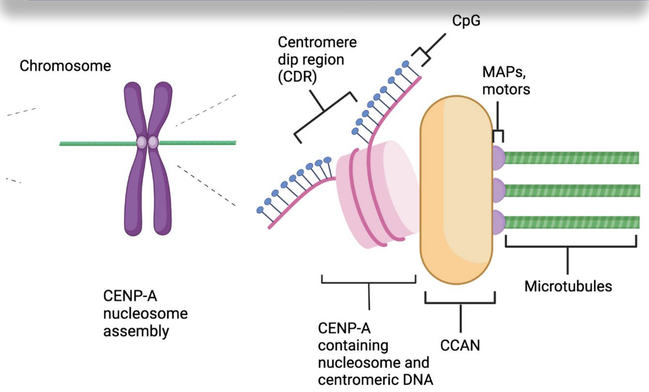
Determination of the centromere protein A landscape at single-molecule resolution

Yuan (Dan) Xu, Julian K. Lucas, Mobin Asri, Mira Mastoras, Brandy McNulty, Sara O' Rourke, Benedict Paten, Nicolas Altemose, Karen H. Miga
University of California, Santa Cruz, Santa Cruz, CA

Centromere regions play a critical role in cell division

- Centromeres are essential for proper chromosome segregation in mitosis and meiosis.
- A T2T entromere reference genome was published, but the detailed chromatin organization of human centromeres is still unknown.

CENP-A centromeric localization directs kinetochore assembly and attachment

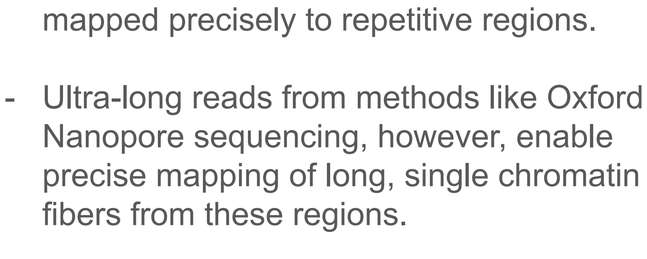


- Centromeres are defined and maintained epigenetically by centromere protein A (CENP-A) containing chromatin.
- CENP-A recruits the constitutive centromere-associated network (CCAN) complex and microtubules to initiate chromosome segregation.
- Active centromeres have high CpG methylation compared to inactive arrays. Inside of these CpG enriched areas, there are local regions showing reduced CpG methylation, identified as centromere dip regions (CDR). Previous research suggests the strongest CENP-A enrichment is within CDRs on all chromosomes.

Objective

- The spacing and density of CENP-A are still unknown, yet very crucial to understand part of the mechanism of chromosome segregation.
- We aim to use Directed-Methylation with Long-read sequencing (DiMeLo-Seq) **to study the epigenetic properties of centromeres, including centromere protein density and positioning.**

Method to study CENP-A



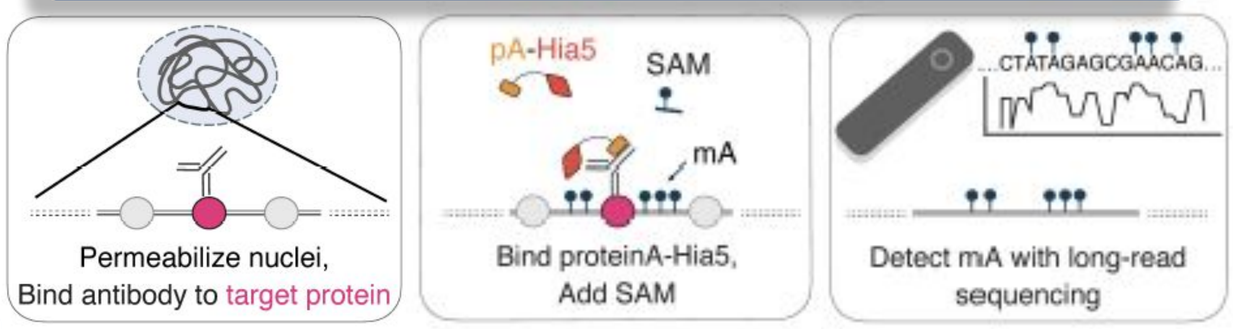
- So far, the field has been studying CENP-A using fluorescence microscopy
- This approach cannot reveal:

1. Exact length of the targeted regions
2. Complete sequence information in these regions
3. Detailed methylation status of the bases in these regions

Ultra-long provides comprehensive sequence info in the centromere regions

- ChIP-seq/CUT&RUN sequencing methods provide short-read sample data on protein binding sequences, which cannot be mapped precisely to repetitive regions.
- Ultra-long reads from methods like Oxford Nanopore sequencing, however, enable precise mapping of long, single chromatin fibers from these regions.

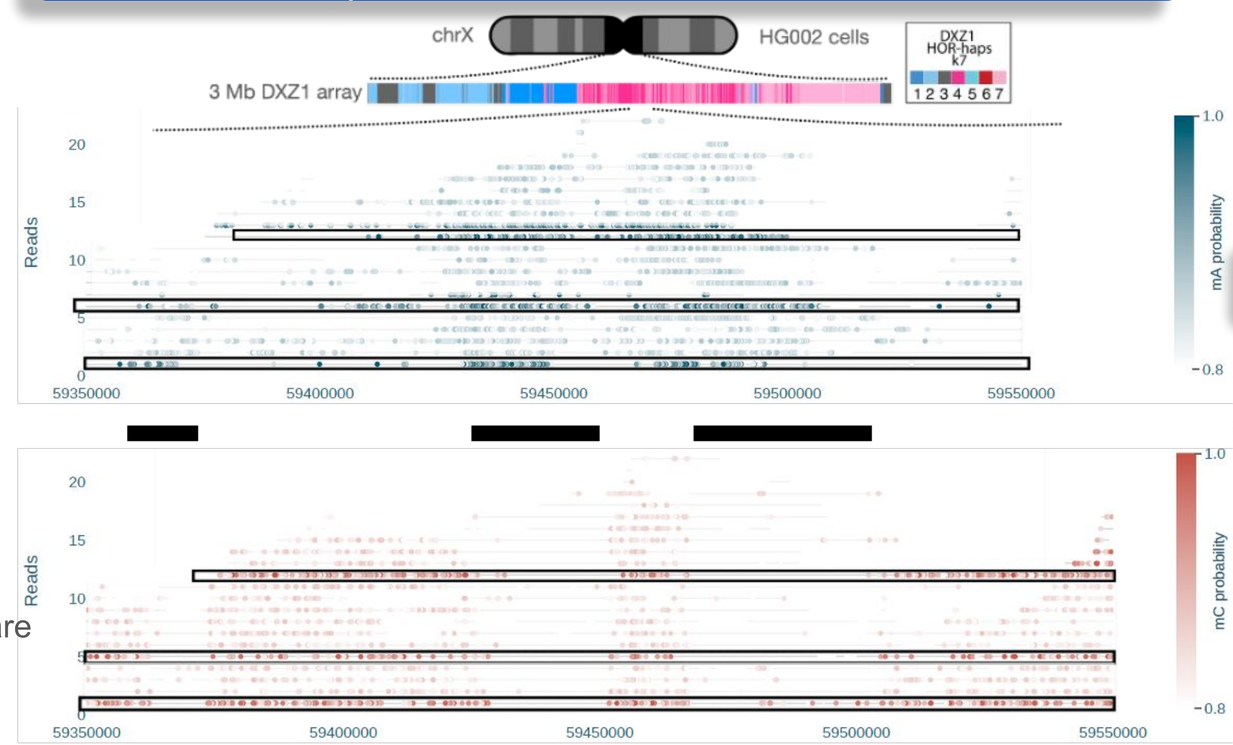
DiMeLo-seq identifies protein positioning and DNA methylation



Altemose, *et al.* DiMeLo-seq: a long-read, single-molecule method for mapping protein−DNA interactions genome wide. *Nat Methods* 19, 711−723 (2022).

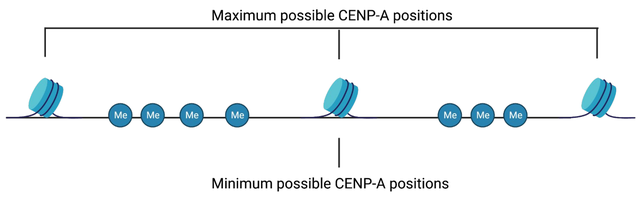
- DiMeLo-seq combines elements of antibody-directed protein-DNA mapping approaches to deposit methylation marks (mA) near a specific target protein.
- Long-read sequencing methods can read out exogenous (mA) and endogenous (mCpG) methylation tags.

DiMeLo-seq data reveal that CENP-A can localize at high density to multiple places on the same chromatin fiber



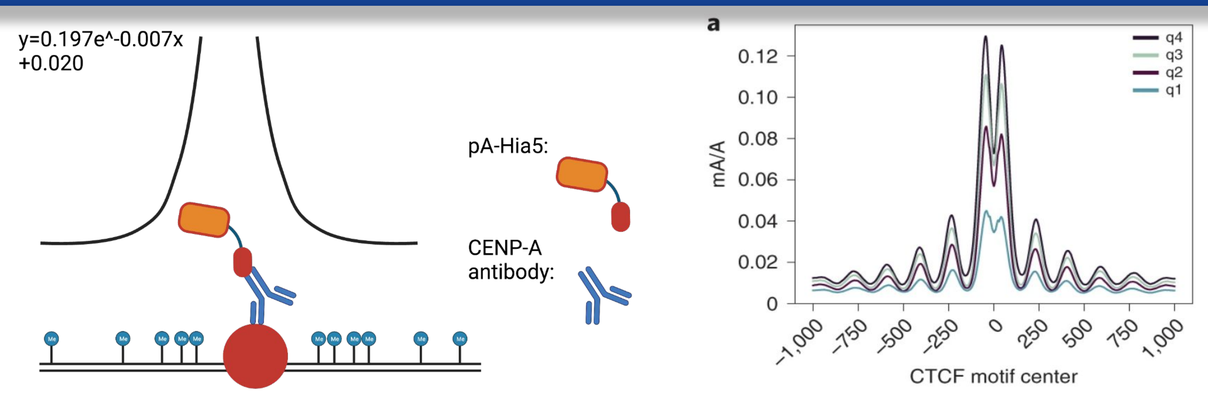
- "sub-CDRs" are clearly showing from mCpG distribution
- mA distribution suggests high-density CENP-A localization
- Multiple reads covering the three sub-CDR regions show that CENP-A can localize to different sub-CDRs simultaneously on the same single fibers

Data analysis approach and challenges



- Ideally, we would like to resolve the exact footprints of CENP-A nucleosomes, but there are many possible CENP-A positions.
- Determining the specific CENP-A distribution requires many assumptions.
- With the information we have in hand, it is more accurate to provide a range of maximum and minimum CENP-A possible positions on each read.

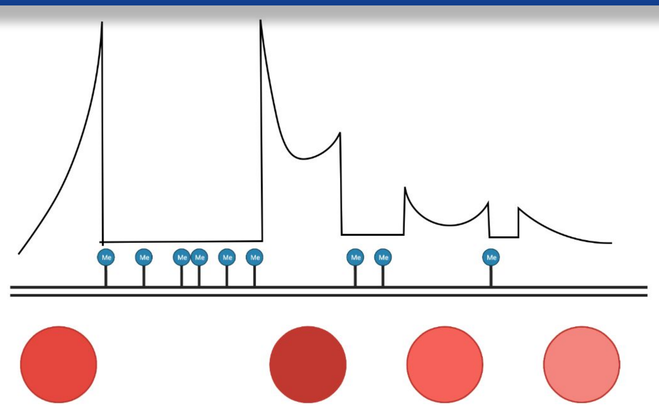
CTCF model suggests the distribution equation



Altemose, *et al.* DiMeLo-seq: a long-read, single-molecule method for mapping protein−DNA interactions genome wide. *Nat Methods* 19, 711−723 (2022).

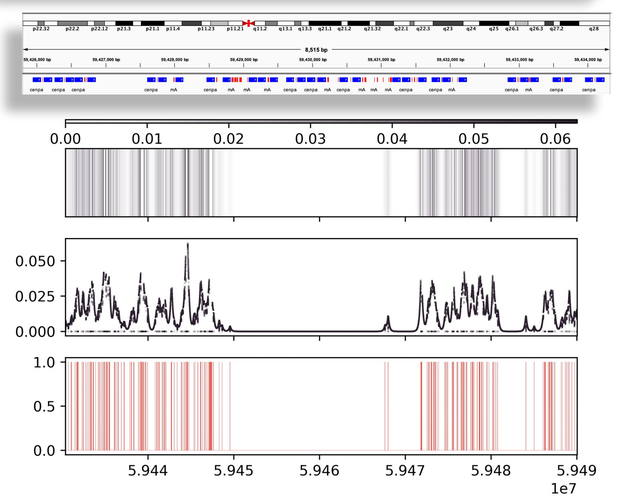
- Using CTCF as a benchmark model, the DiMeLo-seq publication provided the mA methylation decay distribution curve on the left and right side of pA-Hia5

The same model can be applied to CENP-A mA distribution



- The same methylation distribution decay curve can be applied to infer the landscape of possible nucleosome positions for each and every mA.
- All the non-mA bases (possible footprints) then have a sum methylation score from mAs within 1000 bps on both the left and right sides.
- Scores of bases between mAs that are less than 120 apart are forced to zero (nucleosome cannot fit).
- CENP-As are then placed at the highest scored positions in each and every 'gap'.

Sample results



- We were able to infer the maximum possible CENP-A positions.
- Additional work is going into sampling from the solution space to summarize the center and spread of possible position sets.

Future work

Benchmarking:
We will compare CENP-A ChIP-Seq aggregate data with CENP-A aggregate DiMeLo-seq data in a neocentromere

Mapping:
We will re-map the CENP-A DiMeLo-seq reads to the the T2T diploid HG002 assembly

Funneling:
We will then use the same CENP-A ChIP-seq aggregate data to help calibrate our inference of CENP-A positioning.

Future experiments:

1. DiMeLo-seq on CENP-B/C/T, H3K9me3, and H3K4me2
2. DiMeLo-seq on synchronized cells on the same protein panel

Citations

1. Cleveland. et al. Centromeres and kinetochores: From epigenetics to mitotic checkpoint signaling. Cell 112, 407–421
2. Dattoli, et. al. Asymmetric assembly of centromeres epigenetically regulates stem cell fate. *Journal of Cell Biology*, *219*(4)
3. Altemose. et al. DiMeLo-seq:A long-read,single-molecule method for mapping protein-DNA interactions genome-wide. bioRxiv 2021.07.06.451383;
4. Nechemia-Arbely. et al. Human centromeric CENP-A chromatin is a homotypic, octameric nucleosome at all cell cycle points. *J Cell Biol* 6 March 2017; 216 (3): 607–621.
5. Erliandri. et al. Replication of alpha-satellite DNA arrays in endogenous human centromeric regions and in human artificial chromosome, Nucleic Acids Research, Volume 42.
6. Régnier V, Vagnarelli P, Fukagawa T, Zerjal T, Burns E, Trouche D, Earnshaw W, Brown W. CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1. Mol Cell Biol. 2005 May;25(10):3967-81. doi: 10.1128/MCB.25.10.3967-3981.2005.
7. McGovern SL, Qi Y, Pusztai L, Symmans WF, Buchholz TA. Centromere protein-A, an essential centromere protein, is a prognostic marker for relapse in estrogen receptor-positive breast cancer. Breast Cancer Res. 2012 May 4;14(3):R72. doi: 10.1186/bcr3181. PMID: 22559056; PMCID: PMC3446334.