Docket #: S21-140

Methods for Measuring Protein-DNA Interactions with Long-read DNA Sequencing

Researchers at the University of California Berkeley and Stanford have developed a bespoke, high throughput method for determining DNA-protein interactions in highly repetitive portions of the human genome.

Protein-DNA interactions are essential to the reading, regulation, and replication of genomic DNA. Systematic mapping of these interactions provides key insights into vital cellular functions in both healthy and disease states. Current approaches to determining genome-wide interactions with a protein of interest (ChIP-Seq, CUT & RUN) were developed using short-read next generation sequencing (NGS), where short reads (~200bp) are mapped to a reference genome by stitching together uniquely overlapping sequencing reads. However, some highly repetitive regions of the human genome are unable to be correctly assembled using this technology because short repetitive sequences often lack unique overlaps that assembly algorithms rely upon. Long-read sequencing technology has made it possible to assemble some of these genomic regions for the first time as recently as 2020, and has elucidated their role key biological processes including chromosome segregation, nuclear organization, and transcriptional regulation. Additionally, current techniques rely on amplification-based readouts which have intrinsic biases and therefore only provide a semi-quantitative picture of protein-DNA interaction frequencies.

Stage of Development
Research-

in vitro

Stage of Research

The inventors of the new method, DiMeLo-Seq, establish a novel way to quantify DNA-protein interactions in previously unexplored, highly repetitive portions of the human genome. This method combines elements of previously established antibody-directed protein-DNA mapping techniques with nanopore sequencing to elucidate interactions. Briefly, cells are permeabilized and incubated with a primary antibody to target a DNA-interacting protein of interest. Next, a non-specific deoxyadenosine methyltransferase fused with antibody binding protein A (pA-Hia5) is allowed to bind to the primary antibody and all unbound pA-Hia5 is washed away. Nuclei are then incubated with methyl donor S-adenosylmethionine (SAM) which allows for adenosine methylation near the protein interaction of interest. Finally, DNA is isolated and sequenced using modification sensitive long-read sequencing to detect the frequency of adenosine methylation in genomic regions of interest, providing a quantitative readout of genomic sites of protein-DNA interaction.

Technology Reference

CZ Biohub ref. no. CZB-205B
Stanford ref. no. S21-140
UCB ref. no. BK2021-127

Applications

- Map multiple DNA-protein interaction sites on single DNA molecules
- Infer protein interaction frequency at each genomic site across cells
- Map haplotype-specific DNA interactions with proteins of interest
- Identify protective or maladaptive DNA-protein interactions in healthy and disease states

Advantages

- This technology is able to map DNA-protein interactions in previously unexplored, highly repetitive portions of the genome up to 1,000 kilobases in length
- Method allows for fully quantitative measurements of DNA-protein interactions whereas other current methods are only semi-quantitative

- Compatible with both Oxford Nanopore and PacBio high fidelity (HiFi) sequencing technologies

Publications

- Altemose, N., Maslan, A., Smith, O.K. et al. "DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome wide." Nat Methods 19, 711–723 (2022).

Patents

- Published Application: WO2022256469
- Published Application: 20240240234

Innovators

- Aaron Straight
- Kousik Sundararajan
- Owen Smith
- Aaron Streets
- Annie Maslin
- Nicolas Altemose

Licensing Contact

Kimberly Griffin

Technology Licensing and Strategic Alliances Manager

Email