Savana's Prelim:
Sequencing centromeric RNAs (cenRNAs)

Lab Meeting
2/21/25

Objectives

1) Introduce centromeric RNAs
2) Sequencing Approach
3) Preliminary Results
4) Discuss my prelim aims → Feedback Welcome :)

Background - Centromeric Transcription

- Centromeric transcription happens across many eukaryotic species - from yeast to humans
  - From different centromere types and regions
- RNA Pol II Transcription
- Non-coding → both small and large
- Regulation across the cell cycle

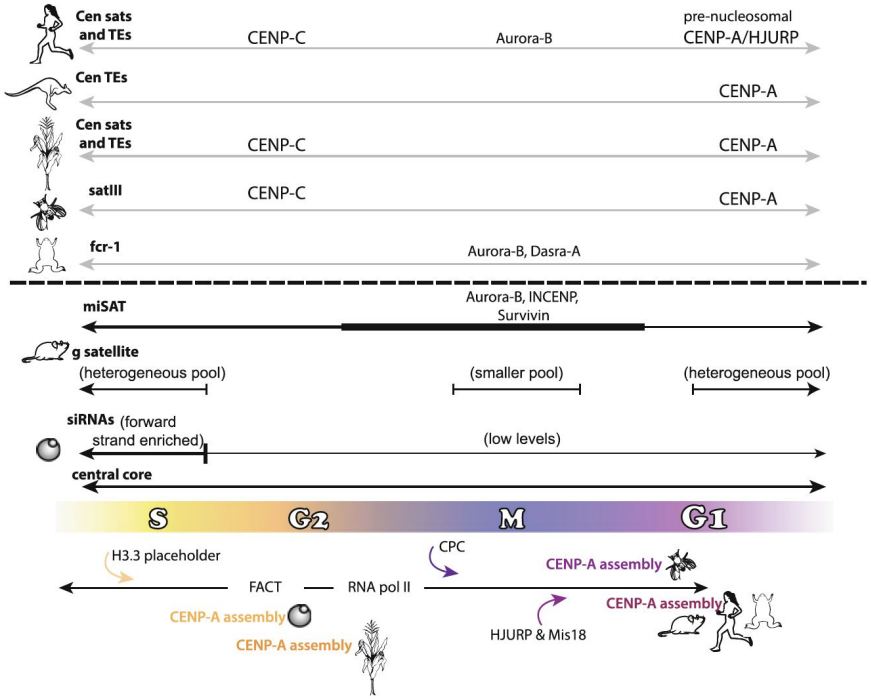


Figure from Duda, Trusiak and O'Neill (2017) Depicting timing, type, and known interactors of cenRNAs in different organisms.

Background - cenRNA regulation

- Human - **?**
  - Both a modulator of cell cycle progression and regulated by the cell cycle → potentially regulated by the proteasome (NFY-A)
- Yeast - Cbf1 transcription factor
  - Transcription begins in pericentromeric regions and stalls at Cbf1 (transcription factor)
  - In S-phase, Cbf1 is lost and transcription extends into the centromere

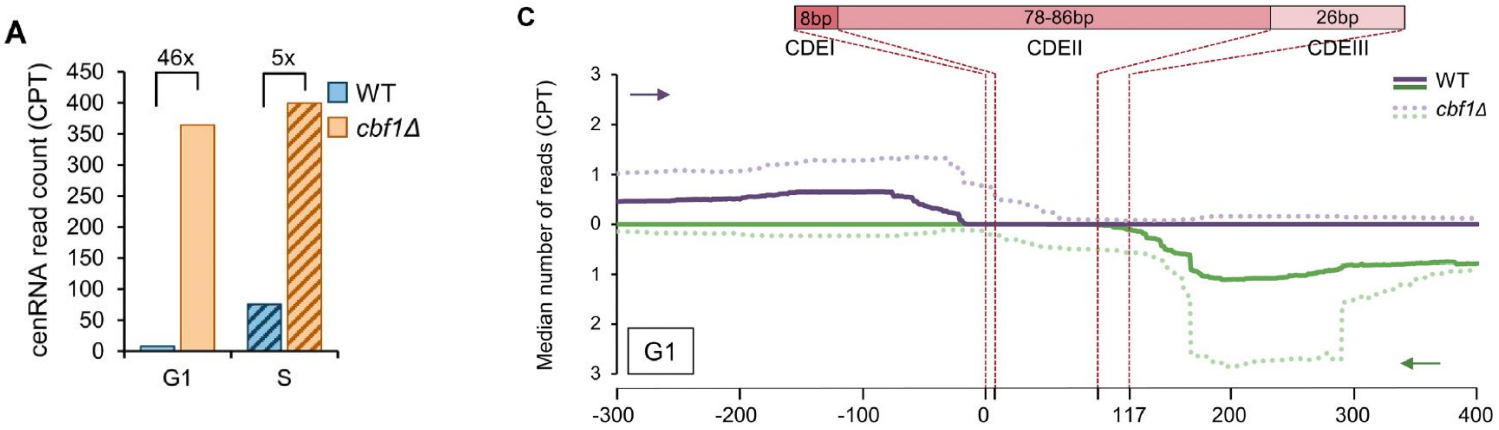


Figure 3 A & C from Hedouin *et al.* (2022) depicting cenRNA expression changes at G1 and S phase in WT and *cbf1* knockout cells.

Background - What do cenRNAs Do?

- Most research indicates roles in *cis*, but some in *trans* (non-human)
- Loss of cenRNAs across species leads to mis-segregation and centromere dysfunction
- Protein targeting studies show potential interactions with CENP-A, CENP-C, and Aurora B
  - CENP-A and CENP-C have RNA binding domains
  - cenRNA depletion interferes with CENP-A loading (human)

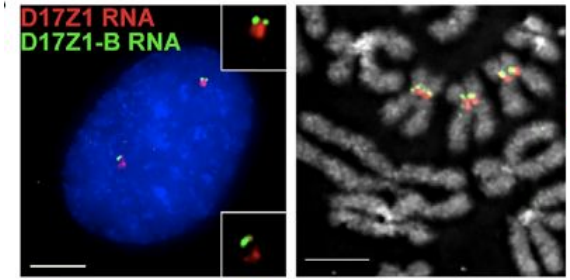


Figure 2B from McNulty *et al.* (2017) showing RNA FISH of cenRNAs on metaphase cells

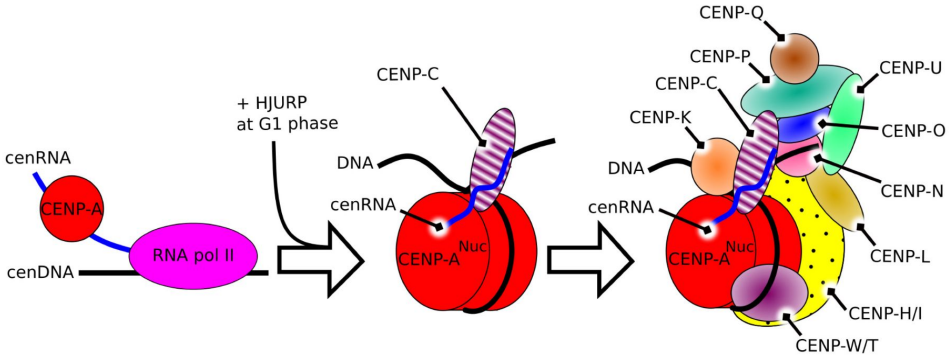


Figure 1 from Leclerc & Kitagawa (2021) showing a model for human cenRNAs' role in kinetochore formation

Background - Are cenRNAs Important?

Basic Biology

- Tight regulation implies yes → functionally necessary, regulatory, or something else?
- Prevalence across species indicates important role in potential centromere determination, function, and maintenance?

Disease Relevance?

- Changes in cenRNA (cenRNA) expression appear to interfere with appropriate chromosome segregation and increase chromosomal instability (CIN)
- Potential Roles in:
  - Cancer
  - Aneuploidies
  - Infertility

Background - What are we missing?

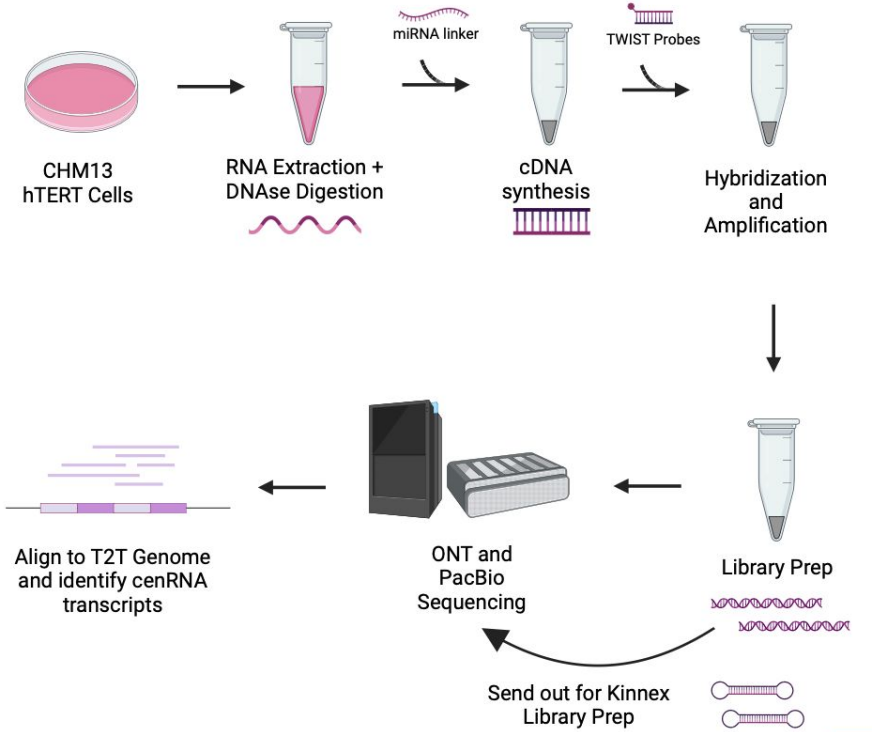
- Previous studies in human cells largely rely on FISH, qPCR, CHIP, or depletion to study cenRNAs → this limits our ability to study different RNA species and isoforms
- Where exactly are cenRNAs coming from and how are they regulated?

Targeted Long-Read Sequencing will allow us to:

- Map transcripts back to the genome to discover where they are transcribed
- Identify different isoforms
- Understand the diversity of cenRNAs within a single human cell type

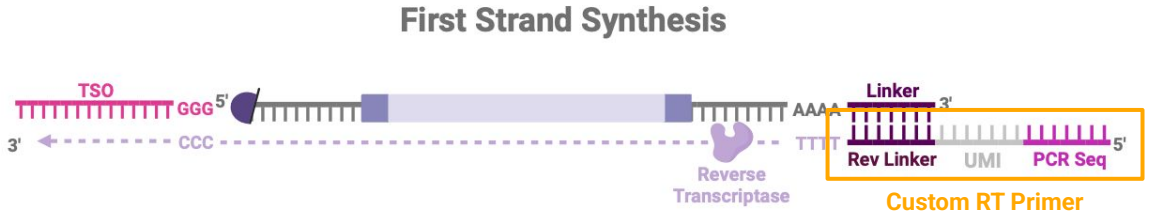
Long-Read Sequencing Approach

Workflow



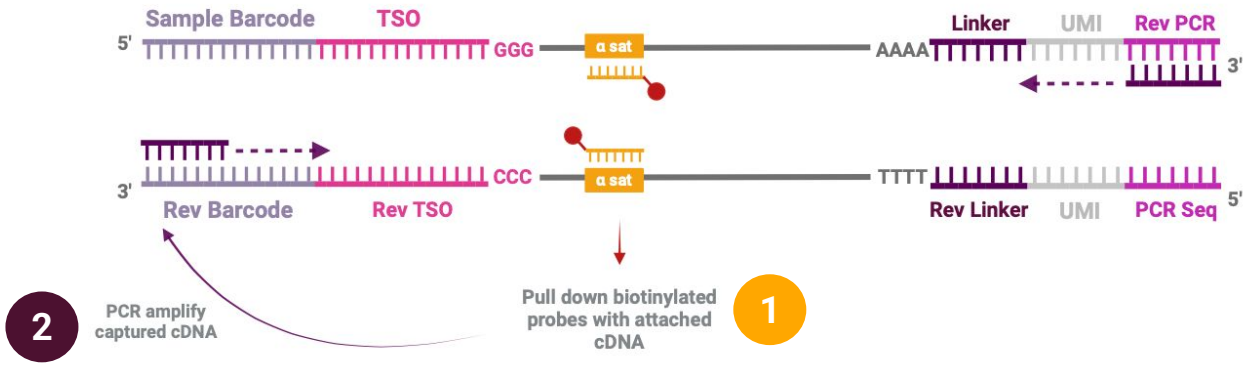
- CHM13 has the T2T reference genome, single haplotype (paternal)
- miRNA linker allows for 3’ amplification of ALL RNA transcripts (polyA and unpackaged)
- Using 95,459 probes covering 83Mb of alpha satellite DNA throughout the genome (100% coverage)

Library Prep - cDNA synthesis



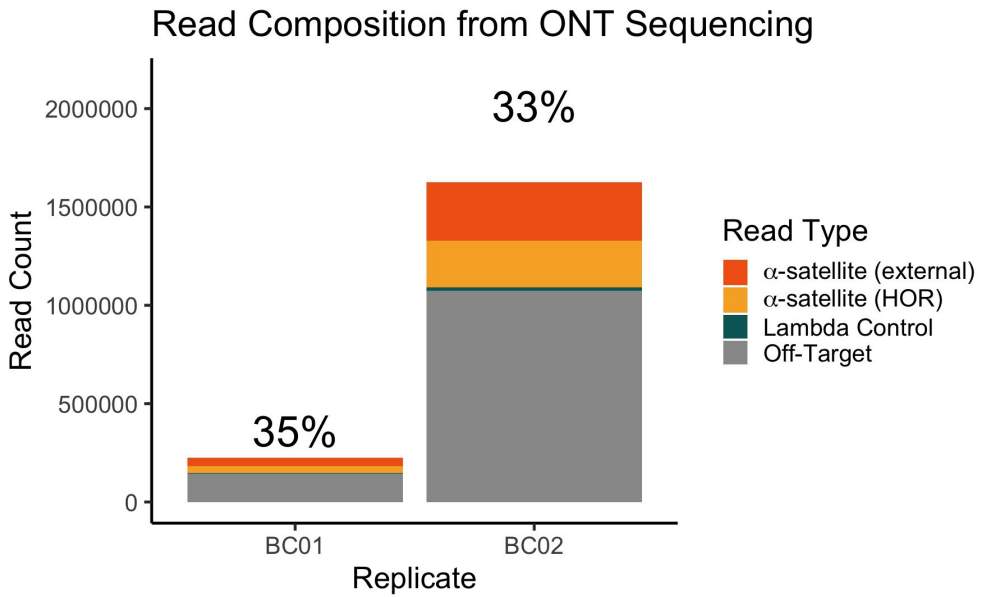
Iso-Seq PCR amplification did not work, used KAPA HiFi instead

Hybridization Capture

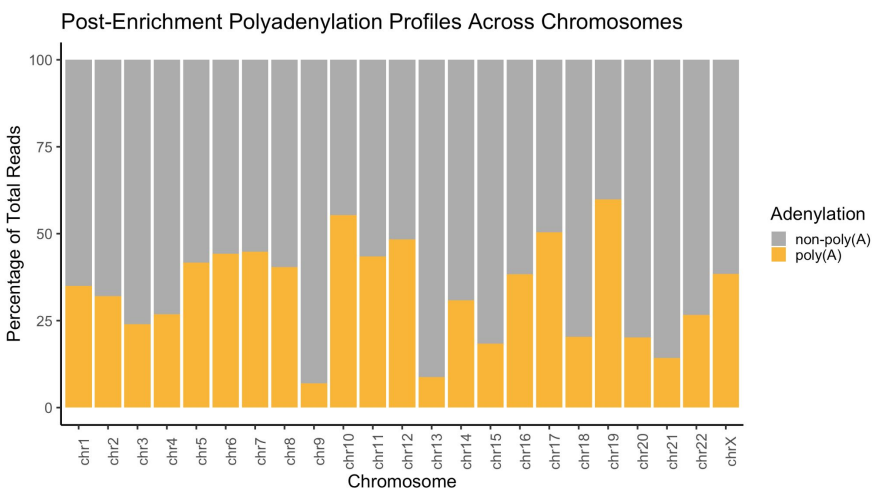


Preliminary Data

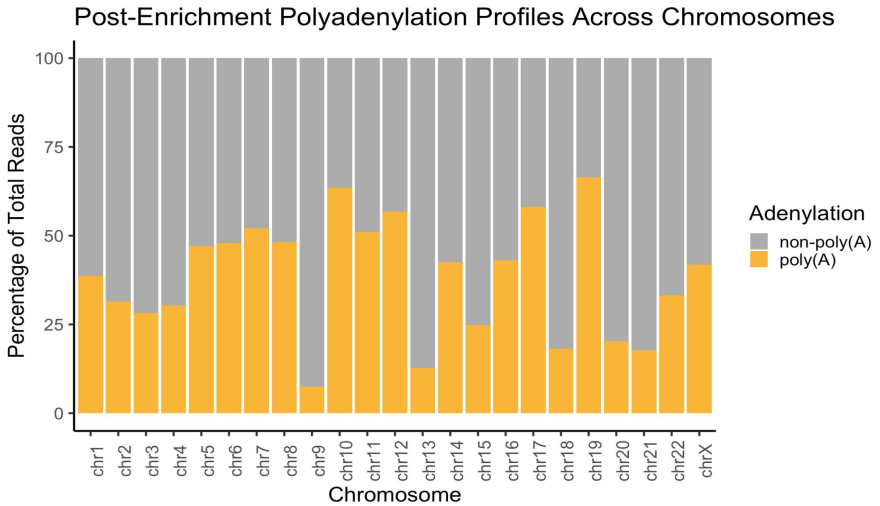
Results - Enrichment



Results - Polyadenylation

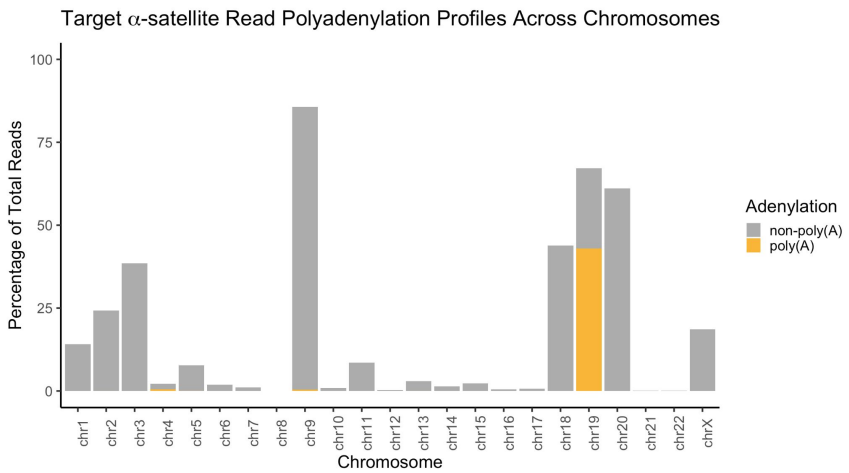


ONT

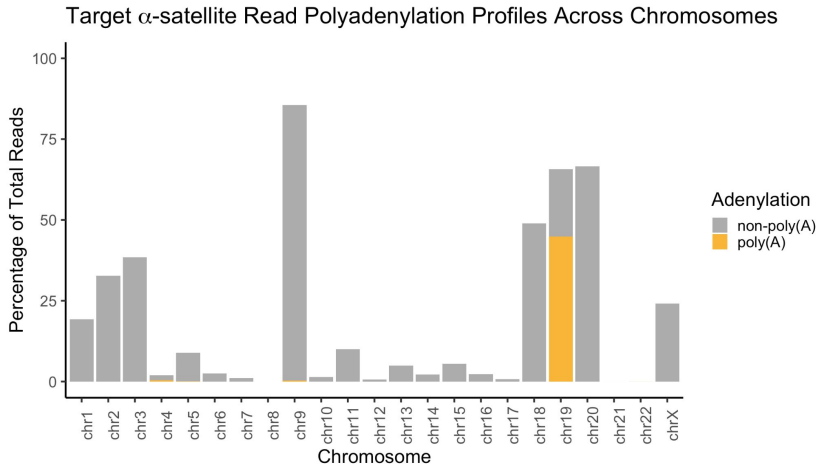


PacBio

Generally, polyA read counts are a little lower in ONT than PacBio
(BC01 36% in ONT and 42% in PacBio, BC02 is 27% in ONT and 31% in PacBio)

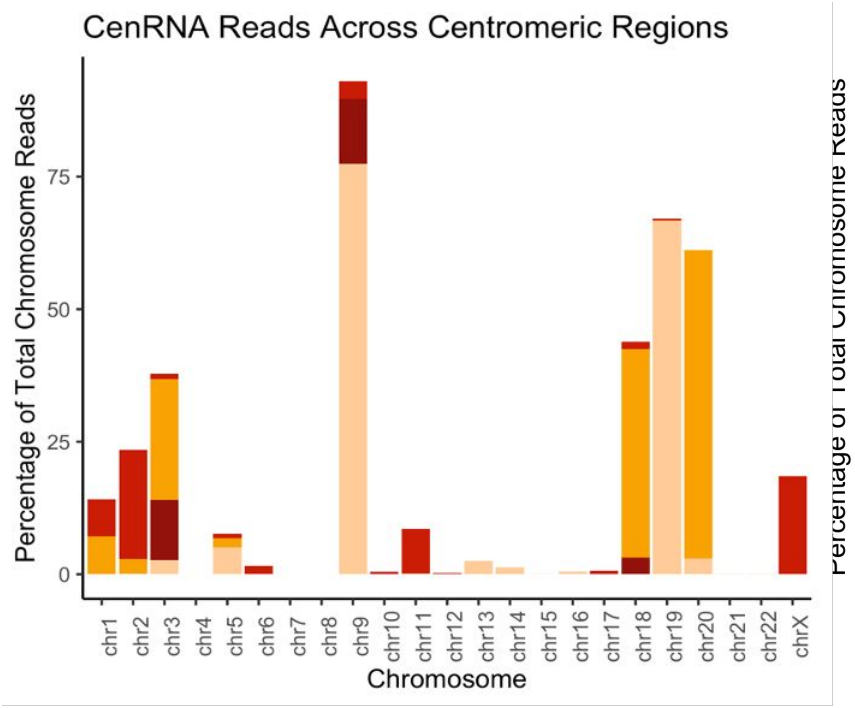

Target α-satellite Read Polyadenylation Profiles Across Chromosomes
Percentage of Total Reads
100
75
50
25
0
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
Chromosome
Adenylation
non-poly(A)
poly(A)

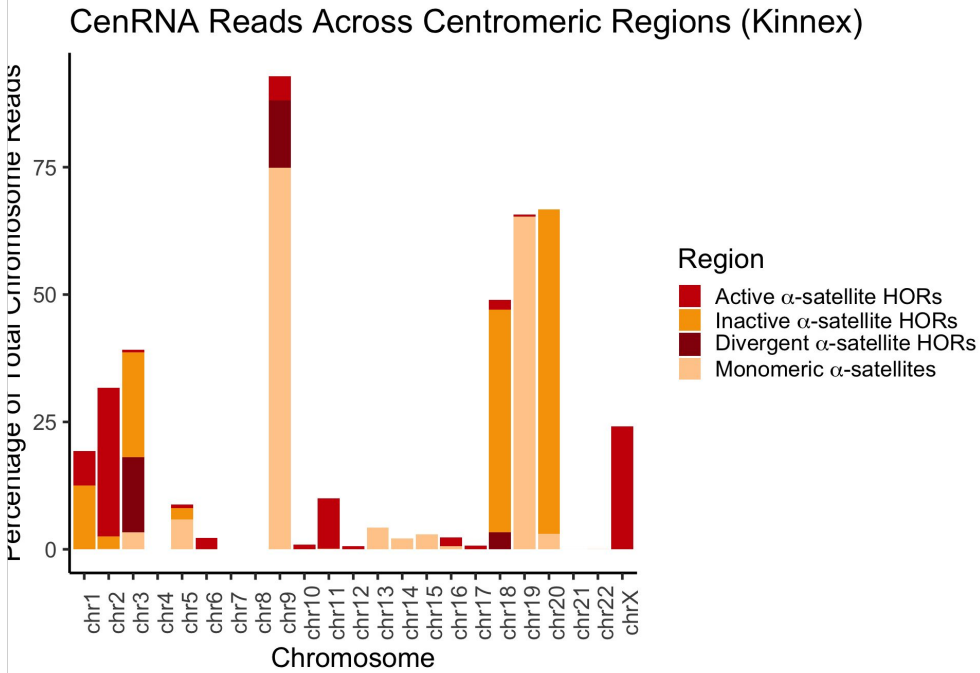
ONT


Target α-satellite Read Polyadenylation Profiles Across Chromosomes
Percentage of Total Reads
100
75
50
25
0
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
Chromosome
Adenylation
non-poly(A)
poly(A)

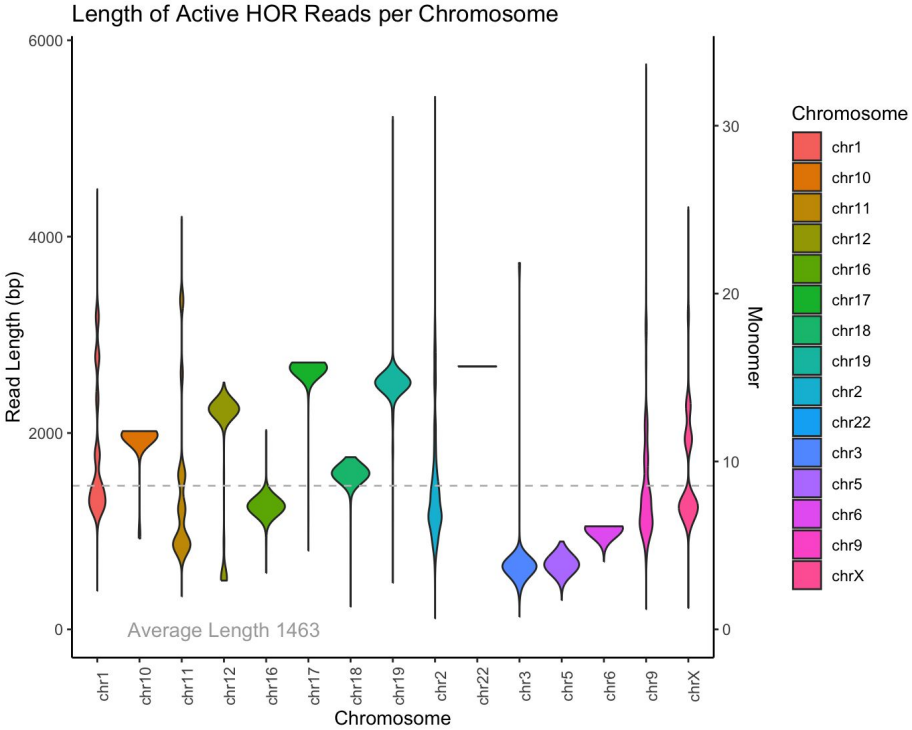
PacBio

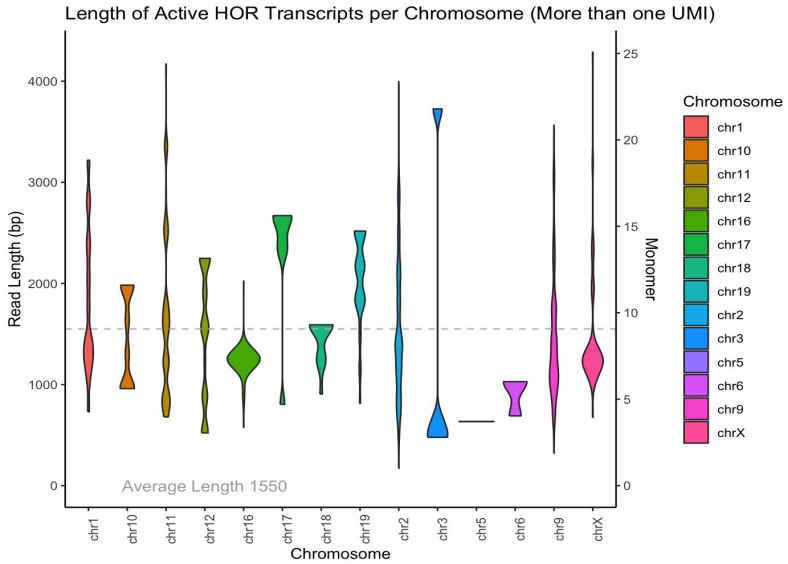
Proportion of reads per chromosome in Centromere





Active HOR Read length per chromosome (Kinnex)





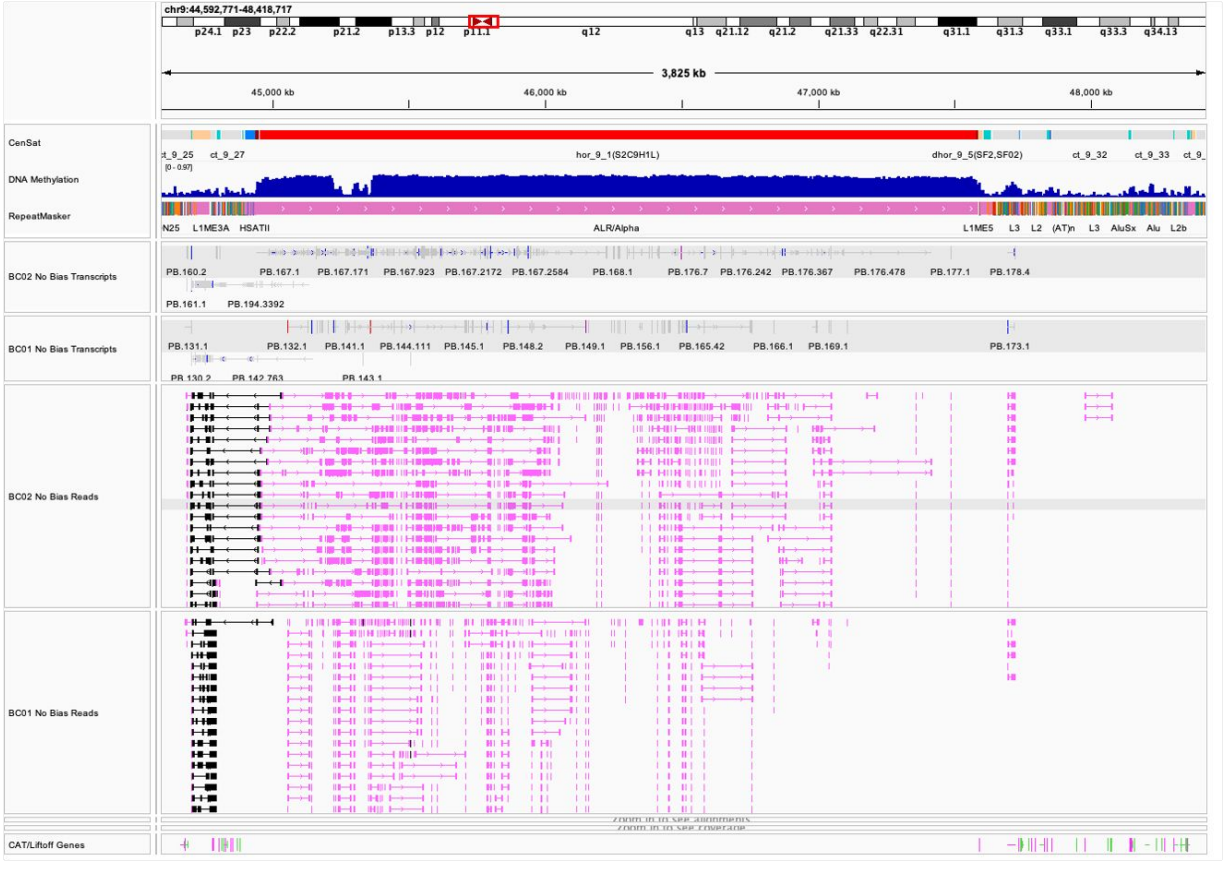
Summary

- Centromeric reads we are able to capture are largely not polyadenylated
- Reads come from different array types in each chromosomes and not all chromosomes produce cenRNAs (that we can capture)
- Despite differences in read depth and technology, read distributions look similar in both ONT and PacBio sequencing
- Average length of active HOR transcripts is 1.5 kb, however this varies across and within chromosomes

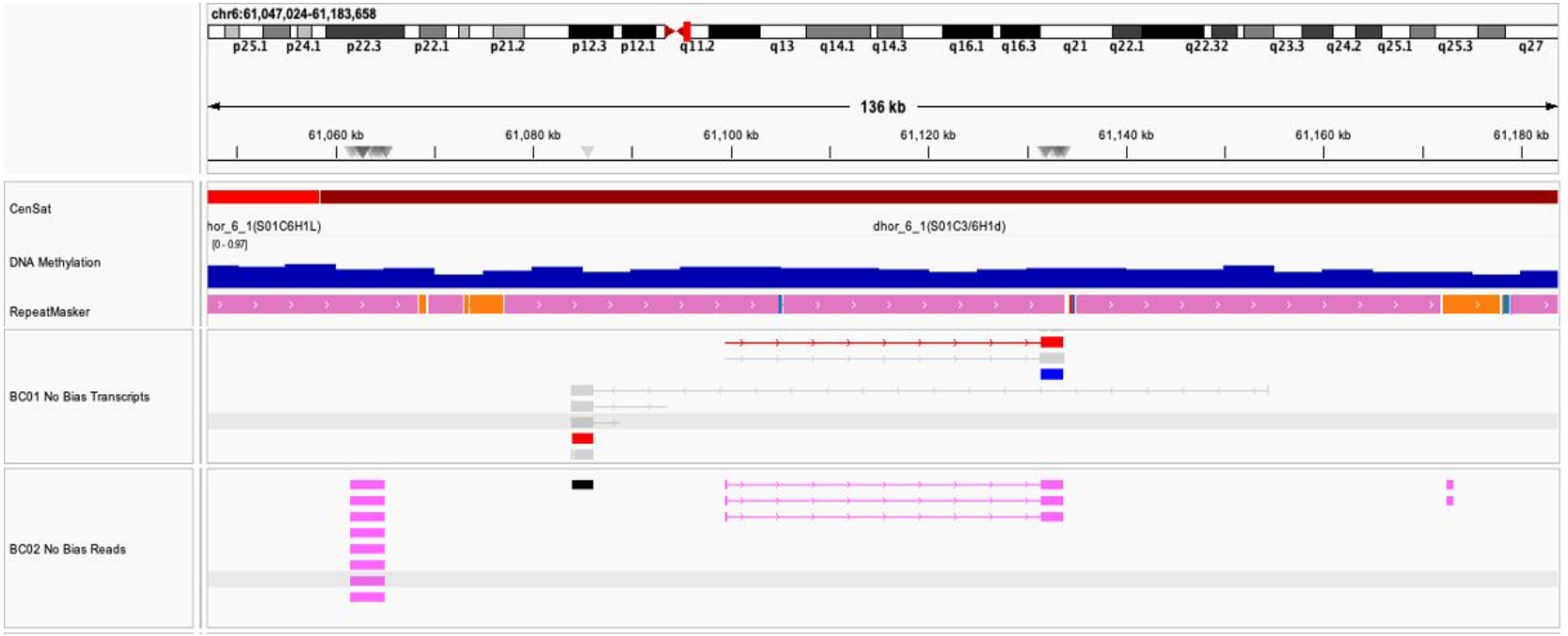
Example Transcripts

Chromosome 1

Chromosome 9



Chromosome 6 (dHOR)



Pericentromeric (monomeric) - Chr 16

Promoted off of L1PA2 - LINE TE

Prelim Discussion

What are prelims?

Qualifying exams involving:

1) Written proposal
    a) Grant style proposal with 2 aims
    b) Preference for hypothesis driven work
    c) Preliminary data is not required
2) Oral presentation on
    a) 5 slides
    b) 15 minutes without interruptions
    c) Anything related to slides/in proposal is fair game for further questioning

Aim 1: Variation of cenRNAs across human genomes

Objective:

To gain a comprehensive understanding of how cenRNAs vary across diverse genomes and between haplotypes

Hypothesis:

cenRNA transcription will vary both in sequence and locus of origin across individuals of diverse ancestries with unique centromeric haplotypes and methylation profiles

Significance:

Potential to elucidate a connection between, sequence, epigenetics, and cenRNA expression across individuals

Approach

1.1: Sequence cenRNAs across genomes

- Use Iso-Seq hyb-cap approach on diverse individuals from the HPRC and HGSVC
- Want to start by choosing 10 individuals across the 5 populations (2 of each)
- Samples with mostly(?) Complete centromeres across chromosomes
    - How many individual genomes do we have where every chromosome has both haplotypes assembled correctly?
    - Any recommendations?

1.2: Evaluate Structure and Modification

- Model potential secondary structures
- Perform native RNA sequencing on cenRNAs to look at modifications on select genomes

Limitations/Concerns

- Worried about effectively being able to map RNAs back to specific haplotypes given size of cenRNAs I've been able to pickup
- Potential that probes miss regions that are in these genomes and not CHM13 since probes were designed to CHM13…
  - Want to map probes to each assembled centromere, then check for large gaps of regions with less than 80% sequence similarity…may need to then add in those probes…
- **Question for you:** Any recommendations on what individuals to choose? Unique haplotypes and structures?

Questions/Suggestions?

Aim 2: Characterize the Function of cenRNAs in CHM13

Objective:

To evaluate the function of specific cenRNA transcripts and deconvolute the role of transcription from the role of cenRNA transcripts themselves

Hypothesis:

Human cenRNAs function in cis through transcriptionally facilitated epigenetic changes to ensure proper chromosome segregation

Significance:

Function of cenRNAs has implications in understanding and identifying disease contexts, as well as centromere determination

Approach

1.1: Perform Functional Knockdown Experiments in CHM13:

- CRISPRi (knockdown) dCas9 - prevent transcription
  - Also operates via KRAB domains, which may not be able to effectively prevent transcription since we don't really understand loci
- Cas13a or LNA knockdown of specific transcripts - test function of transcript themselves
- Test chromosome specific segregation defects and evaluate changes in epigenetic profiles (CENP-A, methylation)

1.2: cenRNA overexpression in CHM13:

- Introduce an overexpression vector to increased expression in *trans*
- Recruit RNA pol II via dCas 9 targeting (*cis* transcription)
- Evaluate chromosome segregation and cell-cycle defects (timing, cell stress, apoptosis, etc)

Validate expression changes with qPCR, sequencing or smFISH, FIBER-FISH, or Oligopaint?, which can provide more localization information

Limitations/Concerns

- CHM13 for functional characterization
  - They’re trophoblast like - generally not representative of epiblast expression
  - Altered expression profile due to lack of imprinting
  - We won’t have multiple haplotypes, which could be functionally relevant…
- dCas9 targeting
  - May run into limitations with sequence or specificity
  - Potential difficulty localizing to centromere due to tight packaging
- For overexpression
  - Promoter?
  - Localization of transcripts
- # of replicates (biological and technical)?

Questions/Suggestions?

Other Possible Experiments and Validation

Other Possible Aims

- Primate cenRNA sequencing
  - Part of the reason I'm not including this is because it's kinda a lot, expensive, and harder to write as a hypothesis driven aims
- Characterizing RNA modifications in cenRNAs
  - Where are they modified? → direct RNA ont-sequencing (limited by probe pulldown though)?
  - When?
  - Are certain transcripts modified while others aren't
  - Do modifications change across the cell cycle or in the presence of cell stress?
  - Does lack (or increase) of modifications via dCas9 or knockout on these RNAs cause chromosome mis-segregation?
- Pol II long-read ChIP (DeMeLo Seq or Timp's) → attempt to do across cell cycle?
  - Serves as a potential indicator of active transcription in the centromere

Supplementary Slides

Kinnex Sequencing Summary

Deconcatenated Reads	Barcoded Reads	%	Replicate	Reads	Replicate Proportion	Refine Command	Refined Reads	noA reads	Adenylation Proportion	Reads w/ Linker	% w/ Linker	% Post Barcoding	Total Yield
77259260	75770397	98.07	BC01	9681535	12.78	flnc all	9676429	5658761	**58.45**	5616153	99.25	99.42	**97.29**
						flnc polyA (min 10)	4017687		**41.50**	4009320	99.79		
			BC02	66088862	87.22	flnc all	66038930	45814692	**69.32**	45372953	99.04	99.17	
						flnc polyA (min 10)	20224467		**30.60**	20170097	99.73		