

Identification of chromatin-associated RNAs at human centromeres.

Kelsey Fryer^{1,2}, Charles Limouse¹, and Aaron F. Straight^{1*}

¹Department of Biochemistry, Stanford University School of Medicine

²Department of Genetics, Stanford University School of Medicine

*correspondence:

astraight@stanford.edu

Department of Biochemistry

Beckman 409A

Stanford University School of Medicine

Stanford, CA 94305

Abstract

Centromeres are a specialized chromatin domain that are required for the assembly of the mitotic kinetochore and the accurate segregation of chromosomes. Non-coding RNAs play essential roles in regulating genome organization including at the unique chromatin environment present at human centromeres. We performed Chromatin-Associated RNA sequencing (ChAR-seq) in three different human cell lines to identify and map RNAs associated with centromeric chromatin. Centromere enriched RNAs display distinct contact behaviors across repeat arrays and generally belong to three categories: centromere encoded, nucleolar localized, and highly abundant, broad-binding RNAs. Most centromere encoded RNAs remain locally associated with their transcription locus with the exception of a subset of human satellite RNAs. This work provides a comprehensive identification of centromere bound RNAs that may regulate the organization and activity of the centromere.

Introduction

Faithful chromosome segregation during cell division is essential for maintaining the genome. In eukaryotes, chromosome segregation is accomplished by attachment of each chromosome to the mitotic spindle through a multiprotein complex, called the kinetochore. The kinetochore is assembled on a specialized chromatin domain, known as the centromere. Kinetochore proteins are conserved across species, however the DNA of the centromere is rapidly evolving. The chromatin of the centromere consists of a core domain that is epigenetically defined by the presence of a centromere specific histone H3 variant, CEntromere Protein A (CENP-A).¹ The chromatin of the regions surrounding the centromere, known as pericentromeric heterochromatin, is characterized by high levels of the repressive histone modification, H3K9me3.²

Human centromeric DNA contains AT rich, tandemly repeated ~171bp monomers of DNA sequence collectively known as alpha satellite DNA.^{3,4} Arrays of tandem repeats within the core centromere are further repeated to form higher order repeats (HORs). CENP-A chromatin occupies a subset of the HORs to generate the active centromere where kinetochores will assemble in mitosis. Recent work has shown that in addition to CENP-A occupancy, the active HOR is also reduced in cytosine DNA methylation compared to the rest of the centromeric DNA.⁵ Satellite repeat arrays flanking HORs, known as pericentromeres, are made up of several different repeat array types, including the major sequence component of classical human satellite fractions I, II, III (HSAT1, HSAT2, HSAT3) originally identified using density centrifugation.⁶⁻⁹ Fully sequenced and assembled centromeres from the Telomere-to-Telomere consortium have revealed the complete set of repeat array types that comprise centromeric and pericentromeric regions.¹⁰ We refer to core centromeric repeats as HORs and the CENP-A occupied region as the active HOR. Pericentromeric repeats denote all other repeat arrays outside of HORs.

In addition to CENP-A, DNA methylation, and histone modification, non-coding RNAs have also been proposed to regulate centromere function. Non-coding RNAs are known to regulate chromatin through several mechanisms. Direct interaction of non-coding RNA with DNA can occur when nascent transcripts hybridize with the template DNA strand forming R-loops or when RNAs form triplex structures with double stranded DNA. R-loops and triplexes have been shown to antagonize or promote DNA methylation by Dnmt3b1 respectively.^{11,12} non-coding RNA also plays important roles in modulating histone modification to repress chromatin through direct regulation of histone methyltransferase complexes, including the HUSH complex,^{13,14} polycomb repressive complex 2 (PRC2),¹⁵⁻¹⁸ and SUV39H1.¹⁹⁻²¹ RNA has also been implicated in long range genomic organization through interactions with chromatin proteins, including HP1 α ^{22,23}, YY1,²⁴⁻²⁶ and CTCF.²⁷⁻³¹ However, our understanding of RNA binding by chromatin associated proteins is confounded by potential nonspecific interactions captured by crosslinking dependent immunoprecipitation techniques as well as the broad effects of global RNA degradation leading to debate in the field over whether these interactions are direct RNA binding.³²⁻³⁴ Regardless of the interaction mechanism between RNA and chromatin modifiers, it is clear that RNA plays a critical role in genome regulation and organization, particularly with respect to heterochromatin formation.

RNA has also been implicated in chromatin maintenance at both pericentromeric heterochromatin and the core centromere. In mice and humans, RNase treatment resulted in decreased pericentromeric H3K9me3, heterochromatin protein 1 (HP1), and loss of centromeric localization of the histone methyltransferase responsible for depositing H3K9me3, SUV39H1^{19,35} as well as mislocalization of centromeric proteins.³⁶ Moreover, SUV39H1 directly binds RNA, and that RNA binding activity is necessary for its localization at pericentromeres.^{37–46} RNA interaction with CENP-C, a necessary component of CENP-A assembly machinery, promotes DNA binding and centromere localization in maize and humans implicating RNA in core centromeric chromatin maintenance.^{36,47,48}

Many studies have suggested that RNAs involved in regulating centromeric chromatin may be derived from centromeric repeats. Centromeric transcription and/or centromere derived RNAs have been observed in several species.^{49–59} However, the functional role of centromeric transcription and/or centromeric RNAs is widely debated and differs across organisms. In fission yeast, disruption of the RNAi pathway led to accumulation of centromere derived RNAs and compromised heterochromatin formation.⁶⁰ Deletion of a centromere binding protein in budding yeast led to overexpression of centromeric RNA and decreased expression and centromeric localization of CENP-A.^{61,62} In humans, centromeric transcription has been shown to promote Sgo1 localization to maintain sister chromatin cohesion,⁶³ but is not required for CENP-A incorporation at the core centromere.⁶⁴ RNAs transcribed from centromeric and pericentromeric repeats have been implicated in maintaining pericentromeric heterochromatin as well as CENP-A chromatin, and remain associated with their centromere of origin.^{65–68} Abundance of centromeric RNAs is variable across cell types and anti-correlated with centromere proximity to the nucleolus.⁶⁹ Together these studies suggest that centromere-associated RNAs, including those derived from centromeric repeats, are an important component of centromeric and pericentromeric chromatin.

Our understanding of centromere-associated RNAs has been limited by both the lack of unbiased methods to comprehensively identify chromatin associated RNAs and the complex repeat structure of human centromeres which poses a challenge for mapping and localizing sequencing reads. To identify and map RNA-chromatin contacts in a genome wide manner, we developed chromatin associated RNA sequencing (ChAR-seq).^{70–73} ChAR-seq does not require *a priori* knowledge of RNA identity and captures genome scale RNA identity and location information. Here we use ChAR-seq in conjunction with an alignment independent classification of ambivalent sequences with k-mers (CASK)⁷⁴ to identify centromere interacting RNAs. The recent generation of complete human reference genomes with fully assembled centromeres provided a clear picture of the complexity and diversity of centromeric repeats.¹⁰ Leveraging the T2T assembled centromere sequences we provide a detailed, repeat array level characterization of centromere-associated RNAs. To understand how the RNA-DNA contact patterns at centromeres change across cell types or differentiation states we performed ChAR-seq in the human erythroleukemia cell line, K562, and utilized data from our previous work in human H9 embryonic stem cells (ES) and definitive endoderm (DE) cells differentiated from the ES cells.⁷³ Together this study provides the first comprehensive picture of RNA-centromere interactions in humans.

Results

Identification and classification of centromere-associated and centromere derived RNAs

We characterized RNA-chromatin contacts in K562 cells using our previously described ChAR-seq approach.^{70–73} Briefly, we generated chimeric molecules capturing RNAs in proximity to DNA by ligating a biotinylated bridge oligo to RNA, digesting the genome with DpnII and then ligating the other end of the bridge to the genome. We next converted the RNA component of each of these chimeric molecules into cDNA. We recovered the bridge containing molecules by precipitation with magnetic streptavidin beads and then amplified and sequenced the molecules (**Figure 1A**). To map each RNA-DNA contact we split each read into its RNA and DNA derived component and aligned each to the human genome (GRCh38). To annotate repetitive reads, we also performed an alignment independent classification of all RNA and DNA derived sequences using CASK (Classification of ambivalent sequences using k-mers).⁷⁴ CASK generates and intersects k-mer representations of all repeat types to identify k-mers that are unique to each repeat type or are present in multiple repeat types. To differentiate between the different types of centromeric repeat arrays (i.e. Human Satellite 3 vs Higher Order Repeat), we used CASK to generate k-mer databases of T2T assembled centromeric sequences and repeat array type annotations (**Figure 1B, Supplemental Figure S1A,B**).¹⁰ Reads

are then classified based on the intersection of all possible repeats in which their k-mers were found. For example, a read that contains k-mers that were only found in a single repeat type, such as HSAT3, can be definitively classified as that repeat type. Reads that contain k-mers that are present in multiple repeat types, and cannot be classified as a single repeat, are annotated with an ambivalence group that encompasses all the k-mer derived repeat types (**Figure 1B**). We used CASK to classify both the RNA and DNA derived sequences of ChAR-seq reads in order to examine

Figure 1: CASK annotation of ChAR-seq identifies centromeric RNA-DNA contacts

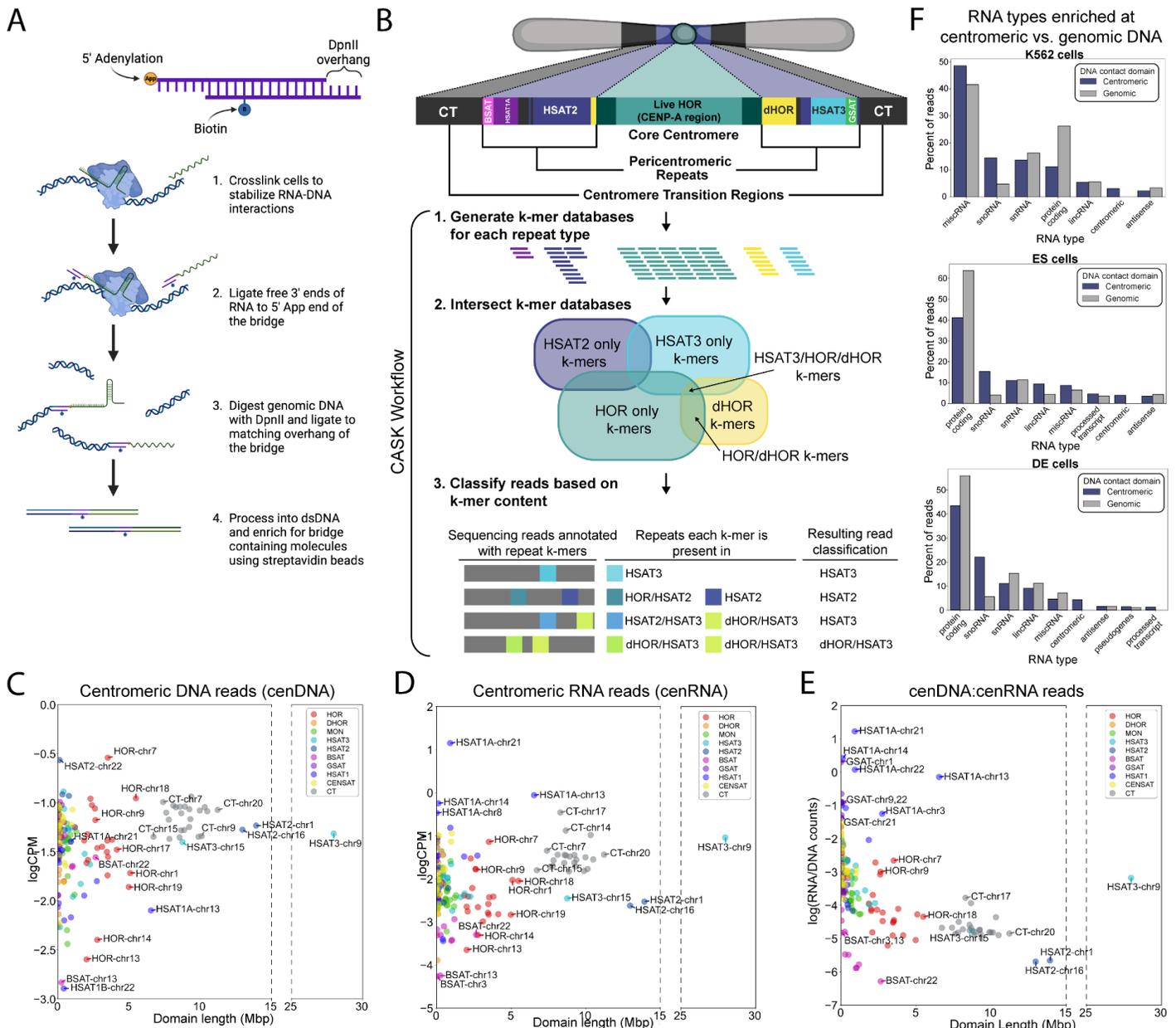


Figure 1: A. Schematic of ChAR-seq bridge molecule (top) and protocol (bottom). **B.** Schematic of Classification of Ambivalent Sequences (CASK). k-mer databases were generated from T2T assembled centromeric repeat sequences and intersected to create groups of k-mers present in each repeat type or combination of repeat types. Reads were then classified into repeat types of ambivalence groups containing multiple repeat types based on their k-mer content. **C&D.** Log of **C.** DNA or **D.** RNA ChAR-seq reads classified by CASK in K562 cells for each repeat domain normalized by the number of DpnII sites in the domain per million reads vs. length of repeat domain in Mbp. Reads classified as rDNA regions were excluded. **E.** Log of ratio of RNA reads to DNA reads classified as each domain normalized by number of DpnII sites in the domain vs. length of domain in Mbp. Reads classified as rDNA regions were excluded. **F.** Proportion of RNA reads enriched at centromeric (purple) or non-centromeric (gray) domains belonging to each RNA type in K562 cells (top), ES cells (middle), and DE cells (bottom).

the contact patterns of 1) RNAs that contact centromeres regardless of their origin and 2) the contact sites across the genome of RNAs that originate from centromeres. To determine how the contact frequency varies

across centromeric repeat domains, we counted the number of reads containing centromere derived DNA sequences (cenDNA) regardless of the identity of the RNA derived portion of the read. Accounting for DpnII site frequency and sequencing depth, we found that the number of cenDNA reads classified as each domain varied widely depending on the specific repeat domain and was not correlated with the length of the domain (**Figure 1C, Supplemental Figure S2 left panels, Supplemental Table S1**). The same phenomenon was observed for centromere derived RNAs - the number of reads classified as centromeric on the RNA portion of the read (cenRNA) varied across domains and was not correlated with the length of the centromeric domain (**Figure 1D, Supplemental Figure S2 right panels, Supplemental Table S1**). For example, the longest repeat array, the human satellite 3 (HSAT3) repeat on chromosome 9, which is 27.9 Mb, contacts roughly the same number of RNAs as the HSAT2 domains that are half the size, and fewer than HORs on chromosome 7 and 18 which are less than one sixth the size. For both cenRNA and cenDNA reads, the relative relationship between most domains remained consistent across all three cell types with the exception of HSAT1A domains. Increased HSAT1A derived RNA reads in K562 cells compared to both ES and DE cells in all replicates may reflect repeat type specific transcriptional regulation (**Figure 1C,D, Supplemental Figure S2**). Overall, these data indicate that both the RNA contact frequency at centromeric domains and the contact frequency of centromere derived RNAs vary depending on the identity of the centromeric domain and not its size or cell type.

We next examined the relationship between the number of RNA contacts and the number of RNAs derived from each domain. In previous studies, we observed a correlation between RNA contact density and ATAC-seq peaks which demarcate regions of open chromatin and tend to be more transcriptionally active.⁷³ To examine how the relationship between the abundance of RNA contacts and chromatin associated RNAs derived from each domain varied across repeat types, we compared cenDNA and cenRNA counts for each repeat array. The ratio of cenRNA:cenDNA counts were moderately correlated with the length of the domain (spearman's rank correlation -0.66) indicating larger domains tend to have fewer RNA reads compared to DNA reads (**Figure 1E, Supplemental Table S1**). Despite the wide range of cenRNA:cenDNA ratios across centromeric repeat arrays, the relative relationship between arrays remained consistent across replicates and cell types (**Supplemental Figure S3**). This is most clearly exemplified by HSAT1A arrays which consistently have the highest RNA:DNA count ratios in all cell types (**Figure 1E, Supplemental Figure S3**). Together, these data depict features of RNA-DNA contact patterns that are shared amongst members repeat families (ie. all HORs) that are not entirely dependent on the length of the domain.

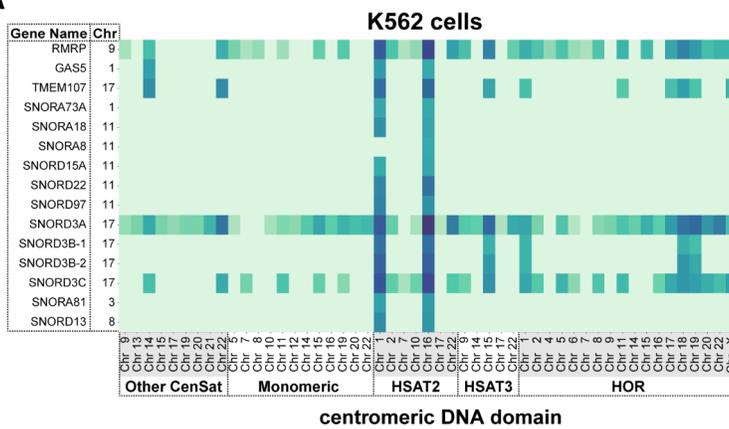
Identification of non-centromere encoded transcripts enriched at centromeric repeats

To identify transcripts encoded by non-centromeric regions that associate with centromeric chromatin, we used CASK to identify ChAR-seq reads in which the genomic DNA was derived from centromeric DNA and genome alignment to annotate the RNA-derived portion of the read. We classified centromeric DNA sequences at the chromosomal level (i.e. all HSAT2 arrays on chromosome 16) and the repeat type level (i.e. the third HSAT2 array on chromosome 16) based on the fully assembled centromeric sequences in the CHM13 T2T genome. The remaining DNA derived portion of the reads that were not classified as centromeric were sorted into 10 kb, 100 kb, and 1000 kb genomic bins. To quantify the enrichment of an RNA (RNAX) at a specific genomic region (DNAX), we designed an enrichment score (E score) that reflects how frequently RNAX contacts DNAX relative to its overall contact frequency across the genome. E scores incorporate the number of DpnII sites in the DNA contact domains as well as its length, to normalize for differences in the number of potential contact sites across different DNA contact domains (**Supplemental Figure S4A**). ChAR-seq also captures non-specific RNA interactions with chromatin as any RNA that is diffusing or being transported within the nucleus will make transient contacts with chromatin. To differentiate between bona fide RNA-DNA interactions and nonspecific interactions, we modeled the background contact behavior of RNAs using a generalized additive model.⁷⁵ Because protein-coding RNAs generally do not make specific contacts with the genome outside of their transcription locus, we trained the background model on all protein coding RNA-chromatin contacts, excluding the source chromosome on which the protein coding gene is present. Predicted enrichment scores for each RNA-DNA contact based on the background model were subtracted from observed enrichment scores to generate a residual score (R score). Contacts with residual scores in the top 5% were considered significantly enriched. (**Supplemental Figure S4B,C**).

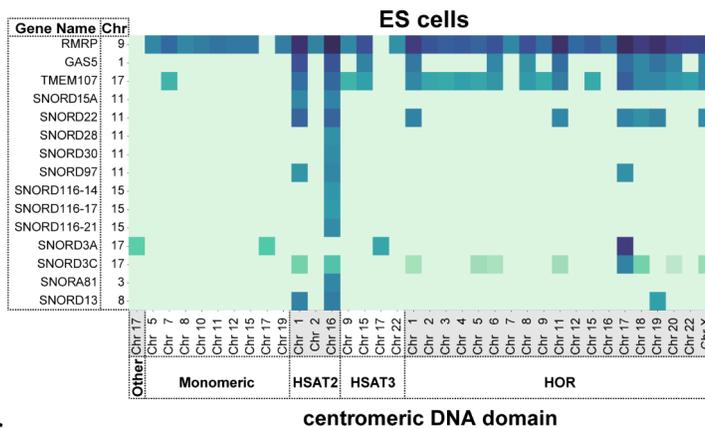
We identified over 40 non-repeat derived RNAs significantly enriched at centromeres belonging to 8 different classes of RNA. The relative proportion of most classes of RNAs, including snRNAs, miscRNAs, lincRNAs, and antisense RNAs, enriched at centromeres compared to all other genomic regions varied across cell types (**Figure 1F**). However, protein-coding RNAs made up a larger proportion of genome enriched RNAs while snoRNAs consistently made up a larger proportion of RNAs at centromeres (**Figure 1F**). Several snoRNAs and snoRNA host genes, including Gas5 and TMEM107, were enriched at centromeric repeats, but most were enriched at only a subset of centromeric repeat domains, specifically the HSAT2 domains on chromosome 1 and 16. Broad enrichment across multiple centromeric repeat types and chromosomes was only observed for two RNAs - SNORD3A (U3) and RMRP (**Figure 2**), both of which have been implicated in processing of other non-coding RNAs, particularly in cleavage of pre-ribosomal RNA.^{76,77} RMRP and SNORD3A enrichment at centromeres was observed in all cell types with the strongest enrichment at HORs and HSAT2 domains (**Figure 2, Supplemental Figure S5**). Despite the genomic proximity of centromeres on acrocentric chromosomes to rDNA clusters, which are contained within the nucleolus when actively transcribing, centromeric enrichment of nucleolar localized RNAs was not limited to acrocentric chromosomes, but may reflect the close proximity of centromeres to the nucleolar periphery^{78,79} and the potential regulatory crosstalk between the two domains.

Figure 2: Nucleolar RNAs enriched at centromeric repeats

A



B



C

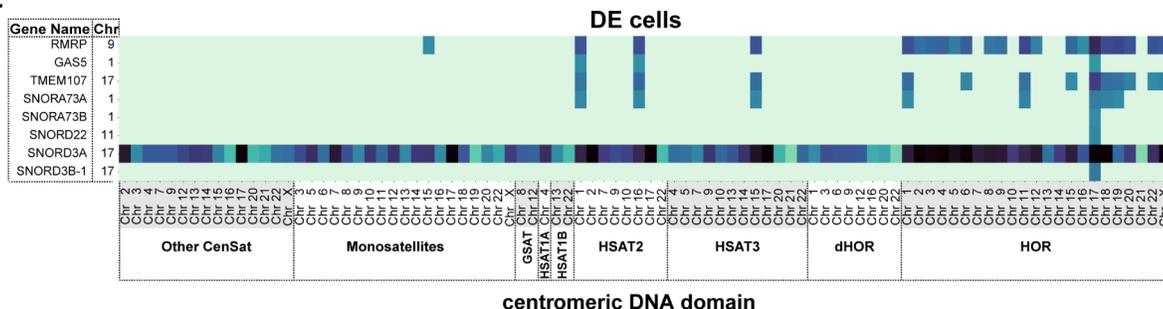
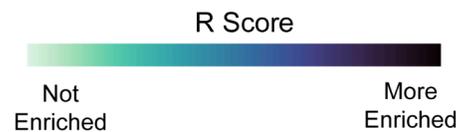
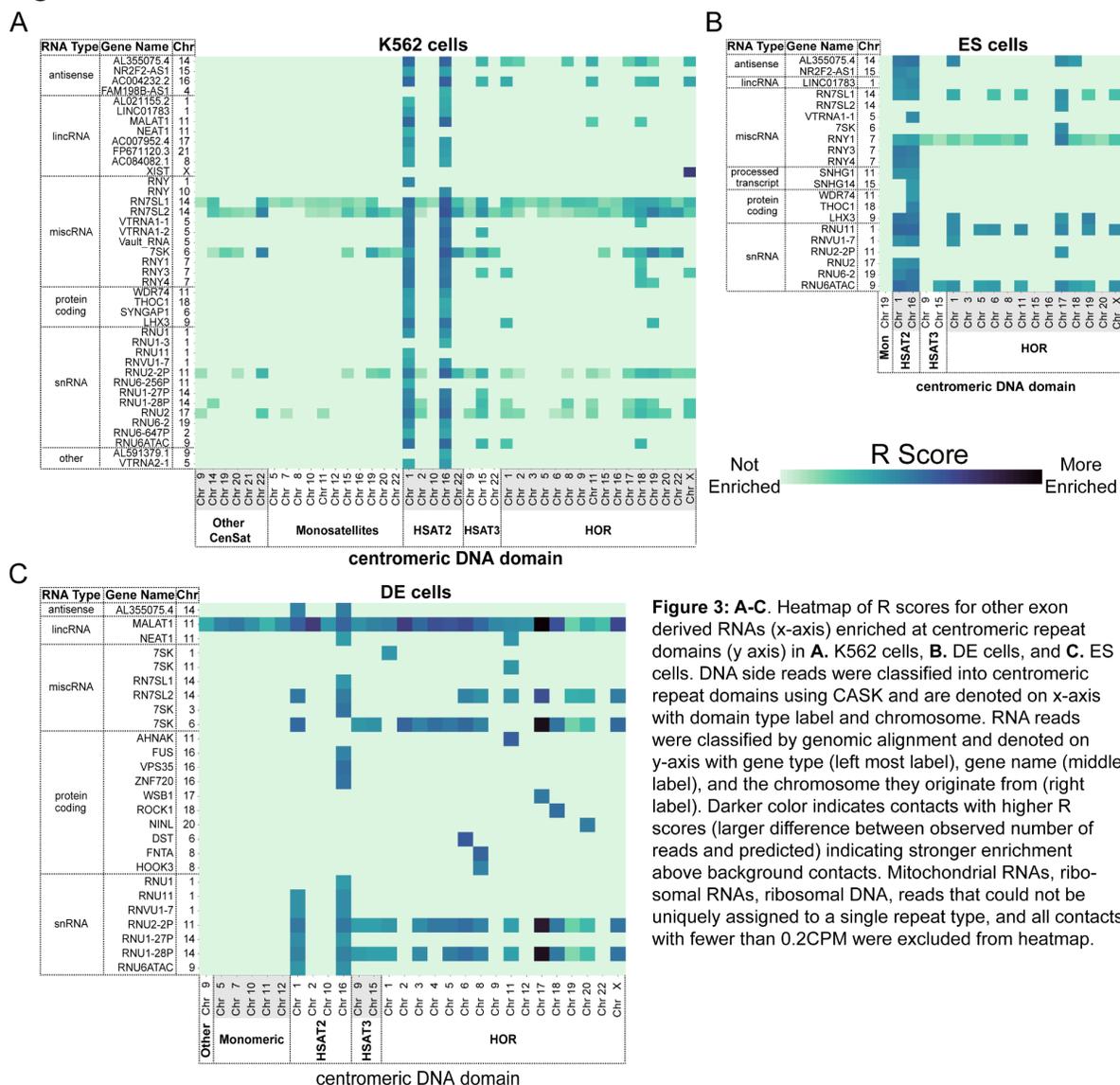


Figure 2: A-C. Heatmap of R scores for nucleolar RNAs (x-axis) enriched at centromeric repeat domains (y axis) in **A.** K562 cells, **B.** DE cells, and **C.** ES cells. DNA side reads were classified into centromeric repeat domains using CASK and are denoted on x-axis with domain type label and chromosome. RNA reads were classified by genomic alignment and denoted with gene name and the chromosome they originate from. Darker color indicates contacts with higher R scores (larger difference between observed number of reads and predicted) indicating stronger enrichment above background contacts. Mitochondrial RNAs, ribosomal RNAs, ribosomal DNA, reads that could not be uniquely assigned to a single repeat type, and all contacts with fewer than 0.2CPM (counts per million total reads) were excluded from heatmap.



Several broad binding RNAs were enriched widely across centromeric and pericentromeric repeat arrays on all chromosomes, including RN7SL1/2 - the RNA component of the signal recognition particle,⁸⁰ several snRNAs involved in regulating splicing, and 7SK - a non-coding RNA that represses RNA polymerase II transcription by binding to the elongation factor P-TEFb (**Figure 3, Supplemental Figure S6**).^{80–84} A subset of previously identified broad binding RNAs, including Y RNAs and Vault RNAs, were also enriched at the centromere, but were limited to HSAT2 domains on chromosome 1 and 16 (**Figure 3, Supplemental Figure S6**). This restricted centromeric enrichment contrasts with their proposed roles in nuclear and cellular processes and widespread nucleoplasmic and cytoplasmic localization. Y RNAs have been shown to regulate RNA stability and DNA replication and have been reported to localize to both the cytoplasm and nucleoplasm as well as directly on chromatin.^{85–91} A small percentage of vault RNAs are associated with vaults - large ribonucleoprotein complexes in the cytoplasm, while the rest are distributed throughout the cytoplasm and are thought to be involved in several processes including nucleocytoplasmic transport and multidrug resistance.⁹² MALAT1, a long non-coding RNA that localizes to nuclear speckles and has been implicated in several processes including transcriptional regulation and splicing,⁹³ was enriched widely across centromeric and centromere transition DNA domains only in DE cells (**Figure 3C, Supplemental Figure S6C**). This cell type specific enrichment pattern is consistent with our previous findings that depict MALAT1 as a broad binding RNA that is upregulated in DE cells.⁷³ The varied centromeric enrichment patterns of broad binding RNAs illustrate the RNA specific contact behaviors that may be mediated by the inherent properties of the RNA or its interaction partners, as well as by differences in the local chromatin environment across centromeric repeats.

Figure 3: Non-nucleolar localized RNAs enriched at centromeric DNA



In addition to the RNA specific contact patterns observed at the centromere, individual centromeric repeat arrays display distinct RNA-DNA contact patterns. The most striking repeat array specific patterns are the Human Satellite 2 (HSAT2) arrays on chromosomes 1 and 16 which were found to have many more enriched RNAs than other domains in all three cell lines (**Figure 2,3, Supplemental Figure S5,S6**). Several broad binding and nucleolar localized ncRNAs were only found enriched at these two HSAT2 domains and virtually every RNA enriched at any of the other centromeric repeats was also found enriched at these HSAT2 domains. Given that the HSAT2 domains on chromosome 1 and 16 make up two of the three largest pericentromeric/centromeric arrays (13.2Mbp and 12.7Mbp respectively), we compared the contact behavior of other repeat arrays that spanned over 5Mbp including the longest array, HSAT3 on chromosome 9 which is over 27Mbp and found that no other repeat array of any size displayed similar promiscuous RNA contact behavior (**Figure 2,3, Supplemental Figure S5,S6**).

DNA contact patterns of centromere-derived RNAs

Centromeres are transcribed at low levels in humans but centromeric transcripts play important functions in local regulation of chromatin state. Relatively few centromere encoded RNAs have been studied in depth but in cases where they have been characterized, the RNAs remain locally associated with their sites of transcription.^{19,67,68} To perform an extended analysis of all centromere-associated RNAs we examined the centromere encoded RNAs and their binding sites in the genome. We carried out classification and enrichment calculations of repeat derived RNAs using CASK. We found that most centromere-derived RNAs were exclusively enriched at their transcript locus and did not extend into neighboring domains in K562 and ES cells with increased enrichment at neighboring domains on the same chromosome in DE cells (**Figure 4A-C, Supplemental Figure S7,S8**). Local enrichment of HOR derived transcripts, consistent with previous reports,^{67,68} was observed at the class level which includes all HORs on a given chromosome, as well as the repeat array specific level which describes individual HOR repeat arrays. HOR derived RNAs enriched at the centromere emanated from a single HOR array on each chromosome, typically the active HOR which tends to also be the longest HOR array, and their enrichment generally did not spread even to other HORs on the same chromosome (**Figure 4B, Supplemental Figure S7B,D,F**). Overall these data suggest that HOR derived transcripts remain locally associated and do not spread to other centromeres.

While most centromere derived RNAs displayed a similar local enrichment pattern to that of HOR RNAs, some HSAT and CenSat RNA enrichment did spread to other centromeric domains. The greatest degree of spreading was observed with HSAT3 and HSAT1A derived transcripts which were enriched at HSAT2 domains on chromosome 1 and 16 and across all centromere transition (CT) regions which are genic regions flanking pericentromeric repeats. Centromeric transcripts that were enriched outside their own transcription locus were often also enriched at HSAT2 domains on chromosomes 1 and 16 (**Figure 4A**) which appear to act as RNA contact hubs for centromere derived transcripts in addition to non-repetitive RNAs (**Figure 2,3,4**). Despite the increased contacts at HSAT2 DNA domains (**Figure 2,3,4A,C**), the abundance of reads with HSAT2 derived RNA sequences was equal to or less than that of most other domains (**Figure 1D**) and enrichment of RNAs derived from HSAT2 domains on chromosomes 1 and 16 generally did not spread outside HSAT2 domains (**Figure 4A,C, Supplemental Figure S7 left panels, S8A-E**). The distinct behavior of different HSAT repeat types illustrates that the RNA contact behavior of a centromeric domain and RNAs derived from it are features specific to the repeat type and not general features of centromeres or pericentromeres.

RNAs derived from genic centromere transition (CT) regions exhibited a distinct pattern with enrichment across CT regions on every chromosome as well as other centromeric repeat domains on the same chromosome (**Figure 4C,D, Supplemental Figure S8,S9**). RNAs derived from centromere transition regions on chromosome 17 were broadly enriched across centromeric domains (**Figure 4D, Supplemental Figure S9**) mimicking the pattern seen with SNORD3A which is encoded within the centromere transition regions on chromosome 17, likely representing the overlap between CASK assignment and genome alignment in non-repetitive regions. This widespread enrichment could be indicative of increased transcriptional activity as well as increased accessibility allowing for more RNA contacts. Additionally, the uniqueness of CT regions results in a larger database of CT specific k-mers increasing the proportion of reads that can be confidently classified as CT (**Supplemental Figure S1A**).

Figure 4: Centromere derived RNAs enriched at centromeric DNA

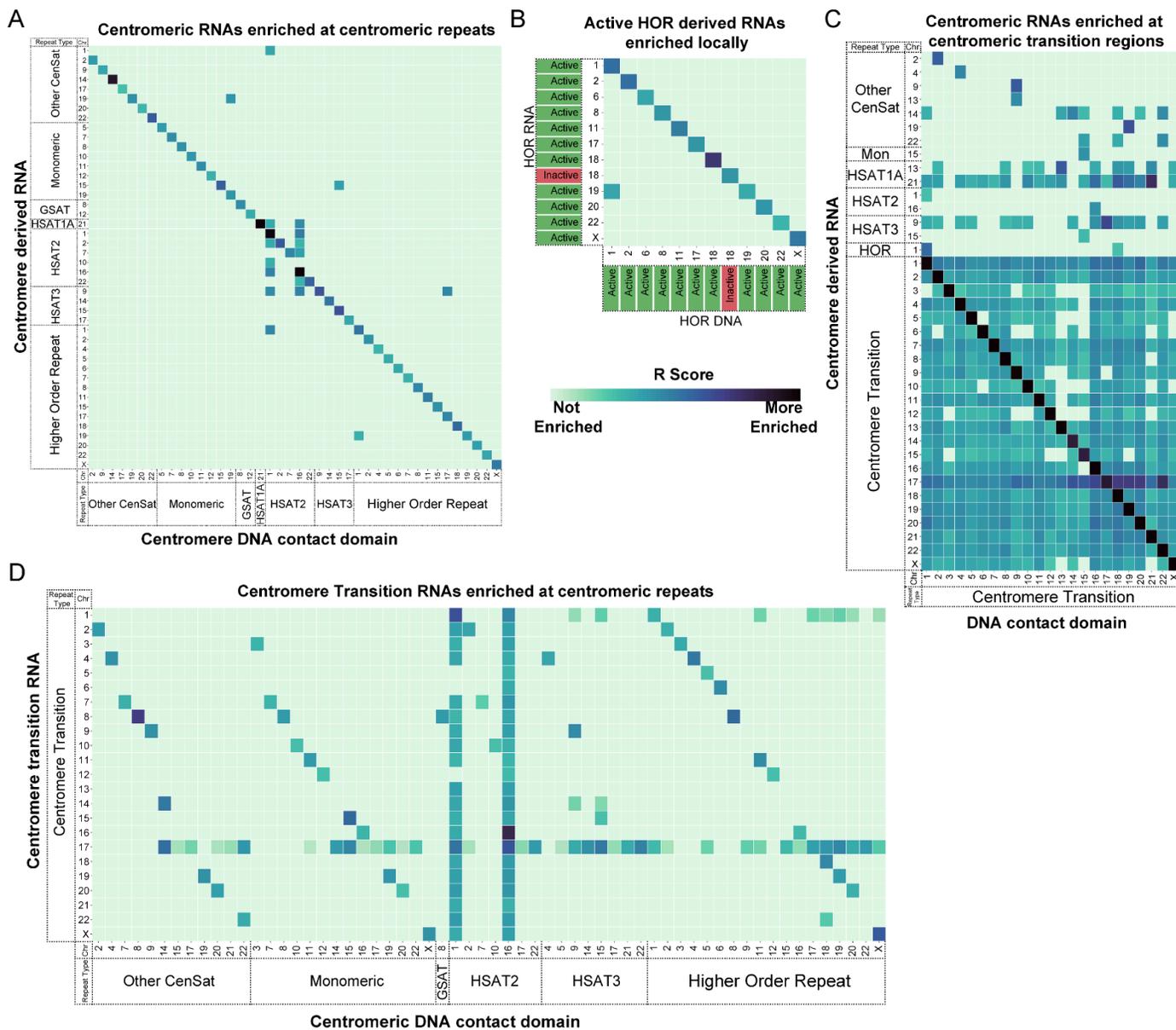


Figure 4: A-D. Heatmap of R scores for RNAs derived from centromere and centromere transition (CT) regions (y-axis) enriched at centromere/CT domains (x-axis) in K562 cells. DNA and RNA reads were classified using CASK and denoted with domain type label and chromosome. Darker color indicates contacts with higher R scores (larger difference between observed number of reads and predicted) indicating stronger enrichment above background contacts. Contacts with R scores below the 95th percentile were considered not significantly enriched and were excluded from heatmaps. Reads classified as ribosomal on either RNA or DNA side, reads that could not be uniquely assigned to a single repeat type, and all contacts with fewer than 0.1CPM were excluded from heatmap. **A.** Class level assigned centromere derived RNAs enriched at centromere DNA domains **B.** Repeat type level HOR derived RNAs enriched at HOR DNA domains. Green label corresponds to T2T active HOR annotation denoting CENP-A containing HOR. Red, inactive label denotes HOR that does not contain CENP-A. **C.** Centromere/CT derived RNAs (y-axis) enriched at CT regions (x-axis). **D.** CT derived RNAs (y-axis) enriched at centromeric domains (x-axis).

Discussion

By analyzing RNA-DNA contacts across all centromeres we provide the first comprehensive description of centromere enriched RNAs in human cells. We used ChAR-seq data that links RNAs to their binding sites in the genome to identify RNAs that bind to centromeres and centromere encoded RNAs that bind across the genome. Applying CASK, which classifies sequences based on their repeat content, to annotate centromeric reads in CHAR-seq data, we demonstrate that RNA-DNA contact patterns differ drastically between repeat array types. Understanding the RNA contact landscape at centromeres expands our definition of the complex and specialized chromatin environments at centromeres beyond what is described by differences in DNA sequence and histone content.

We hypothesized that RNAs necessary for establishing core centromeric chromatin would be enriched at every centromere. We identified a subset of nucleolar localized ncRNAs, including SNORD3A and RMRP, that were enriched at centromeres on virtually every chromosome. SNORD3A (U3) is a highly abundant box C/D snoRNA that, in contrast to most other box C/D snoRNAs, does not mediate RNA methylation, but instead has been shown to base pair with pre-ribosomal RNA to direct site specific cleavage.^{94–97} SNORD3A has also been reported to competitively bind an miRNA responsible for down regulating uridine monophosphate synthetase (UMPS) expression and has also been shown to encode several of its own miRNAs.^{98,99} RMRP, the RNA component of the mitochondrial RNA processing endoribonuclease, has been shown to cleave mitochondrial RNA as well as ribosomal RNA.^{77,100} Mutations in RMRP have been linked to changes in gene expression and defects in cell cycle progression in the disease state cartilage-hair hypoplasia.¹⁰¹ This enrichment pattern echoes previous reports of centromere and peri/centromeric RNA localization at the nucleolar periphery.^{59,68,78,79} The bias for snoRNA enrichment at HORs compared to other centromeric repeats may be further evidence of a regulatory role for the interaction between centromeres and the nucleolus.⁶⁸ In addition to their broad centromere enrichment, the proposed multifunctionality of SNORD3A and RMRP make them compelling candidates for centromere regulation. The only other RNAs that were enriched across most centromeres were abundant broad-binding RNAs involved in transcriptional regulation, splicing, and RNA modification. We did not detect CCTT RNA which was previously reported to localize at centromeres and bind CENP-C.¹⁰² In light of these observations, we propose that centromere enriched nucleolar RNAs, especially SNORD3A and RMRP, represent the most compelling candidates for deeper investigation into their potential roles in chromatin regulation.

The vast majority of enriched RNAs at centromere are found contacting the HSAT2 domains on chromosomes 1 and 16 which seem to act as RNA contact hubs. Each array spans over 12 Mbp making them the largest HSAT2 arrays and among the largest repeat arrays across all centromeres. The other largest arrays only have a fraction of the enriched RNAs indicating that the increased enriched contacts cannot be simply explained by array length. In contrast to their promiscuous RNA contact behavior, RNAs derived from these HSAT2 domains were mostly enriched at their transcription locus with some spreading that mimics the contact pattern of other HSAT and CenSat derived RNAs in that cell type. The RNA-DNA contact pattern at these massive HSAT2 arrays depict a chromatin environment that is unique from all other repeat domains and may be evidence of repeat array specific epigenetic regulation.

Despite differences in relative abundances, HOR derived RNAs are almost exclusively enriched at their own transcription locus. While the function and importance of active transcription at the centromere is debated and differs between organisms, the presence of HOR derived transcripts and their *cis* localization is in agreement with previous work^{19,67,68} and implies that if HOR derived RNAs play a role in centromeric chromatin regulation, it is mediated by RNA *in cis* and not by a single or subset of HOR RNAs diffusing between centromeres. This does not appear to be a centromere wide phenomenon as HSAT1A, HSAT3, and some monomeric satellites (Mon), and other centromeric satellites (CenSat), and especially CT derived RNAs do spread to other repeat arrays both on their own chromosome and to others.

The different RNA contact patterns observed across repeat types in conjunction with the increased spreading of centromere derived RNAs seen in DE cells illustrates the repeat array and cell type specific contact patterns. Locus specific RNA-DNA contact patterns may be a component of the distinct local chromatin environments that exist with the centromere and surrounding pericentromeric regions. Differences in RNA-DNA contacts across cell types may also reflect the changes in chromatin environment that occur through differentiation. ES and many cancer cell lines, including K562, have been shown to have less constitutive heterochromatin, particularly at pericentromeres.^{103–107} Increased heterochromatin in DE cells may impact 3D genome organization either through compaction or proximity to the nucleolus or other RNA hubs and would result in the increase in *cis* and *trans* RNA contacts observed at the centromere. Further study in cells with compromised heterochromatin maintenance is necessary to delineate the role of compaction and 3D genome organization in centromeric RNA-DNA contacts.

This study describes an RNA-chromatin environment that is rarely uniform across centromeric repeat types or cell types. It is clear from this work that to advance our understanding of the role of RNA at the centromere, it is

necessary to consider each subdomain of the centromere or pericentromere and the unique environment they occupy individually. The comprehensive list of centromere-associated RNAs and the specific domains they are enriched at has provided the groundwork for further investigation into the role of RNA in regulating centromeric chromatin. Future work to selectively perturb these centromere-associated RNAs and/or centromere derived RNAs will provide insight into their functional contribution to centromeres.

Methods

Cell harvest and fixation

Human K562 cells were cultured in RPMI 1640 media supplemented with 50 U/mL penicillin and streptomycin to a concentration of 0.8×10^6 cells/mL. For each sample, 10×10^6 cells were collected in 50 mL conical tubes, centrifuged at 500 xg for 5 min, liquid media was aspirated, and resuspended in 16.25 mL serum free media. Cells were fixed by adding 3.75 mL fresh 16% formaldehyde while shaking to a final concentration of 3% and incubated on a rocker for 10 minutes at room temperature. Formaldehyde was quenched by adding 6.25 mL of 2.5 M glycine and incubating on rocker at room temperature for 5 minutes and on ice for 15 minutes with intermittent mixing. Fixed cells were then pelleted and resuspended in ~1 mL ice cold PBS. Using a hemocytometer, cells were counted and aliquoted to 10 million cells each.

ChAR-seq library preparation

Two replicate ChAR-seq libraries were generated for K562 cells as in (Limouse et al 2023) with the following modifications: 1) After decrosslinking, DNA was isolated by phenol:chloroform extraction. Briefly, 300 μ L phenol:chloroform:isoamylalcohol was added to each sample and vigorously mixed. The entire volume was transferred to a heavy phase lock tube and centrifuged at 13,000 x g for 3 minutes at room temperature. 300 μ L chloroform was added to the top layer while still in the phase lock tube, samples were inverted several times to mix and then centrifuged again at 13,000 x g for 3 minutes at room temperature. The top layer was transferred to a fresh 1.5 mL microcentrifuge tube with 900 μ L 100% ice cold ethanol, incubated at -20°C for 20 minutes, and centrifuged at max speed for 20 minutes at 4°C . Ethanol was aspirated, and DNA pellet was washed with 80% ethanol, spun again at max speed for 5 minutes at 4°C . All ethanol was removed and samples were allowed to air dry before the DNA pellet was resuspended in 132 μ L TE (10 mM Tris-HCl pH 8, 1 mM EDTA pH 8). 2) Two additional wash steps were performed after initial binding to magnetic streptavidin beads (for a total of 4 washes), one at 50°C for 2 min, and one at room temperature. 3) The optional PacI digest was performed as described. 4) Libraries were quantified using high sensitivity Qubit Fluorometer 3.0 and size distribution was determined using high sensitivity Agilent TapeStation 4200. Libraries still containing high or low molecular weight fragments underwent an additional size selection using SPRIselect paramagnetic beads (BeckmanCoulter B23317) prior to sample pooling. 5) Libraries were first sequenced at a low depth using an illumina MiSeq to assess bridge content and duplication rate prior to pooling and high depth sequencing using illumina NovaSeq X Plus. After sequencing, raw data files for each replicate were combined. All downstream processing was performed on combined files.

ChAR-seq data processing

Raw reads were processed using a snakemake pipeline described in Limouse et al. 2023.⁷² Briefly, PCR duplicates were removed, sequencing adapters were trimmed, and reads lacking any bridge sequence were filtered out. The remaining reads were split into RNA and DNA fastq files corresponding to the sequences of the RNA-derived and DNA-derived portions of the chimeric molecules and aligned to hg38 transcriptome and genome respectively. RNA reads aligned to ribosomal sequences and corresponding DNA sides were filtered out.

Repeat classification

CASK (classification of ambivalent sequences using k-mers). First, repeat sequence k-mer databases were generated using KMC (<https://github.com/refresh-bio/KMC>) for each repeat class and repeat type based on repeat sequences extracted from genome fasta file (T2T CHM13v2; <https://github.com/marbl/CHM13>) defined by genomic coordinates (Cen/Sat bedfile from T2T CHM13v2) using bedtools. A reference repeat k-mer database was then generated by aggregating all repeat k-mers, subtracting k-mers present outside of defined repeat regions, and assigning each k-mer a unique identifier using BBDuk (part of BBTools suite). Reference

repeat k-mers were then identified and annotated in both RNA and DNA ChAR-seq reads using BBDuk. Reads were then classified into ambivalence groups based on the intersection of all possible repeats defined by k-mer content. For example, a read containing 3 different repeat k-mers, where the first is present in repeat_type1 and repeat_type2, the second is present in repeat_type1 and repeat_type3, and the third is present in repeat_type1 and repeat_type4, would be classified as repeat_type1. Reads were then counted for each RNA-DNA pair. Non-repetitive RNA reads were annotated according to their alignment to hg38 and non-repetitive DNA reads were annotated with a genomic bin identifier. Normalized CPM (counts per million reads) were calculated as the number of RNA or DNA reads annotated as each repeat class (N), normalized by the number of DpnII (D) sites present in the repeat domain as well as by the total number of reads in the dataset after filtering (R) per million reads ($R/10^6$): $(N/D)/R/10^6$.

Enrichment calculation

To quantify RNA contacts at repetitive regions, we calculated an enrichment (E score) for each RNA-DNA contact as the number of times that RNA-DNA combination was observed in the dataset normalized by the number of DpnII restriction sites in the DNA contact domain. To account for the fact that more highly expressed RNAs and larger DNA domains will make more contacts, these counts were also normalized by the total number of times that RNA was found contacting anywhere in the genome, and the length of the contact domain (**Supplemental Figure S4A**).

The background contact behavior of RNAs non specifically interacting with the genome was established using a generalized additive model of all trans contacting protein coding RNAs. For each RNA-DNA contact, a predicted enrichment score was calculated based on the background model specific for the contact chromosome which includes all protein coding RNAs contacting that chromosome and transcribed from any other chromosome (trans contacting). For DNA contacts including ambivalence groups, the background model included all trans contacting protein coding RNAs for all chromosomes in the ambivalence group. Predicted enrichment scores for each RNA-DNA were then subtracted from observed enrichment scores to generate a residual score (R score). The top 5% of residual scores were considered significantly enriched and plotted in heat maps (**Supplemental Figure S4B,C**).

Data and software availability

ChAR-seq data previously generated in embryonic stem cells and definitive endoderm cells are available as [GSE240435](#). K562 ChAR-seq data generated as a part of this study are available through SRA as PRJNA1270001. Processed data supporting the findings of this study for all cell types are available as GSE298896.

All software packages and code released as part of this study are available as public repositories <https://github.com/straightlab/>. Software packages as well as Snakemake pipelines used to process raw ChAR-seq data are available at <https://github.com/straightlab/chartools> and https://github.com/straightlab/charseq_dynamics_paper. CASK repeat classification tool is available at <https://github.com/straightlab/cask>. Enrichment calculation pipeline and all code used to generate figures in this study are available at <https://github.com/straightlab/centrochar>.

Acknowledgements

We thank the members of the Straight lab for meaningful discussion and feedback, and Pragya Sidhwani for comments on data analysis and figures. The computing for this project was performed on the Stanford Sherlock HPC cluster and we thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. The development of the ChAR-seq method and CASK tools was supported by NIH/NHGRI R01HG009909 to AFS. This work was supported by NIH R01GM074728 to AFS and NIH Training Grant T32-GM007790-40 to KAF.

Author Contributions

K.A.F: generated ChAR-seq libraries for K562 cells, conducted data analysis, prepared this manuscript, and generated figures. C.L. designed and wrote ChAR-seq data processing pipeline and k-mer classification software. A.F.S. supervised the work and contributed to manuscript preparation and editing.

Declarations of interest

The authors declare no competing interests.

Declaration of generative AI and AI-assisted technology

During the preparation of this manuscript, the authors used ChatGPT to suggest alternative words and phrases in some sentences to improve clarity. The authors reviewed and edited all content and take full responsibility for the content of the publication.

Supplemental Information

Document S1: Supplemental Figures S1-S9

Table S1: excel file containing all correlation statistics for CPM comparisons.

References

1. Earnshaw, W.C., and Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* *91*, 313–321. <https://doi.org/10.1007/BF00328227>.
2. Sullivan, B.A., and Karpen, G.H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol* *11*, 1076–1083. <https://doi.org/10.1038/nsmb845>.
3. Sullivan, B.A., and Karpen, G.H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol* *11*, 1076–1083. <https://doi.org/10.1038/nsmb845>.
4. Manuelidis, L. (1978). Chromosomal localization of complex and simple repeated human DNAs. *Chromosoma* *66*, 23–32. <https://doi.org/10.1007/BF00285813>.
5. Gershman, A., Sauria, M.E.G., Guitart, X., Vollger, M.R., Hook, P.W., Hoyt, S.J., Jain, M., Shumate, A., Razaghi, R., Koren, S., et al. (2022). Epigenetic patterns in a complete human genome. *Science* *376*, eabj5089. <https://doi.org/10.1126/science.abj5089>.
6. Kit, S. (1961). Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. *J Mol Biol* *3*, 711–716. [https://doi.org/10.1016/s0022-2836\(61\)80075-2](https://doi.org/10.1016/s0022-2836(61)80075-2).
7. Corneo, G., Ginelli, E., and Polli, E. (1968). Isolation of the complementary strands of a human satellite DNA. *J. Mol. Biol.* *33*, 331–335. [https://doi.org/10.1016/0022-2836\(68\)90301-x](https://doi.org/10.1016/0022-2836(68)90301-x).
8. Corneo, G., Ginelli, E., and Polli, E. (1970). Repeated sequences in human DNA. *J. Mol. Biol.* *48*, 319–327. [https://doi.org/10.1016/0022-2836\(70\)90163-4](https://doi.org/10.1016/0022-2836(70)90163-4).
9. Prosser, J., Frommer, M., Paul, C., and Vincent, P.C. (1986). Sequence relationships of three human satellite DNAs. *J. Mol. Biol.* *187*, 145–155. [https://doi.org/10.1016/0022-2836\(86\)90224-x](https://doi.org/10.1016/0022-2836(86)90224-x).
10. Altemose, N., Logsdon, G.A., Bzikadze, A.V., Sidhwani, P., Langley, S.A., Caldas, G.V., Hoyt, S.J., Uralsky, L., Ryabov, F.D., Shew, C.J., et al. (2022). Complete genomic and epigenetic maps of human centromeres. *Science* *376*, eabl4178. <https://doi.org/10.1126/science.abl4178>.
11. Chen, P.B., Chen, H.V., Acharya, D., Rando, O.J., and Fazzio, T.G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat Struct Mol Biol* *22*, 999–1007. <https://doi.org/10.1038/nsmb.3122>.
12. Schmitz, K.-M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of non-coding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* *24*, 2264–2269. <https://doi.org/10.1101/gad.590910>.
13. Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M.A., and Lehner, P.J. (2015). GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* *348*, 1481–1485. <https://doi.org/10.1126/science.aaa7227>.
14. Seczynska, M., Bloor, S., Cuesta, S.M., and Lehner, P.J. (2022). Genome surveillance by HUSH-mediated silencing of intronless mobile elements. *Nature* *601*, 440–445. <https://doi.org/10.1038/s41586-021-04228-1>.
15. Kaneko, S., Son, J., Bonasio, R., Shen, S.S., and Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check. *Genes Dev.* *28*, 1983–1988. <https://doi.org/10.1101/gad.247940.114>.

16. Tsai, M.-C., Manor, O., Wan, Y., Mosammamparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693. <https://doi.org/10.1126/science.1192002>.
17. Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. *Mol. Cell* 55, 171–185. <https://doi.org/10.1016/j.molcel.2014.05.009>.
18. Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076. <https://doi.org/10.1038/nature08975>.
19. Johnson, W.L., Yewdell, W.T., Bell, J.C., McNulty, S.M., Duda, Z., O'Neill, R.J., Sullivan, B.A., and Straight, A.F. (2017). RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. <https://doi.org/10.7554/eLife.25299>.
20. Shirai, A., Kawaguchi, T., Shimojo, H., Muramatsu, D., Ishida-Yonetani, M., Nishimura, Y., Kimura, H., Nakayama, J.-I., and Shinkai, Y. (2017). Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. *Elife* 6. <https://doi.org/10.7554/elife.25317>.
21. Velazquez Camacho, O., Galan, C., Swist-Rosowska, K., Ching, R., Gamalinda, M., Karabiber, F., De La Rosa-Velazquez, I., Engist, B., Koschorz, B., Shukeir, N., et al. (2017). Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. *Elife* 6, e25293. <https://doi.org/10.7554/eLife.25293>.
22. Muchardt, C., Guillemé, M., Seeler, J., Trouche, D., Dejean, A., and Yaniv, M. (2002). Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1 α . *EMBO reports*. <https://doi.org/10.1093/embo-reports/kvf194>.
23. Maison, C., Bailly, D., Roche, D., Montes de Oca, R., Probst, A.V., Vassias, I., Dingli, F., Lombard, B., Loew, D., Quivy, J.-P., et al. (2011). SUMOylation promotes de novo targeting of HP1 α to pericentric heterochromatin. *Nat Genet* 43, 220–227. <https://doi.org/10.1038/ng.765>.
24. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. *Science* 350, 978–981. <https://doi.org/10.1126/science.aad3346>.
25. Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. *Cell* 171, 1573–1588.e28. <https://doi.org/10.1016/j.cell.2017.11.008>.
26. Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146, 119–133. <https://doi.org/10.1016/j.cell.2011.06.026>.
27. Saldaña-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E.P., Bruneau, B.G., Tsirigos, A., Furlan-Magaril, M., Skok, J., et al. (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol Cell* 76, 412–422.e5. <https://doi.org/10.1016/j.molcel.2019.08.015>.
28. Huber, J., Tanasie, N.-L., Zernia, S., and Stigler, J. (2024). Single-molecule imaging reveals a direct role of CTCF's zinc fingers in SA interaction and cluster-dependent RNA recruitment. *Nucleic Acids Res* 52, 6490–6506. <https://doi.org/10.1093/nar/gkae391>.
29. Hansen, A.S., Hsieh, T.-H.S., Cattoglio, C., Pustova, I., Saldaña-Meyer, R., Reinberg, D., Darzacq, X., and Tjian, R. (2019). Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Mol Cell* 76, 395–411.e13. <https://doi.org/10.1016/j.molcel.2019.07.039>.
30. Oh, H.J., Aguilar, R., Kesner, B., Lee, H.-G., Kriz, A.J., Chu, H.-P., and Lee, J.T. (2021). Jpx RNA

regulates CTCF anchor site selection and formation of chromosome loops. *Cell* **184**, 6157–6173.e24. <https://doi.org/10.1016/j.cell.2021.11.012>.

31. Gavrilov, A.A., Sultanov, R.I., Magnitov, M.D., Galitsyna, A.A., Dashinimaev, E.B., Lieberman Aiden, E., and Razin, S.V. (2022). RedChIP identifies noncoding RNAs associated with genomic sites occupied by Polycomb and CTCF proteins. *Proc Natl Acad Sci U S A* **119**. <https://doi.org/10.1073/pnas.2116222119>.
32. Healy, E., Zhang, Q., Gail, E.H., Agius, S.C., Sun, G., Bullen, M., Pandey, V., Das, P.P., Polo, J.M., and Davidovich, C. (2024). The apparent loss of PRC2 chromatin occupancy as an artifact of RNA depletion. *Cell Rep*. **43**, 113858. <https://doi.org/10.1016/j.celrep.2024.113858>.
33. Nielsen, M., and Ulitsky, I. (2024). The links are still missing: Revisiting the role of RNA as a guide for chromatin-associated proteins. *Mol. Cell* **84**, 1178–1179. <https://doi.org/10.1016/j.molcel.2024.03.005>.
34. Guo, J.K., Blanco, M.R., Walkup, W.G., 4th, Bonesteele, G., Urbinati, C.R., Banerjee, A.K., Chow, A., Ettl, O., Strehle, M., Peyda, P., et al. (2024). Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA in vivo. *Mol. Cell* **84**, 1271–1289.e12. <https://doi.org/10.1016/j.molcel.2024.01.026>.
35. Maison, C., Bailly, D., Peters, A.H.F.M., Quivy, J.-P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., and Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat Genet* **30**, 329–334. <https://doi.org/10.1038/ng843>.
36. Wong, L.H., Brettingham-Moore, K.H., Chan, L., Quach, J.M., Anderson, M.A., Northrop, E.L., Hannan, R., Saffery, R., Shaw, M.L., Williams, E., et al. (2007). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res*. **17**, 1146–1160. <https://doi.org/10.1101/gr.6022807>.
37. Ferri, F., Bouzinba-Segard, H., Velasco, G., Hubé, F., and Francastel, C. (2009). Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. *Nucleic Acids Res* **37**, 5071–5080. <https://doi.org/10.1093/nar/gkp529>.
38. Carone, D.M., Longo, M.S., Ferreri, G.C., Hall, L., Harris, M., Shook, N., Bulazel, K.V., Carone, B.R., Oberfell, C., O'Neill, M.J., et al. (2009). A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. *Chromosoma* **118**, 113–125. <https://doi.org/10.1007/s00412-008-0181-5>.
39. Choi, E.S., Strålfors, A., Castillo, A.G., Durand-Dubief, M., Ekwall, K., and Allshire, R.C. (2011). Identification of noncoding transcripts from within CENP-A chromatin at fission yeast centromeres. *J Biol Chem* **286**, 23600–23607. <https://doi.org/10.1074/jbc.M111.228510>.
40. Ohkuni, K., and Kitagawa, K. (2011). Endogenous transcription at the centromere facilitates centromere activity in budding yeast. *Curr Biol* **21**, 1695–1703. <https://doi.org/10.1016/j.cub.2011.08.056>.
41. Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H.A., and Wong, L.H. (2012). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proc Natl Acad Sci U S A* **109**, 1979–1984. <https://doi.org/10.1073/pnas.1108705109>.
42. Blower, M.D. (2016). Centromeric Transcription Regulates Aurora-B Localization and Activation. *Cell Rep* **15**, 1624–1633. <https://doi.org/10.1016/j.celrep.2016.04.054>.
43. Ideue, T., Cho, Y., Nishimura, K., and Tani, T. (2014). Involvement of satellite I noncoding RNA in regulation of chromosome segregation. *Genes Cells* **19**, 528–538. <https://doi.org/10.1111/gtc.12149>.
44. May, B.P., Lippman, Z.B., Fang, Y., Spector, D.L., and Martienssen, R.A. (2005). Differential regulation of strand-specific transcripts from Arabidopsis centromeric satellite repeats. *PLoS Genet* **1**, e79. <https://doi.org/10.1371/journal.pgen.0010079>.

45. Koo, D.-H., Zhao, H., and Jiang, J. (2016). Chromatin-associated transcripts of tandemly repetitive DNA sequences revealed by RNA-FISH. *Chromosome Res* 24, 467–480. <https://doi.org/10.1007/s10577-016-9537-5>.
46. Bouzinba-Segard, H., Guais, A., and 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, 8709–8714. <https://doi.org/10.1073/pnas.0508006103>.
47. Du, Y., Topp, C.N., and Dawe, R.K. (2010). DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet* 6, e1000835. <https://doi.org/10.1371/journal.pgen.1000835>.
48. Du, Y., Topp, C.N., and Dawe, R.K. (2010). DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet* 6, e1000835. <https://doi.org/10.1371/journal.pgen.1000835>.
49. Ferri, F., Bouzinba-Segard, H., Velasco, G., Hubé, F., and Francastel, C. (2009). Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. *Nucleic Acids Res* 37, 5071–5080. <https://doi.org/10.1093/nar/gkp529>.
50. Carone, D.M., Longo, M.S., Ferreri, G.C., Hall, L., Harris, M., Shook, N., Bulazel, K.V., Carone, B.R., Obergfell, C., O'Neill, M.J., et al. (2009). A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. *Chromosoma* 118, 113–125. <https://doi.org/10.1007/s00412-008-0181-5>.
51. Choi, E.S., Strålfors, A., Castillo, A.G., Durand-Dubief, M., Ekwall, K., and Allshire, R.C. (2011). Identification of noncoding transcripts from within CENP-A chromatin at fission yeast centromeres. *J Biol Chem* 286, 23600–23607. <https://doi.org/10.1074/jbc.M111.228510>.
52. Ohkuni, K., and Kitagawa, K. (2011). Endogenous transcription at the centromere facilitates centromere activity in budding yeast. *Curr Biol* 21, 1695–1703. <https://doi.org/10.1016/j.cub.2011.08.056>.
53. Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H.A., and Wong, L.H. (2012). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proc Natl Acad Sci U S A* 109, 1979–1984. <https://doi.org/10.1073/pnas.1108705109>.
54. Blower, M.D. (2016). Centromeric Transcription Regulates Aurora-B Localization and Activation. *Cell Rep* 15, 1624–1633. <https://doi.org/10.1016/j.celrep.2016.04.054>.
55. Ideue, T., Cho, Y., Nishimura, K., and Tani, T. (2014). Involvement of satellite I noncoding RNA in regulation of chromosome segregation. *Genes Cells* 19, 528–538. <https://doi.org/10.1111/gtc.12149>.
56. May, B.P., Lippman, Z.B., Fang, Y., Spector, D.L., and Martienssen, R.A. (2005). Differential regulation of strand-specific transcripts from Arabidopsis centromeric satellite repeats. *PLoS Genet* 1, e79. <https://doi.org/10.1371/journal.pgen.0010079>.
57. Koo, D.-H., Zhao, H., and Jiang, J. (2016). Chromatin-associated transcripts of tandemly repetitive DNA sequences revealed by RNA-FISH. *Chromosome Res* 24, 467–480. <https://doi.org/10.1007/s10577-016-9537-5>.
58. Bouzinba-Segard, H., Guais, A., and 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, 8709–8714. <https://doi.org/10.1073/pnas.0508006103>.
59. Lopes, M., Louzada, S., Ferreira, D., Veríssimo, G., Eleutério, D., Gama-Carvalho, M., and Chaves, R. (2023). Human Satellite 1A analysis provides evidence of pericentromeric transcription. *BMC Biol* 21, 28. <https://doi.org/10.1186/s12915-023-01521-5>.
60. Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of

heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297, 1833–1837.
<https://doi.org/10.1126/science.1074973>.

61. Ling, Y.H., and Yuen, K.W.Y. (2019). Point centromere activity requires an optimal level of centromeric noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A.* 116, 6270–6279. <https://doi.org/10.1073/pnas.1821384116>.
62. Chen, C.-F., Pohl, T.J., Chan, A., Slocum, J.S., and Zakian, V.A. (2019). *Saccharomyces cerevisiae* centromere RNA is negatively regulated by Cbf1 and its unscheduled synthesis impacts CenH3 binding. *Genetics* 213, 465–479. <https://doi.org/10.1534/genetics.119.302528>.
63. Liu, H., Qu, Q., Warrington, R., Rice, A., Cheng, N., and Yu, H. (2015). Mitotic transcription installs Sgo1 at centromeres to coordinate chromosome segregation. *Mol. Cell* 59, 426–436.
<https://doi.org/10.1016/j.molcel.2015.06.018>.
64. Watanabe, R., Perea-Resca, C., and Blower, M.D. (2025). Robust CENP-A incorporation in human cells is independent of transcription and cohesin components. *bioRxiv*, 2025.04.29.651290.
<https://doi.org/10.1101/2025.04.29.651290>.
65. Johnson, W.L., Yewdell, W.T., Bell, J.C., McNulty, S.M., Duda, Z., O'Neill, R.J., Sullivan, B.A., and Straight, A.F. (2017). RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin.
<https://doi.org/10.7554/eLife.25299>.
66. Kitagawa, R., Niikura, Y., Becker, A., Houghton, P.J., and Kitagawa, K. (2023). EWSR1 maintains centromere identity. *Cell Rep* 42, 112568. <https://doi.org/10.1016/j.celrep.2023.112568>.
67. McNulty, S.M., Sullivan, L.L., and Sullivan, B.A. (2017). Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. *Dev Cell* 42, 226–240.e6. <https://doi.org/10.1016/j.devcel.2017.07.001>.
68. Bury, L., Moodie, B., Ly, J., McKay, L.S., Miga, K.H.H., and Cheeseman, I.M. (2020). Alpha-satellite RNA transcripts are repressed by centromere–nucleolus associations. <https://doi.org/10.7554/eLife.59770>.
69. Bury, L., Moodie, B., Ly, J., McKay, L.S., Miga, K.H.H., and Cheeseman, I.M. (2020). Alpha-satellite RNA transcripts are repressed by centromere–nucleolus associations. <https://doi.org/10.7554/eLife.59770>.
70. Bell, J.C., Jukam, D., Teran, N.A., Risca, V.I., Smith, O.K., Johnson, W.L., Skotheim, J.M., Greenleaf, W.J., and Straight, A.F. (2018). Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *Elife* 7. <https://doi.org/10.7554/eLife.27024>.
71. Jukam, D., Limouse, C., Smith, O.K., Risca, V.I., Bell, J.C., and Straight, A.F. (2019). Chromatin-associated RNA sequencing (ChAR-seq). *Curr. Protoc. Mol. Biol.* 126, e87.
<https://doi.org/10.1002/cpmb.87>.
72. Limouse, C., Jukam, D., Smith, O.K., Fryer, K.A., and Straight, A.F. (2020). Mapping transcriptome-wide and genome-wide RNA-DNA contacts with chromatin-associated RNA sequencing (ChAR-seq). *Methods Mol. Biol.* 2161, 115–142. https://doi.org/10.1007/978-1-0716-0680-3_10.
73. Limouse, C., Smith, O.K., Jukam, D., Fryer, K.A., Greenleaf, W.J., and Straight, A.F. (2023). Global mapping of RNA-chromatin contacts reveals a proximity-dominated connectivity model for ncRNA-gene interactions. *Nat Commun* 14, 6073. <https://doi.org/10.1038/s41467-023-41848-9>.
74. Hoyt, S.J., Storer, J.M., Hartley, G.A., Grady, P.G.S., Gershman, A., de Lima, L.G., Limouse, C., Halabian, R., Wojenski, L., Rodriguez, M., et al. (2022). From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science* 376, eabk3112. <https://doi.org/10.1126/science.abk3112>.
75. Hastie, T.J., and Tibshirani, R.J. (1990). *Generalized Additive Models* (CRC Press).

76. Bachellerie, J.P., Michot, B., Nicoloso, M., Balakin, A., Ni, J., and Fournier, M.J. (1995). Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem Sci* 20, 261–264. [https://doi.org/10.1016/s0968-0004\(00\)89039-8](https://doi.org/10.1016/s0968-0004(00)89039-8).
77. Topper, J.N., and Clayton, D.A. (1990). Characterization of human MRP/Th RNA and its nuclear gene: full length MRP/Th RNA is an active endoribonuclease when assembled as an RNP. *Nucleic Acids Res* 18, 793–799. <https://doi.org/10.1093/nar/18.4.793>.
78. Stahl, A., Hartung, M., Vagner-Capodano, A.M., and Fouet, C. (1976). Chromosomal constitution of nucleolus-associated chromatin in man. *Hum. Genet.* 35, 27–34. <https://doi.org/10.1007/bf00295616>.
79. Ochs, R.L., and Press, R.I. (1992). Centromere autoantigens are associated with the nucleolus. *Exp. Cell Res.* 200, 339–350. [https://doi.org/10.1016/0014-4827\(92\)90181-7](https://doi.org/10.1016/0014-4827(92)90181-7).
80. Walter, P., and Blobel, G. (1982). Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* 299, 691–698. <https://doi.org/10.1038/299691a0>.
81. Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* 414, 322–325. <https://doi.org/10.1038/35104581>.
82. Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* 414, 317–322. <https://doi.org/10.1038/35104575>.
83. Yik, J.H.N., Chen, R., Nishimura, R., Jennings, J.L., Link, A.J., and Zhou, Q. (2003). Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol Cell* 12, 971–982. [https://doi.org/10.1016/s1097-2765\(03\)00388-5](https://doi.org/10.1016/s1097-2765(03)00388-5).
84. Michels, A.A., Fraldi, A., Li, Q., Adamson, T.E., Bonnet, F., Nguyen, V.T., Sedore, S.C., Price, J.P., Price, D.H., Lania, L., et al. (2004). Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J* 23, 2608–2619. <https://doi.org/10.1038/sj.emboj.7600275>.
85. Farris, A.D., Puvion-Dutilleul, F., Puvion, E., Harley, J.B., and Lee, L.A. (1997). The ultrastructural localization of 60-kDa Ro protein and human cytoplasmic RNAs: association with novel electron-dense bodies. *Proc Natl Acad Sci U S A* 94, 3040–3045. <https://doi.org/10.1073/pnas.94.7.3040>.
86. Christov, C.P., Gardiner, T.J., Szüts, D., and Krude, T. (2006). Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. *Mol Cell Biol* 26, 6993–7004. <https://doi.org/10.1128/MCB.01060-06>.
87. Kheir, E., and Krude, T. (2017). Non-coding Y RNAs associate with early replicating euchromatin in concordance with the origin recognition complex. *J Cell Sci* 130, 1239–1250. <https://doi.org/10.1242/jcs.197566>.
88. Gardiner, T.J., Christov, C.P., Langley, A.R., and Krude, T. (2009). A conserved motif of vertebrate Y RNAs essential for chromosomal DNA replication. *RNA* 15, 1375–1385. <https://doi.org/10.1261/rna.1472009>.
89. Chen, X., Smith, J.D., Shi, H., Yang, D.D., Flavell, R.A., and Wolin, S.L. (2003). The Ro autoantigen binds misfolded U2 small nuclear RNAs and assists mammalian cell survival after UV irradiation. *Curr Biol* 13, 2206–2211. <https://doi.org/10.1016/j.cub.2003.11.028>.
90. O'Brien, C.A., and Wolin, S.L. (1994). A possible role for the 60-kD Ro autoantigen in a discard pathway for defective 5S rRNA precursors. *Genes Dev* 8, 2891–2903. <https://doi.org/10.1101/gad.8.23.2891>.
91. Xiao, J. (2020). *Non-coding RNAs in Cardiovascular Diseases* (Springer Nature).
92. Aghajani Mir, M. (2024). Vault RNAs (vtRNAs): Rediscovered non-coding RNAs with diverse

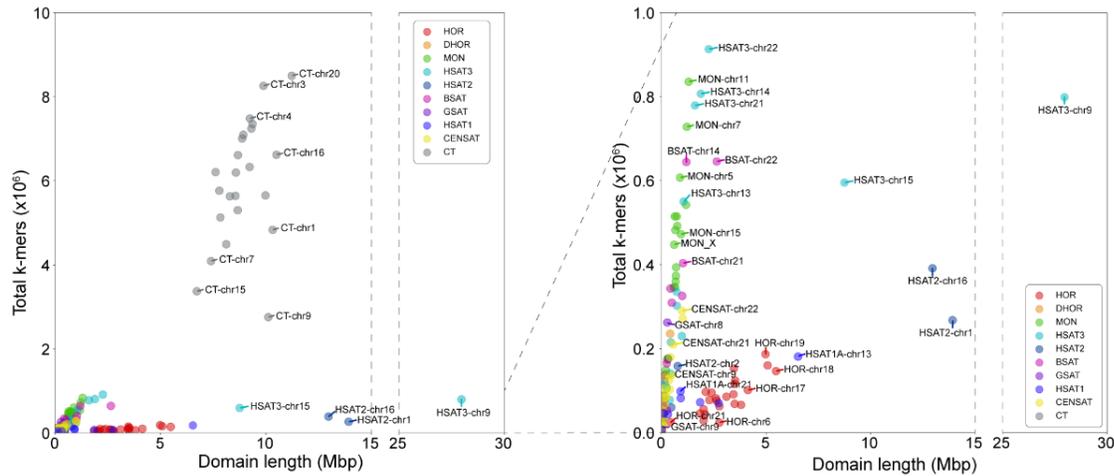
- physiological and pathological activities. *Genes Dis* 11, 772–787. <https://doi.org/10.1016/j.gendis.2023.01.014>.
93. Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39, 925–938. <https://doi.org/10.1016/j.molcel.2010.08.011>.
94. Jeppesen, C., Stebbins-Boaz, B., and Gerbi, S.A. (1988). Nucleotide sequence determination and secondary structure of *Xenopus* U3 snRNA. *Nucleic Acids Res* 16, 2127–2148. <https://doi.org/10.1093/nar/16.5.2127>.
95. An in vivo and in vitro structure-function analysis of the *Saccharomyces cerevisiae* U3A snoRNP: protein-RNA contacts and base-pair interaction with the pre-ribosomal RNA (1997). *Journal of Molecular Biology* 273, 552–571. <https://doi.org/10.1006/jmbi.1997.1320>.
96. Parker, K.A., and Steitz, J.A. (1987). Structural analysis of the human U3 ribonucleoprotein particle reveal a conserved sequence available for base pairing with pre-rRNA. *Mol Cell Biol* 7, 2899–2913. <https://doi.org/10.1128/mcb.7.8.2899-2913.1987>.
97. Sharma, K., and Tollervey, D. (1999). Base pairing between U3 small nucleolar RNA and the 5' end of 18S rRNA is required for pre-rRNA processing. *Mol Cell Biol* 19, 6012–6019. <https://doi.org/10.1128/MCB.19.9.6012>.
98. Luo, L., Zhang, J., Tang, H., Zhai, D., Huang, D., Ling, L., Wang, X., Liu, T., Zhang, Q., Zhang, Z., et al. (2020). LncRNA SNORD3A specifically sensitizes breast cancer cells to 5-FU by sponging miR-185-5p to enhance UMPS expression. *Cell Death Dis* 11, 329. <https://doi.org/10.1038/s41419-020-2557-2>.
99. Lemus-Diaz, N., Ferreira, R.R., Bohnsack, K.E., Gruber, J., and Bohnsack, M.T. (2020). The human box C/D snoRNA U3 is a miRNA source and miR-U3 regulates expression of sortin nexin 27. *Nucleic Acids Res* 48, 8074–8089. <https://doi.org/10.1093/nar/gkaa549>.
100. Tollervey, D., and Kiss, T. (1997). Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.* 9, 337–342. [https://doi.org/10.1016/s0955-0674\(97\)80005-1](https://doi.org/10.1016/s0955-0674(97)80005-1).
101. Ridanpää, M., van Eenennaam, H., Pelin, K., Chadwick, R., Johnson, C., Yuan, B., vanVenrooij, W., Pruijn, G., Salmela, R., Rockas, S., et al. (2001). Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. *Cell* 104, 195–203. [https://doi.org/10.1016/s0092-8674\(01\)00205-7](https://doi.org/10.1016/s0092-8674(01)00205-7).
102. Zhang, C., Wang, D., Hao, Y., Wu, S., Luo, J., Xue, Y., Wang, D., Li, G., Liu, L., Shao, C., et al. (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. <https://doi.org/10.1016/j.molcel.2022.09.022>.
103. Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2012). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* 22, 233–234. <https://doi.org/10.1016/j.devcel.2011.12.005>.
104. Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat. Genet.* 41, 246–250. <https://doi.org/10.1038/ng.297>.
105. Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152, 642–654. <https://doi.org/10.1016/j.cell.2012.12.033>.

106. Chen, T., and Dent, S.Y.R. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat. Rev. Genet.* 15, 93–106. <https://doi.org/10.1038/nrg3607>.

107. Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., and Verma, I.M. (2011). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* 477, 179–184. <https://doi.org/10.1038/nature10371>.

Supplemental Figure S1:

A Total number of k-mers in each repeat class



B Total number of single class k-mers in each repeat class

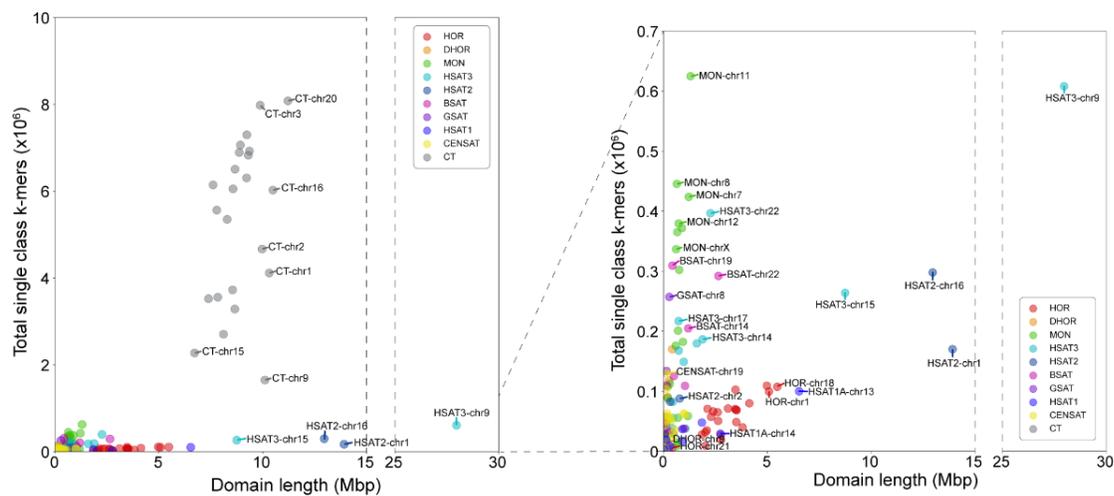
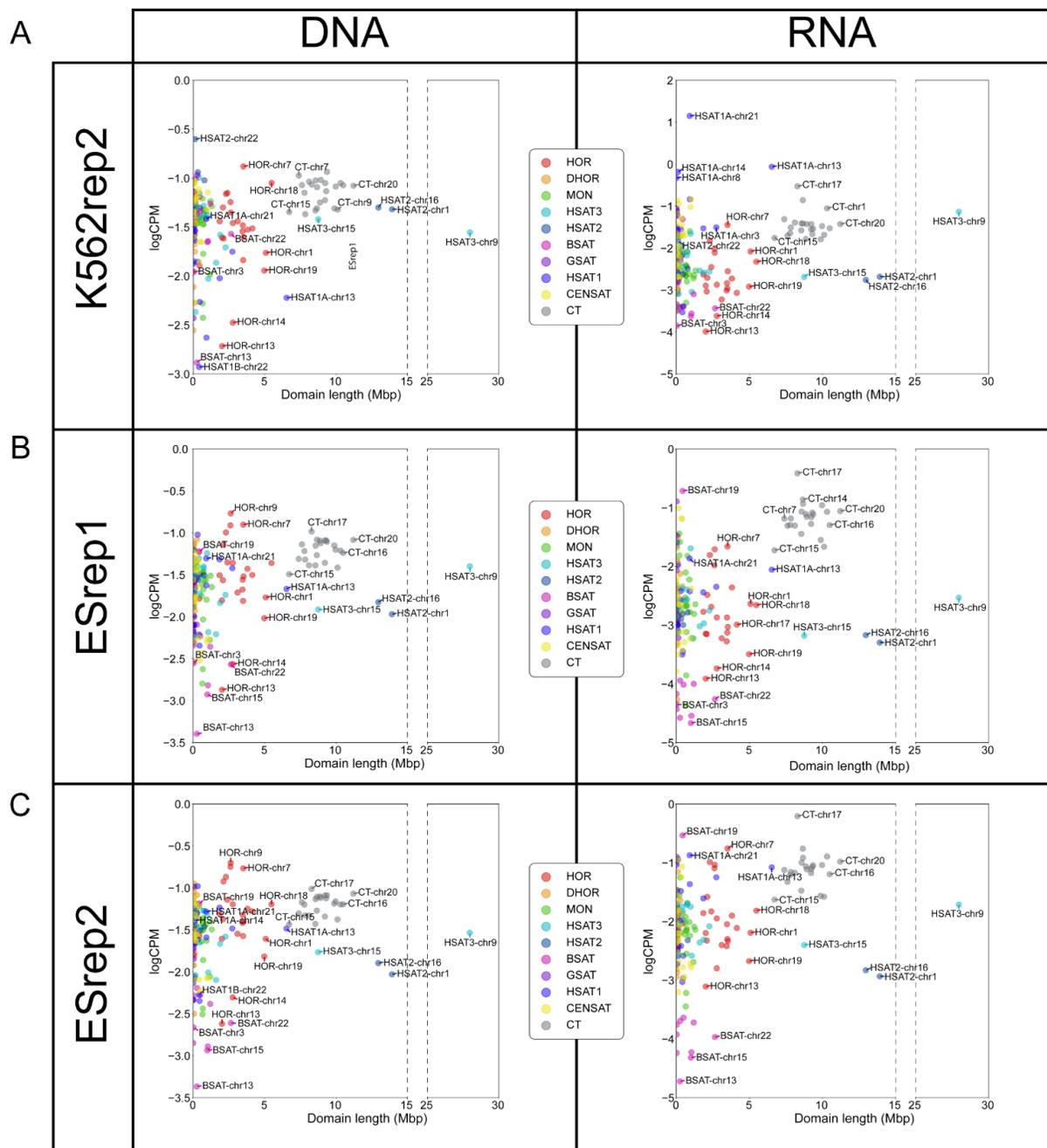


Figure S1:

A. Scatter plot depicting the total number of different k-mers in each repeat class (y-axis) vs. class length in Mbp. Includes k-mers that are shared between multiple repeat classes as well as k-mers that are class specific. Right plot shows zoomed in view.

B. Scatter plot depicting the total number of different k-mers unique to each repeat class (y-axis) vs. class length in Mbp. Includes k-mers unique to single repeat arrays (i.e. HOR-chr1-1: the first HOR domain on chr1) and unique to classes (i.e. HOR-chr1). Right plot shows zoomed in view.

Supplemental Figure S2: Normalized cenDNA and cenRNA read counts vs. domain length



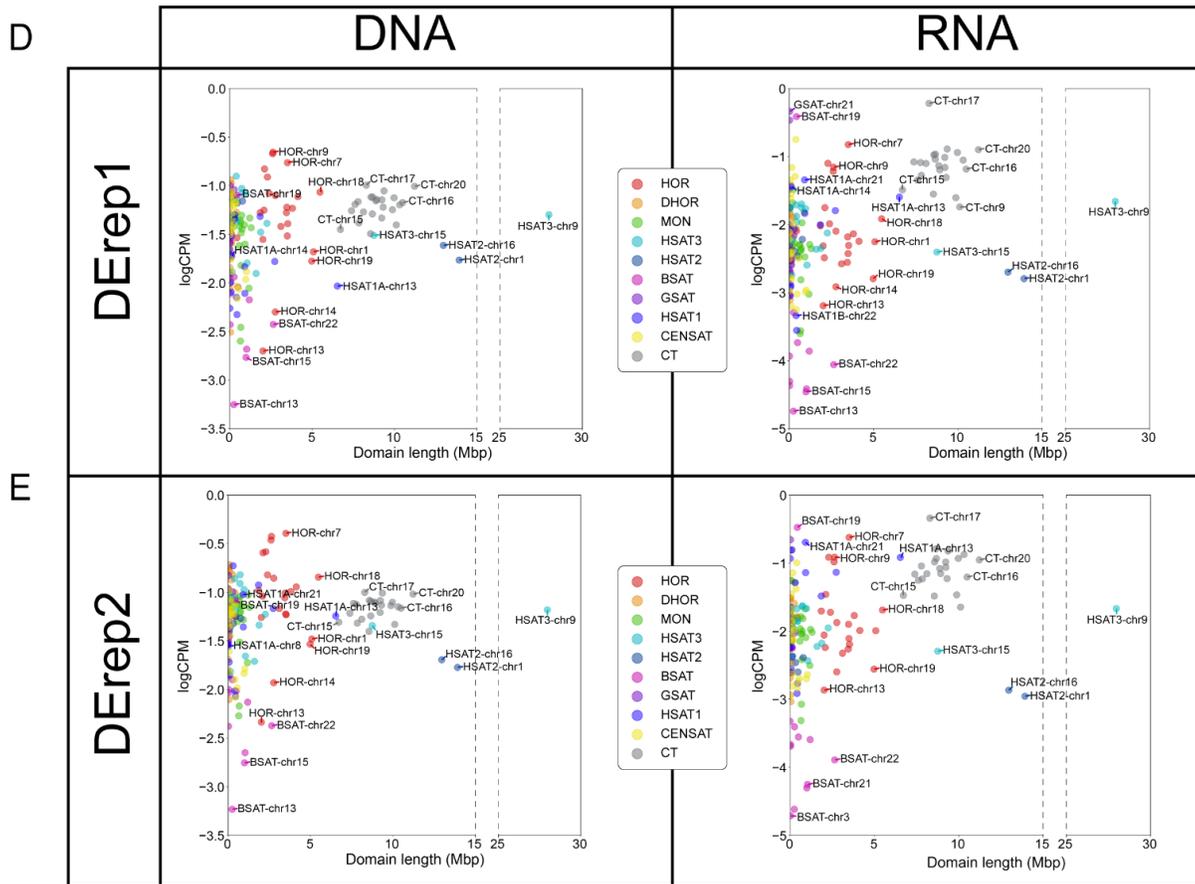


Figure S2: A-E. Similar to Figure 4.1C&D: Log of DpnII normalized DNA and RNA read counts classified by CASK for each repeat domain per million reads. Reads that could not be uniquely classified or were classified as rDNA were excluded.

- A. K562 replicate 2 DNA reads (left) and RNA reads (right)
- B. ES replicate 1 DNA reads (left) and RNA reads (right)
- C. ES replicate 2 DNA reads (left) and RNA reads (right)
- D. DE replicate 1 DNA reads (left) and RNA reads (right)
- E. DE replicate 2 DNA reads (left) and RNA reads (right)

Supplemental Figure S3: Normalized cenRNA:cenDNA counts vs. domain length

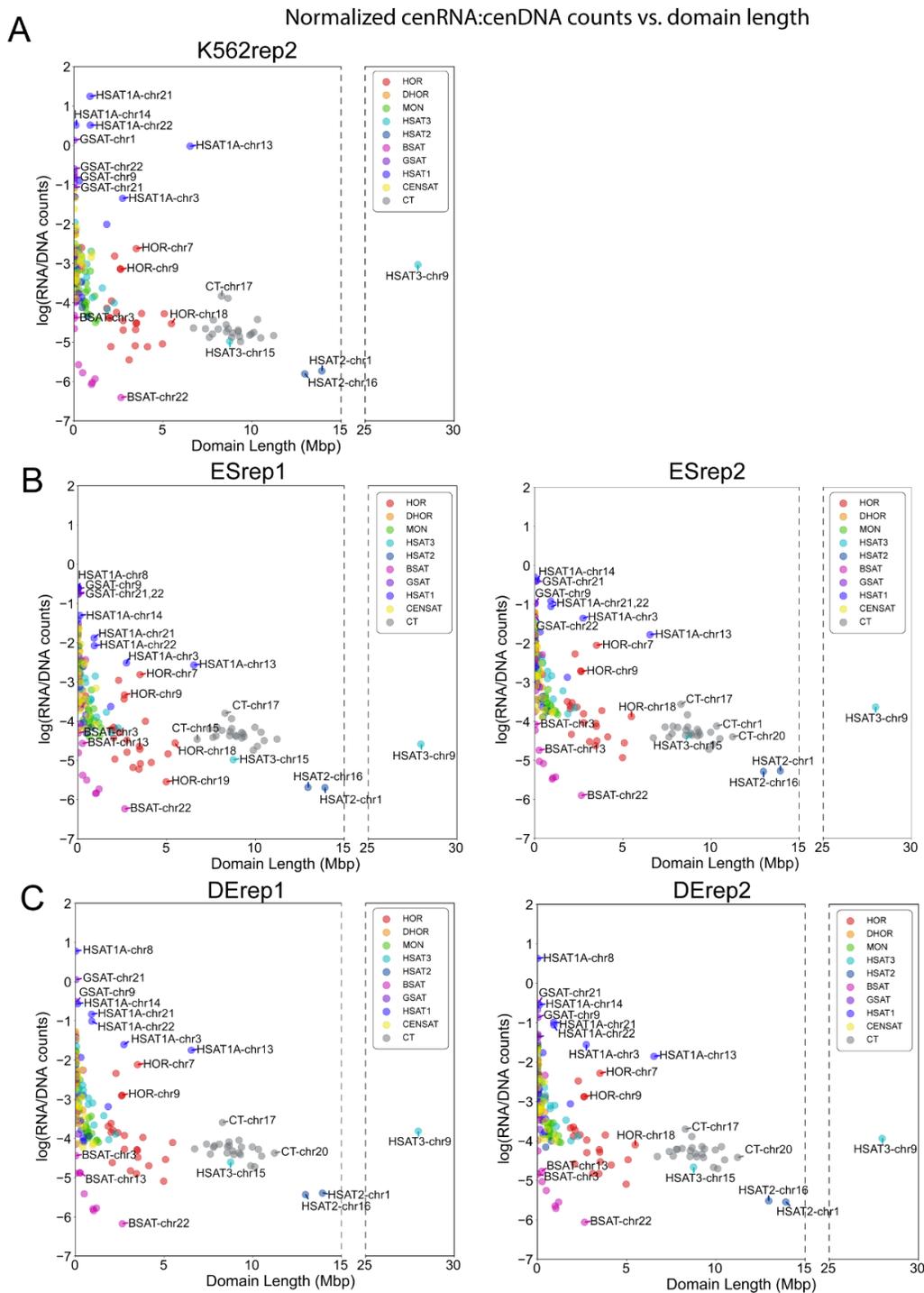


Figure S3:

A-C. Log of ratio of RNA to DNA read counts classified by CASK for each repeat domain divided by the number of DpnII sites in the domain. Reads that could not be uniquely classified or were classified as rDNA were excluded.

A. log(RNA:DNA) for K562 replicate 2

B. log(RNA:DNA) for ES replicate 1 (left) and ES replicate 2 (right)

C. log(RNA:DNA) for DE replicate1 (left) and DE replicate 2 (right)

Supplemental Figure S4: Enrichment score calculation and enriched RNA types at centromeres vs. genomic regions

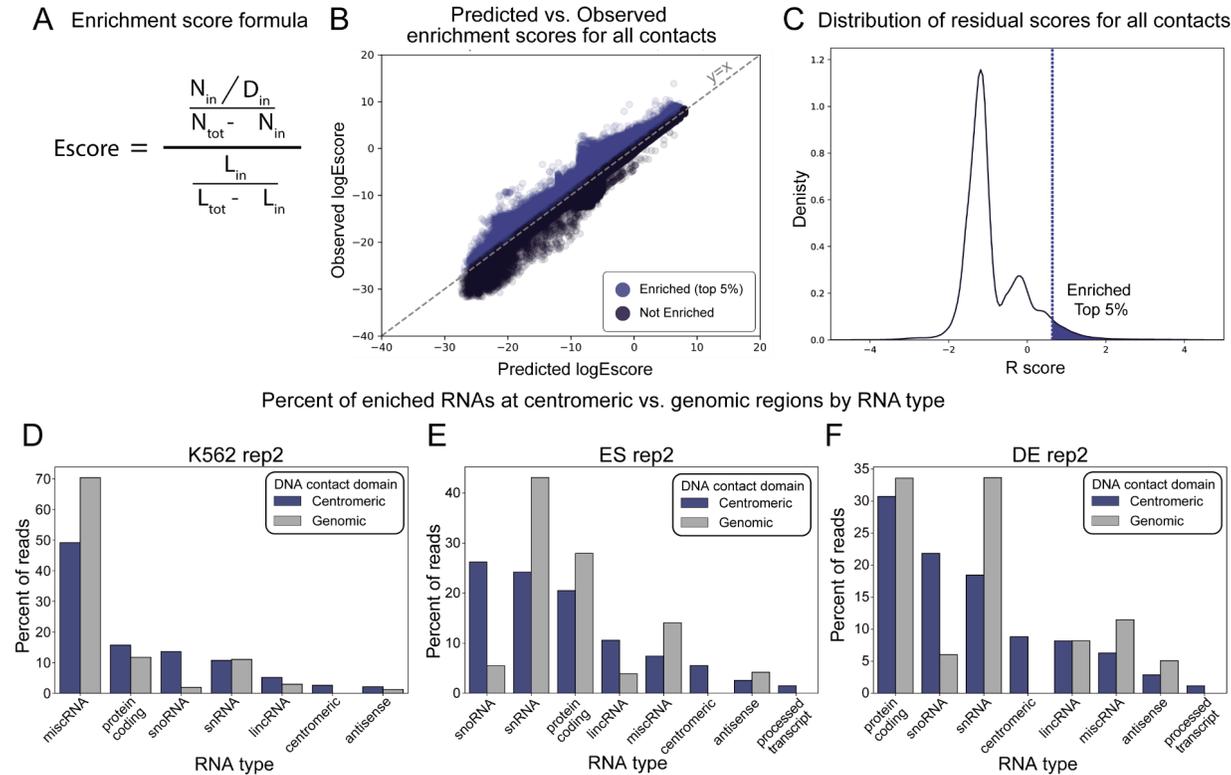


Figure S4:

A. Enrichment score calculation for each RNA-DNA contact. N_{in} represents the number of reads where a given RNA was found contacting a given DNA domain. N_{tot} is the total number of times that RNA was found in the entire dataset. D_{in} is the total number of DpnII sites in the DNA domain. L_{in} is the length of the DNA contact domain and L_{tot} is the total length of the genome.

B. Scatter plot depicting observed (y-axis) and predicted (x-axis) enrichment scores (E scores) for all RNA-DNA contacts from K562 rep1. Mitochondrial and ribosomal RNA contacts were excluded. Residual scores were calculated as the difference between predicted and observed E scores and the top 5% were considered significantly enriched (purple)

C. Distribution of residual scores. Top 5% (purple section denoted by vertical dotted line) defined as significantly enriched.

D-F. Proportion of RNA reads enriched at centromeric (purple) or non-centromeric (gray) domains belonging to each RNA type in **D.** K562 replicate 2, **E.** ES replicate 2, and **F.** DE replicate 2.

Supplemental Figure S5: Nucleolar localized RNAs enriched at centromere DNA domains

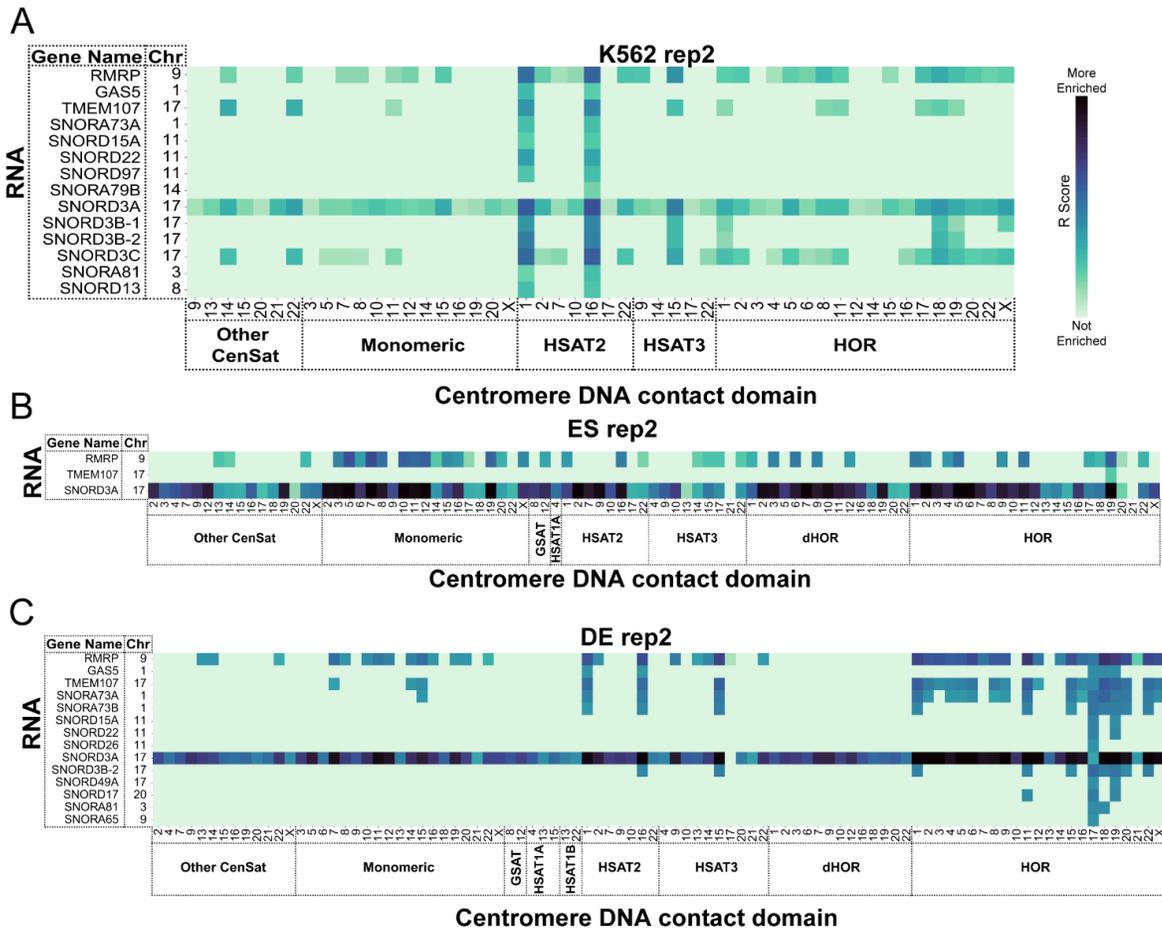


Figure S5: Heatmap of R scores for nucleolar localized RNAs (x-axis) enriched at centromeric repeat domains (y axis) in **A.** K562 replicate 2, **B.** ES replicate 2, and **C.** DE replicate 2. DNA side reads were classified into centromeric repeat domains using CASK and are denoted on x-axis with domain type label and chromosome. RNA reads were classified by genomic alignment and denoted with gene name and the chromosome they originate from. Darker color indicates contacts with higher R scores (larger difference between observed number of reads and predicted) indicating stronger enrichment above background contacts. Mitochondrial RNAs, ribosomal RNAs, ribosomal DNA, reads that could not be uniquely assigned to a single repeat type, and all contacts with fewer than 0.2CPM were excluded from heatmap.

Supplemental Figure S6: Non-nucleolar localized RNAs enriched at centromere DNA domains

Non-nucleolar RNAs enriched at centromeric DNA domains

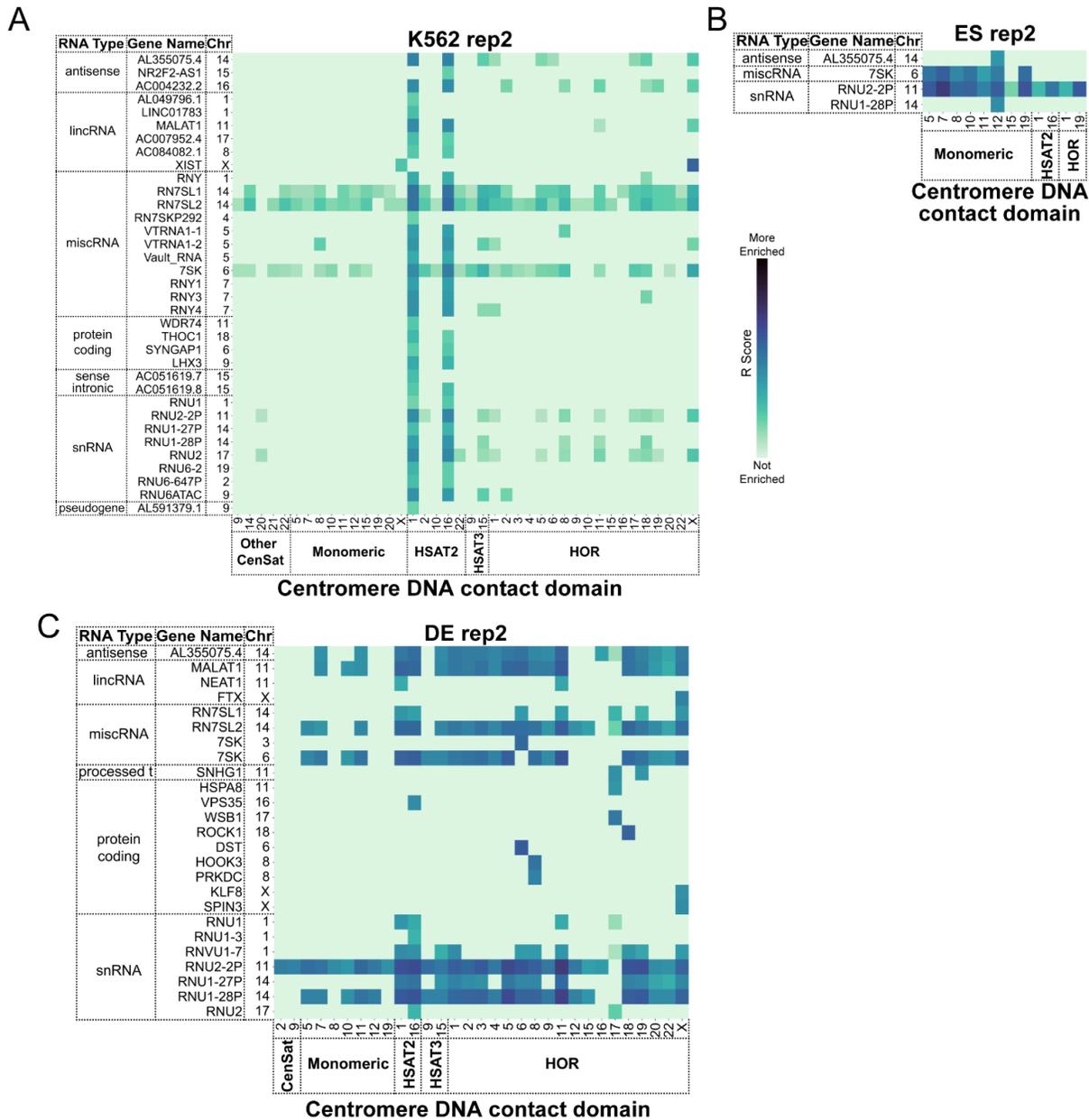
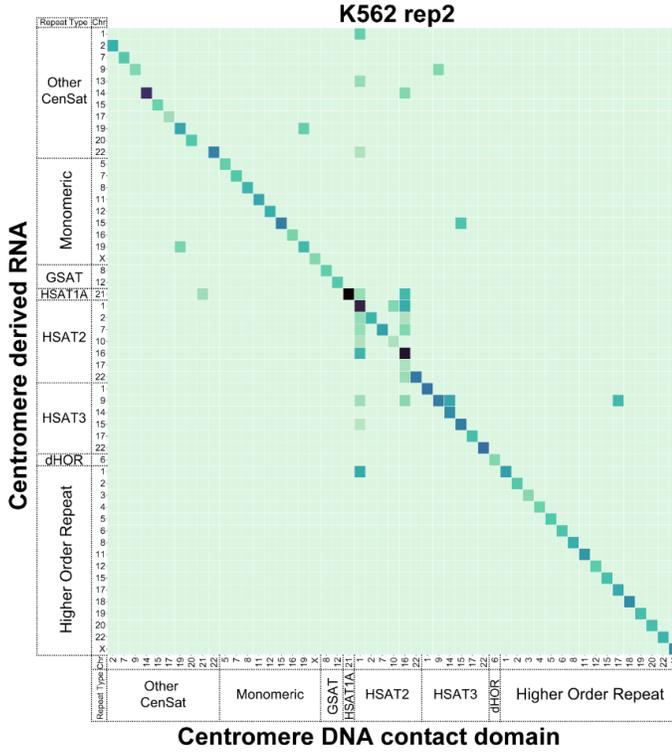


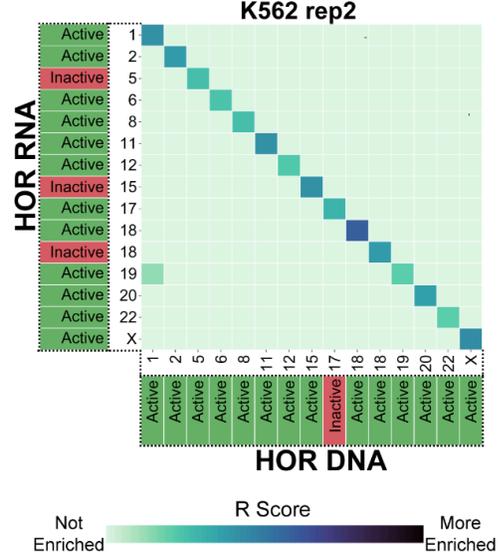
Figure S6: Heatmap of R scores for all non-nucleolar localized RNAs (x-axis) enriched at centromeric repeat domains (y axis) in **A.** K562 replicate 2, **B.** ES replicate 2, and **C.** DE replicate 2. DNA side reads were classified into centromeric repeat domains using CASK and are denoted on x-axis with domain type label and chromosome. RNA reads were classified by genomic alignment and denoted with gene name and the chromosome they originate from. Darker color indicates contacts with higher R scores (larger difference between observed number of reads and predicted) indicating stronger enrichment above background contacts. Mitochondrial RNAs, ribosomal RNAs, ribosomal DNA, reads that could not be uniquely assigned to a single repeat type, and all contacts with fewer than 0.2CPM were excluded from heatmap.

Supplemental Figure S7: Centromere derived RNAs enriched at centromere DNA domains

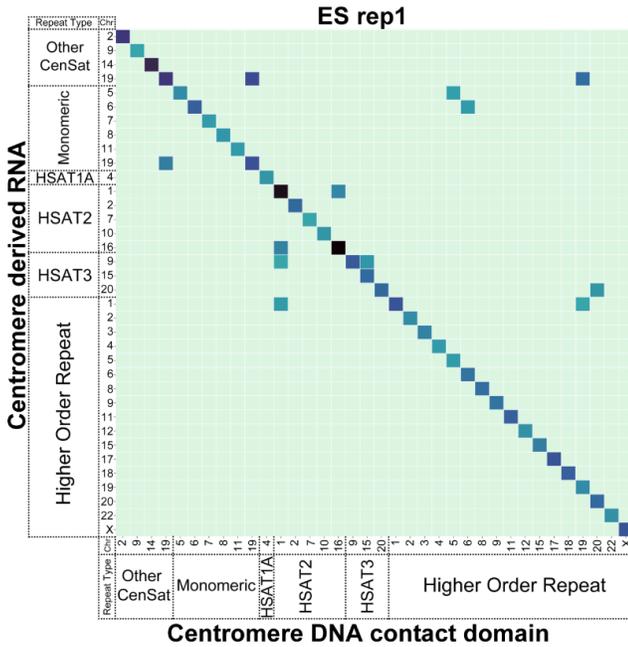
A Centromeric RNAs enriched at centromeric repeats



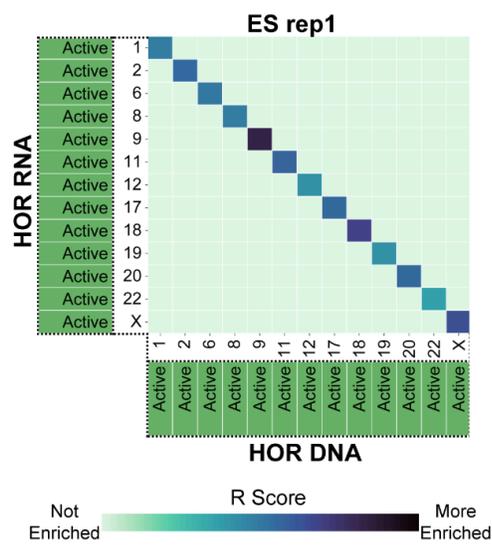
B Active HOR derived RNAs enriched locally



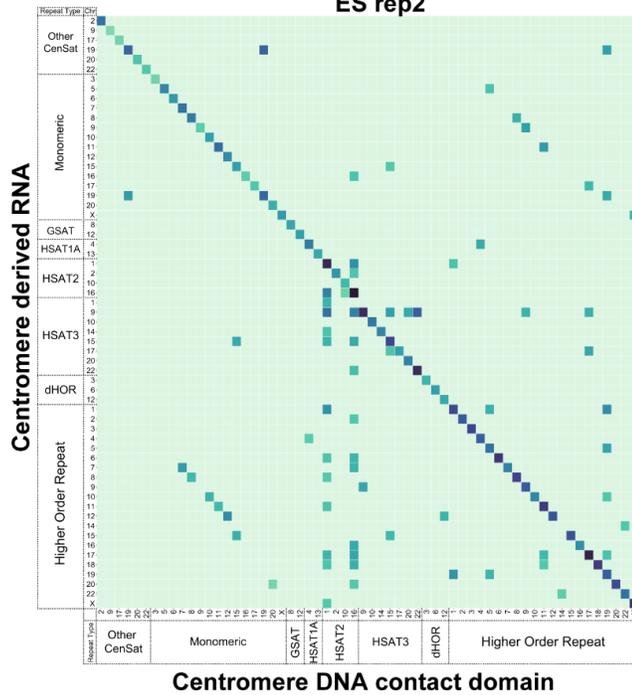
C Centromeric RNAs enriched at centromeric repeats



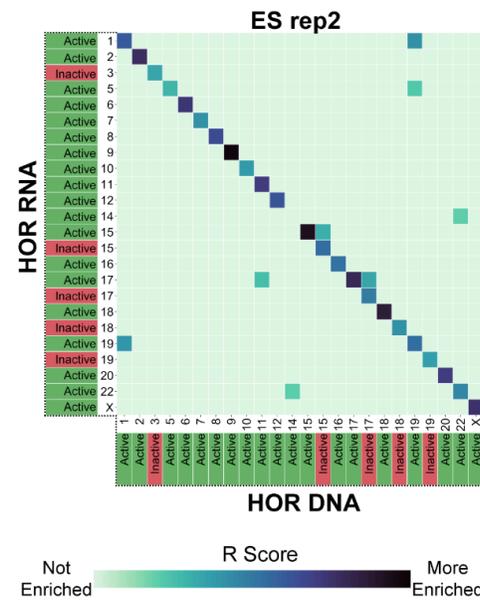
D Active HOR derived RNAs enriched locally



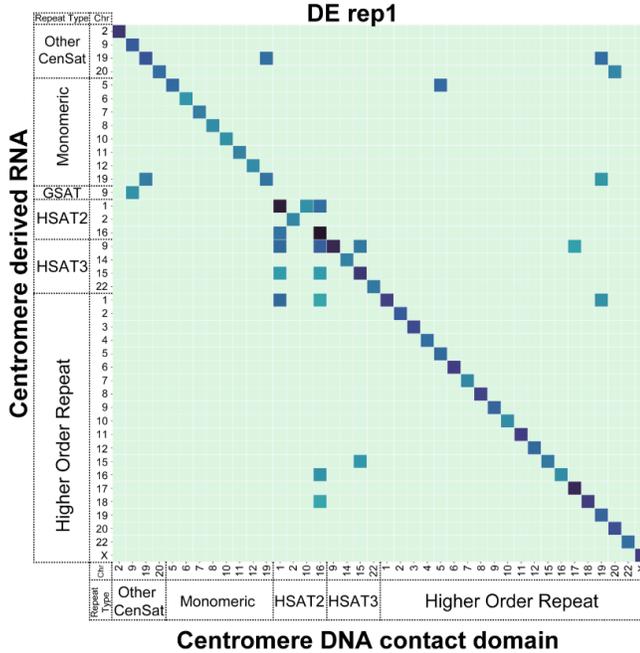
E Centromeric RNAs enriched at centromeric repeats ES rep2



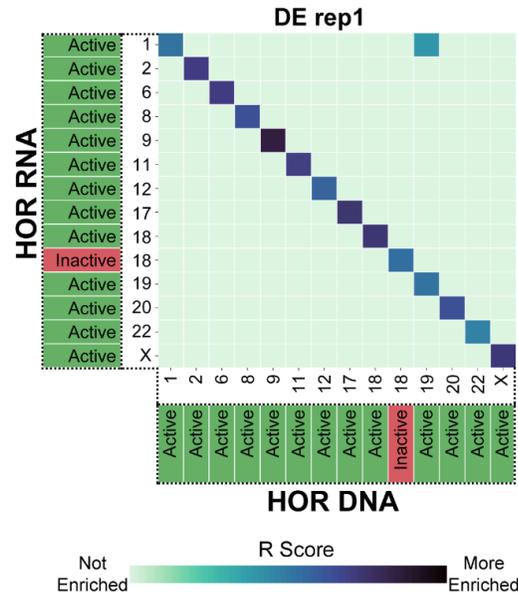
F Active HOR derived RNAs enriched locally ES rep2



G Centromeric RNAs enriched at centromeric repeats DE rep1



H Active HOR derived RNAs enriched locally DE rep1



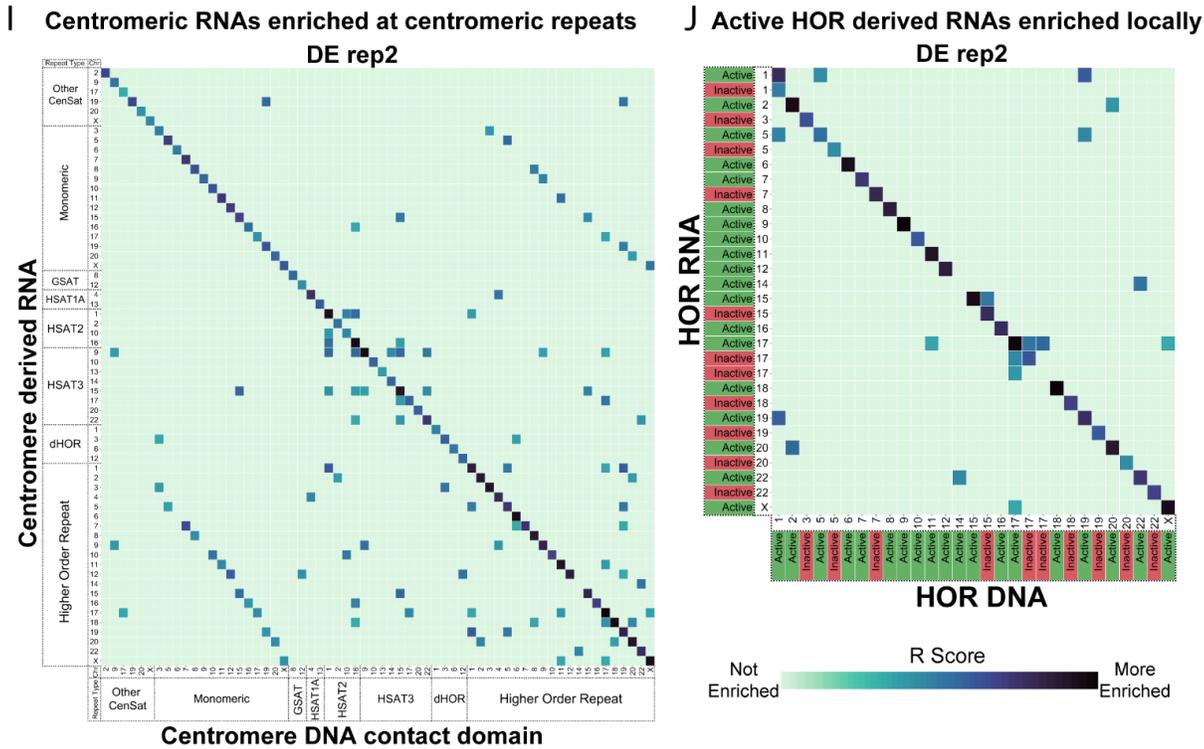
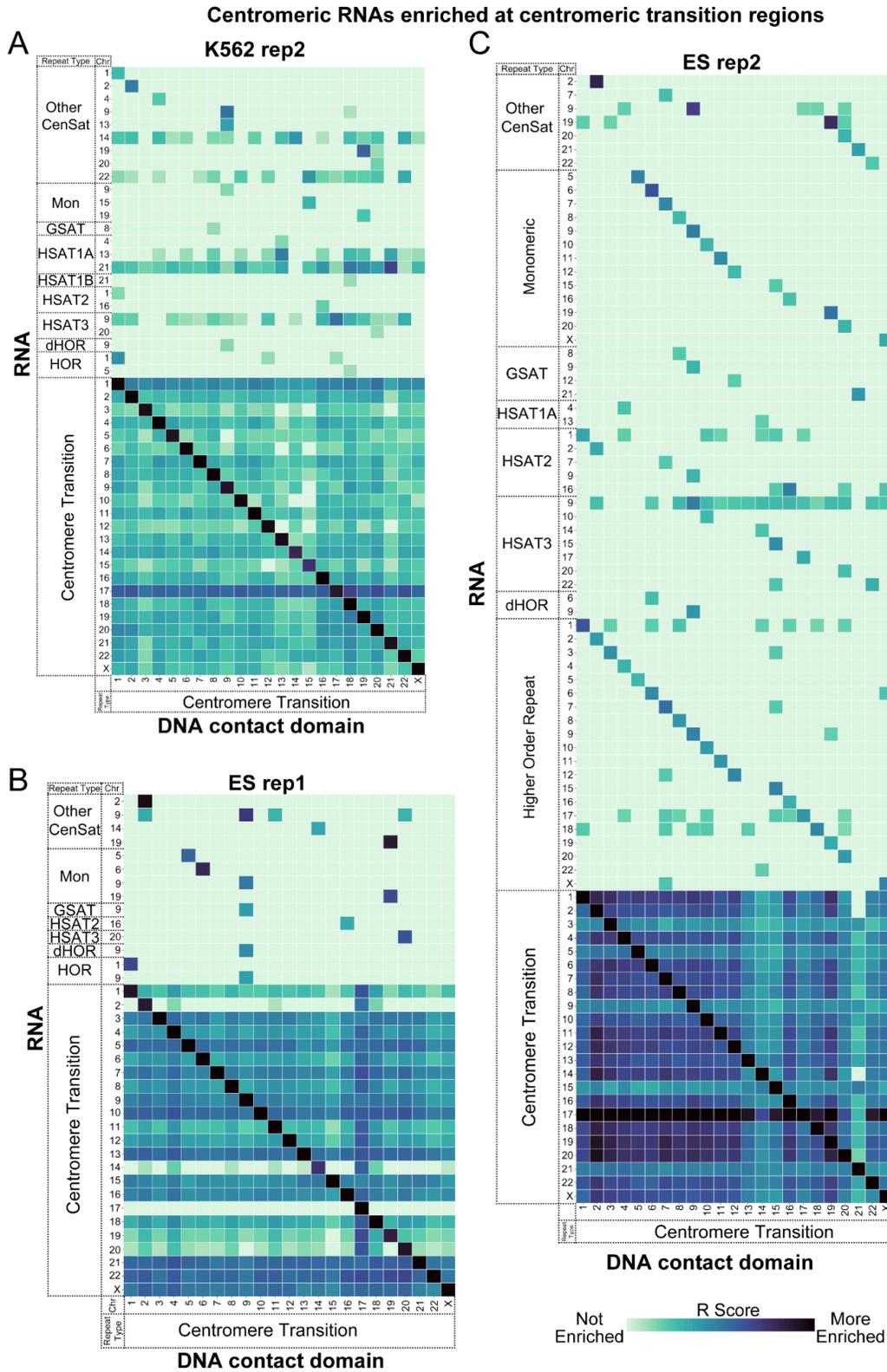


Figure S7: As in Figure 4A,B, heatmaps of R scores for centromere derived RNAs (y-axis) enriched at centromere DNA domains (x-axis) in **A-B**.K562 replicate 2, **C-F** ES replicates 1 and 2, and **G-J**. DE replicates 1 and 2. DNA and RNA reads were classified using CASK and denoted with domain type label and chromosome. Darker color indicates contacts with higher R scores (larger difference between observed number of reads and predicted) indicating stronger enrichment above background contacts. Contacts with R scores below the 95th percentile were considered not significantly enriched and were excluded from heatmaps. Reads classified as ribosomal on either RNA or DNA side, reads that could not be uniquely assigned to a single repeat type, and all contacts with fewer than 0.1CPM of total reads were excluded from heatmap.

A,C,E,G,I. Heatmap of R scores for class level assigned centromere derived RNAs enriched at centromere DNA domains in **A.** K562 replicate 2 **C.** ES replicate 1 **E.** ES replicate 2 **G.** DE replicate 1 **I.** DE replicate 2

B,D,F,H,J. Heatmap of R scores for repeat type level HOR derived RNAs enriched at HOR DNA domains in **B.** K562 replicate 2 **D.** ES replicate 1 **F.** ES replicate 2 **H.** DE replicate 1 **J.** DE replicate 2. Green label corresponds to T2T active HOR annotation denoting CENP-A containing HOR. Red, inactive label denotes HOR that does not contain CENP-A.

Supplemental Figure S8: Centromere derived RNAs enriched at centromere transition DNA domains



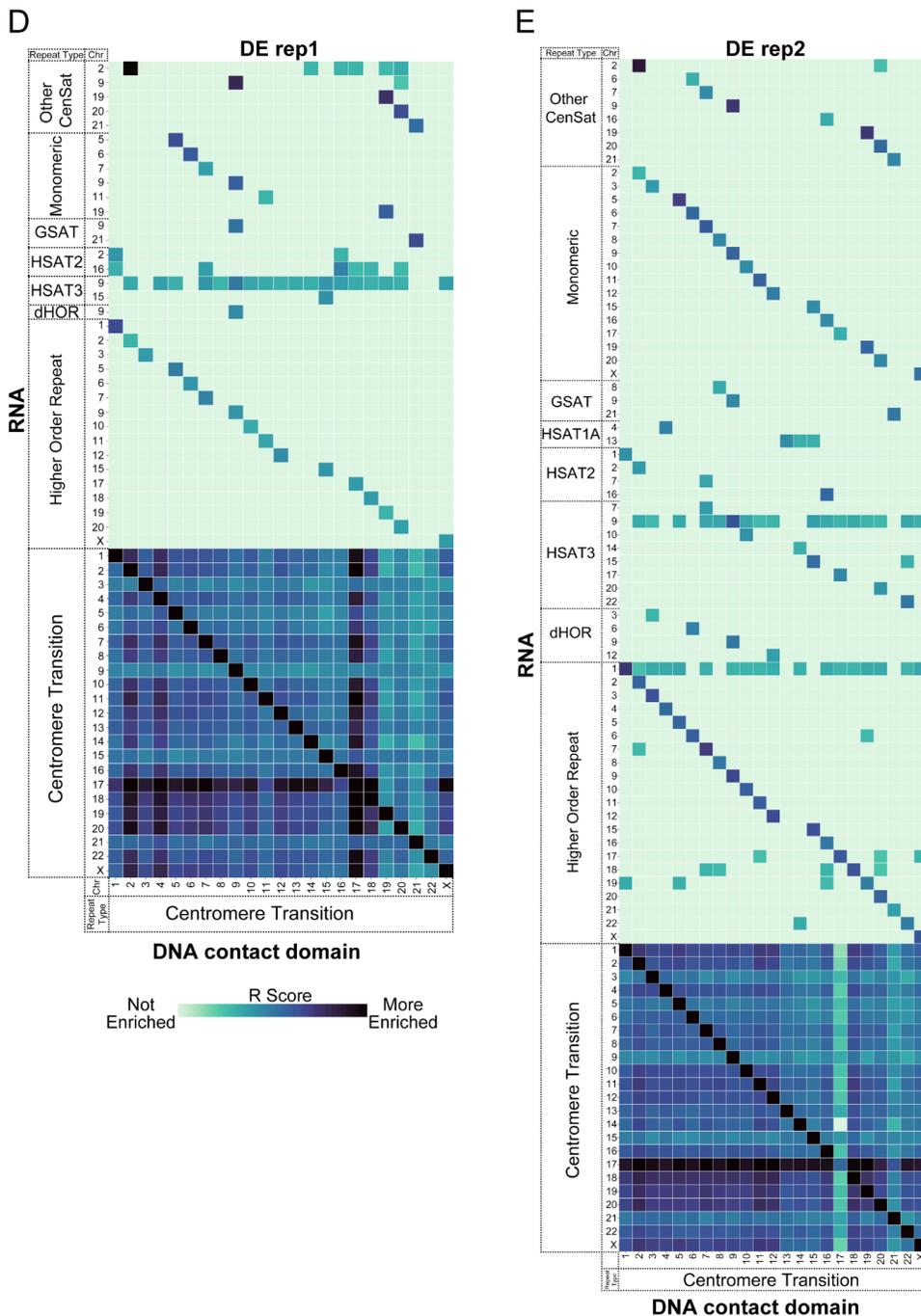
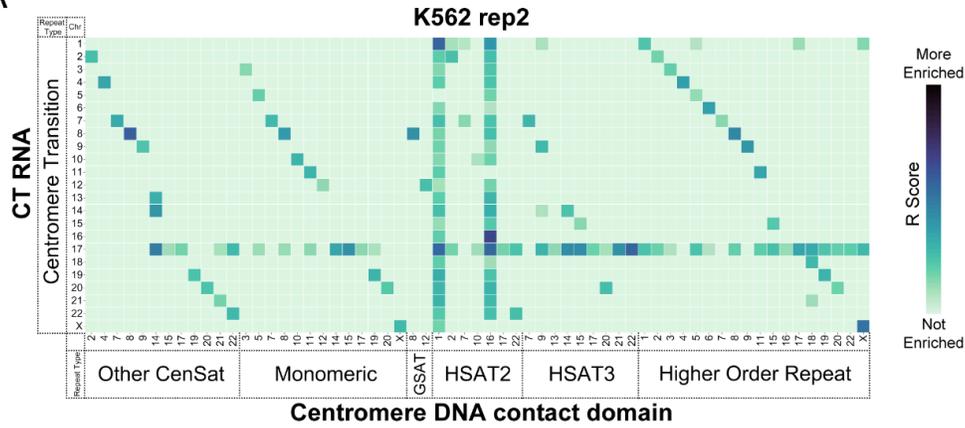


Figure S8: As in Figure 4C: heatmaps of R scores for centromere derived and centromere transition (CT) derived RNAs (y-axis) enriched at CT DNA domains (x-axis) in **A.** K562 replicate 2 **B.** ES replicate 1 **C.** ES replicate 2 **D.** DE replicate 1 **E.** DE replicate 2. DNA and RNA reads were classified using CASK and denoted with domain type label and chromosome. Darker color indicates contacts with higher R scores (larger difference between observed number of reads and predicted) indicating stronger enrichment above background contacts. Contacts with R scores below the 95th percentile were considered not significantly enriched and were excluded from heatmaps. Reads classified as ribosomal on either RNA or DNA side, reads that could not be uniquely assigned to a single repeat type, and all contacts with fewer than 0.1CPM of total reads were excluded from heatmap.

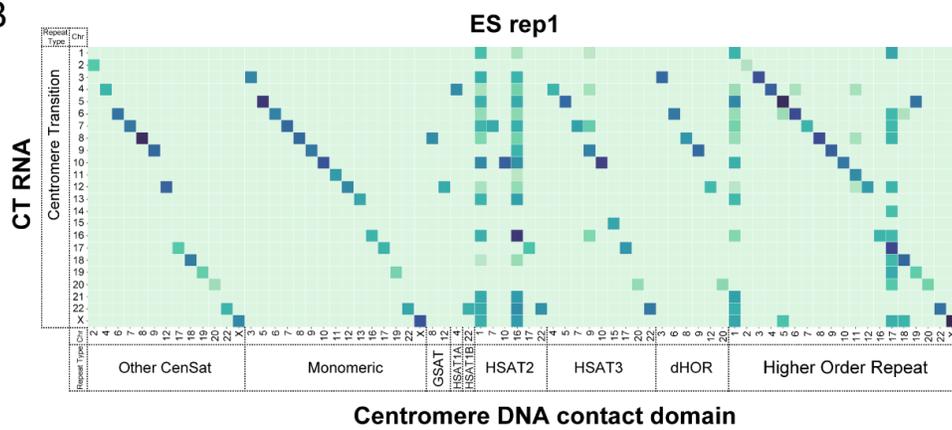
Supplemental Figure S9: Centromere transition derived RNAs enriched at centromere DNA domains

Centromere transition RNAs enriched at centromeric DNA domains

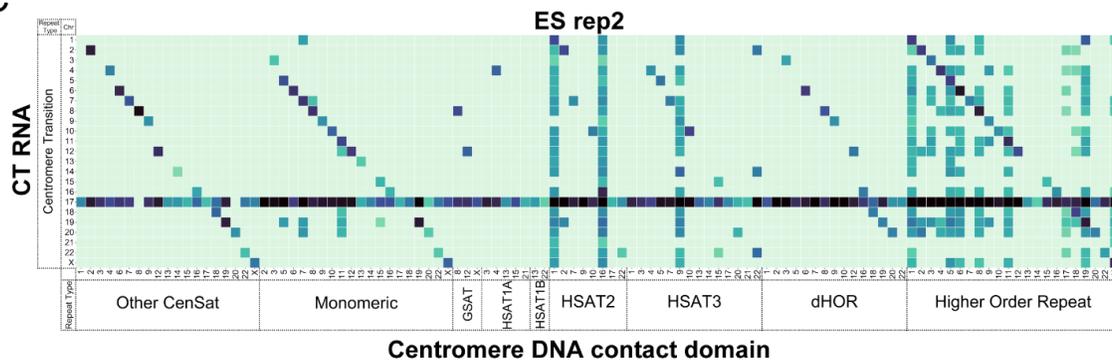
A



B



C



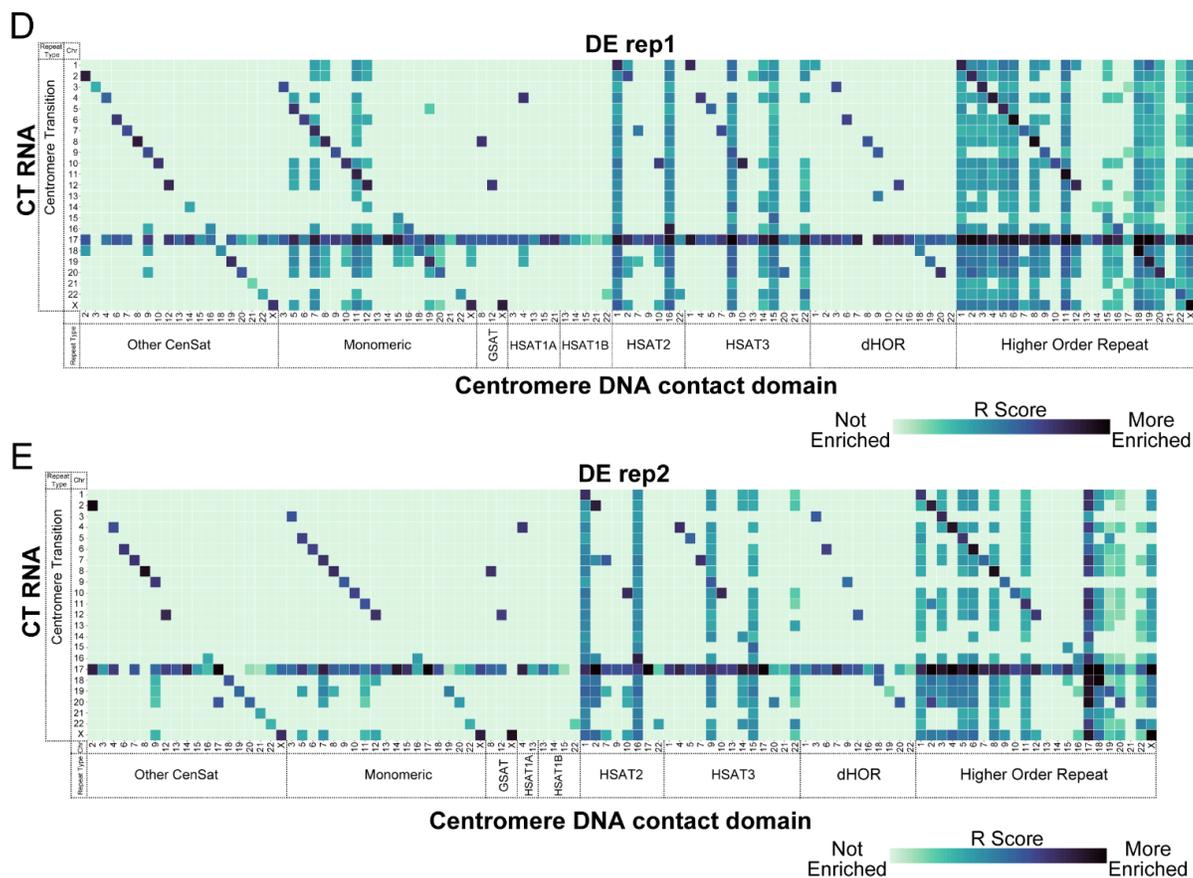


Figure S9: As in Figure 4D, heatmaps of R scores for centromere transition region (CT) derived RNAs (y-axis) enriched at centromere DNA domains (x-axis) in **A.** K562 replicate 2 **B.** ES replicate 1 **C.** ES replicate 2 **D.** DE replicate 1 **E.** DE replicate 2. DNA and RNA reads were classified using CASK and denoted with domain type label and chromosome. Darker color indicates contacts with higher R scores (larger difference between observed number of reads and predicted) indicating stronger enrichment above background contacts. Contacts with R scores below the 95th percentile were considered not significantly enriched and were excluded from heatmaps. Reads classified as ribosomal on either RNA or DNA side, reads that could not be uniquely assigned to a single repeat type, and all contacts with fewer than 0.1CPM of total reads were excluded from heatmap.