semin Cell Dev Biol 22: 946-52

McCown PJ, Wang MC, Jaeger L, Brown JA (2019) Secondary Structural Model of Human MALAT1 Reveals Multiple Structure-Function Relationships. Int J Mol Sci 20

Meluh PB, Yang P, Glowczewski L, Koshland D, Smith MM (1998) Cse4p is a component of the core centromere of Saccharomyces cerevisiae. Cell 94: 607-13

Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc 127: 4223-31

Mishra PK, Chakraborty A, Yeh E, Feng W, Bloom KS, Basrai MA (2021) R-loops at centromeric chromatin contribute to defects in kinetochore integrity and chromosomal instability in budding yeast. Mol Biol Cell 32: 74-89

Moree B, Meyer CB, Fuller CJ, Straight AF (2011) CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. J Cell Biol 194: 855-71

Morl M, Schmelzer C (1993) A simple method for isolation of intact RNA from dried

polyacrylamide gels. Nucleic Acids Res 21: 2016

Musacchio A, Salmon ED (2007) The spindle-assembly checkpoint in space and time. Nat Rev Mol Cell Biol 8: 379-93

Nakaseko Y, Adachi Y, Funahashi S, Niwa O, Yanagida M (1986) Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. EMBO J 5: 1011-21

Nakatani T, Lin J, Ji F, Ettinger A, Pontabry J, Tokoro M, Altamirano-Pacheco L, Fiorentino J, Mahammadov E, Hatano Y, Van Rechem C, Chakraborty D, Ruiz-Morales ER, Arguello Pascualli PY, Scialdone A, Yamagata K, Whetstine JR, Sadreyev RI, Torres-Padilla ME (2022) DNA replication fork speed underlies cell fate changes and promotes reprogramming. Nat Genet 54: 318-327

Neumann P, Navratilova A, Schroeder-Reiter E, Koblizkova A, Steinbauerova V, Chocholova E, Novak P, Wanner G, Macas J (2012) Stretching the rules: monocentric chromosomes with multiple centromere domains. PLoS Genet 8: e1002777

Noviello TMR, Di Liddo A, Ventola GM, Spagnuolo A, D'Aniello S, Ceccarelli M, Cerulo L (2018) Detection of long non-coding RNA homology, a comparative study on alignment and alignment-free metrics. BMC Bioinformatics 19: 407

O'Doherty A, Ruf S, Mulligan C, Hildreth V, Errington ML, Cooke S, Sesay A, Modino S, Vanes L, Hernandez D, Linehan JM, Sharpe PT, Brandner S, Bliss TV, Henderson DJ, Nizetic D, Tybulewicz VL, Fisher EM (2005) An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. Science 309: 2033-7

Ohkuni K, Kitagawa K (2012) Role of transcription at centromeres in budding yeast. Transcription 3: 193-7

Ottoz DSM, Berchowitz LE (2020) The role of disorder in RNA binding affinity and

specificity. Open Biol 10: 200328

Pegueroles C, Iraola-Guzman S, Chorostecki U, Ksiezopolska E, Saus E, Gabaldon T (2019) Transcriptomic analyses reveal groups of co-expressed, syntenic lncRNAs in four species of the genus Caenorhabditis. RNA Biol 16: 320-329

Perea-Resa C, Blower MD (2018) Centromere Biology: Transcription Goes on Stage. Mol Cell Biol 38

Pesenti ME, Raisch T, Conti D, Walstein K, Hoffmann I, Vogt D, Prumbaum D, Vetter IR, Raunser S, Musacchio A (2022) Structure of the human inner kinetochore CCAN complex and its significance for human centromere organization. Mol Cell 82: 2113-2131 e8

Ponjavic J, Ponting CP, Lunter G (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res 17: 556-65

Quenet D, Dalal Y (2014) A long non-coding RNA is required for targeting centromeric protein A to the human centromere. Elife 3: e03254

Raisch T, Ciossani G, d'Amico E, Cmentowski V, Carmignani S, Maffini S, Merino F, Wohlgemuth S, Vetter IR, Raunser S, Musacchio A (2022) Structure of the RZZ complex and molecular basis of Spindly-driven corona assembly at human kinetochores. EMBO J 41: e110411

Ray D, Laverty KU, Jolma A, Nie K, Samson R, Pour SE, Tam CL, von Krosigk N, Nabeel-Shah S, Albu M, Zheng H, Perron G, Lee H, Najafabadi H, Blencowe B, Greenblatt J, Morris Q, Hughes TR (2023) RNA-binding proteins that lack canonical RNA-binding domains are rarely sequence-specific. Sci Rep 13: 5238

Rieder CL (1979) Ribonucleoprotein staining of centrioles and kinetochores in newt lung cell spindles. J Cell Biol 80: 1-9

Rosic S, Kohler F, Erhardt S (2014) Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. J Cell Biol 207: 335-49

Round EK, Flowers SK, Richards EJ (1997) Arabidopsis thaliana centromere regions: genetic map positions and repetitive DNA structure. Genome Res 7: 1045-53

Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, Osawa H, Kashiwagi S, Fukami K, Miyata T, Miyoshi H, Imamura T, Ogawa M, Masai H, Miyawaki A (2008) Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. Cell 132: 487-98

Sakuno T, Tada K, Watanabe Y (2009) Kinetochore geometry defined by cohesion within the centromere. Nature 458: 852-8

Sanchez de Groot N, Armaos A, Grana-Montes R, Alriquet M, Calloni G, Vabulas RM, Tartaglia GG (2019) RNA structure drives interaction with proteins. Nat Commun 10: 3246

Santaguida S, Musacchio A (2009) The life and miracles of kinetochores. EMBO J 28: 2511-31

Shen MH, Yang J, Loupart ML, Smith A, Brown W (1997) Human mini-chromosomes in mouse embryonal stem cells. Hum Mol Genet 6: 1375-82

Shinohara T, Tomizuka K, Miyabara S, Takehara S, Kazuki Y, Inoue J, Katoh M, Nakane H, Iino A, Ohguma A, Ikegami S, Inokuchi K, Ishida I, Reeves RH, Oshimura M (2001) Mice containing a human chromosome 21 model behavioral impairment and cardiac anomalies of Down's syndrome. Hum Mol Genet 10: 1163-75

Smurova K, De Wulf P (2018) Centromere and Pericentromere Transcription: Roles and Regulation ... in Sickness and in Health. Front Genet 9: 674

Steiner FA, Henikoff S (2014) Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. Elife 3: e02025

Sun X, Wahlstrom J, Karpen G (1997) Molecular structure of a functional Drosophila centromere. Cell 91: 1007-19

Suntronpong A, Kugou K, Masumoto H, Srikulnath K, Ohshima K, Hirai H, Koga A (2016) CENP-B box, a nucleotide motif involved in centromere formation, occurs in a New World monkey. Biol Lett 12: 20150817

Sweta S, Dudnakova T, Sudheer S, Baker AH, Bhushan R (2019) Importance of Long Non-coding RNAs in the Development and Disease of Skeletal Muscle and Cardiovascular Lineages. Front Cell Dev Biol 7: 228

Szczesniak MW, Kubiak MR, Wanowska E, Makalowska I (2021) Comparative genomics in the search for conserved long noncoding RNAs. Essays Biochem 65: 741-749

Tanaka K, Hirota T (2016) Chromosomal instability: A common feature and a therapeutic target of cancer. Biochim Biophys Acta 1866: 64-75

Tatchell K, Van Holde KE (1979) Nucleosome reconstitution: effect of DNA length on nuclesome structure. Biochemistry 18: 2871-80

Thakur J, Packiaraj J, Henikoff S (2021) Sequence, Chromatin and Evolution of Satellite DNA. Int J Mol Sci 22

Thompson SL, Compton DA (2011) Chromosome missegregation in human cells arises through specific types of kinetochore-microtubule attachment errors. Proc Natl Acad Sci U S A 108: 17974-8

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: 15986-91

Tsuiko O, Jatsenko T, Parameswaran Grace LK, Kurg A, Vermeesch JR, Lanner F, Altmae S, Salumets A (2019) A speculative outlook on embryonic aneuploidy: Can molecular pathways be involved? Dev Biol 447: 3-13

Ulitsky I, Bartel DP (2013) lincRNAs: genomics, evolution, and mechanisms. Cell 154: 26-46

UniProt C (2023) UniProt: the Universal Protein Knowledgebase in 2023. Nucleic

Acids Res 51: D523-D531

Vader G, Cruijsen CW, van Harn T, Vromans MJ, Medema RH, Lens SM (2007) The chromosomal passenger complex controls spindle checkpoint function independent from its role in correcting microtubule kinetochore interactions. Mol Biol Cell 18: 4553-64

Valgardsdottir R, Chiodi I, Giordano M, Rossi A, Bazzini S, Ghigna C, Riva S, Biamonti G (2008) Transcription of Satellite III non-coding RNAs is a general stress response in human cells. Nucleic Acids Res 36: 423-34

Vissel B, Nagy A, Choo KH (1992) A satellite III sequence shared by human chromosomes 13, 14, and 21 that is contiguous with alpha satellite DNA. Cytogenet Cell Genet 61: 81-6

Wang J, Liu X, Dou Z, Chen L, Jiang H, Fu C, Fu G, Liu D, Zhang J, Zhu T, Fang J, Zang J, Cheng J, Teng M, Ding X, Yao X (2014) Mitotic regulator Mis18beta interacts with and specifies the centromeric assembly of molecular chaperone holliday junction recognition protein (HJURP). J Biol Chem 289: 8326-36

Webster A, Schuh M (2017) Mechanisms of Aneuploidy in Human Eggs. Trends Cell Biol 27: 55-68

Weinreb C, Riesselman AJ, Ingraham JB, Gross T, Sander C, Marks DS (2016) 3D RNA and Functional Interactions from Evolutionary Couplings. Cell 165: 963-75

Wilkinson KA, Merino EJ, Weeks KM (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc 1: 1610-6

Wong LH, Brettingham-Moore KH, Chan L, Quach JM, Anderson MA, Northrop EL, Hannan R, Saffery R, Shaw ML, Williams E, Choo KH (2007) Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res 17: 1146-60

Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A et al. (2002) The genome sequence of Schizosaccharomyces pombe. Nature 415: 871-80

Wu T, Lane SIR, Morgan SL, Tang F, Jones KT (2021) Loss of centromeric RNA activates the spindle assembly checkpoint in mammalian female meiosis I. J Cell Biol 220

Yam CQX, Lim HH, Surana U (2022) DNA damage checkpoint execution and the rules of its disengagement. Front Cell Dev Biol 10: 1020643

Yamagishi Y, Honda T, Tanno Y, Watanabe Y (2010) Two histone marks establish the inner centromere and chromosome bi-orientation. Science 330: 239-43

Yeo G, Burge CB (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 11: 377-94

Yi Q, Chen Q, Liang C, Yan H, Zhang Z, Xiang X, Zhang M, Qi F, Zhou L, Wang F (2018) HP1 links centromeric heterochromatin to centromere cohesion in mammals. EMBO Rep 19

Yoon S, Kim J, Hum J, Kim H, Park S, Kladwang W, Das R (2011) HiTRACE: high-throughput robust analysis for capillary electrophoresis. Bioinformatics 27: 1798-805

Zeke A, Schad E, Horvath T, Abukhairan R, Szabo B, Tantos A (2022) Deep structural insights into RNA-binding disordered protein regions. Wiley Interdiscip Rev RNA 13: e1714

Zhang C, Wang D, Hao Y, Wu S, Luo J, Xue Y, Wang D, Li G, Liu L, Shao C, Li H, Yuan J, Zhu M, Fu XD, Yang X, Chen R, Teng Y (2022) LncRNA CCTT-mediated RNA-DNA and RNA-protein interactions facilitate the recruitment of CENP-C to centromeric DNA during kinetochore assembly. Mol Cell 82: 4018-

4032 e9

Zickler D, Kleckner N (1999) Meiotic chromosomes: integrating structure and function. Annu Rev Genet 33: 603-754

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-15

Appendix

Figure 1. Plasmid map of pCAGpuro

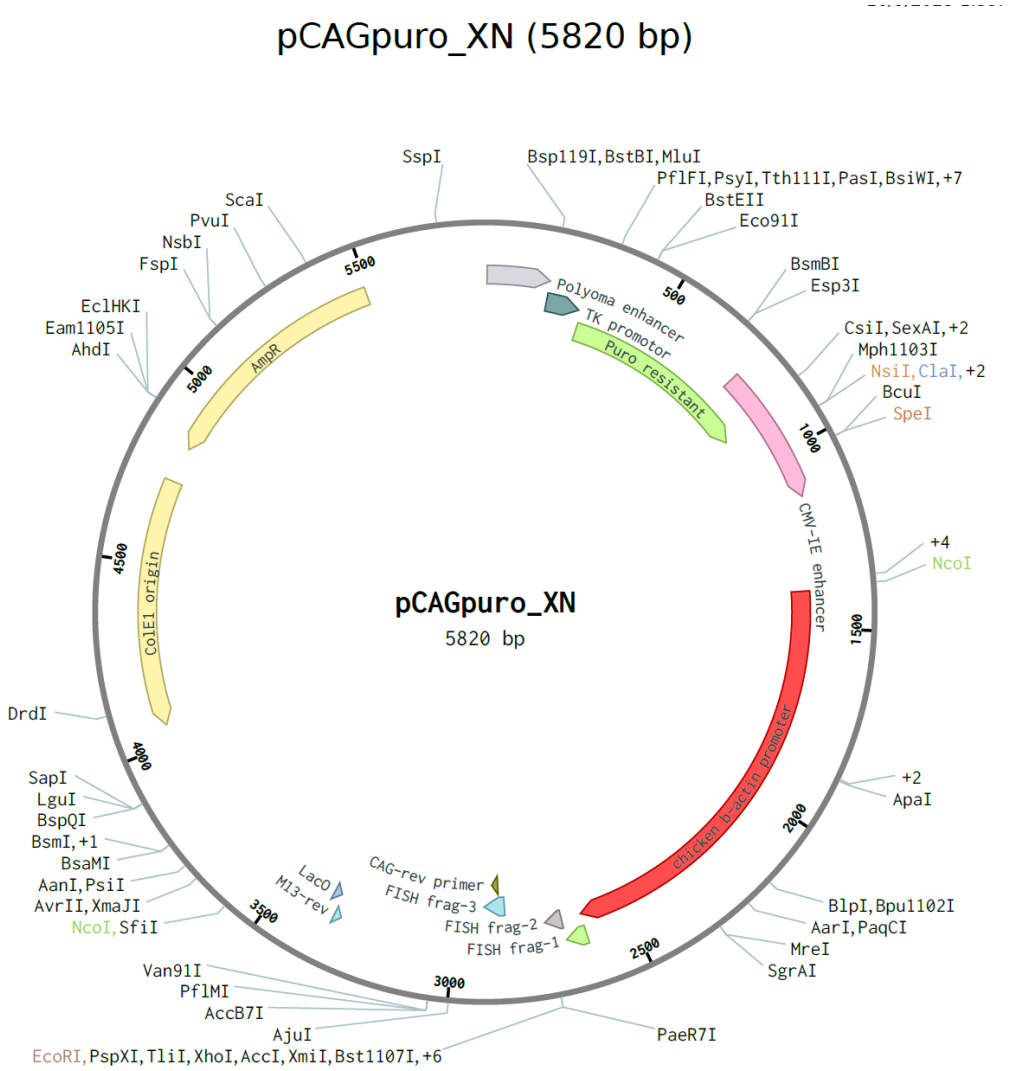


Figure 2. Plasmid map of pETM-11

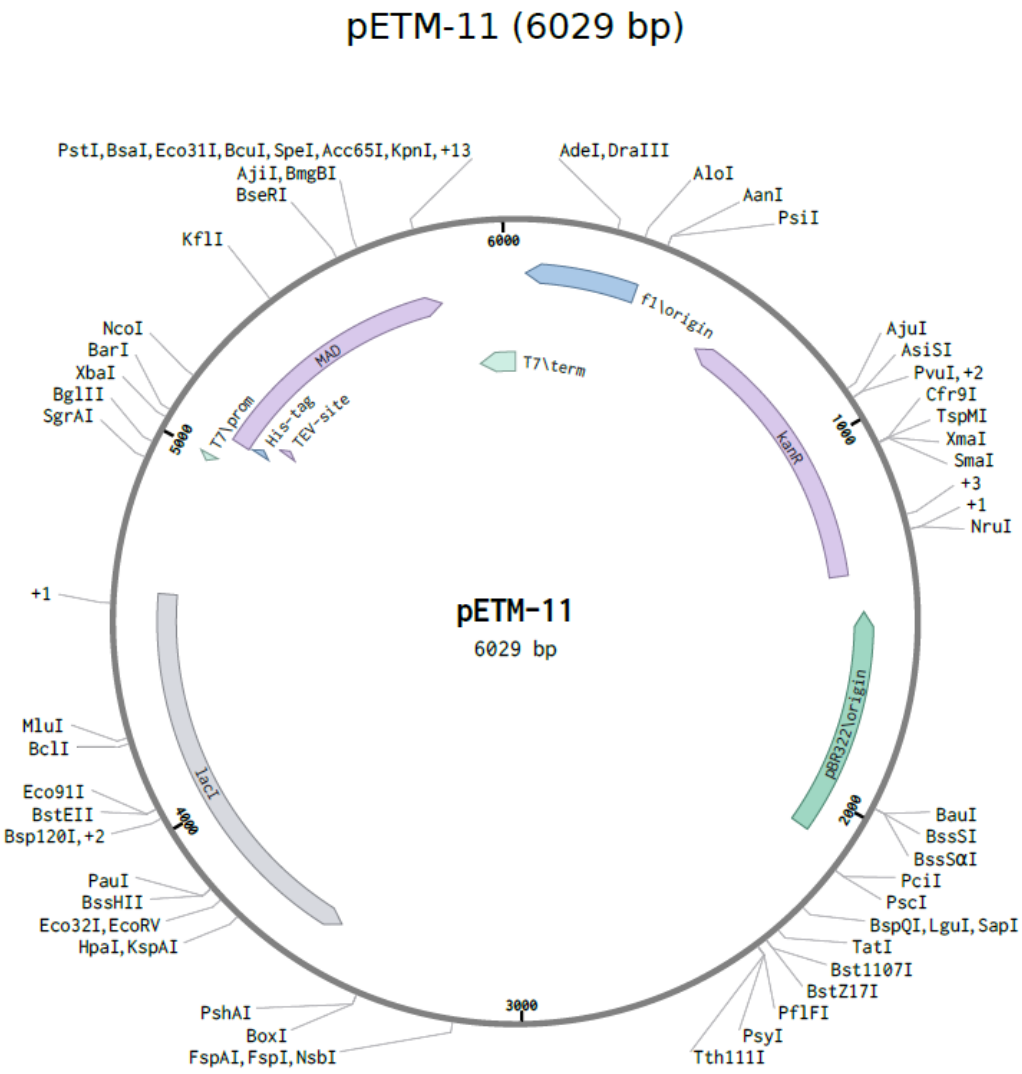


Table 1. Strand-specific and non-strand-specific qPCR primers were applied for quantification of MinSat expression level.

Primers	**Sequences (5'→ 3')**
MinSat forward-specific-RT	CTGATCTAGAGGTACCGGATCCGACTCGAGTCGACATCGTCATCTAATATGTTCTACAGTGTGG
MinSat reverse-specific-RT	CTGATCTAGAGGTACCGGATCCGACTCGAGTCGACATCGCATGGAAAATGATAAAAACCACACT
ACTB gene-specific-RT	TGTTTGTGTAAGGTAAGGTGTGCACTTTTA
F-Minsat forward strand	GTGAGTTACACTGAAAAACACATTCGTTG
F-Minsat reverse strand	TGATATACACTGTTCTACAAAGCCCG
R-MinSat adapter sequence	AGAGGTACCGGATCCGACTCGAGTCGACATC

Table 2. siRNA predesigned by Sigma-Aldrich for CENPC knockdown.

siRNAs	**Strand**	**Sequences (5'→ 3')**
siCENPC-mouse-1 (SASI_Mm02_00312214)	Sense	CACAGAAGAUAUUGCCGGU[dT][dT]
	Antisense	ACCGGCAAUAUCUUCUGUG[dT][dT]
siCENPC-mouse-2 (SASI_Mm02_00312215)	Sense	GACAUCAUCAGAAAUCACU[dT][dT]
	Antisense	AGUGAUUUCUGAUGAUGUC[dT][dT]

Table 3. 2'-MOE antisense oligos for MinSat RNA knockdown.

Antisense oligos	**Sequences (5'→ 3')**	**Reference**
Scramble ASO	mU*mC*mA*mC*mC*T* T*C*A* C*C*C* T*C*T*mC*mC*mA*mC*mU	(Ideue et al., 2014)
MinSat forward ASO	mU*mG*mU*mU*mU*T*T*C*A*G*T*G* T*A*A*mC*mU*mC*mA*mC	(Ideue et al., 2014)
MinSat reverse ASO	mU*mG*mU*mA*mG*A*A*C*A*G*T*G *T*A*T*A*T*C*A*mA*mU*mG*mA*mG	(Mallm & Rippe, 2015)

Table 4. 3' end biotinylated DNA oligos to enrich MinSat RNAs

DNA oligos	Sequences (5'→ 3'[Biotin])
MinSat-35-biotin	TCAGTGTAACTCACTCATCT
MinSat-74-biotin	ACACTGTTCTACAAATCCCG
MinSat-99-biotin	GTTTCTCATTGTAACTCATT
MinSat-R-1-biotin	AAAAACCACACTGTAGAACA
MinSat-R-2-biotin	AAAACACATTCGTTGGAAAC
MinSat-R-3-biotin	CATGGAAAATGATAAAAACC

Table 5. RT primers used for Nanopore cDNA sequencing. Red colored sequence represent adapter sequence applied from Oxford Nanopore protocol

Primers	Sequences (5'→ 3')
Specific RT-MinSat (+)	ACTTGCCTGTCGCTCTATCTTCCATCTAATATGT TCTACAGTGTGGTTTTTATC
Specific RT-MinSat (-)	ACTTGCCTGTCGCTCTATCTTCGATAAAAACCA CACTGTAGAACATATTAGATG

Table 6. PCR primers used for the amplification of cenRNAs from *Homo sapiens, Xenopus laevis* and *Bos Taurus.*

Primers	Sequences (5'→ 3')
F-S-human-α-Sat-XhoI	ATGCATGCCTCGAGAGGGAATGTCTTCCCATAAAAAC
R-S-human-α-Sat-NotI	ATGCATGCGCGGCCGCGTCTACCTTTTATTTGAATTCCCG
F-AS-human-α-Sat-XhoI	ATGCATGCCTCGAGGTCTACCTTTTATTTGAATTCCCG
R-AS-human-α-Sat-NotI	ATGCATGCGCGGCCGCAGGGAATGTCTTCCCATAAAAAC
F-S-bovine-Sat1.723-XhoI	ATGCATGCCTCGAGCTGTGCTAGGGAGCCCAAAC
R-S-bovine-Sat1.723-NotI	ATGCATGCGCGGCCGCAGTGAGCTCGGGCAGTTACT

F-AS-bovine-Sat1.723-XhoI	ATGCATGCCTCGAGAGTGAGCTCGGGCAGTTACT
R-AS-bovine-Sat1.723-NotI	ATGCATGCGCGGCCGCCTGTGCTAGGGAGCCCAAAC
F-S-Xenopus-fcr1-XhoI	ATGCATGCCTCGAGTGGCAGAGTGCTTTTGCAAG
R-S-Xenopus-fcr1-NotI	ATGCATGCGCGGCCGCAAATGCATTATAAGTCTATGGGAG
F-AS-Xenopus-fcr1-XhoI	ATGCATGCCTCGAGAAATGCATTATAAGTCTATGGGAG
R-AS-Xenopus-fcr1-NotI	ATGCATGCGCGGCCGCTGGCAGAGTGCTTTTGCAAG

Table 7. Antibodies used for immunofluorescence.

Proteins	**Antibodies catalog number**	**Dilution**
CENPA	Cell signaling (#2048)	1:1000 (WB), 1:500 (IF)
INCENP	Abcam (ab12183)	1:1000 (WB)
CENPC	Abcam (ab193666)	1:1000 (WB)
CREST	Europa Bioproducts (FZ90C-CS1058)	1:2000 (IF)
GAPDH	Millipore (MAB374)	1:1000 (WB)
histone H3	Abcam (ab1791)	1:10000 (WB)
α-Tubulin (DM1A)	Sigma (T9026)	1:1000 (IF)

WB: Western Blot IF: Immunofluorescence

Contributions

In this thesis, some parts of the research were done in collaborations. For the biochemical and structural biology analysis, Dr. Alisha Jones was in charge of doing in vitro gel shift assay (Figure 11D and 13D), SHAPE analysis (Figure 12A,B,D and 13B,C) and NMR (Figure 12C). For Oxford Nanopore sequencing analysis, Dr. Anna Danese helped to perform mapping of MinSat sequences (Figure 18 and 19).

Curriculum Vitae

Name: YUNG-LI CHEN

E-mail: yung-li.chen@riken.jp

■ Education

- **Ludwig-Maximilians-University Munich**

  Sept. 2019 – now

  PhD student, Faculty of Biology

- **Helmholtz Zentrum München**

  Sept. 2019 – now

  PhD student, Institute of Epigenetics and Stem cells (IES), Helmholtz Zentrum München

- **Chang Gung University**

  Sept. 2017 – Aug. 2018

  Master of Science (MS), Graduate Institute of Biomedical Sciences, Division of Microbiology (CGPA: 4.0/4.0)

- **Chang Gung University**

  Sept. 2013 – June 2017

  Bachelor of Science (BS), College of Medicine, Department of Biomedical Sciences (CGPA: 3.41/4.0)

  Certificate Program: Biomedical Sciences Research Scholar Program

■ Research Experiences

- **Institute of Epigenetics and Stem cells (IES), Helmholtz Zentrum München**

  Sept. 2019 – now

  I am currently a fourth-year PhD student working in **Maria-Elena Torres-Padilla**'s lab, working on the project titled "Expression and function analysis of minor satellite RNA during mitosis in mouse embryonic stem cells". I am under co-supervision of both Dr. Andreas Ettinger and Prof. Dr. Maria-Elena Torres-Padilla.

- **Tumor Virus and Cancer Research Lab, Chang Gung University**

  Oct. 2014 –Sept. 2018

  When I was a senior undergraduate student Dr. **Hsin-Pai Li**'s lab, I worked on a project

"Epigenetic aberration and epigenomic editing research in nasopharyngeal cancer" which was supported by **Ministry of Science and Technology Undergraduate Research Fellowship**. (2016 July –2017 Feb)

- Involved in research of hypermethylated tumor suppressor gene CLDN11 in nasopharyngeal carcinoma and performed cell-based migration assay to prove the inhibition of tubulin polymerization can repress cancer cell migration ability.
- Performed co-immunoprecipitation assay and mass spectrometry to identify CDC20B interacting proteins.
- Established a CRISPR/Cas9 based DNA demethylation system (dCas9-TET1) to specifically target and activate a hypermethylated tumor suppressor gene UCHL1 in nasopharyngeal cancer. **Master thesis**: **"CRISPR/Cas9-Based DNA Demethylation Reactivates Tumor Suppressor Gene in Nasopharyngeal Carcinoma"**

- **Chang Gung University International Genetically Machine (iGEM) Team, CGU_Taiwan**
  Dec. 2014 – Oct. 2015
  One of the co-founders and team leaders who organized the first Chang Gung University iGEM team. Forging an interdisciplinary cooperation opportunity for participating in a global synthetic biology competition.
  - **Wet lab supervisor** of the CGU_TAIWAN team who was responsible for the project "Using Toehold Switch to Detect mRNA Biomarkers of Oral Cavity Cancer".
  - Recruiting team members, designing and performing experiments for the team project. All results were presented and published at 2015 iGEM Giant Jamboree in Boston.
- **Dr. Wen-Chang Lin's lab, Institute of Biomedical Science, Academia Sinica – Summer internship**
  July 2014 – Sept. 2014
- **Tumor and Molecular Medicine lab, Chang Gung University – Student trainee**
  Nov. 2013 – June 2014

■ Honors and Awards

- **Excellent work Winner, Annual Research Poster Contest** - Department of Biomedical Sciences
  May 2017
- **Silver Medal Award at International Genetically Machine (iGEM) Competition**
  Sept. 2015

■ Leadership & Extracurricular Activities

- Wet lab supervisor - **International Genetically Machine (iGEM) Team, CGU_Taiwan**
  Dec. 2014 – Oct. 2015

- Leader of Activities Department - **Student Association of Biomedical Sciences Department**
  Sept. 2013 – June 2014

■ Teaching Experiences

- Teaching Assistant - single cell analysis course
  Every Feb. 2020-2023
- Teaching Assistant in Physiology (LS3018) - Dept. of Biomedical Sciences, C.G.U..
  Feb. 2018 – June 2018
- Teaching Assistant in Evolutionary Biology (LS2007) - Dept. of Biomedical Sciences, C.G.U.
  Feb. 2017 – June 2017

■ REFERENCES

Prof. Dr. Maria-Elena Torres-Padilla (PhD thesis supervisor)

Director of the Stem Cell Center, Director of the Institute for Epigenetics and Stem Cells, Group Leader

Helmholtz Center Munich.

Feodor-Lynen-Straße 21, 81375 München

Tel: +49 89 3187-3317

E-mail: torres-padilla@helmholtz-munich.de

Dr. Andreas Ettinger (PhD thesis supervisor)

Head of Microscopy Core, Institute of Epigenetics and Stem cells, Helmholtz Center Munich.

Feodor-Lynen-Straße 21, 81375 München

Tel: +49 89 3187-1504

E-mail: andreas.ettinger@helmholtz-munich.de

Acknowledgments

Time flies! In the blink of an eye, my journey of PhD has come to an end. During this journey, I would like to thank all the people who have supported or accompanied me, especially…

Maria-Elena: Thank you for being supportive throughout my entire PhD. It's really my pleasure to join this lab. The experience I have gained from this lab is fruitful. The attitude and logical training from you are going to be carved into my mind. Besides, I really appreciate the atmosphere and interaction that you create in this lab. This allows me to have a wonderful PhD life working here.

Andreas: Undoubtedly, without you, I will not be in this lab and working on this beautiful project. It's really fun to work with you as you are always the one who encourages me to operate any cool idea. From the training, I did learn a lot from you no matter about microscopy or perspective to basic research. I believe this will benefit me a lot in my future career.

Michael and Jonesy: Many thanks to this collaboration, it not only brings this project to another level of impact but also teaches me a lot about RNA structures. This inspires me a lot and opens a possibility to keep working on RNA biology.

Stephan: Thank you a lot for showing me many different kinds of techniques for example Northern and Southern blot. I really learn a lot from you in terms of molecular biology and trouble shooting. You are always very helpful and being there to answer my question.

Next, I would also like to thank the amazing lab members in the Torres-Padilla lab. **Fede-** Thank you for no matter scientific discussion or humorous conversation with you. Lab with you is never boring. **Tsune-** Thanks for all the help in experiments and it's always fun to talk about comics and Japanese culture with you. **Antoine-** Your passion for science is always what I admire. Thanks for all the inspiring ideas. **Marlies-** Thanks for giving all the advice regarding the job application and encouraging discussion. **Mrinmoy-** Thanks for being the best company during the weekend in the lab and a helpful peer in the lab. **Clara-** Thanks for being positive all the time in small lab. The lab is always full of laughter. **Adam-** Thanks for being helpful for organizing the lab and events to gather all the lab members. **Yicong-** Thanks a lot for the encouragement no matter related to work or daily life. **Ilya-** Thanks for being a cool guy that it always

fun talking to you. **Tamas-** Thank you for bringing your expertise into all our discussion and it's nice chatting with you.

You are the best and coolest labmates I have ever had. As a group of people coming from different places, I did enjoy a lot the company of all of you. The events we have done together are such a relief from our stressful timing. I wish you all good luck and a successful future career.

Finally, I would like to thank my family who is the backup force all the time. Although working in another continent is not really what you wish at the first place, the care and concern I received from all of you have never stopped. Thank you for the mental support and physical care when I went back home. No matter what, the end is finally coming and I want to proudly say to you that I make it.

最後，我想謝謝我的家人，我知道你們一直是我強力的後盾。我知道儘管你們一開始並沒有很支持出國念博士班，也許是充滿太多未知的原因，但我還是感受到滿滿的關心。謝謝你們每次回家總是給予我不管精神上或是身體上的照顧。但無論如何，我的博士班生涯即將告一個段落，我想很驕傲地跟你們說我堅持到最後，我做到了。

我想在這本論文中也留下一些話給過去及未來的自己: 謝謝當時候的你做下了出國念博士班的決定。還記得當時離開台灣前內心的不安與恐懼，那種賭下一切籌碼的惶恐依然歷歷在目，來到德國後，即使在這裡的一切都不簡單，不管是生活上或工作上的挑戰都是既陌生又充滿障礙，但走到現在，你克服了這一切，不管是心理上或做人處事都比之前成熟很多，這絕對是你一生當中最難忘的一段時光，大概也是一輩子成長速度最快的時候，我深信這四年半的時間會讓你的人生變得與別人很不一樣，希望你能牢牢記住在這裡發生的一切，讓它變成未來人生中的養分。帶著這些經歷，希望在未來的職涯或人生中能夠更從容且不畏懼地應對挑戰，相信你可以做得更好更漂亮。