

Reviewed Preprint

v1 • October 29, 2024

Not revised

Reviewed Preprint

v2 • March 23, 2026

Revised by authors

✉ For correspondence:

beliveau@uw.edu

dkschwep@uw.edu

* Author notes: See [page 31](#)

Funding: See [page 31](#)

Reviewing editor: Nicolas Altemose,
Stanford University, United States

© 2024, Liu et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

DNA O-MAP uncovers the molecular neighborhoods associated with specific genomic loci

Yuzhen Liu^{1,2,3,*}, Christopher D McGann^{1,*}, Conor P Herlihy^{1,2,*}, Mary Krebs^{1,2}, Thomas A Perkins¹, Rose Fields¹, Conor K Camplisson^{1,2}, David Z Nwizugbo^{1,2}, Qiaoyi Lin^{1,2}, Nicolas J Longhi^{1,2}, Chris Hsu^{1,2}, Shayan C Avanesian^{1,2,3}, Ashley F Tsue^{2,4,5}, Evan E Kania^{2,4,5}, David M Shechner^{2,4,5}, Brian J Beliveau^{1,2,5} ✉, Devin K Schweppe^{1,2,5} ✉

¹Department of Genome Sciences, University of Washington, Seattle, United States • ²Brotman Bay Institute for Precision Medicine, Seattle, United States • ³Molecular and Cellular Biology Program, University of Washington, Seattle, United States • ⁴Department of Pharmacology, University of Washington, Seattle, United States • ⁵Institute of Stem Cell and Regenerative Medicine, University of Washington, Seattle, United States

eLife Assessment

This study presents an **important** new method for probing the DNA and proteins associated with targeted chromatin domains in cells. The authors present **solid** evidence that the method can map DNA-DNA interactions for individual loci and can detect proteins enriched near repetitive DNA loci or targeted gene clusters. The methodological details of this study will be of particular interest and utility to chromatin biologists.

<https://doi.org/10.7554/eLife.102489.2.sa4>

Abstract

The accuracy of crucial nuclear processes such as transcription, replication, and repair, depends on the local composition of chromatin and the regulatory proteins that reside there. Understanding these DNA-protein interactions at the level of specific genomic loci has remained challenging due to technical limitations. Here, we introduce a method termed “DNA O-MAP”, which uses programmable peroxidase-conjugated oligonucleotide probes to biotinylate nearby proteins. We show that DNA O-MAP can be coupled with label-free or sample multiplexed quantitative proteomics, targeted chemical perturbations, and next-generation sequencing to quantify DNA-proximal proteins and DNA-DNA interactions at specific genomic loci. Furthermore, we establish that DNA O-MAP is applicable to both repetitive and unique genomic loci of varying sizes, from kilobase *HOX* gene clusters to megabase alpha-satellite repeats, and that DNA O-MAP can measure proximal molecular effectors in a homolog-specific manner.

Introduction

Eukaryotic cells store their genetic material in the form of chromatin, a DNA-protein complex. The function of a eukaryotic DNA locus is executed through the cooperation between its nucleotide sequence and the hundreds of protein factors assembled around it. DNA-protein interactions thus play a fundamental role in regulating both the genome’s structure and message storing functions¹. Therefore, developing methods to decipher DNA-protein interactions in cells has been a focus of technology development efforts for decades². For instance, chromatin immunoprecipitation followed by sequencing (ChIP-seq³), which has emerged as a core technology for epigenomics⁴, surveys the genome-wide binding profile of a target DNA-associated protein. ChIP-seq and related technologies (e.g., DamID⁵, CUT&Tag⁶) have produced an abundance of high-quality datasets that enabled the establishment of database consortia such as ENCODE^{7,8} and IHEC⁹, and significantly

accelerated chromatin state annotation efforts^{10,11}. Such methods, which profile DNA-protein interactions through a protein-centric lens, require the *a priori* knowledge of which protein(s) to target, and rely on the availability of suitable reagents such as antibodies or genetically engineered cell lines. By targeting a single protein at a time, these methods also inherently ignore the context of protein complexes or transient interactions that may be present at a given locus.

In addition to methods that profile the DNA bound by specific proteins, efforts have been dedicated to addressing the inverse problem—identifying the full collection of proteins assembled on a given DNA locus^{12–15}. Such methods include the foundational proteomics of isolated chromatin segment (PICH) technology, which uses a biotinylated oligonucleotide (oligo) probe to affinity label specific genomic DNA intervals via *in situ* hybridization (ISH)¹⁶. To enhance the stability of probe-chromatin interactions throughout the purification workflow, PICH utilizes oligos containing locked nucleic acid residues¹⁷, which are highly efficient as hybridization probes against repetitive DNA targets but cost-prohibitive to use to target non-repetitive intervals that require dozens to hundreds of probes to properly target a region of interest¹⁸. As noted in follow-up work, PICH was effective for repeat sequences but would require significant additional work to extend to more complex genomic sequences¹⁹. Additionally, even with the increased stability gained from the use of locked nucleic acid probes, the probe-chromatin hybrids can be difficult to maintain when coupled with stringent purification washes¹⁹. As a consequence, an input of one trillion cells was used for a single purification and identification of proteins interacting with telomeres¹⁶.

To reach a higher degree of enrichment, which is critical for lower abundance DNA targets, an alternative strategy is to directly biotinylate the proteins that occupy a target DNA locus. This biotinylation can be achieved via targeted proximity labeling using promiscuous biotin ligases^{20,21} or the engineered ascorbate peroxidase (APEX/APEX2) enzymes^{22,23}. Since the development of APEX, several methods including C-BERST¹² and GLoPro¹³, have combined APEX with CRISPR genome targeting to endow it with locus specificity. This involves fusing APEX to a catalytically dead RNA-guided nuclease, Cas9 (dCas9) and directing the fusion enzyme to a specific locus of interest by single guide RNAs (sgRNAs). The locus-docked dCas9-APEX biotinylates the neighboring proteins on electrophilic amino acid side chains, such as tyrosine, enabling protein purification and subsequent identification by mass spectrometry. In the case of GLoPro, APEX-based proximity labeling reduced the input required for each replicate analysis to ~300 million cells—a 10-fold reduction in cell input compared to PICH. Most recently, an approach termed TurboCas²⁴ introduced the combination of dCas9 fused to the miniTurbo²⁵ biotin ligase, enabling detection of locus-specific proteins from 50 million cells per replicate. Nevertheless, a notable limitation of CRISPR-guided proximity labeling is the requirement of the fusion dCas9-APEX or dCas9-miniTurbo enzyme and sgRNAs in a suitable host cell line. Since a successful locus purification canonically requires tens to hundreds of millions of cells, if not more, most current methods aim to create stable cell lines for this purpose. These requirements limit the use of previous locus proteomics methods since efficient and well-tolerated gene delivery remains a major challenge and considerable effort in primary cells²⁶. In addition, the labeling reagents necessary for APEX-based proximity labeling—hydrogen peroxide and biotin phenoxyl radicals—are toxic to cells and living organisms, limiting the use of CRISPR-based peroxidase labeling to cell lines amenable to genetic engineering. Thus, an unmet need exists for extensible methods capable of scaling and profiling multiple genomic loci.

We address these technical limitations by introducing DNA O-MAP, a locus purification method that uses oligo-based ISH probes to recruit peroxidase activity to specific DNA intervals. DNA O-MAP builds on our previously introduced RNA O-MAP²⁷ and pSABER²⁸ techniques, which target peroxidase activity to specific RNAs and RNA/DNA intervals for purification or visualization, respectively. Here, we describe a cost-effective and scalable bulk hybridization and biotinylation workflows capable of processing millions of cells in parallel in just a few days, and demonstrate that the recovered material is compatible with sample multiplexed proteomics²⁹ and drug perturbation. We benchmark our approach by recovering telomere-specific DNA binding proteins after targeting telomeric DNA. We further showcase the scalability of DNA O-MAP by

distinguishing the DNA-associated proteomes around human pericentromeric alpha-satellite repeats, telomeres, and mitochondrial genomes in quadruplicates using tandem mass tags²⁹. We go on to demonstrate that DNA O-MAP can capture functionally relevant DNA-DNA interactions, read out by DNA sequencing, from 20 kilobase intervals. Additionally, we show that DNA O-MAP can measure the local proteome of non-repetitive elements such as the *HOXA* and *HOXB* gene clusters as well as differential proteomes at these gene clusters before and after chemical inhibition of chromatin regulation. Finally, we show that DNA O-MAP can be applied to discern homolog-specific local proteomes of the active and inactive X-chromosome. We anticipate that the flexible targeting, scalable protocol, and robust labeling capabilities provided by DNA O-MAP will lead to its adoption as a platform technology for uncovering locus-proximal chromatin proteomes.

Results

Design of DNA O-MAP

DNA O-MAP is a molecular profiling methodology that combines the targeting flexibility of oligo-based ISH with the ability of horseradish peroxidase (HRP) to catalyze the localized deposition of small biomolecules at sites where it is bound. DNA O-MAP works by recruiting a ‘secondary’ HRP-conjugated oligo to sites where the primary ISH probes are bound. HRP-mediated deposition of biotin at targeted loci enables the pull-down and purification of proximal and chromatin associated proteins and DNA from *trans*-interacting genomic loci. DNA O-MAP reports on both direct DNA-protein interactions and proteins in spatial proximity to a target locus and allows for identification and quantitative comparison of proximal proteins between target genomic loci. As in RNA O-MAP²⁷, the specificity of ISH and/or biotinylation can be assessed by microscopy using a small sample of cells immobilized on solid support before the cells enter affinity purification. Importantly, the HRP-conjugated oligo is available via several commercial sources, allowing researchers without the expertise to perform their own conjugations to utilize DNA O-MAP.

Establishing a scalable in-solution hybridization-biotinylation workflow for DNA O-MAP

During the development of DNA O-MAP, we refined an in-solution hybridization workflow on cells in suspension for cost-efficient genomic labeling in parallel with an in-dish workflow used for RNA O-MAP²⁷ (Figure 1A [↗](#)). We began with adherent cells grown on multi-layer flasks, each yielding 90-120 million cells, and subsequently released and fixed (4% PFA) in order to be compatible with DNA ISH. Samples can be processed in parallel, thereby increasing the number of samples that could be handled at once. We note that the in-solution version of the protocol reduces reagent costs by ~1,000-fold relative to conventional ISH protocols performed on solid substrates to further enhance the scalability of DNA O-MAP.

DNA O-MAP reveals the organization of the telomeric proteome

To demonstrate that O-MAP can successfully purify proteins from genomic viewpoints, we selected human telomeres for initial testing (Figure 1A [↗](#)). Mammalian telomeres are several kilobases of tandemly repeated arrays of 5'-TTAGGG-3' hexamers with terminal 3' single-stranded overhangs at the ends of chromosomes³⁰. Telomeric DNA is specifically bound by a proteinaceous cap that protects the natural chromosome ends from being recognized as damaged DNA—the shelterin complex^{31,32}. Shelterin is a six-subunit complex, which is comprised of the telomeric repeat-binding factor 1 (TERF1), telomeric repeat-binding factor 2 (TERF2), protection of telomeres protein 1 (POT1), adrenocortical dysplasia protein homolog (ACD), TERF2-interacting protein 1 (TERF2IP), and TERF1-interacting nuclear factor 2 (TINF2). Due to the unique telomeric sequence and characteristic DNA structure, the shelterin proteins accumulate at the ends of the chromosomes. Accordingly, this well-defined set of proteins has been widely accepted as goalposts for a successful locus-proximal enrichment experiment^{12,13,16}. In the near-diploid HCT-116 cells, telomeres have an average length of 5.6 kb and their cumulative length approximates 0.017%

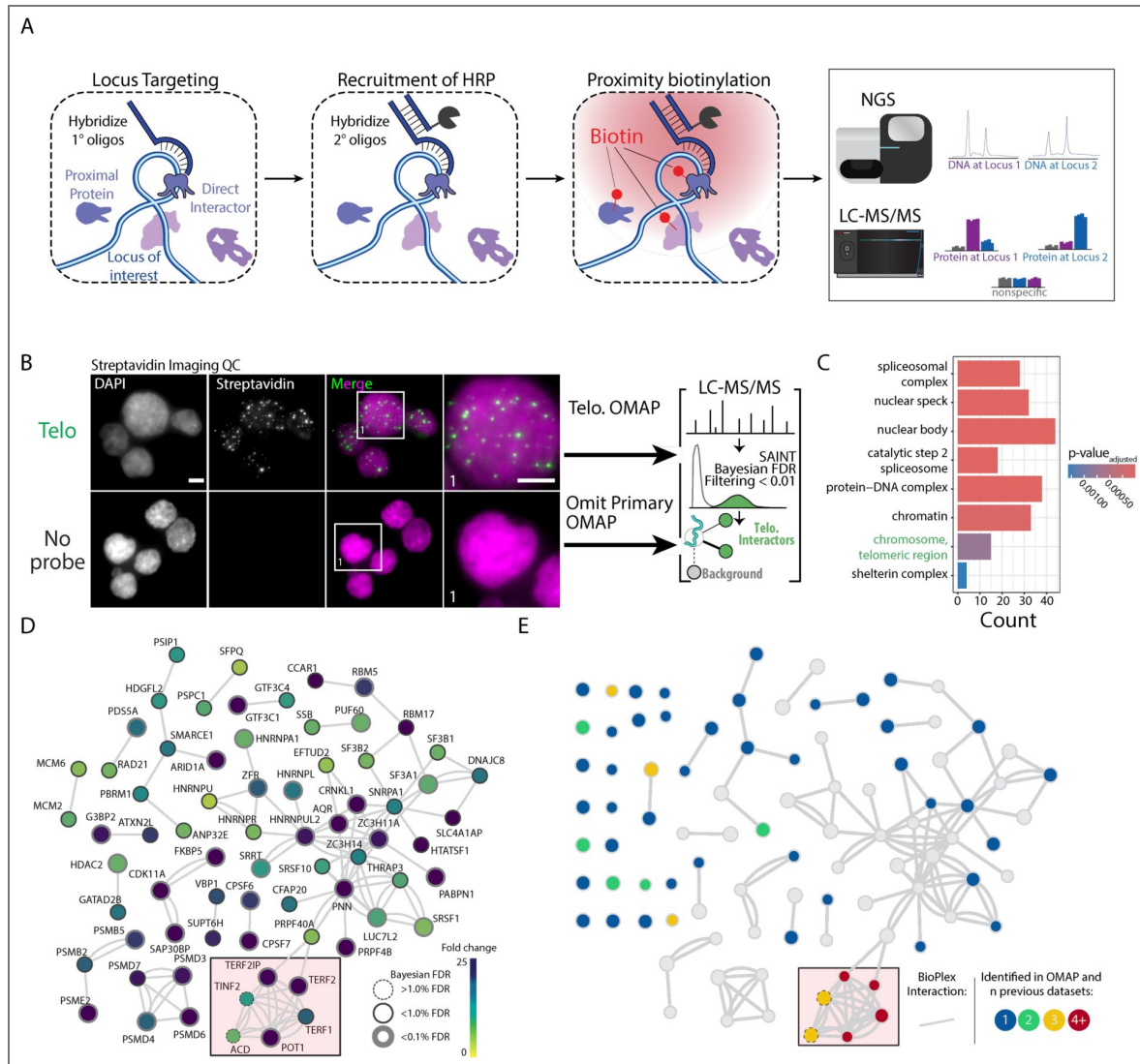


Figure 1. Overview of DNA O-MAP workflow and label-free quantitative proteomics analysis of telomeres.

A) Schematic of DNA O-MAP. B) Fluorescent microscopy data showing the observed patterns of DNA (DAPI, left) and *in situ* biotinylation detected by staining with fluorescent streptavidin conjugates (middle, left) and overview of telomere targeted DNA O-MAP experiment. C) Significant gene sets identified by the Gene Set Enrichment Analysis of the proteins enriched by the telomere probe. D) DNA O-MAP telomeric proteins mapped onto the BioPlex interaction network^{32,33}. The red box highlights shelterin complex proteins. Nodes are colored by the fold-enrichment compared to a no-primary-probe control shown in B, excluding unconnected nodes. E) Telomeric proteins observed in five previous datasets (PICH, C-BERST, CAPLOCUS, CAPTURE, BioID) superimposed onto Figure 1E, colored by the number of prior datasets where the protein was present and including unconnected nodes. Scale bars, 5 μ m.

(~500kb) of the human genome (Table S4 [33](#)). Compared to other repetitive elements in the human genome, telomeres are relatively short in HCT-116 cells and thus serve as a rigorous test case for DNA viewpoints of around 500 kb in aggregate across the genome.

We performed a DNA O-MAP experiment in which we either targeted telomeric DNA or omitted the primary hybridization probe (negative control). We purified biotinylated proteins from <60 million cells in three technical replicates followed by imaging of biotinylation and identification of proteins using label-free, MS1-based quantitative proteomics. By streptavidin staining, the punctate fluorescence pattern of biotin-labeled biomolecules closely mimicked telomere FISH, whereas we did not observe patterns of these puncta in the negative control samples (Figure 1B [34](#)). From our label-free proteomics analysis, we identified 163 proteins as significantly enriched at telomeres. As expected, gene set enrichment analysis [34](#) identified significant enrichment of telomeric chromosomal components, chromatin, and protein-DNA complexes (Figure 1C-D [34](#)). Importantly, we identified all six shelterin proteins in the telomere sample and these proteins were completely absent from the control samples. Of the six shelterin proteins, four (TERF1, TERF2, TERF2IP, POT1) passed stringent false-discovery rate control while ACD and TINF2 did not due to low spectral intensity.

To benchmark DNA O-MAP, we compared the full set of telomeric proteins to proteins observed in five established telomeric datasets (PICH, C-BERST, CAPLOCUS, CAPTURE, BioID) [12,14,16,35,36](#) (Figure 1E [34](#)). DNA O-MAP captured both previously observed telomeric interacting proteins (shelterins) as well as telomere associated proteins (ribonucleoproteins). We identified multiple heterogeneous nuclear ribonucleoproteins (hnRNPs) previously annotated as telomere-associated, including HNRNPA1 and HNRNPU. HNRNPA1 has been demonstrated to displace replication protein A (RPA) and directly interact with single-stranded telomeric DNA to regulate telomerase activity [37–39](#). HNRNPU belongs to the telomerase-associated proteome [40,41](#) where it binds the telomeric G-quadruplex to prevent RPA from recognizing chromosome ends. We mapped DNA O-MAP enriched telomeric proteins to the BioPlex protein interactome [42,43](#) and observed that in addition to capturing proteins from previously observed telomeric datasets (Figure 1E [34](#)), DNA O-MAP enriched for protein-protein interactors with previously observed telomeric proteins. Previous data found RBM17 and SNRPA1 at telomeres, and in BioPlex these proteins interact with three SF3 proteins (SF3A1, SF3B1, SF3B2). Though they were not identified in previous telomeric proteome datasets, all three of these SF3 proteins were enriched in the DNA O-MAP telomeric data. Furthermore, through interactions with G-quadruplex binding factors, these SF3 proteins are regulators of telomere maintenance [44](#).

DNA O-MAP quantitatively compares nuclear- and mitochondrial-targeted DNA proximal proteomes

We next evaluated the utility of DNA O-MAP to quantitatively measure proteins associated with specific genomic loci. We integrated sample multiplexing quantitative [29,45–47](#) proteomics downstream of DNA O-MAP to enable spectral quantification of all samples simultaneously (Figure 2A [34](#)). In our experimental design, we measured the proteomes at three well-characterized DNA loci with distinct protein occupants in the human genome and a no-primary probe negative control: 1) telomeres, 2) peri-centromeric alpha satellite repeats, 3) the mitochondrial genome, 4) no primary probe negative control (Figure 2B [34](#)). Centromeres are epigenetically defined chromosomal loci where kinetochore proteins assemble for spindle microtubule attachment to ensure equal chromosome segregation during cell division [48,49](#). Human centromeres are located within the AT-rich alpha satellite repeats, which are higher-order repeats composed of 171-base-pair monomeric units [50,51](#). Due to the sequence independence of centromeres, we utilized a previously described probe [28,52](#) that targets a subset of alpha satellite repeats to represent centromeres, hereafter denoted as the ‘Pan Alpha’ probe (Table S4 [34](#)). The predicted genome-wide binding profile [53](#) of the pan-alpha probe closely overlaps with centromeres and covers an estimated 35 Mb (Figure S1 [34](#)). Mitochondria are intracellular organelles of eukaryotic cells with

their own genome (mtDNA). The mtDNA is a circular double-stranded DNA molecule of about 16.6 kb, located in the mitochondrial matrix associated with the inner membrane^{54,55}. In HCT-116 cells, mtDNA copy number has been reported to range from 310–677⁵⁶.

To determine the localization of biotinylation using the new oligos and oligo pools, we performed DNA O-MAP in human HCT-116 cells with a co-hybridization of both fluorescent oligos and HRP oligos in order to observe fluorescent *in situ* hybridization (FISH) and *in situ* biotinylation signals in the same cell. Biotinylation patterns of the pan-alpha, telomere, and mtDNA probes showed strong concordance with FISH (Figure 2C). To quantify the local proteomes corresponding to each of these biotinylated patterns, we prepared replicate (n=4) samples for each probe and control (Figure S2). Comparison of FISH and biotin signal revealed that DNA O-MAP labeling was highly specific to each of the telomeric, centromeric, and mitochondrial FISH signals (Figure S3). After *in situ* HRP-mediated labeling, we performed thermal reversal of fixation of cells prior to lysis, enrichment of biotinylated proteins⁵⁷, tryptic digestion, and labeling with isobaric TMTpro barcodes²⁹. We confirmed that artifactual lysine acylation due to cellular fixation with PFA did not widely affect TMTpro labeling of peptides as only 1.38% of lysine-containing peptides were acylated with fixative. Notably, the fixative was not present during protease digestion, and thus all peptide N-termini were available for TMTpro labeling and quantitation, irrespective of lysine modification status.

In total we quantified 3,055 proteins across all four targeted and control samples (Figure 2D–E). We observed consistent proteome enrichment for replicate analyses with O-MAP by principal component analysis and correlation analyses (Figure 2D–E, S2). Based on Human Protein Atlas annotations⁵⁹, we observed significant enrichment of mitochondrial proteins with the mtDNA-probe proteomes and proteins from nuclear locations such as nuclear speckles, nucleoplasm, and nucleoli enriched by the telomere and pan-alpha probes (Figure 2F, S4). Notably, the pan-alpha probe enriched proteins from the nucleoli, consistent with the known nucleoli-centromere associations⁶⁰ and chromosomal passenger complex member AURKB, consistent with the centromeric localization of AURKB in early mitosis to ensure faithful chromosome segregation^{61,62} and the localization of chromosomal passenger complex members to pericentromeric heterochromatin^{63,64}. We also observed pericentromeric enrichment of spindle and chromosomal segregation associated proteins TPX2⁶⁵ and KIF20A⁶⁶ (Figure S4, S5).

Next, we explored the enrichment of several multi-unit protein complexes across the examined loci. To dissect the differences between enriched proteomes for each probe, we chose a subset of proteins of interest and measured the fold change of the two nuclear targets compared to mitochondria. RNA Polymerase I,II,III subunits were all higher in the nuclear probes than mitochondria, however in contrast to RNA Polymerase II and III, POLR1 proteins are significantly enriched in pan-alpha compared to telomere (Figure 2G). This enrichment is likely due to clustering of centromeres around nucleoli^{67,68}, the location of ribosomal RNA synthesis by RNA Polymerase I. Conversely, mitochondrial RNA Polymerase POLRMT abundance was significantly lower in the nuclear probe proteomes compared to the mitochondrial probe proteome ($\log_2$ Pan-Alpha Sat./Mito. = -2.51; $\log_2$ Telomere/Mito. = -1.88). Similarly, we observed enrichment of mtDNA-packaging nucleoid components⁵⁸ with the mtDNA probes (TFAM, SSBP1, POLG, POLRMT, Lon, ATAD3A/B, and PHB/PHB2; Figure 2G–H). As above, we observed consistent enrichment of shelterin components at telomeres (Figure 2I). We also observed CENP-A nucleosomal complexes enriched in the pan-alpha proteomes (Figure 2J). Histones were enriched with our nuclear probes and a subset (H2A1C, H2AX, and H4C1) were significantly enriched by the pan-alpha probe compared to the telomere probe (Figure S5). We also observed enrichment of catenins CTNNB1 and CTNND1 at telomeres (Figure S4). The transcription factor CTNNB1 has been observed at the transcriptional start site of *hTERT* where it regulates *hTERT* expression⁶⁹. The *hTERT* gene is located in the subtelomeric region of chromosome 5 (chr5:1,253,167-1,295,068) and expressed in HCT-116 cells⁷⁰. Collectively, these results highlight the subcompartment sensitivity of DNA O-MAP to distinguish differential compartment components even for ubiquitous chromatin constituents like histones.

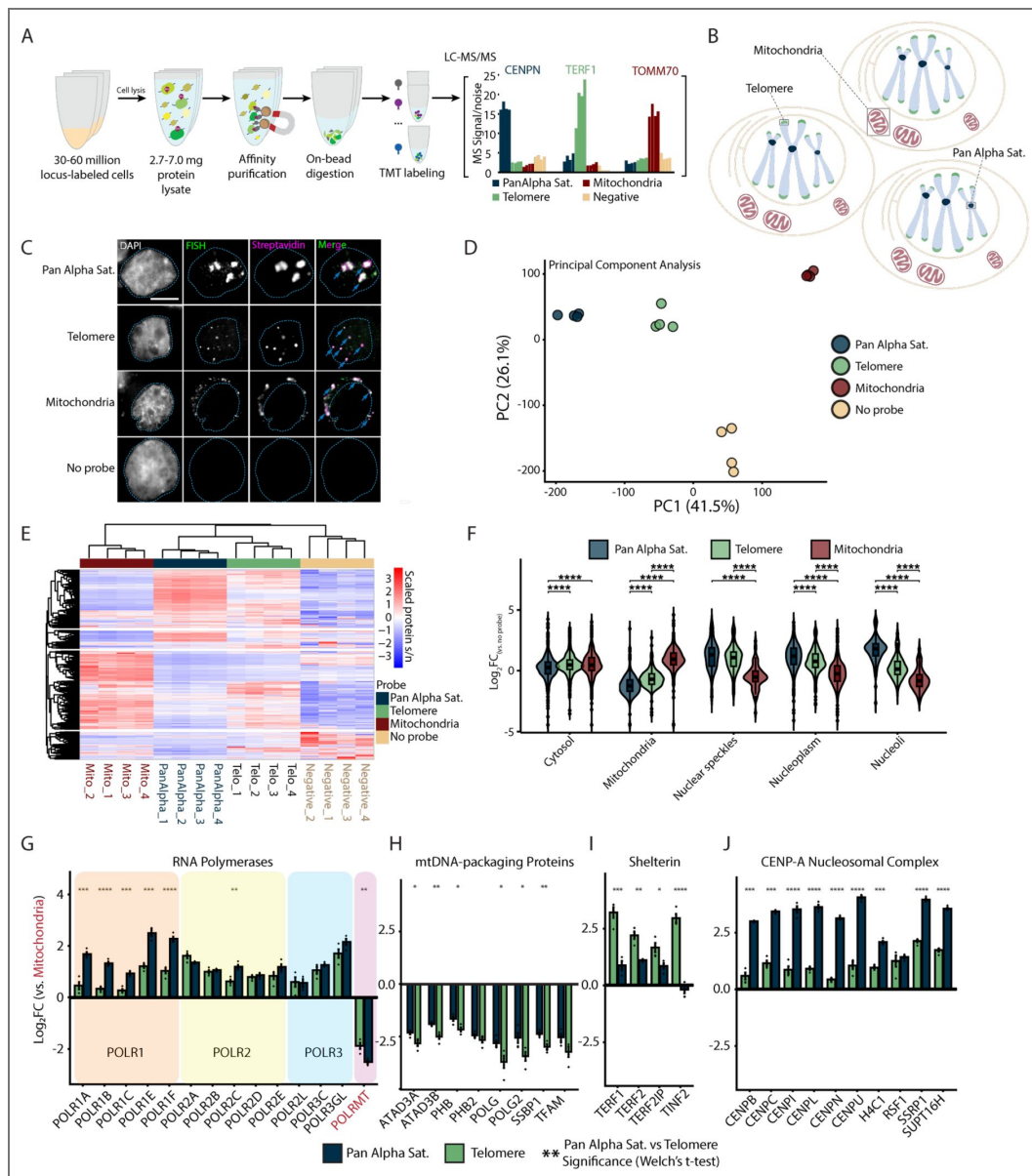


Figure 2. DNA O-MAP reveals distinct features of the sub-proteomes at peri-centromeric alpha satellites, telomeres, and the mitochondrial genome.

A) Workflow of DNA O-MAP integrated with sample multiplexing quantitative proteomics. B) Schematic of the three DNA loci examined in the TMT16plex experiment: peri-centromeric alpha satellites, telomeres, and mitochondrial genomes. C) Co-localization of DNA FISH and the streptavidin staining of the proteins biotinylated by DNA O-MAP targeting the peri-centromeric alpha satellites, telomeres, and mitochondrial genomes. Scale bar: 5 μ m. D) Principal component analysis of scaled intensities of proteins enriched by the pan-alpha probe, telomere probe, mitochondrial genome oligo pool, and no-primary-probe control. E) Unsupervised hierarchical clustering of scaled intensities of proteins enriched by the pan-alpha probe, telomere probe, mitochondrial genome oligo pool, and no-primary-probe control. F) Log2 fold change of proteins compared to no-primary-probe control, grouped by HPA subcellular location. Significance calculated based on Welch's t-test for pairwise comparisons (****: p-value < 0.0001). G–J) Log2 fold change of proteins compared to mitochondrial probe enriched proteins for the RNA Polymerases (G), mtDNA nucleoid packaging proteins⁵⁸ (H), Shelterin (I), and CENP-A nucleosomal complexes (J). Significance calculated based on Welch's t-test for pairwise comparisons (p-value: * < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001).

DNA O-MAP uncovers DNA-DNA interactions from non-repetitive DNA loci

Beyond repetitive regions in the human genome, we explored whether DNA O-MAP can recover material from single-copy DNA intervals. To this end, we designed an experiment in which we performed *in situ* biotinylation followed by chromatin extraction, affinity purification, and sequencing (Figure 3A). The human genome is folded into thousands of chromatin loops where two loci on the same chromosome are tethered to each other (Figure 3B). The anchors of the loops are bound by the insulator protein CTCF. The ring-shaped cohesin protein complex is thought to often stall at CTCF-bound sites while dynamically moving along the genome, creating contact domains of preferential DNA-DNA interaction⁷¹. These contacts between chromatin loop anchors have been captured genome-wide with *in situ* Hi-C⁷². Normally present in two copies per genome, these 20–25 kb loop anchor intervals are considerably less abundant than telomeres.

We first evaluated whether DNA O-MAP can specifically biotinylate loop anchors with microscopy by a co-hybridization of both fluorescent oligos and HRP oligos at four anchors: chr3 left (chr3:187,729,712-187,749,712), chr3 right (chr3:188,939,711-188,964,711), chr19 left-2 (chr19:33,425,000-33,450,000), and chr19 right (chr19:33,750,000-33,775,000). DNA O-MAP specifically biotinylated the biomolecules proximal to these small DNA intervals, as observed in the co-localizing patterns of FISH and streptavidin staining in the same cells (Figure 3C). We next evaluated whether DNA O-MAP could recover the DNA interactions originally discovered by Hi-C. We targeted a pair of intervals with high contact frequency—chr3 left and chr3 right anchors, one non-looping interval (chr10:123,187,984-123,207,984), and a no-primary-probe control. We performed DNA O-MAP to biotinylate these DNA intervals, subjected the labeled cells to chromatin solubilization and desthiobiotin purification, and sequenced the eluate DNA. As expected, all three probed DNA intervals were highly enriched compared with other genomic regions, indicating efficient purification of the loci (Figures 3D, S6A). Furthermore, chr3 left and chr3 right anchors reciprocally recovered each other, indicating that DNA O-MAP was able to recover known DNA interactions mediated by proteins. In contrast, the non-looping chr10 anchor did not enrich any other peak other than itself (Figure S6B). Lastly, in the cells that received no primary oligos, no pronounced enrichment was observed genome wide (Figure S6B).

To examine the multiplexability and reproducibility of DNA O-MAP, we simultaneously targeted three chromatin loop anchors: chr3 left, chr10 right (chr10:123,957,984-123,977,984), and chr19 right anchors in duplicates and subjected the cell pellets to purification and DNA sequencing. All three targeted anchors, chr3 left, chr10 right, and chr19 right anchors were successfully enriched (Figures 3E–F, S7A), whereas no pronounced enrichment was observed in the no-primary-probe controls genome-wide (Figure S7B). Furthermore, chr10 left (contacting chr10 right), chr19 left-1, and chr19 left-2 (both contacting chr19 right) were also efficiently recovered, accurately matching the Hi-C contact maps and the signals from two replicates was consistent (Figure 3E–F, S8). We confirmed this by distance-dependent normalization previously developed for proximity-labeling methods (Figure S8)⁷³. These imaging and genomics data demonstrate that DNA O-MAP is capable of labeling small, single-copy DNA intervals with high specificity.

DNA O-MAP uncovers differential protein enrichment from non-repetitive loci

Given our success with recovering DNA-DNA contact information from unique loci, we explored whether DNA O-MAP can recover the local proteome from single-copy DNA intervals as we did earlier with repetitive intervals targeting telomeres, alpha satellite repeats, and the mitochondrial genome. To investigate this, we utilized the *HOXA* and *HOXB* gene clusters as targets. The *HOX* subgroup of the homeobox family of genes has been well described for their roles in determining body plan formation, cell identity, and has been implicated in susceptibility to oncogenesis^{74–78}. Here we designed probes to target 83kb and 81kb for *HOXA* and *HOXB*, respectively. We performed DNA O-MAP targeting either *HOXA*, *HOXB*, or no primary hybridization probes from 50

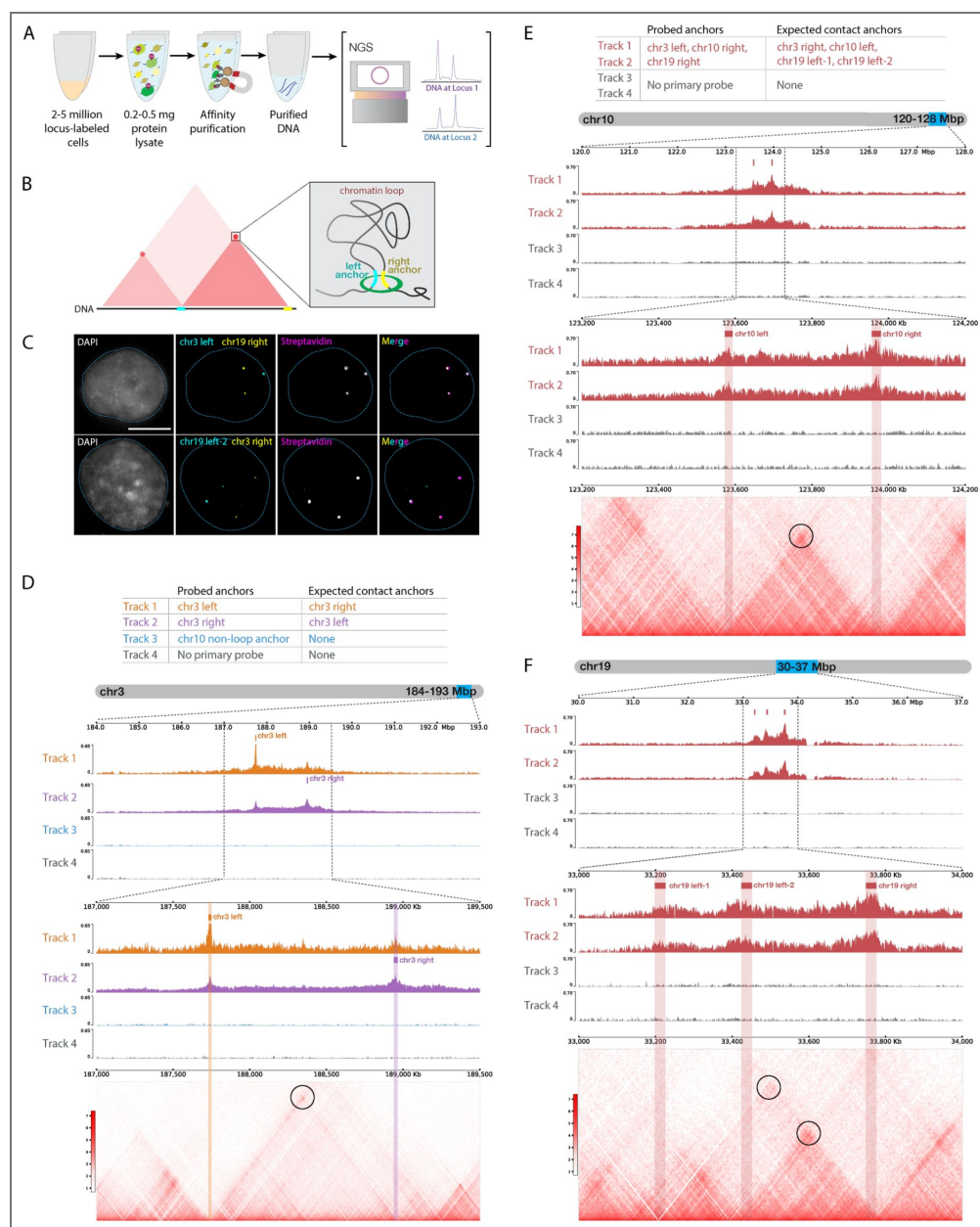


Figure 3. DNA O-MAP efficiently labels single-copy chromatin loop anchors.

A) Workflow of DNA O-MAP integrated with biotin purification sequencing. B) Schematic of a pair of chromatin loop anchors on a hypothetical Hi-C map and 3-dimensional space. C) DNA FISH and the streptavidin staining of the proteins biotinylated by DNA O-MAP targeting anchors of chromatin loops on chromosome 3 and chromosome 19. D) Table listing the three anchors (Track 1-3) and no-primary-probe control (Track 4) biotinylated by DNA O-MAP and their expected anchors in contact in each track (top). Desthiobiotin purification sequencing signals across the 9-Mb region on chromosome 3 corresponding to the chr3 chromatin loop (middle). Desthiobiotin purification sequencing signals and pairwise contact map at 5-kb resolution across the 2.5-Mb region on chromosome 3 corresponding to the chr3 chromatin loop. Black circle on the contact map indicates the presence of a loop (bottom). E) Table listing the three chromatin loop anchors (Track 1-2) and no-primary-probe controls (Track 3-4) biotinylated by DNA O-MAP in duplicates and their expected anchors in contact in each track (top). Desthiobiotin purification sequencing signals across the 8-Mb region on chromosome 10 corresponding to the chr10 chromatin loop targeted (middle). Desthiobiotin purification sequencing signals and pairwise contact map at 5-kb resolution across the 1-Mb region on chromosome 10 corresponding to the chr10 chromatin loop. Black circle on the contact map indicates the presence of a loop (bottom). F) Desthiobiotin purification sequencing signals across the 7-Mb region on chromosome 19 corresponding to the chr19 chromatin loops targeted (top). Desthiobiotin purification sequencing signals and pairwise contact map at 5-kb resolution across the 1-Mb region on chromosome 19 corresponding to the chr19 chromatin loops. Black circles on the contact map indicate the presence of loops (bottom).

million human K562 cells across six replicates for *HOXA* and *HOXB* and five replicates for no primary (Figure 4A). We also performed DNA O-MAP in human K562 cells with co-hybridization of both fluorescent oligos and HRP oligos to demonstrate locus-targeted biotinylation and FISH signal within the same cell. Biotinylation patterns for both *HOXA* and *HOXB* showed strong concordance with the FISH signal (Figure 4B). Identification of biotinylated proteins was carried out using label-free quantitative proteomics. From our label-free proteomics analysis, we identified 42 proteins that were significantly differentially enriched between the *HOXA* and *HOXB* loci (Figure 4C, S9A). Of the differentially enriched proteins, both HDAC3 and TCF12 were scored as enriched in the *HOXB* labeled sample compared to *HOXA*. This agrees with ENCODE⁷⁹ ChIP-seq data in K562 cells, with more called peaks for these proteins at *HOXB* than *HOXA* (Figure 4D). Additionally, SMARCB1 was scored as enriched in the *HOXA* labeled sample compared to *HOXB* and ENCODE⁷⁹ ChIP-seq data showed enrichment of SMARCB1 at *HOXA* over *HOXB* (Figure 4D). SWI/SNF factors such as SMARCB1 have been described to interact with the *HOX* genes⁸⁰, but differential enrichment has not been well documented.

To further test the sensitivity and application of our non-repetitive DNA O-MAP approach at the *HOX* genes, we performed a chemical perturbation analysis targeting *HOXA* and *HOXB* in the presence or absence of enhancer of zeste homolog 2 (EZH2) inhibition with GSK126. In differentiated cells, such as K562s, the *HOX* genes are silenced by the Polycomb Repressive Complex 2 (PRC2). The main catalytic component of PRC2 is EZH2, which is responsible for trimethylating histone 3 at lysine 27 (H3K27me3)^{81,82}. Deposition of H3K27me3 has been well characterized as a mark of transcriptional repression associated with PRC2 activity. Additionally, EZH2 inhibition has become a prominent target in cancer therapies^{83,84}. For the perturbation analyses, we used sample multiplexing quantitative proteomics^{29,45,46} and DNA O-MAP to quantify the local proteomes in response to drug inhibition (Figure 4E). We prepared replicate (n=3) samples for each probe following treatment with either the EZH2 inhibitor GSK126 or DMSO control. In total, we identified 11 proteins at *HOXA* and 8 proteins from *HOXB* with significantly altered enrichment following EZH2 inhibition (Figure 4F, G, S9B), many of which are RNA-interacting proteins, such as ELAV1, RMB14, ILF3, HNRNPK, HNRNPA1, PRPF40A, and PRPF8. These data are consistent with recent evidence that suggests that long non-coding RNA-interacting proteins are critical for the establishment of PRC2 repressed loci^{85,86}. Several proteins we identified also have established connections to PRC2 repression of the *HOX* genes, such as GATA1 which has been suggested to mediate the switch to non-canonical PRC2 function⁸⁷. These data demonstrate that DNA O-MAP has sufficient sensitivity to detect proteomes at non-repetitive genomic loci and measure inhibitor-induced changes to the proximal proteome at sub-megabase, non-repetitive genomic regions (~80 kb).

Establishing an on-plate hybridization-biotinylation workflow for DNA O-MAP

During the development of DNA O-MAP, we refined an on-plate workflow on cells adhered to glass bottom plates (Figure 5A). We began with adherent cells grown in 6-well plates, using 3 wells per replicate per condition with each well yielding 0.75 - 1 million cells. The cells are subsequently fixed (4% PFA) in order to be compatible with DNA ISH. Multiple plates can be processed in parallel to enable multiple replicates per experiment. We note that the on-plate version of the protocol further reduces the total number of cells required per replicate by ~10-fold compared to the in-solution protocol.

DNA O-MAP detection of homolog specific proteomes

To further understand the level of sensitivity we can achieve when investigating the local proteome of non-repetitive loci, we explored whether DNA O-MAP can recover material in a homolog specific manner. To this end, we designed primary hybridization probes that differentially target either the active or inactive X-Chromosome homolog (Xa and Xi, respectively). The Xa-Xi paradigm is a particularly interesting variable homolog target as the Xi is largely transcriptionally silenced through cooperative activity of the non-coding RNA Xist and H3K27me3

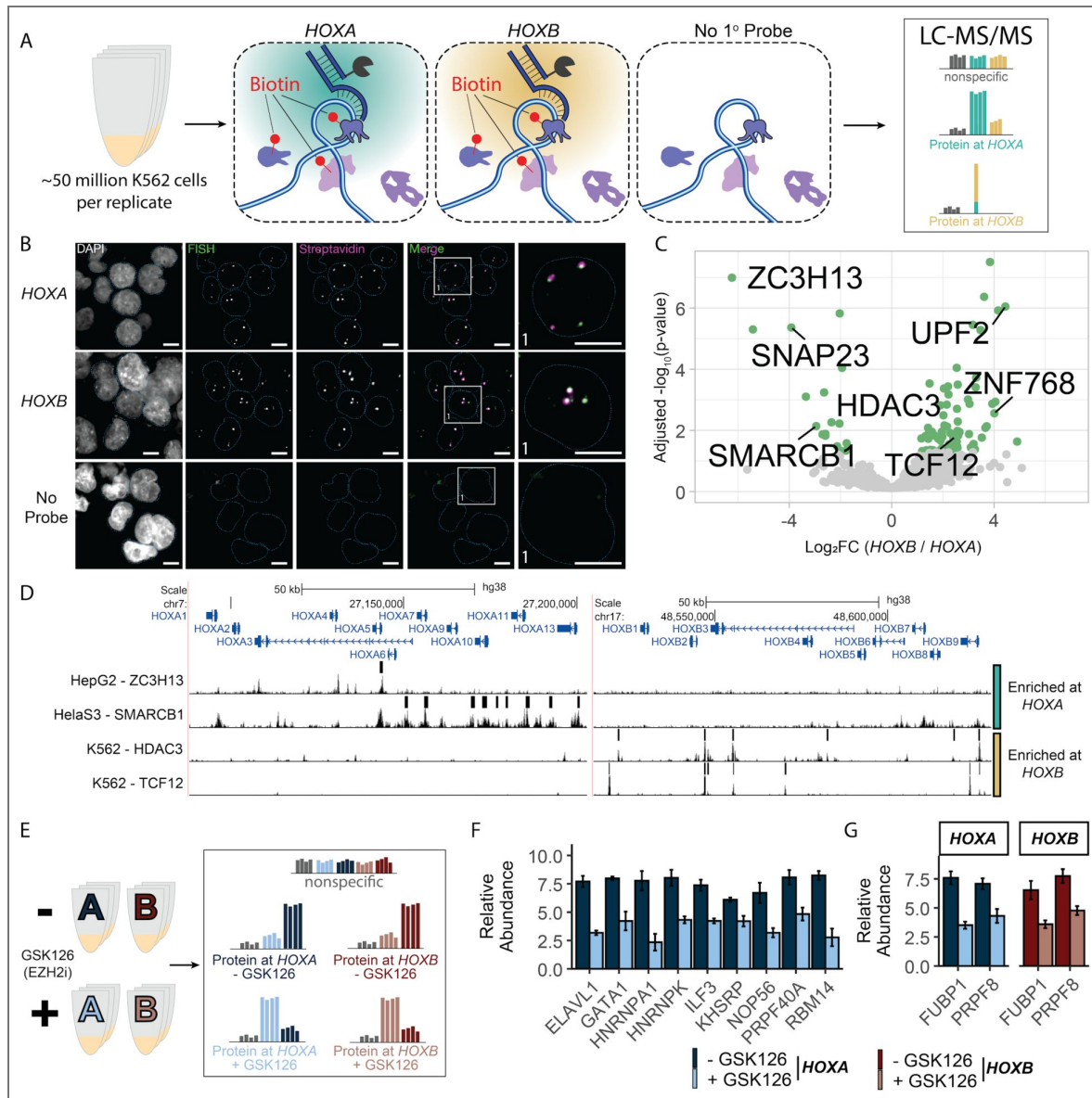


Figure 4. DNA O-MAP efficiently identifies the local proteome of the *HOXA* and *HOXB* gene clusters.

A) Schematic of DNA O-MAP being applied to the *HOXA* and *HOXB* gene clusters for identification of differentially enriched proteins. B) Representative images depicting overlap of FISH and Streptavidin labeling at *HOXA* and *HOXB* loci. C) Volcano plot of proteins identified at *HOXA* and *HOXB* loci. Each dot represents a single protein with proteins of interest called out in black. Green dots indicate proteins that passed significant enrichment thresholds with an absolute $\log_2$ Fold Change greater than 1 (2-fold change) and corrected p-value < 0.05 . D) ENCODE ChIP-seq data showing peak calls and p-values at *HOXA* and *HOXB* loci for selected enriched proteins, ZC3H13, SMARCB1, HDAC3, and TCF12. E) Schematic depicting the use of GSK126 with DNA O-MAP. F) Bar chart showing proteins with significantly altered abundance following treatment with GSK126 at *HOXA*. G) Bar chart showing proteins with significantly altered abundance at both *HOXA* and *HOXB* following treatment with GSK126 (Welch's t-test, corrected p-value < 0.05).

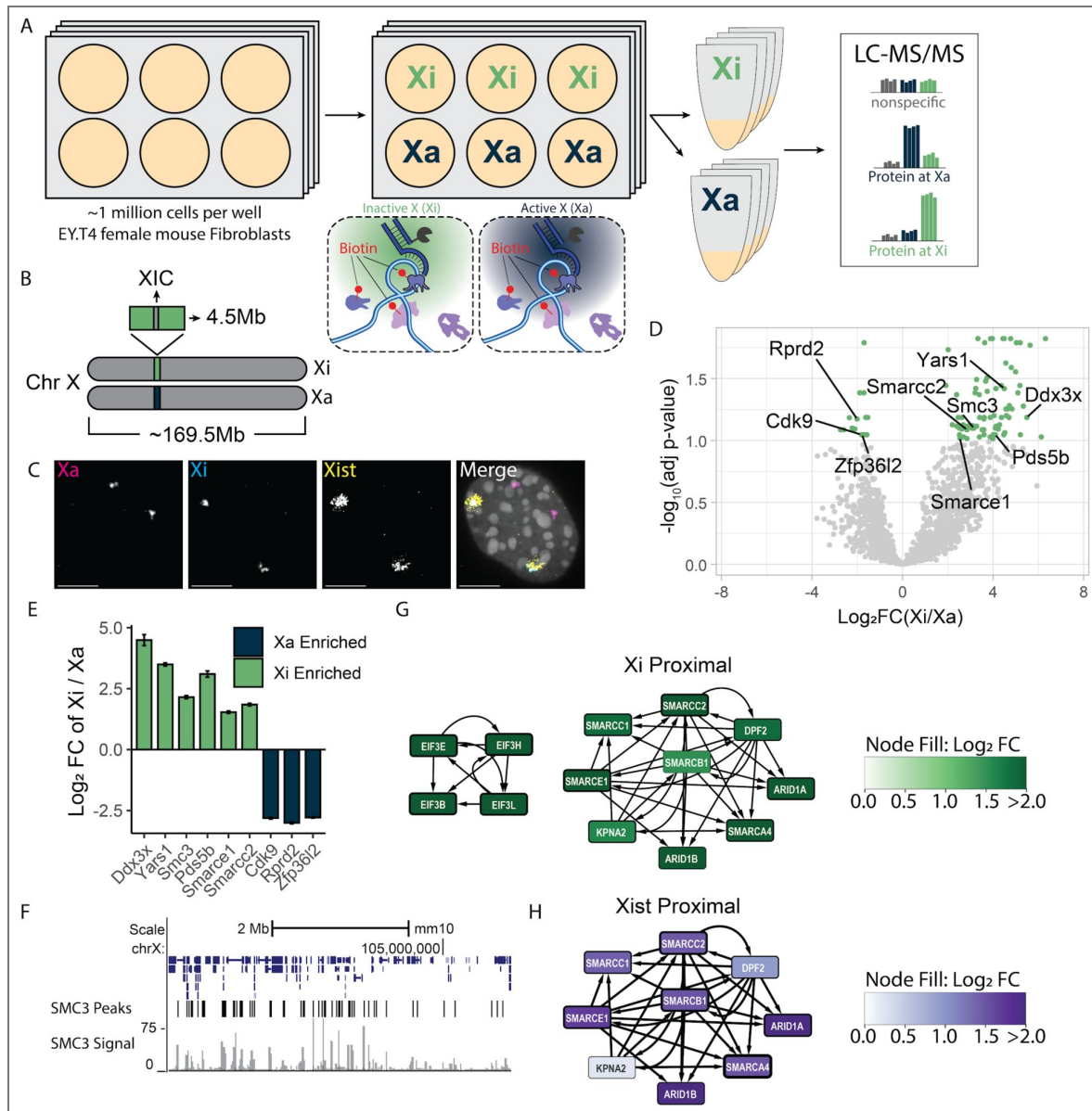


Figure 5. DNA O-MAP elucidates the homolog-resolved chromosome X proteome.

A) Schematic of DNA O-MAP being applied to Xi and Xa for identification of differentially enriched proteins. B) Schematic showing the region of the X chromosome targeted by our primary hybridization probes. C) Representative images depicting overlap of Xist FISH and Xi Streptavidin labeling while spatially differentiated from Xa FISH. Scale bars are 10 μ m. D) Volcano plot of proteins identified at Xa and Xi. Each dot represents a single protein with proteins of interest called out in black text. Green dots indicate proteins that passed significant enrichment thresholds with an absolute Log2 Fold Change greater than 1 (2-fold change) and corrected p-value < 0.1. E) Bar chart showing example proteins with significant enrichment at Xi in green and at Xa in blue. Corrected p-value < 0.1. F) ENCODE ChIP-seq data in mouse fibroblast cells at our targeted region of chromosome X for SMC3. G) Protein interaction networks of EIF and SWI/SNF complexes enriched at Xi. Node width is a function of corrected p-value and node color is a function of enrichment (Log2 Fold Change). H) Protein interaction network of the SWI/SNF complex from previously published RNA O-MAP of Xist. Node width is a function of $-\log_{10}(\text{p-value})$ and node color is a function of enrichment (Log2 Fold Change).

by the PRC, along with other factors, in direct opposition of the Xa^{85,86,88–90}. Thus, these two homologs present a vastly different chromatin environment. To individually target Xi and Xa, we used differential single nucleotide polymorphisms (SNPs) between the maternal and paternal homologs in EY.T4 female mouse fibroblast cells that were identified by homolog resolved genome sequencing^{91,92}. In doing so we were able to design two sets of primary hybridization probes to target the same 4.5Mb region on either Xi or Xa straddling the X inactivation center (Figure 5B⁴). The approach of leveraging SNPs for homolog specific probe binding has been previously validated for use with other ISH-based methods^{92,93}, but has not been previously applied to proximity labeling or proteomics approaches. For validation of our differential labeling approach we performed FISH for Xist and Xa alongside streptavidin labeling of Xi. Labeling patterns of Xi and Xist overlapped as expected while remaining spatially distinct from Xa FISH (Figure 5C⁴). We performed a DNA O-MAP experiment in which we targeted either Xa or Xi across four replicates. Each replicate consisted of three wells of a 6-well plate each containing 0.75 - 1 million cells (Figure 5A⁴). Of note, in contrast to the previous experiments, here we did not release our adherent cells and all steps were carried out with cells fixed to glass-bottom 6-well plates prior to lysis. From our label-free proteomics analysis, we identified 96 proteins that were significantly differentially enriched at either Xi or Xa (Figure 5D⁴, S10⁴). The Xa showed enrichment of proteins associated with a more active transcriptional pattern in comparison to Xi, such as RPRD2, ZFP3612, and CDK9 (Figures 5D⁴ & E⁴). For the Xi in comparison to Xa, we observed enrichment of protein interactors and complexes^{42,43}, including the EIF3 family, members of the cohesin complex Smc3 and Pds5b, and SWI/SNF factors. Both the cohesin and SWI/SNF complexes have been previously shown to interact with the inactive X chromosome⁹⁴ and enrichment of SWI/SNF factors was consistent with recent work measuring the proteome of Xist, the Xi-resident long non-coding RNA²⁷ (Figure 5G⁴). Interestingly, we also see enrichment of Ddx3x, an X-linked gene previously implicated as a mediator of sex-based differences in neurodevelopment and disease⁹⁵. Here we show that DNA O-MAP was able to obtain homolog-specific, differentially enriched local proteomes from a 4.5Mb, non-repetitive, region on the X chromosome from less than 3 million cells per replicate.

Discussion

By combining the versatility of hybridization-based genome targeting with the robustness of proximity biotinylation, DNA O-MAP offers a scalable approach to study DNA proximal proteomes of a specific locus. The hybridization-biotinylation workflow allows for efficient processing of samples and is compatible with both proteomic and genomic readouts. Integration with multiplexed quantitative proteomics enables simultaneous analysis of multiple loci or conditions, increasing data completeness and throughput. Label-free analysis of the telomeres shows strong concordance of labeling with in-situ hybridization and recapitulates previous similar proteomic datasets. Our tri-locus experiment was able to differentiate proteins with a quantitative profile suggesting general nuclear location from those specifically associated with telomeres and pericentromeres. DNA O-MAP can target single-copy loci, as evidenced by the chromatin loop anchor, *HOXA* and *HOXB* gene cluster, and X-chromosome targeting experiments, which enables the study of a wide range of DNA proximal proteomes. The ability to detect differences in these proteomes is exemplified by the application of DNA O-MAP to the *HOXA* and *HOXB* loci following EZH2 inhibition and to the X-chromosome in a homolog specific manner.

O-MAP has now been shown to be a highly flexible technology for the exploration of biomolecular interactions with RNAs²⁷ and DNA loci. Using oligos to target the DNA locus, DNA O-MAP can be theoretically adapted for use in any sample type amenable to *in situ* hybridization, including cultured cells, tissue sections, and primary tissue samples^{28,53,96}. As the purification tag is decoupled from the probe oligos, labeled chromatin fragments can undergo stringent washes to achieve efficient purification with minimal background. Moreover, without the need to genetically modify the biological system at hand, the probes in this dataset alone could be used to explore telomeric remodeling in cancer cells³⁶, spindle-associated proteome dynamics at the pericentromere⁹⁷, and molecular drivers of hetero- or euchromatin formation⁹⁸ at nearly any

locus in the human genome (O-MAP probes can feasibly cover >99% of the human genome)^{53,96}. Additionally, we show that this method is able to detect DNA-DNA contacts through detection of biotinylated loop anchors. Our approach functions similarly to 4C⁹⁹, however, the biotin labeling of contacts does not rely on pairwise ligation events. Thus, detection of contacts through DNA O-MAP will vary in the sampling of DNA-DNA contacts in comparison.

DNA O-MAP has several current limitations. First, fixation can modify proteins and must be accounted for when labeling is performed in fixed cells. Second, DNA O-MAP captures proximal proteins for a given genomic locus. This means measured proteins will be both direct interactors and also proteins that are reproducibly within close physical proximity to the locus. This distinction has been described previously^{27,100}, but is generally important for interpretation of DNA O-MAP proteomics results and design of future DNA O-MAP experiments. This is due, at least in part, to the long-lived nature of the biotin-phenoxyl radical species generated by peroxidase-based proximity labeling methods. Third, DNA O-MAP may miss proteins at specific loci if the protein is especially low abundance, highly post-translationally modified, or otherwise challenging to detect by mass-spectrometry-based proteomics. Finally, DNA O-MAP reports more proteins than those that are ‘specific’ to a given locus. For example, all nuclear targeting probes identified nuclear, but non-specific, proteins like histones. An important aspect of the quantitative approaches used here is that proteins are measured as differentially abundant at a given loci in comparison to other loci or across treatments.

By taking a comparative quantitative approach, we remove the need to pre-define the local context of probe localization, but experimental design is critical and novel interactors may require further validation to confirm their co-localization at a given locus (e.g., with imaging/FISH). With developments in automation and instrument sensitivity, DNA O-MAP has the potential to expand to post-translational modifications and be used for large-scale chromatin perturbation screens. We anticipate that DNA-OMAP will have broad utility for research questions seeking to understand the intricate relationships between DNA sequence, chromatin structure, and cellular function.

Methods

Cell culture and fixation

Colorectal cancer HCT-116 cells were grown in ATCC-formulated McCoy’s 5A Medium Modified (ATCC 30-2007). EY.T4 mouse fibroblasts were grown in ATCC-formulated DMEM (Gibco 11965-092). Human lymphoblast K562 cells were grown in ATCC-formulated RPMI (Gibco 11875-093). All cell lines were supplemented with 10% fetal bovine serum and 100 U/ml Penicillin-Streptomycin and grown at 37°C in a humidified atmosphere of 5% CO₂. For each HCT 116 purification, 20 million cells were seeded into one T-500 flask (Thermo Scientific 132867) to culture for 36–48 hours to reach 90–120 million cells. Before collection, cells were briefly rinsed once with Dulbecco’s phosphate buffered saline (DPBS) and then incubated with 25 ml of TrypLE Express Enzyme (Gibco 12604-021) at 37°C for two minutes or until loosely attached. For each K562 purification, 25 million cells were seeded into one T-500 flask to culture for 48 hours to reach 100 million cells. Before collection, cells were briefly rinsed once with DPBS. For inhibition of EZH2, K562 cells were grown and harvested, as described above, in medium supplemented with either 5uM GSK126 (Tocris 6790) or DMSO for 48 hours and then collected. For both HCT 116 and K562 purifications, the cell suspension was collected into two 50 ml conical tubes and the T-500 flask was rinsed with DPBS. The wash was combined with the cell suspension and centrifuged at 300 G for 5 minutes. After a DPBS wash to remove remaining TrypLE, cells were fixed in 4% paraformaldehyde (wt/vol) (Electron Microscopy Sciences 15710) in PBS in suspension at room temperature for 10 minutes with rotation, followed by 125 mM Glycine quenching for 5 minutes at room temperature with rotation and 15 minutes on ice. Fixed cells were collected by centrifugation at 350G for 5 minutes, and stored in fresh DPBS at 4°C until liquid-phase hybridization. Fixed cells were used within 3–5 days.

For each EY.T4 purification, 300,000 cells were seeded into each well of a glass-bottom 6-well plate (Cellvis P06-1.5H-N) to culture for 24 hours to reach 750,000 - 1 million cells per well. Cells were briefly rinsed once with DPBS then fixed in 4% paraformaldehyde (wt/vol) (Electron Microscopy Sciences 15710) in PBS in at room temperature for 10 minutes. Fixed cells were then washed with DPBS three times for 5 minutes each. Fixed cells were stored in fresh DPBS at 4°C until solid-phase hybridization. Fixed cells were used within 5 days.

Primary oligo probes

Primary oligos targeting the human alpha satellite repeat and telomere were purchased as individually column-synthesized DNA oligos from Integrated DNA Technologies. Probe sets targeting *HOXA* (chr7:27092311 - 27175959), *HOXB* (chr17:48527413 - 48608584), mtDNA (chrM:1-16,569), chr3 left anchor (chr3:187,729,712-187,749,712), chr3 right anchor (chr3:188,939,711-188,964,711), chr10 non-looping anchor (chr10:123,187,984-123,207,984), chr10 right anchor (chr10:123,957,984-123,977,984), and chr19 right anchor (chr19:33,750,000-33,775,000) were designed using PaintSHOP⁹⁶ and ordered in oPool format from Integrated DNA Technologies. Homolog specific chromosome X (chrX:100,254,241 - 102,428,950, 102,601,850 - 104777629) targeting probes were designed using an in-house computational pipeline and ordered in oPool format from Integrated DNA Technologies. More than 300 primary oligos were designed to cover each single-copy DNA interval to ensure a sufficient number of probes at the locus for FISH. The sequences of the oligo and oligo pools used are listed in Supplementary Dataset 1.

Oligo library synthesis

Amplification of the *HOXA*, *HOXB*, Xi, and Xa targeting primary oligo libraries was performed as follows. 20 ng μl^{-1} of oligo library in 10 mM Tris, pH 8.0 was amplified by PCR. The reaction mix contained 34 μl of dH₂O, 10 μl 5 $\times$ Phusion HF Buffer, 1.5 μl of 10 mM dNTP Mix, 1.5 μl of 10 μM F primer, 1.5 μl of 10 μM R primer, 1.0 μl of resuspended oligo pool and 0.5 μl of Phusion DNA Polymerase (2 U μl^{-1}). The thermal cycler program was as follows; 95 °C for 3 min, followed by 12 cycles of 98 °C for 20 s, 60 °C for 15 s, 72 °C for 15 s then 72 °C for 1 min followed by a 4 °C hold. This PCR product was purified using a Zymo DNA Clean and Concentrator-5 (DCC-5) kit according to the manufacturer's protocol. 20 pg μl^{-1} of the first PCR product was prepared as a template for the second PCR. The second PCR mix contained 27 μl of dH₂O, 10 μl of 5 $\times$ Phusion HF Buffer, 1.5 μl of 10 mM dNTP Mix, 5.0 μl of 10 μM F Primer, 5.0 μl of 10 μM R Primer, 1.0 μl of diluted DNA template and 0.5 μl of Phusion DNA Polymerase (2 U μl^{-1}). The thermal cycler program was repeated, but for 18 cycles instead of 12. The second PCR product was purified as before. RNA was synthesized using the NEB HiScribe T7 Quick High Yield RNA Synthesis Kit with a modified reaction mix containing 8 μl of dH₂O, 2.5 μl of diluted DNA template, 15.0 μl of NTP Buffer Mix, 3 μl of T7 RNA Polymerase Mix and 1.5 μl of RNaseOUT. The reaction was carried out for 16 hours at 37 °C followed by a 12°C hold. The DNA template was then digested using DNase I (M0303) by adding 2 μL (4 units) of DNase I and 50 μL of Ultra Pure Water (UPW) (10977-015) to each reaction. The reverse transcription reaction contained 45 μl of synthesized RNA, 15 μl of 5 $\times$ RT Buffer, 10.5 μl of 10 mM dNTP Mix, 1 μl of 100 μM RT Primer, 1.5 μl of RNaseOUT and 2 μl of Maxima H Minus Reverse Transcriptase (200 U μl^{-1}) and 20 μl of UPW for a total volume of 75 μl . The reactions were incubated at 50 °C for 2 h, 85 °C for 5 min, then a 4°C hold. RNA templates were degraded by adding 37.5 μl of 0.5M EDTA and 37.5 μl of 1M NaOH to each reaction and incubated at 95°C of 10min followed by a 4°C hold. Final ssDNA probe was purified using a Zymo DNA Clean and Concentrator-100 (DCC-100) kit according to the manufacturer's protocol for oligo purification.

Primer exchange reaction (PER)

To extend primary oligos with PER concatemers, reactions were set up as previously described¹⁰¹ in 100 μl -volume containing 10 mM MgSO₄, 300 μM dATP/dCTP/dTTP mix, 100 nM Clean.G hairpin, 80 U/ml Bst DNA Polymerase, Large Fragment (NEB M0275L), 1 μM hairpin, and 1 μM primary oligos in PBS. To verify the length of primary oligos, the reactions were assessed with denaturing

polyacrylamide gel electrophoresis. Primary oligos extended to 300-500 nucleotides were used in hybridizations downstream. Unpurified reactions were dehydrated using vacuum concentrators and stored dry at -20°C until hybridization.

In-solution hybridization and biotinylation of cell pellets

HCT 116 and K562 oligo hybridizations were performed on cells in solution for the cost-effectiveness of primary and secondary oligos. Fixed cells were split into 6e7 cell aliquots in 1.5 ml microcentrifuge tubes. All washes and buffer exchanges were performed as follows: centrifuging at 350G for 3.5 minutes or until pelleted, pouring away used buffers from the pellets, adding new buffers, and gentle shaking or low speed vortexing to dislodge cell pellets into tiny clusters or cell suspensions for incubations or washes. Cells in fresh wash buffer were rotated on a low speed nutator for 5 minutes.

Cells were rinsed once with fresh phosphate buffer saline (PBS), and permeabilized in PBS-0.5% TritonX-100 (Sigma T8787) for 10 minutes with nutation. After a PBS-0.1% Tween20 (PBS-T) (Sigma T2287) wash, permeabilized cells were incubated in 0.1 N hydrochloric acid (HCl) for 5 minutes. After a PBS-T wash to remove acid, cells were incubated in PBS-T-0.5% hydrogen peroxide to block endogenous peroxidases. After a 2X saline sodium citrate-0.1% Tween20 (2X SSC-T) wash to remove acid, cells were incubated in 2X SSC-T-50% formamide for 20 minutes at 60°C on a Thermomixer C dry block (Eppendorf 2231001005). Cells were exchanged into primary hybridization buffer (Hyb1) comprising 2X SSC-T, 50% (vol/vol) formamide, 10% (wt/vol) dextran sulfate, 0.4 µg/ul RNase A, and ~1 µM extended primary oligos (resuspended dry, unpurified PER reactions). The cell-Hyb1 mixture was distributed into PCR strip tubes at 1e7-1.5e7 cells in 100 µL volumes. The cells were denatured and primary oligos were hybridized to the genome in the PCR strip tubes in a thermocycler using the cycling protocol: 78°C 3 minutes, 37°C ∞ incubating overnight for more than 18 hours. The next day, cells were rinsed with 60°C 2X SSC-T into 1.5 ml microcentrifuge tubes, followed by two 2X SSC-T buffer exchanges to remove residual Hyb1. Cell pellets were then washed in 1 ml 2X SSC-T at 60°C, followed by two two-minute washes in 2X SSC-T at room temperature. Fully washed cell pellets were exchanged into 1 ml PBS, and then exchanged into 100 nM secondary HRP oligo that map to the PER concatemer sequence on the primary oligo (custom synthesis by Integrated DNA Technologies or Bio-Synthesis Inc) in PBS. Secondary hybridization was performed at 37°C with nutation for one hour. Cell pellets underwent three 5-minute washes in 1 ml PBS-T at 37°C with nutation. Fully washed cells were incubated in 5 µM desthiobiotin tyramide (Iris Biotech LS-1660) and 1 mM hydrogen peroxide in PBS-T for 5 minutes at room temperature with nutation. To quench the HRP activity, biotinylated cells were washed twice in 10 mM sodium ascorbate and 10 mM sodium azide in PBS-T for 5 minutes at room temperature with nutation. Quenched cells were washed with PBS to remove residual sodium azide. After sampling cells for quality control, the cell pellets were stored dry in -80°C until chromatin solubilization and affinity purification.

On-plate hybridization and biotinylation

EY.T4 oligo hybridizations were performed on cells fixed to glass bottom 6-well plates, with all washes being performed in 1mL. Cells were rinsed once with fresh phosphate buffered saline (PBS), and permeabilized in PBS-0.5% TritonX-100 (Sigma T8787) for 10 minutes. After a PBS-0.5% Tween20 (PBS-T) (Sigma T2287) wash, permeabilized cells were incubated in 0.1 N hydrochloric acid (HCl) for 5 minutes. After a PBS-T wash to remove acid, cells were incubated in PBS-T-0.5% hydrogen peroxide to block endogenous peroxidases. After a 2X saline sodium citrate-0.1% Tween20 (2X SSC-T) wash to remove acid, cells were incubated in 2X SSC-T-50% formamide for 20 minutes on a heat block in a 60°C water bath. Cells were exchanged into 500µL per well of primary hybridization buffer (Hyb1) comprising 2X SSC-T, 50% (vol/vol) formamide, 10% (wt/vol) dextran sulfate, 0.4 µg/ul RNase A, and ~250 nM extended primary oligos. The cells were denatured on a heat block in a water bath at 78°C for 3 minutes. The cells were then hybridized with primary oligo at 37°C overnight with nutation for about 24 hours. The next day, cells were rinsed with 60°C 2X SSC-T four times for five minutes each, followed by two two minute 2X SSC-T buffer exchanges at

room temperature to remove residual Hyb1. Fully washed cells were then washed with 1 ml PBS followed by 100 nM secondary HRP oligo that map to the concatemer sequence on the primary oligo in PBS. Secondary hybridization was performed at 37°C with nutation for one hour. Cells underwent three 5-minute washes in PBS-T at 37°C. Fully washed cells were incubated in 5 uM desthiobiotin tyramide (Iris Biotech LS-1660) and 1 mM hydrogen peroxide in PBS-T for 5 minutes at room temperature. To quench the HRP activity, biotinylated cells were washed twice in 10 mM sodium ascorbate and 10 mM sodium azide in PBS-T for 5 minutes at room temperature. Quenched cells were washed with PBS to remove residual sodium azide. The cells were stored dry at 4°C until chromatin solubilization and affinity purification.

Microscopy-based quality control assays for hybridization and biotinylation

We routinely sample cells along the workflow of preparing AP-MS or NGS samples to monitor the specificity of primary oligo hybridization. To assess the quality of primary oligo hybridization for the liquid-phase prepared cells, we sampled roughly 5% of fully washed cells from primary hybridization to a new 1.5 ml tube. Cells were incubated with 400 nM fluorescent oligos in PBS at 37°C for an hour with nutation. Hybridized cells underwent three washes in 1 ml PBS-T at 37°C with nutation to remove unbound fluorescent oligos. Washed cells were immobilized on glass slides with Slowfade Gold Antifade Mountant with DAPI (Thermo Fisher S36938) and coverslips for confocal imaging of FISH signal.

We assessed the quality of biotinylation specificity for all samples entering the proteomics or genomics workflow. For liquid-phase, roughly 5% of fully quenched cells were sampled into a new 1.5 ml tube and incubated with 0.5-1 µg/ml Alexa Fluor 647-streptavidin (Thermo Fisher S32357) in PBS-T, 1% bovine serum albumin at 37°C for 30 minutes with nutation. Stained cells underwent four washes in 1 ml PBS-T at 37°C with nutation to remove unbound Alexa Fluor 647-streptavidin conjugate. Washed cells were immobilized on glass slides with Slowfade Gold Antifade Mountant with DAPI and coverslips for confocal imaging of Alexa-Fluor 647-streptavidin signals.

For the solid-phase workflow experiments in EY.T4 cells, we assessed primary oligo hybridization and biotinylation specificity simultaneously. A separate 6-well plate was prepared alongside the others that was reserved for quality control with one well per condition. After biotinylation and subsequent washes this plate was incubated with 0.5-1 µg/ml Alexa Fluor 647-streptavidin (Thermo Fisher S32357) in PBS-T, 1% bovine serum albumin at 37°C for 30 minutes with nutation. Stained cells underwent four washes in 1 ml PBS-T at 37°C with nutation to remove unbound Alexa Fluor 647-streptavidin conjugate. These cells were then incubated with 400 nM fluorescent oligos in PBS at 37°C for an hour. Hybridized cells underwent three washes in 1 ml PBS-T at 37°C with nutation to remove unbound fluorescent oligos. Washed cells were immobilized on glass slides with Slowfade Gold Antifade Mountant with DAPI and coverslips for confocal imaging of Alexa-Fluor 647-streptavidin and FISH signal.

Confocal microscopy

Confocal imaging was performed using a Yokogawa CSU-W1 SoRa spinning disc confocal device attached to a Nikon ECLIPSE Ti2 microscope. Excitation light was emitted at 30% of maximal intensity from 405 nm, 488 nm, 561 nm, or 640 nm lasers housed inside of a Nikon LU-NF laser unit. Laser excitation was delivered via a single-mode optical fiber into the CSU-W1 SoRa unit. Excitation light was directed through a microlens array disk and a SoRa spinning disk containing 50 µm pinholes to the rear aperture of a 100x N.A. 1.49 Apo TIRF oil immersion objective lens by a prism in the base of Ti2. Emission light was collected by the same objective and directed by a prism in the base of Ti2 back into the SoRa unit, where it was relayed by a 1x lens (conventional imaging) or 2.8x lens (super-resolution imaging) through the pinhole disk and then directed to the emission path by a quad-band dichroic mirror (Semrock Di01-T405/488/568/647-13X15X0.5). Emission light was then spectrally filtered by one of four single-bandpass filters (DAPI: Chroma ET455/50M; ATTO488: Chroma ET525/36M; ATTO565: Chroma ET605/50M; Alexa Fluor 647: Chroma ET705/72M) and focused by a 1x relay lens onto an Andor Sona 4.2B-11 camera with a physical

pixel size of 11 μm , resulting in an effective resolution of 110 nm (conventional), or 39.3 nm (super-resolution). The Sona was operated in 16-bit mode with rolling shutter readout and exposure times of 70–300 ms.

FISH-biotinylation co-localization experiment

Fixed cells were split into 5e6 cell aliquots in 1.5 ml microcentrifuge tubes. Primary hybridization and washes were performed similarly to described in the in-solution hybridization and biotinylation of cell pellets with fewer cells. Fully washed cell pellets were exchanged into a secondary co-hybridization buffer containing 30 nM of fluorescent oligos and 100 nM of HRP-oligos in PBS, instead of solely HRP-oligos, for simultaneous hybridization of both species. After washes and biotinylation, the pellets were stained with 0.5–1 $\mu\text{g}/\text{ml}$ Alexa-Fluor 647-streptavidin. Cells were immobilized on glass slides with Slowfade Gold Antifade Mountant with DAPI and coverslips for confocal imaging of both FISH and Alexa-Fluor 647-streptavidin signals.

FISH-biotinylation co-localization quantification

To quantify the colocalization of signal in the FISH and Streptavidin channels, binary masks were generated for each image using a Gaussian blur (`scipy.ndimage.gaussian_filter`) and power-law transformation. The preprocessed image was then binarized using Otsu's method (`skimage.filters.threshold_otsu`). Parameter optimization for the sigma value of the Gaussian blur and the exponent values of the power-law transformation were done via grid search. Each parameter set was used to generate a binary mask and scored based on the percentage of pixels labeled as foreground in the mask that exceeded the 90th percentile of pixel intensity values in the input image. Parameters were optimized for each image and channel independently. These binary masks were then used to calculate the labeling efficiency and specificity of the Streptavidin signal relative to the ground truth FISH signal and the Jaccard index of both masks. Labeling efficiency is defined as the percentage of pixels labeled as foreground in the FISH mask that were also labeled as foreground in the Streptavidin mask. Specificity is defined as the percentage of pixels labeled as foreground in the Streptavidin mask that were also labeled as foreground in the FISH mask. The Jaccard index was calculated by dividing the number of pixels in the intersection of the two masks by the number of pixels in their union.

Affinity Purification and sample preparation for proteomics

For liquid-phase prepared samples, biotinylated cell pellets were removed from -80°C to thaw at room temperature. Each cell pellet was resuspended in roughly 0.9 ml of lysis buffer consisting of 1% SDS and 200 mM EPPS with protease inhibitors (Roche 11836170001). For the EY.T4 solid phase samples, 0.5 ml of lysis buffer was added directly to each well and cells were scrapped and collected in 1.5 ml tubes. The cell mixture was boiled at 95°C for 30 minutes. The boiled cell mixture was sonicated at 4°C using a Covaris LE-220 focused ultrasonicator with the following protocol: 300W peak incident power, 50% duty factor, 200 cycles per burst, with a treatment time of 420 seconds in 1-ml milliTUBEs with AFA fiber (Covaris 520135). The sonicated cell mixture was boiled for a second time at 95°C for 30 minutes. The boiled lysates were cleared by centrifuging at 21130 G for 30 minutes in an Eppendorf 5424 Microcentrifuge at room temperature. The supernatants were transferred to a fresh 1.5-ml tube. To prevent any remnants of cell debris, the supernatants were cleared for a second time by centrifuging at 21130 G for 30 minutes and the supernatants were transferred to a fresh 1.5-ml tube. The supernatants were stored in -80°C until protein quantification.

The cleared cell lysates were quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher 23225). Pierce Streptavidin Magnetic Beads (Thermo Fisher 88817) were washed using 1% SDS, 200 mM EPPS lysis buffer three times before use. From each labeled cell pellet, 2.17 milligrams of protein was used to couple with 500 μg of streptavidin beads in a Protein Lo-Bind tube (Eppendorf EP022431081). The lysates were incubated with the bead slurry for one hour at room temperature with nutation allowing biotinylated proteins to bind. The coupled beads were collected and separated from the flow-through using a magnetic rack (Sergi Lab Supplies 1005a). After the flow-

through was removed, the beads underwent the following washes: 2% SDS with 20 mM EPPS twice, 0.1 M Na₂CO₃, 2 M urea, and 1 M KCl with 20 mM EPPS twice. All washes were performed as follows: after immobilizing the beads on a magnetic rack for 5 minutes, the supernatant was removed, and the beads were resuspended in the new wash buffer and incubated for 5 minutes with nutation. Finally, the beads were rinsed once with 20 mM EPPS to remove the excess salt.

The washed streptavidin beads were resuspended in 50 µl of 5 mM TCEP, 200 mM EPPS, pH 8.5 for a 20-minute on-bead protein reduction. The proteins were alkylated on-bead using 10 mM iodoacetamide for one hour in the dark. Then DTT was added to the final concentration of 5 mM to quench the alkylation for 15 minutes. The beads were rinsed twice with 200 mM EPPS for on-bead digest. For liquid phase samples, assuming 20 µg of eluate protein, 200 ng LysC (Wako) was added to the beads in a 50-µl volume and incubated for 16 hours with vortexing. The next day, 200 ng of trypsin (Promega V5113) was added to the beads and incubated for six hours at 37°C at 200 rpm. For solid phase samples 20 ng of LysC and Trypsin was added and the samples were processed in the same way. After digestion, the peptide-containing supernatant was collected in a fresh 0.5-ml Protein Lo-Bind tube. The beads were rinsed once with 100 µl 50% acetonitrile, 5% formic acid and the wash was combined with the peptides. Peptides were desalted via the stop and go extraction (StageTip)¹⁰² method and dried in a vacuum concentrator. For label-free analysis of telomere-enriched samples, one sample consisted of HCT-116-Rad21-mAID cells¹⁰³. For samples intended to be multiplexed, dried, desalted peptides were reconstituted in 4 µl of 200 mM EPPS, pH 8.5. The peptides were labeled using 25 µg of TMTpro 16plex Label Reagents (Thermo Fisher A44520) at 33.3% acetonitrile for one hour at room temperature. The labeling reaction was quenched with the addition of 1 µl of 5% hydroxylamine and incubated at room temperature for 15 minutes. The pooled sample was acidified using formic acid and peptides were desalted using a StageTip cartridge. Peptides were eluted in 70% acetonitrile, 1% formic acid and dried by vacuum centrifugation.

Mass Spectrometry Data Acquisition Methods and Analysis

Samples were resuspended in 5% acetonitrile/2% formic acid prior to being loaded onto an in-house pulled C18 (Thermo Accucore, 2.6 Å, 150 µm) 30 cm column. Peptides were eluted over 180 min gradients running from 96% Buffer A (5% acetonitrile, 0.125% formic acid) and 4% buffer B (95% acetonitrile, 0.125% formic acid) to 30% buffer B. Sample eluate was electrosprayed (2700 V) into a Thermo Scientific Orbitrap Eclipse mass spectrometer for analysis. High field asymmetric waveform ion mobility spectrometry (FAIMS) was set at “standard” resolution, 4.6 L/min gas flow, and 3 CVs: -40/-60/-80 were used. Briefly, generally for both Label-free and TMTpro analyses, MS1 scans were collected at 120,000 resolving power with a 50 ms max injection time, and the AGC target set to 100%. Peaks from the MS1 scans were filtered by intensity (minimum intensity >5 × 10³), charge state (2 ≤ z ≤ 6), and detection of a monoisotopic mass (monoisotopic precursor selection for peptides, MIPS). Dynamic exclusion was used with a duration of 90 s, repeat count of 1, mass tolerance of 10 ppm, and the “exclude isotopes” option checked. For each MS1, 8 data-dependent MS/MS scans were collected. MS/MS scans were conducted in the linear ion trap with the “rapid” scan rate, 50 ms max injection time, AGC target set to 200%, CID collision energy of 35% with 10 ms activation time (TMTpro) or HCD at 30% collision energy (Label-free), and 0.5 m/z (TMTpro) or 0.7 m/z (Label-free) isolation window. For TMTPro labelled samples, an MS3 scan was also included in the method. Unless otherwise noted in the methods, the real-time search filter was enabled⁴⁵. Using a human fasta downloaded from Uniprot, fixed modifications for the TMTpro mass (+304.207146) were added to n-terminal residues and lysines. Carbamidomethyl (+57.021464) was added for cysteines. Oxidation (+15.9949) was added as a variable modification on methionines. Missed cleavages were set to maximum of 1. “TMT mode” was enabled and thresholds of 1 and 0.05 for Xcorr and dCn respectively were used as minimums to trigger SPS-MS3 scans. SPS ions were set to 10 and MS3 scans were performed at a resolving power of 50,000, with an HCD collision energy of 45%, AGC of 200%, with a maximum injection time of 200 ms.

Label-free mass spectrometry data for MS1-based quantitation was analyzed with MSFragger¹⁰⁴ search algorithm searched against a full human protein database for HCT 116 and K562 samples with forward and reverse protein sequences. EY.T4 samples were searched against a full mouse protein database with forward and reverse protein sequences. Fixed modifications included Carbamidomethyl (+57.021464) on cysteines. Variable modifications included oxidation (+15.9949) on methionine and formylation (+27.994915) on lysines. Peptides up to 2 missed cleavages were included. Peptide spectral matches and proteins were filtered to a 1% false discovery rate using Percolator¹⁰⁵.

Multiplexed raw mass spectrometry data was analyzed using the Comet¹⁰⁶ search algorithm, searched against a full human protein database with forward and reverse protein sequences (Uniprot 10/2020). Precursor monoisotopic peaks were estimated using the Monocle package. Fixed modifications included TMTpro (+304.207146) on n-terminal residues and lysines and Carbamidomethyl (+57.021464) on cysteines. Variable modifications included methionine oxidation (+15.9949) and lysine formylation (+27.994915). Peptides with up to 2 missed cleavages were included. Peptide spectral matches and proteins were filtered to a 1% false discovery rate using the rules of parsimony and protein picking. Protein quantification was done using signal-to-noise estimates of reporter ions. Samples were column normalized for total protein concentration. After filtering for contaminants, we performed a two-sided t-test comparing each O-MAP condition using Benjamini-Hochberg adjusted p values (i.e. q-values). Log₂ fold changes of the mean of the biological replicates were also calculated for each biological condition. For the GSK treatment experiment, prior to graphing, the data were run through a custom R script using ComBat to correct for batch effects encountered when processing the samples. Human Protein Atlas⁵⁹ subcellular locations were downloaded and the “main location” was assigned to each protein with a supported or enhanced reliability level. SAINT scores and interaction false discovery rates were calculated with the SAINTexpress software^{107,108}. Significant hits were those with a SAINT calculated FDR less than 1%¹⁰⁹. BioPlex interaction networks were accessed through the online BioPlex Explorer¹¹⁰. Networks were imaged using Cytoscape 3.10.02¹¹¹. Protein complex members were accessed through CORUM¹¹². Gene set enrichment analysis was performed with clusterProfiler¹¹³ and fgsea¹¹⁴ packages.

Preparation of soluble chromatin for affinity purification followed by next generation sequencing

For confirmation of single-copy O-MAP labeling, loop anchor-biotinylated pellets of 10-20 million cells were removed from -80°C to thaw at room temperature. Each cell pellet was resuspended in an SDS lysis buffer consisting of 1% SDS and 200 mM EPPS with protease inhibitors. The cell mixture was sonicated at 4°C using a Covaris LE-220 focused ultrasonicator with the following protocol: 300W peak incident power, 15% duty factor, 200 cycles per burst, with a treatment time of 20-30 minutes in 130-µl microTUBEs with AFA fiber (Covaris 520077). After the samples had returned to room temperature, the sheared fixed chromatin was transferred to fresh 1.5-ml Protein Lo-Bind tubes and centrifuged at 21130 G for 10 minutes to pellet cellular debris. The supernatants were transferred to a new set of tubes. The cleared chromatin samples were quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher 23225). Next, 50 µl of sheared chromatin was sampled for reverse crosslinking, DNA extraction, and gel electrophoresis to verify that a significant amount of DNA had been sheared to <700 base pairs. A sample of 10 µg sheared chromatin was reserved and stored at -20°C as immunoprecipitation input. 200 µg of chromatin was used to couple with 200 µg of streptavidin beads for one hour in a Protein Lo-Bind tube at room temperature with nutation. The coupled beads were collected and separated from the flow-through using a magnetic rack. After the flow-through was removed, the beads underwent the following washes:

- 2% SDS with 20 mM EPPS
- 2% SDS with 20 mM EPPS
- High Salt Buffer containing 500 mM NaCl, 1 mM EDTA, 50 mM of HEPES pH7.5, 0.1% sodium deoxycholate, and 1% TritonX-100

- LiCl Buffer containing 250 mM LiCl, 1 mM EDTA, 10 mM Tris-HCl pH 8.0, and 0.5% of IGEPAL CA-630
- TE Buffer with 10 mM Tris and 1 mM EDTA
- TE Buffer with 10 mM Tris and 1 mM EDTA

The washes were performed as follows: briefly spin and immobilize the beads on a magnetic rack, pipette out the supernatant as much as possible, resuspend the beads in 0.8 ml of wash buffer, and incubate for 5 minutes with nutation. The washed beads were resuspended in 300 µl of reverse crosslinking buffer containing 300 mM NaCl, 300 mM Tris-HCl pH 8.0, and 1 mM EDTA. Both the eluate beads and the input chromatin were incubated at 65°C for 16 hours for reverse crosslinking. The next day, 4 µl of 20 mg/ml proteinase K (Roche 3115836001) was added to the eluates and inputs and incubated at 50°C for 2 hours to cleave away proteins. The DNA was isolated from the mixture using phenol chloroform extraction followed by ethanol precipitation. Before sequencing library generation, the precipitated DNA was further purified using SPRI beads. The purified DNA was used to generate next-generation sequencing libraries using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB E7645S) and NEBNext Multiplex Oligos for Illumina Index Primers Set 1 and 3 (NEB E7335S, E7710S) and PCR-amplified for 15 cycles. The sequencing libraries were quantified using the Qubit 4 fluorometer and library sizes were quantified using the D1000 ScreenTape assay (Agilent 5067-5582) on the TapeStation 4200 automated electrophoresis platform.

DNA sequencing and data analysis

The libraries were mixed and sequenced pair-ended at 50-bp read length on an Illumina NextSeq 2000 sequencer to depths of 14.1-351.8 million reads per eluate sample and 3.14-16.45 millions reads per input sample using the NextSeq 1000/2000 P2 Reagents (100 Cycles) kit (Illumina 20046811). Reads were demultiplexed and adapters were removed using Cutadapt¹¹⁵. Trimmed reads were mapped to the reference genome (GRCh38) using Bowtie2 version 2.5.3 with the parameter -X 1000 keeping reads with a MAPQ>=30¹¹⁶. Duplicate reads were removed using Picard 3.1.1¹¹⁷. Eluate reads were normalized to input reads using DeepTools¹¹⁸ bamCompare with the following parameters: -binSize 20 -normalizeUsing BPM -smoothLength 60 -extendReads 150. Normalized data were visualized using Coolbox 0.3.9¹¹⁹.

To assess the targeting specificity of DNA O-MAP, O-MAP ChIP sequencing reads were aligned to a reference genome using Bowtie2. Duplicate reads were removed with Picard and Samtools. Enrichment scores were calculated using a bin-based method adapted from the TSA-seq protocol. The genome is divided into 100kb windows, and the enrichment score for each window is determined by the following formula: Enrichment Score = (Input reads in bin/Sum of input reads)/(Pull-down reads in bin/Sum of pull-down reads). To normalize the data and prevent division by zero in bins with no reads, a pseudocount equal to 10% of the total reads was added to each bin. Finally, the calculated enrichment scores were plotted against the log-scaled distance from the target regions.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange¹²⁰ Consortium via the MassIVE with the data set identifier PXD054080. Sequencing data will be deposited to Gene Expression Omnibus before formal acceptance for publication. All primary data associated with the manuscript will be made available upon request.

Supplemental figures

Acknowledgements

We would like to thank members of the Shechner, Beliveau, and Schweppe labs for constructive feedback and technical assistance in assembling this work. We would also like to thank Drs. Jay Shendure, Shao-En Ong, Christine Quietsch, Emily Hatch, Gavin Ha, Celeste Berg, Christine

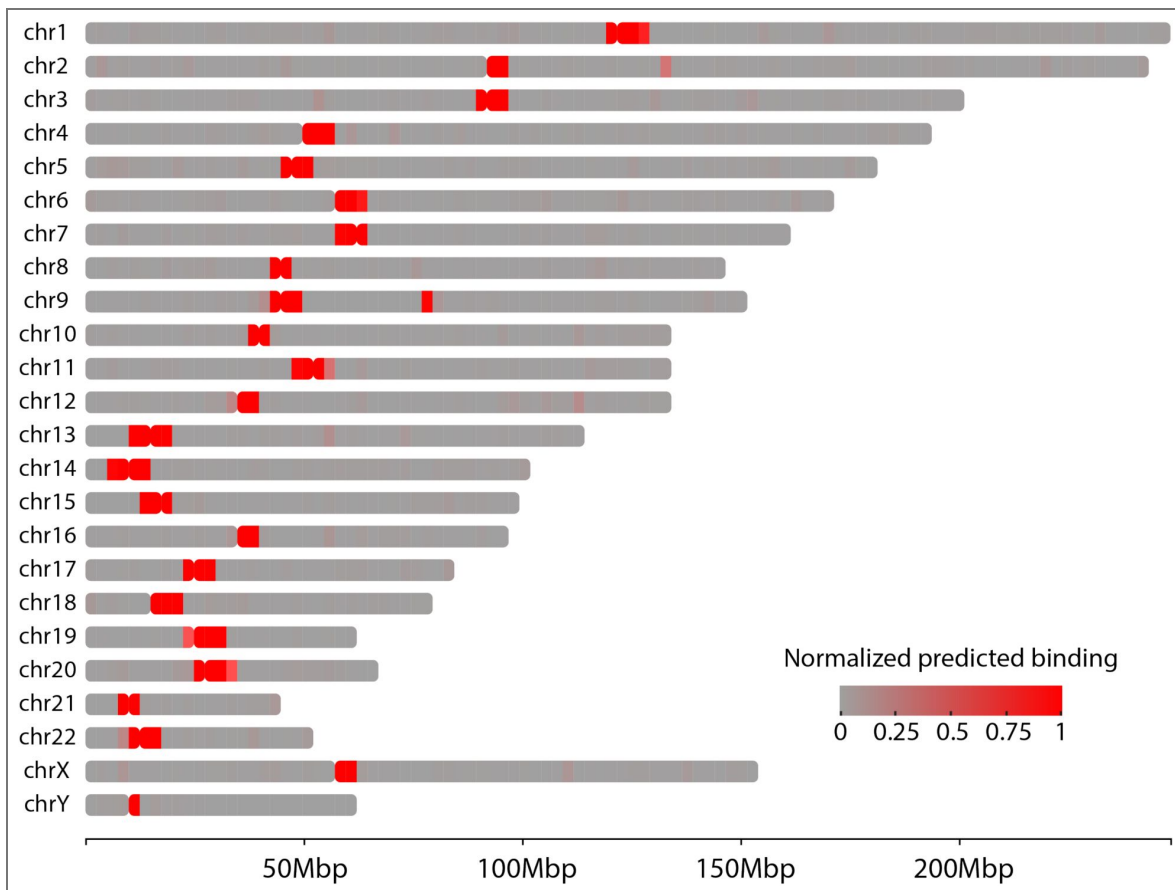


Figure S1. Predicted genome-wide binding profile of the pan-alpha probe.

The intensity of red indicates the amount of predicted probe binding.

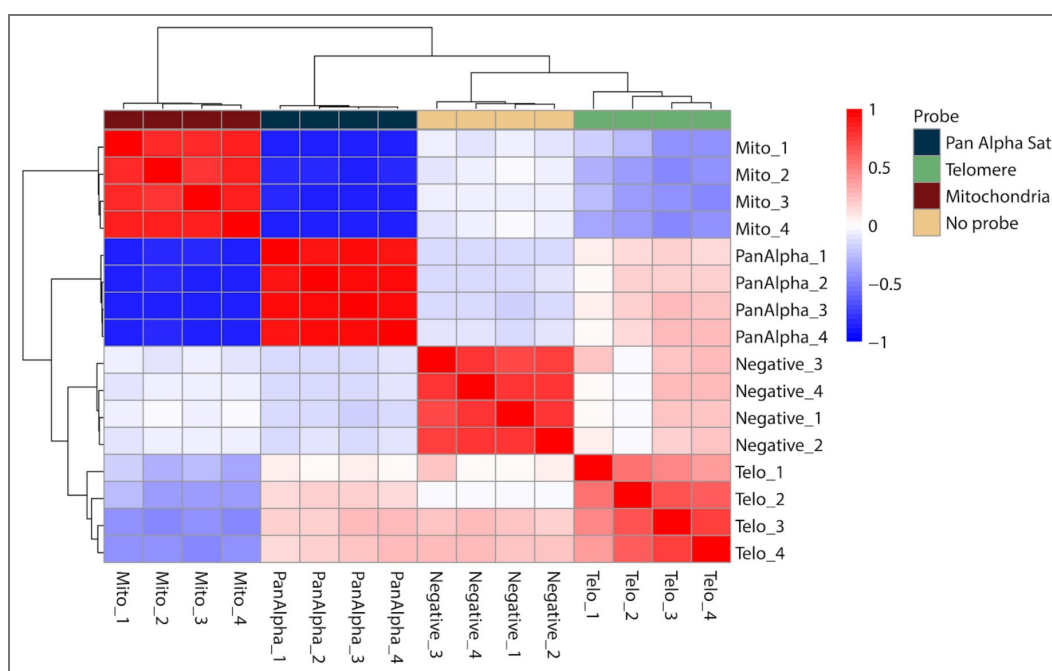


Figure S2. Replicate analysis of multi-target DNA O-MAP proteomics experiment.

A) Pearson correlation coefficient of the raw protein intensity values for each replicate of the analysis with hierarchical clustering on the rows and columns.

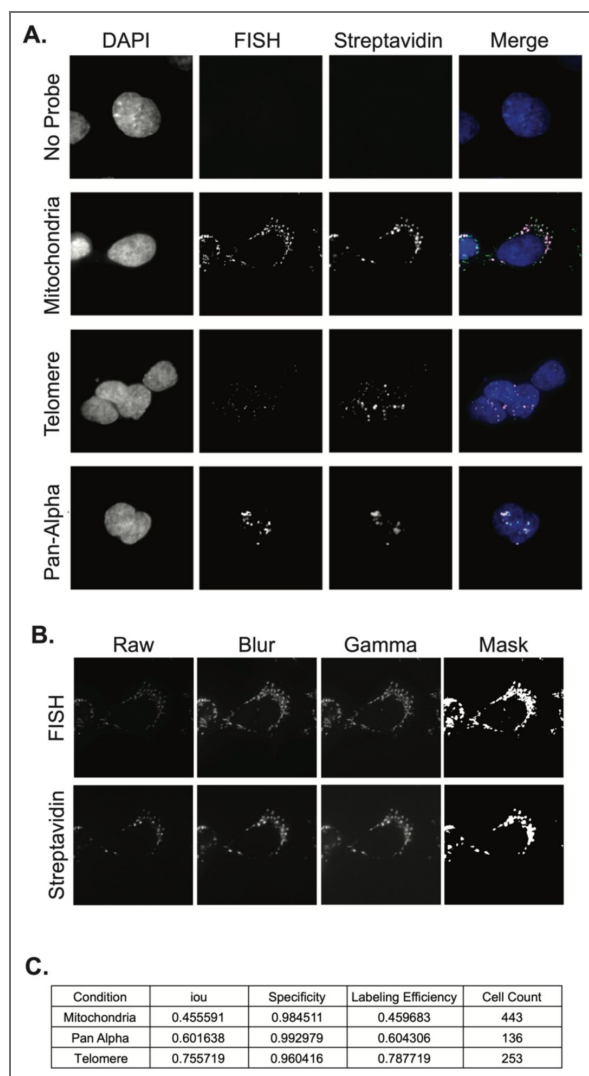


Figure S3. Quantification of DNA O-MAP labeling specificity and efficiency for pan-alpha, telomere, and mitochondria probes.

A) Fluorescent microscopy data showing co-localization of DNA FISH and streptavidin staining of the proteins biotinylated by DNA O-MAP using no primary probe or targeting the mitochondrial genome, telomeres, or alpha satellites. B) Representative images for both DNA FISH and streptavidin stain from each stage of quantifying specificity and labeling efficiency for probes targeting the mitochondrial genome. C) Table of results for the mitochondrial genome, telomeres, and alpha satellites showing iou, specificity, labelling efficiency, and cell count of images quantified for DNA FISH and streptavidin staining.

Figure S4. Relative quantitation for the multi-target DNA O-MAP proteomics experiment compared to no-probe control and mtDNA datasets.

Volcano plots from multiplexed proteomics experiments with proteins of interest highlighted. A-C) Fold-changes and significance calculated compared to no probe. D-F) Fold-changes and significance calculated compared to mtDNA probe.

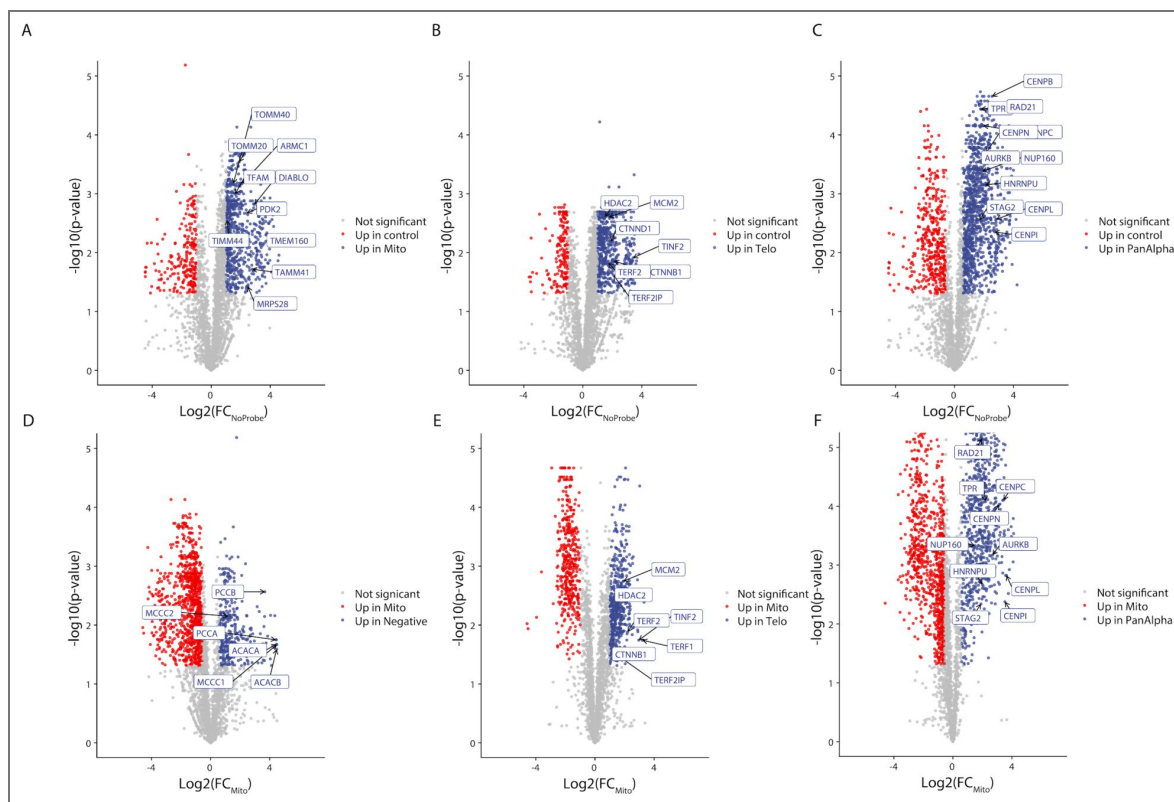
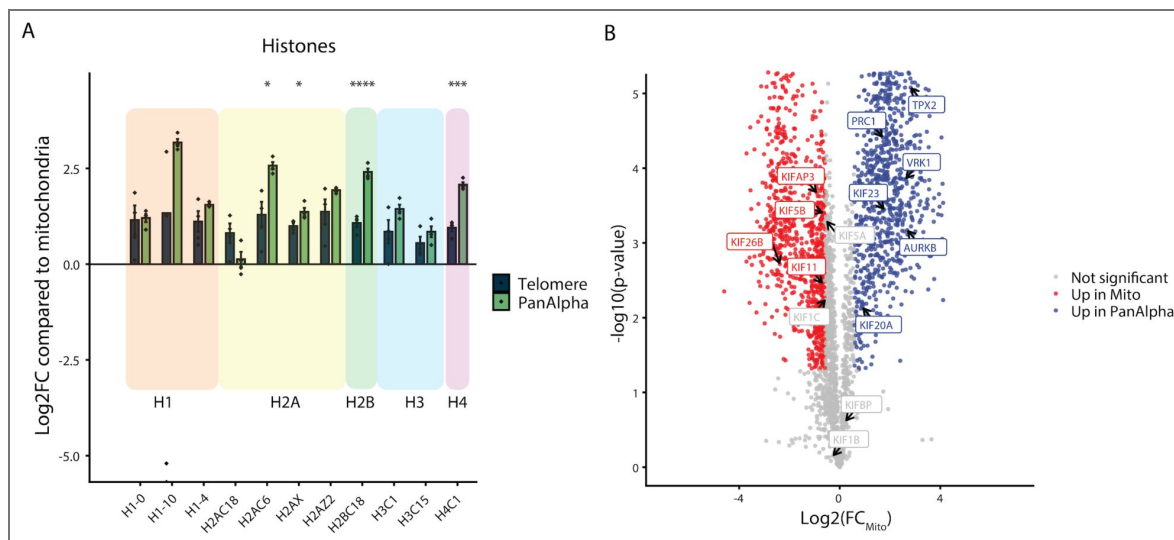


Figure S5. Comparison of histone proteins between telomere and pan-alpha probes.

A) Log2 fold change of proteins compared to mitochondrial probe enriched histone complex proteins. Significance calculated based on Welch's t-test for pairwise comparisons (p-value: $^* < 0.05$, $^{**} < 0.01$, $^{***} < 0.001$, $^{****} < 0.0001$). B) Volcano plot comparing the fold change of pan-alpha to the mtDNA probe with spindle proteins highlighted.



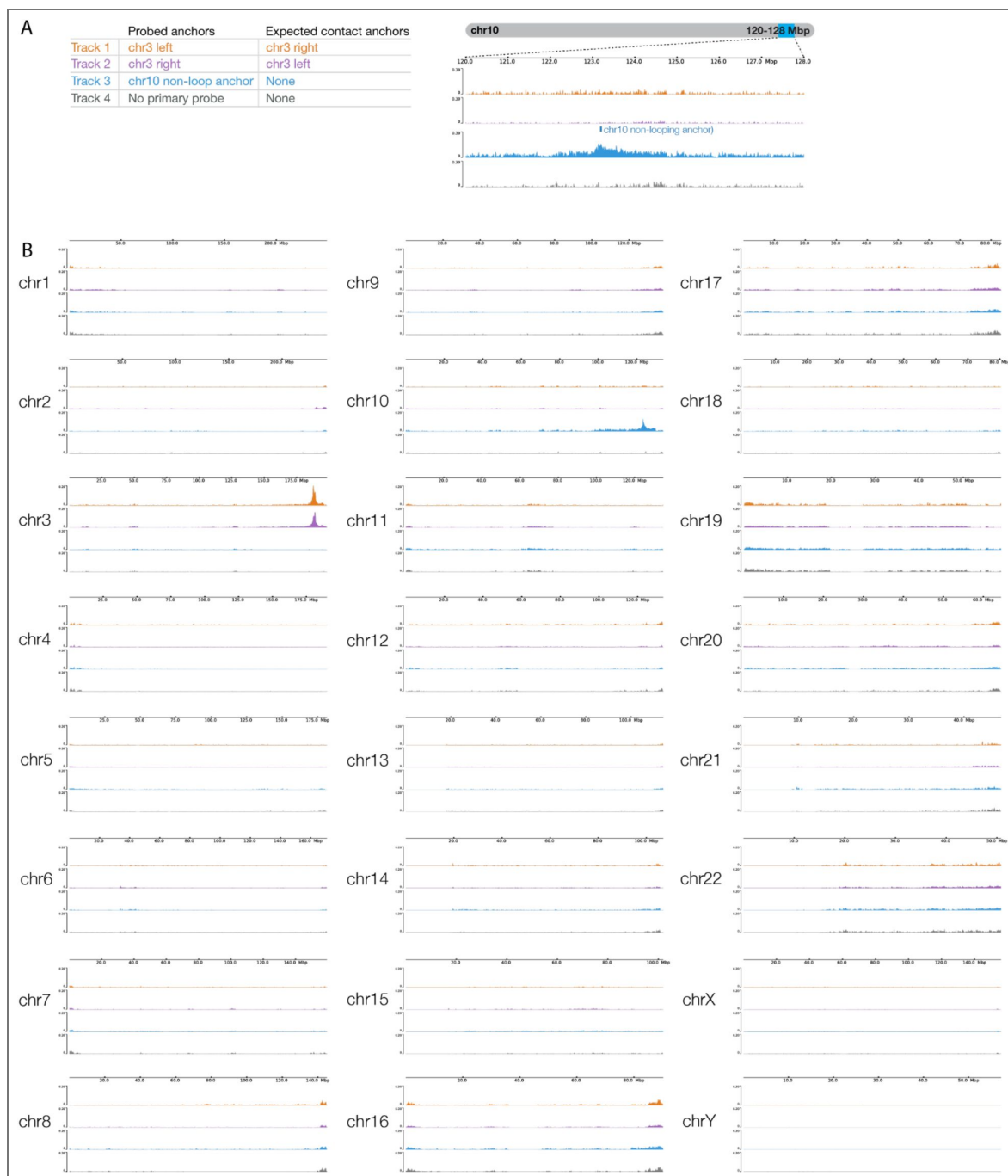


Figure S6. DNA O-MAP biotin purification sequencing of chr3 left, chr3 right, chr10 non-loop anchors, and no-primary-probe control.

A) Table listing the three anchors (Track 1-3) and no-primary-probe control (Track 4) biotinylated by DNA O-MAP and their expected contact anchors (left). Biotin purification sequencing signals across the 8-Mb region on chromosome 10 corresponding to the chr10 non-loop anchor targeted (right). B) Biotin purification sequencing signals across every chromosome in the genome for this experiment.

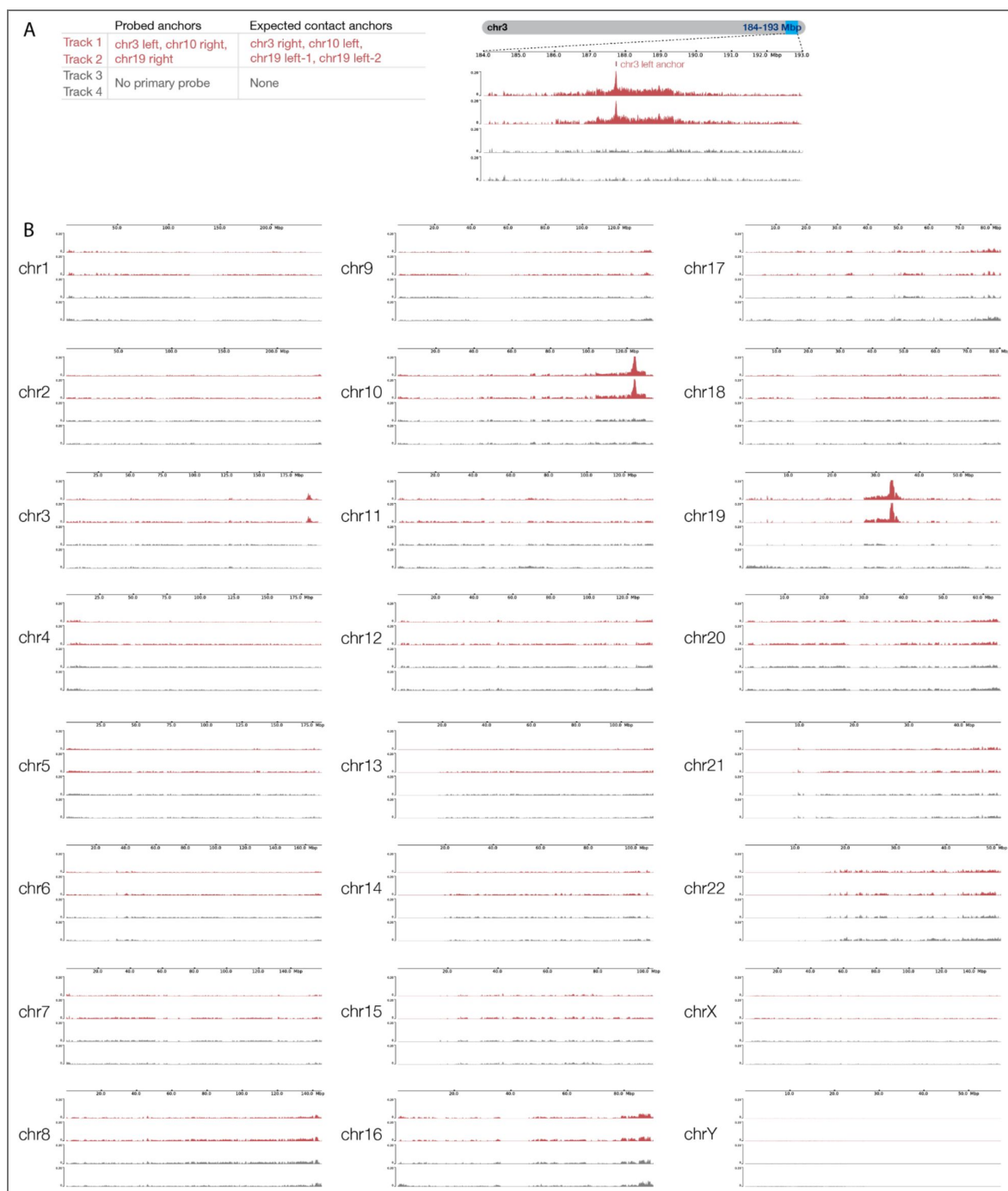


Figure S7. DNA O-MAP biotin purification sequencing of multiplexed targeting of chr3 left, chr10 right, chr19 right anchors, and no-primary-probe control in duplicates.

A) Table listing the three anchors (Track 1-2) and no-primary-probe control (Track 3-4) biotinylated by DNA O-MAP and their expected contact anchors (left). Biotin purification sequencing signals across the 9-Mb region on chromosome 3 corresponding chr3 left anchor targeted in Track 1-2 (right). B) Biotin purification sequencing signals across every chromosome in the genome for this experiment.

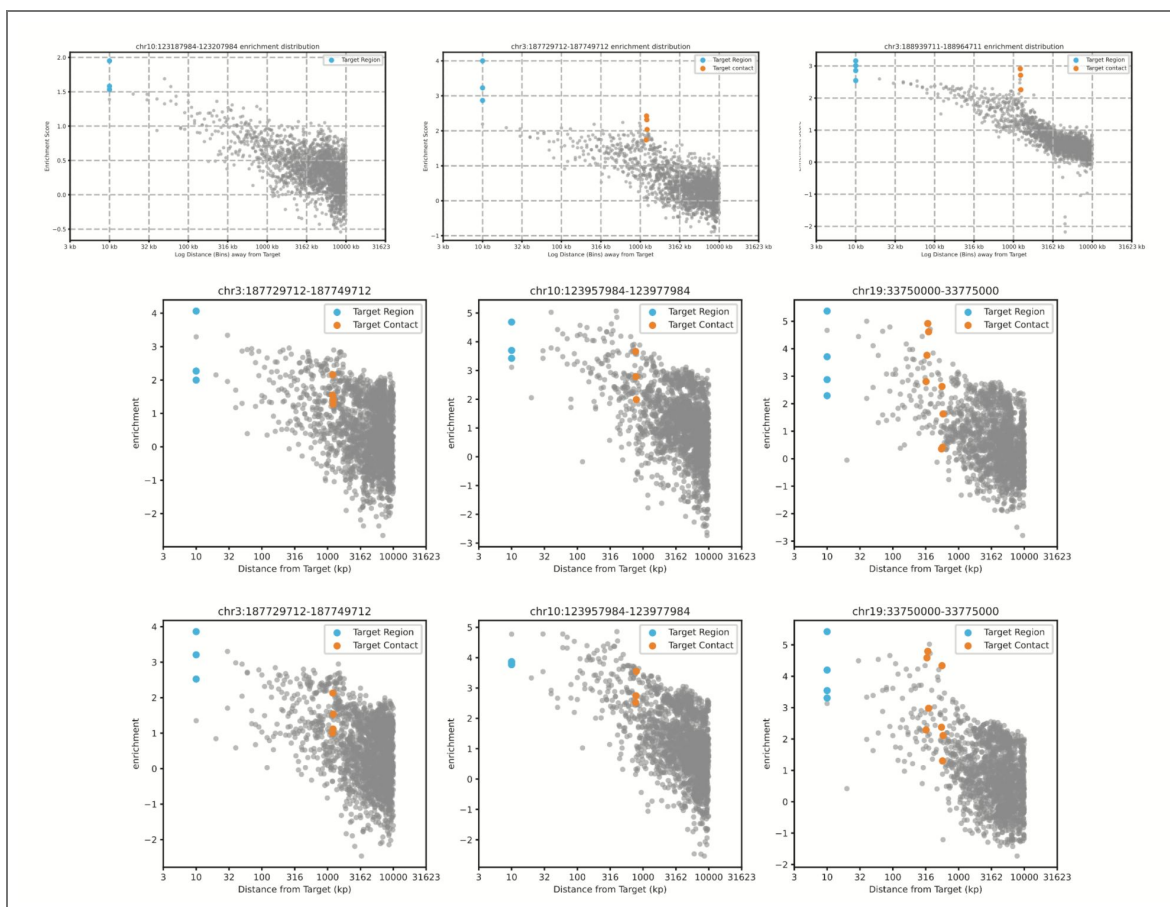


Figure S8. DNA O-MAP biotin purification sequencing enrichment score distribution of multiplexed targeting of chr3 left, chr10 right, chr19 right anchors.

The scatter plots illustrate the enrichment score distribution for experiments targeting specific regions of chromosome 3, 10, and 19. The x-axes show the log distance from the target in kilobase pairs (kbp), and the y-axes show the enrichment scores. Data points highlighted in blue represent the primary target region and the data points highlighted in orange represent the contact regions.

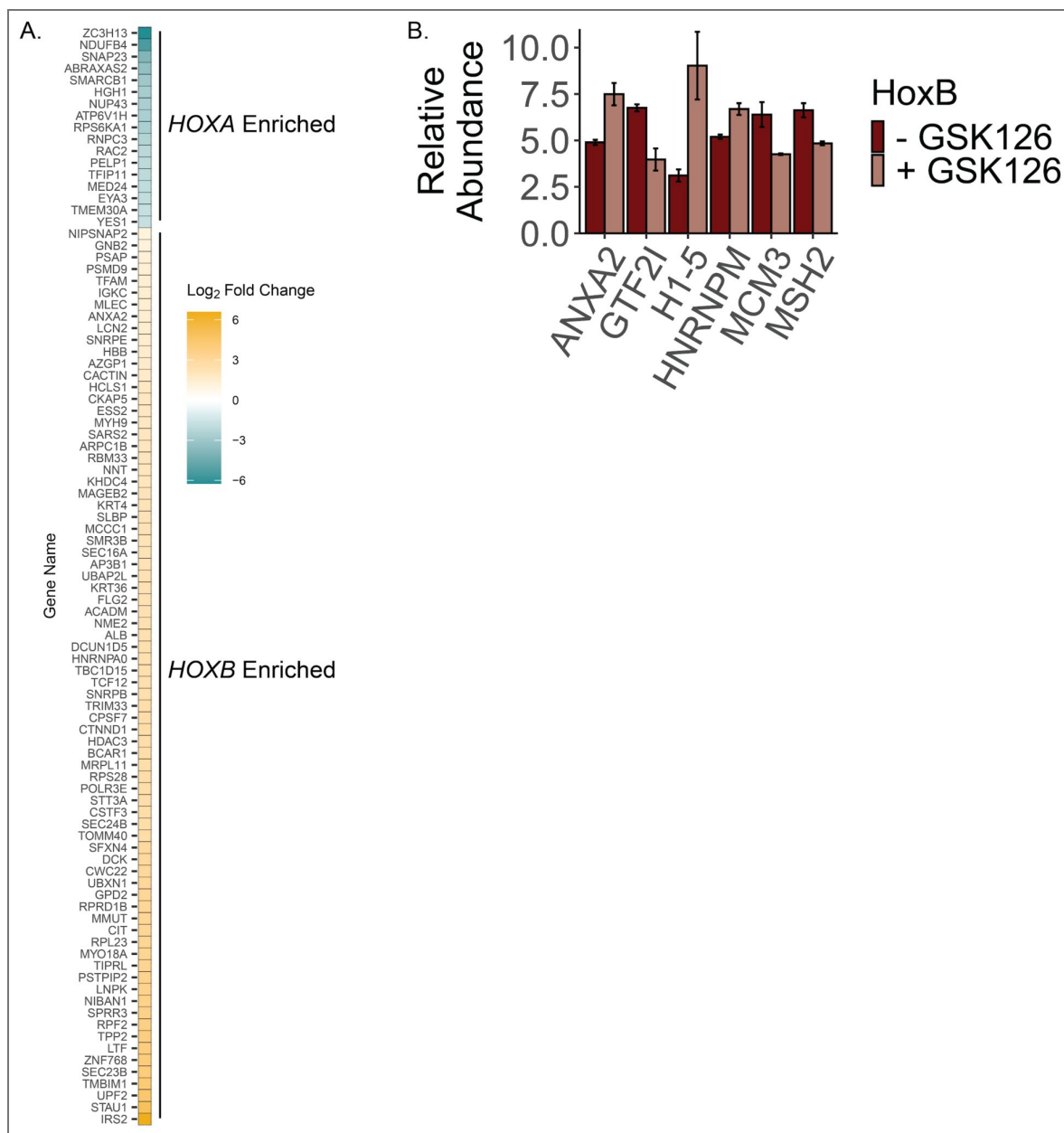


Figure S9. DNA O-MAP elucidates differences between the *HOXA* and *HOXB* proximal proteomes and changes to their proteomes following inhibition of EZH2 with GSK126.

A) Heat map showing the Log₂ fold change of all proteins that were scored as significantly enriched when comparing *HOXA* and *HOXB* proximal proteomes in K562 cells by DNA-O-MAP. Adjusted p-value < 0.05. B) Bar chart showing proteins with significantly altered abundance at *HOXB*, unique from *HOXA*, following treatment with GSK126 (Welch's t-test, corrected p-value < 0.05).

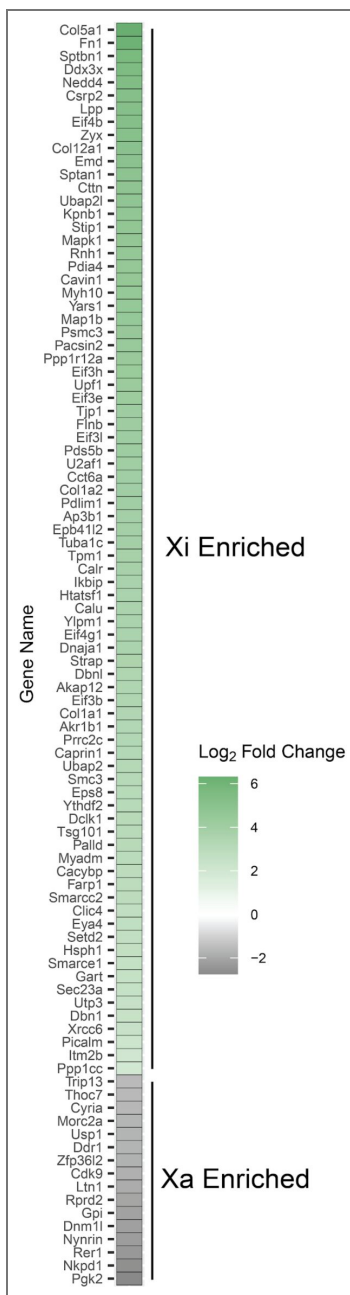


Figure S10. DNA O-MAP elucidates the differential proteomes of the X chromosome at homolog resolution.

Heat map showing the Log₂ fold change of all proteins that were scored as significantly enriched when comparing the active and inactive X chromosome homolog proximal proteomes in EY.T4 cells by DNA-OMAP. Adjusted p-value < 0.1.

Disteche, Andrew Stergachis, and Stanley Fields for helpful discussions of this work. We would like to acknowledge the following sources of support: R35GM137916 (BJB), R35GM150919 (DKS), the W.M. Keck Foundation (BJB, DKS), an Andy Hill CARE Distinguished Researcher Award (DKS), a Damon Runyon Dale Frey Award (BJB), a Cancer Consortium New Investigator Award (DKS), the Pew Charitable Trusts (DKS), 1R01GM138799-01 and 1R01HL160825-01 (DMS), T32GM007750 (to AFT and EEK), AHA 902616 (to EEK). Research reported in this publication was supported by the NHLBI under award number T32HL007093 (to CPH). This work was also supported by a Research and Education Training Fund Award (to CH) from the Center for the Multiplex Assessment of Phenotype at UW.

Additional information

Author Contributions

Y.L., C.D.M., C.P.H., B.J.B., and D.K.S. conceived and designed the project. Y.L., C.D.M., C.P.H., M.K., T.A.P., R.F., C.H., S.A., A.F.T., and E.K. performed experiments. Y.L., C.D.M., C.P.H., N.J.L., and C.K.C. performed computational analyses. Y.L., C.D.M., C.P.H., B.J.B., and D.K.S. wrote the manuscript. All authors edited and reviewed the manuscript. D.M.S., B.J.B., and D.K.S. supervised the work.

Funding

Funder	Grant reference number	Author
HHS NIH National Institute of General Medical Sciences (NIGMS)	R35GM137916	Brian J Beliveau
HHS NIH National Institute of General Medical Sciences (NIGMS)	R35GM150919	Devin Schweppe

Author ORCID iDs

Yuzhen Liu: <https://orcid.org/0000-0002-7299-4708>
Conor P Herlihy: <https://orcid.org/0000-0001-9818-4204>
Conor K Camplisson: <https://orcid.org/0000-0001-5085-7924>
Shayan C Avanesian: <https://orcid.org/0000-0002-8883-0070>
Ashley F Tsue: <https://orcid.org/0000-0001-6935-6601>
Evan E Kania: <https://orcid.org/0000-0002-0266-6610>
David M Shechner: <https://orcid.org/0000-0002-0574-256X>
Brian J Beliveau: <https://orcid.org/0000-0003-1314-3118>
Devin K Schweppe: <https://orcid.org/0000-0002-3241-6276>

Author notes

Yuzhen Liu, Christopher D McGann, Conor P Herlihy: These authors contributed equally: Yuzhen Liu, Christopher D. McGann, Conor P. Herlihy.

Competing interests: D.K.S. is a collaborator with Thermo Fisher Scientific, Genentech, Calico Labs, and AI Proteins. C.K.C., A.F.T., E.K., D.M.S., and B.J.B. have filed a patent application covering aspects of this work. B.J.B. is listed as an inventor on patent applications related to the SABER technology related to this work.

Additional files

Supplemental Table 1. [Oligonucleotide probe sequences used in this work.](#)

Supplemental Table 2. [Proteomic data for the enrichment of telomere probe associated proteins.](#)

Supplemental Table 3. [Quantitative proteomic data for the multi-target DNA O-MAP proteomics experiment.](#)

Supplemental Table 4. [Summary of oligonucleotide probe target information.](#)

Supplemental Table 5. [↗](#) Proteomic data for the differential enrichment of *HOXA* and *HOXB* associated proteins.

Supplemental Table 6. [↗](#) Quantitative proteomic data for the GSK126 treated *HOXA* proteomics experiment.

Supplemental Table 7. [↗](#) Quantitative proteomic data for the GSK126 treated *HOXB* proteomics experiment.

Supplemental Table 8. [↗](#) Proteomic data for the differential enrichment of Xi and Xa associated proteins.

References

1. **Bickmore W. A.,** van Steensel B (2013) Genome architecture: domain organization of interphase chromosomes. *Cell* **152**:1270-1284 <https://doi.org/10.1016/j.cell.2013.02.001> | [PubMed](#)
2. **Jerkovic I.,** Cavalli G (2021) Understanding 3D genome organization by multidisciplinary methods. *Nat. Rev. Mol. Cell Biol* **22**:511-528 <https://doi.org/10.1038/s41580-021-00362-w> | [PubMed](#)
3. **Johnson D. S.,** Mortazavi A., Myers R. M., Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**:1497-1502 <https://doi.org/10.1126/science.1141319> | [PubMed](#)
4. **Ho J. W. K.,** Alekseyenko A. A., Kuroda M. I., Park P. J (2012) *Genome-wide mapping of protein-DNA interactions by ChIP-seq. in Tag-Based Next Generation Sequencing* Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. pp. 139-151
5. **van Steensel B.,** Henikoff S (2000) Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol* **18**:424-428 <https://doi.org/10.1038/74487> | [PubMed](#)
6. **Kaya-Okur H. S.,** et al. (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun* **10**:1930 <https://doi.org/10.1038/s41467-019-09982-5> | [PubMed](#)
7. **Luo Y.,** et al. (2020) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res* **48**:D882-D889 <https://doi.org/10.1093/nar/gkz1062> | [PubMed](#)
8. **Project Consortium ENCODE** (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**:636-640
9. **Bujold D.,** et al. (2016) The International Human Epigenome Consortium Data Portal. *Cell Syst* **3**:496-499.e2 <https://doi.org/10.1016/j.cels.2016.10.019> | [PubMed](#)
10. **Ernst J.,** Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**:215-216 <https://doi.org/10.1038/nmeth.1906> | [PubMed](#)
11. **Hoffman M. M.,** et al. (2012) Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* **9**:473-476 <https://doi.org/10.1038/nmeth.1937> | [PubMed](#)
12. **Gao X. D.,** et al. (2018) C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nat. Methods* **15**:433-436 <https://doi.org/10.1038/s41592-018-0006-2> | [PubMed](#)
13. **Myers S. A.,** et al. (2018) Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nat. Methods* **15**:437-439 <https://doi.org/10.1038/s41592-018-0007-1> | [PubMed](#)
14. **Qiu W.,** et al. (2019) Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system. *Nucleic Acids Res* **47**:e52 <https://doi.org/10.1093/nar/gkz134> | [PubMed](#)
15. **Ugur E.,** Bartoschek M. D., Leonhardt H (2020) Locus-specific chromatin proteome revealed by mass spectrometry-based CasID. *Methods Mol. Biol* **2175**:109-121 https://doi.org/10.1007/978-1-0716-0763-3_9 | [PubMed](#)
16. **Déjardin J.,** Kingston R. E (2009) Purification of proteins associated with specific genomic Loci. *Cell* **136**:175-186 <https://doi.org/10.1016/j.cell.2008.11.045> | [PubMed](#)

17. Silahatoglu A. N., Tommerup N., Vissing H (2003) FISHing with locked nucleic acids (LNA): evaluation of different LNA/DNA mixmers. *Mol. Cell. Probes* **17**:165-169 [https://doi.org/10.1016/s0890-8508\(03\)00048-3](https://doi.org/10.1016/s0890-8508(03)00048-3) | PubMed
18. Beliveau B. J., et al. (2012) Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A* **109**:21301-21306 <https://doi.org/10.1073/pnas.1213818110> | PubMed
19. Ide S., Dejardin J (2015) End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. *Nat. Commun* **6**:6674 <https://doi.org/10.1038/ncomms7674> | PubMed
20. Roux K. J., Kim D. I., Raida M., Burke B (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol* **196**:801-810 <https://doi.org/10.1083/jcb.201112098> | PubMed
21. Cho K. F., et al. (2020) Proximity labeling in mammalian cells with TurboID and split-TurboID. *Nat. Protoc* **15**:3971-3999 <https://doi.org/10.1038/s41596-020-0399-0> | PubMed
22. Lam S. S., et al. (2015) Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**:51-54 <https://doi.org/10.1038/nmeth.3179> | PubMed
23. Martell J. D., et al. (2012) Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol* **30**:1143-1148 <https://doi.org/10.1038/nbt.2375> | PubMed
24. Cenik B. K., et al. (2024) TurboCas: A method for locus-specific labeling of genomic regions and isolating their associated protein interactome. *Mol. Cell* **84**:4929-4944.e8 <https://doi.org/10.1016/j.molcel.2024.11.007> | PubMed
25. Branon T. C., et al. (2018) Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol* **36**:880-887 <https://doi.org/10.1038/nbt.4201> | PubMed
26. Mangeot P. E., et al. (2019) Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat. Commun* **10** <https://doi.org/10.1038/s41467-018-07845-z> | PubMed
27. Tsue A. F., et al. (2024) Multiomic characterization of RNA microenvironments by oligonucleotide-mediated proximity-interactome mapping. *Nat. Methods* **21**:2058-2071 <https://doi.org/10.1038/s41592-024-02457-6> | PubMed
28. Attar S., et al. (2024) Efficient and highly amplified imaging of nucleic acid targets in cellular and histopathological samples with pSABER. *Nature Methods* 1-10 <https://doi.org/10.1038/s41592-024-02512-2> | PubMed
29. Li J., et al. (2020) TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nat. Methods* **17**:399-404 <https://doi.org/10.1038/s41592-020-0781-4> | PubMed
30. Chakravarti D., LaBella K. A., DePinho R. A (2021) Telomeres: history, health, and hallmarks of aging. *Cell* **184**:306-322 <https://doi.org/10.1016/j.cell.2020.12.028> | PubMed
31. Sfeir A., de Lange T (2012) Removal of shelterin reveals the telomere end-protection problem. *Science* **336**:593-597 <https://doi.org/10.1126/science.1218498> | PubMed
32. de Lange T. (2018) Shelterin-mediated telomere protection. *Annu. Rev. Genet* **52**:223-247 <https://doi.org/10.1146/annurev-genet-032918-021921> | PubMed
33. Myung K., et al. (2004) Regulation of telomere length and suppression of genomic instability in human somatic cells by Ku86. *Mol. Cell. Biol* **24**:5050-5059 <https://doi.org/10.1128/mcb.24.11.5050-5059.2004> | PubMed
34. Subramanian A., et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A* **102**:15545-15550 <https://doi.org/10.1073/pnas.0506580102> | PubMed
35. Liu X., et al. (2017) In situ capture of chromatin interactions by biotinylated dCas9. *Cell* **170**:1028-1043.e19 <https://doi.org/10.1016/j.cell.2017.08.003> | PubMed

36. Garcia-Exposito L., et al. (2016) Proteomic profiling reveals a specific role for translesion DNA polymerase η in the alternative lengthening of telomeres. *Cell Rep* **17**:1858-1871 <https://doi.org/10.1016/j.celrep.2016.10.048> | PubMed
37. LaBranche H., et al. (1998) Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat. Genet* **19**:199-202 <https://doi.org/10.1038/575> | PubMed
38. Zhang Q.-S., Manche L., Xu R.-M., Krainer A. R (2006) hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA* **12**:1116-1128 <https://doi.org/10.1261/rna.58806> | PubMed
39. Flynn R. L., et al. (2011) TERRA and hnRNP A1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature* **471**:532-536 <https://doi.org/10.1038/nature09772> | PubMed
40. Fu D., Collins K (2007) Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation. *Mol. Cell* **28**:773-785 <https://doi.org/10.1016/j.molcel.2007.09.023> | PubMed
41. Izumi H., Funa K (2019) Telomere Function and the G-Quadruplex Formation are Regulated by hnRNP U. *Cells* **8** <https://doi.org/10.3390/cells8050390> | PubMed
42. Schweppe D. K., Huttlin E. L., Harper J. W., Gygi S. P (2018) BioPlex Display: An Interactive Suite for Large-Scale AP-MS Protein-Protein Interaction Data. *J. Proteome Res* **17**:722-726 <https://doi.org/10.1021/acs.jproteome.7b00572> | PubMed
43. Huttlin E. L., et al. (2021) Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**:3022-3040.e28 <https://doi.org/10.1016/j.cell.2021.04.011> | PubMed
44. Wang L., et al. (2016) Transcriptomic characterization of SF3B1 mutation reveals its pleiotropic effects in chronic Lymphocytic leukemia. *Cancer Cell* **30**:750-763 <https://doi.org/10.1016/j.ccell.2016.10.005> | PubMed
45. Schweppe D. K., et al. (2020) Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *J. Proteome Res* **19**:2026-2034 <https://doi.org/10.1021/acs.jproteome.9b00860> | PubMed
46. Navarrete-Perea J., Yu Q., Gygi S. P., Paulo J. A (2018) Streamlined tandem mass tag (SL-TMT) protocol: An efficient strategy for quantitative (phospho)proteome profiling using tandem mass tag-synchronous precursor selection-MS3. *J. Proteome Res* **17**:2226-2236 <https://doi.org/10.1021/acs.jproteome.8b00217> | PubMed
47. Schweppe D. K., et al. (2019) Characterization and Optimization of Multiplexed Quantitative Analyses Using High-Field Asymmetric-Waveform Ion Mobility Mass Spectrometry. *Anal. Chem* **91**:4010-4016 <https://doi.org/10.1021/acs.analchem.8b05399> | PubMed
48. McKinley K. L., Cheeseman I. M (2016) The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol* **17**:16-29 <https://doi.org/10.1038/nrm.2015.5> | PubMed
49. Talbert P. B., Henikoff S (2022) The genetics and epigenetics of satellite centromeres. *Genome Res* **32**:608-615 <https://doi.org/10.1101/gr.275351.121> | PubMed
50. McNulty S. M., Sullivan B. A (2018) Alpha satellite DNA biology: finding function in the recesses of the genome. *Chromosome Res* **26**:115-138 <https://doi.org/10.1007/s10577-018-9582-3> | PubMed
51. Altemose N., et al. (2022) Complete genomic and epigenetic maps of human centromeres. *Science* **376**:eabl4178 <https://doi.org/10.1126/science.abl4178> | PubMed
52. Deng Z., Beliveau B. J. (2022) An open source 16-channel fluidics system for automating sequential fluorescent in situ hybridization (FISH)-based imaging. *HardwareX* **12**:e00343 <https://doi.org/10.1016/j.ohx.2022.e00343> | PubMed
53. Aguilar R., et al. (2024) Tigerfish designs oligonucleotide-based in situ hybridization probes targeting intervals of highly repetitive DNA at the scale of genomes. *Nat. Commun* **15**:1027 <https://doi.org/10.1038/s41467-024-45385-x> | PubMed
54. Anderson S., et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**:457-465 <https://doi.org/10.1038/290457a0> | PubMed

55. Rackham O., Filipovska A. (2022) Organization and expression of the mammalian mitochondrial genome. *Nat. Rev. Genet* **23**:606-623 <https://doi.org/10.1038/s41576-022-00480-x> | PubMed
56. Zhou K., et al. (2020) A novel next-generation sequencing-based approach for concurrent detection of mitochondrial DNA copy number and mutation. *J. Mol. Diagn* **22**:1408-1418 <https://doi.org/10.1016/j.jmoldx.2020.09.005> | PubMed
57. Paek J., et al. (2017) Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell* **169**:338-349.e11 <https://doi.org/10.1016/j.cell.2017.03.028> | PubMed
58. Matilainen O., Quirós P. M., Auwerx J (2017) Mitochondria and Epigenetics - Crosstalk in Homeostasis and Stress. *Trends Cell Biol* **27**:453-463 <https://doi.org/10.1016/j.tcb.2017.02.004> | PubMed
59. Thul P. J., et al. (2017) A subcellular map of the human proteome. *Science* **356** <https://doi.org/10.1126/science.aal3321> | PubMed
60. Bersaglieri C., et al. (2022) Genome-wide maps of nucleolus interactions reveal distinct layers of repressive chromatin domains. *Nat. Commun* **13**:1483 <https://doi.org/10.1038/s41467-022-29146-2> | PubMed
61. Liang C., et al. (2020) Centromere-localized Aurora B kinase is required for the fidelity of chromosome segregation. *J. Cell Biol* **219** <https://doi.org/10.1083/jcb.201907092> | PubMed
62. Broad A. J., DeLuca K. F., DeLuca J. G (2020) Aurora B kinase is recruited to multiple discrete kinetochore and centromere regions in human cells. *J. Cell Biol* **219** <https://doi.org/10.1083/jcb.201905144> | PubMed
63. Rangasamy D., Berven L., Ridgway P., Tremethick D. J (2003) Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. *EMBO J* **22**:1599-1607 <https://doi.org/10.1093/emboj/cdg160> | PubMed
64. Ono T., Fang Y., Spector D. L., Hirano T (2004) Spatial and temporal regulation of Condensins I and II in mitotic chromosome assembly in human cells. *Mol. Biol. Cell* **15**:3296-3308 <https://doi.org/10.1091/mbc.e04-03-0242> | PubMed
65. Kufer T. A., et al. (2002) Human TPX2 is required for targeting Aurora-A kinase to the spindle. *J. Cell Biol* **158**:617-623 <https://doi.org/10.1083/jcb.200204155> | PubMed
66. Khongkow P., et al. (2016) Paclitaxel targets FOXM1 to regulate KIF20A in mitotic catastrophe and breast cancer paclitaxel resistance. *Oncogene* **35**:990-1002 <https://doi.org/10.1038/onc.2015.152> | PubMed
67. Politz J. C. R., Scalzo D., Groudine M (2013) Something silent this way forms: the functional organization of the repressive nuclear compartment. *Annu. Rev. Cell Dev. Biol* **29**:241-270 <https://doi.org/10.1146/annurev-cellbio-101512-122317> | PubMed
68. Rodrigues A., et al. (2023) Nucleoli and the nucleoli-centromere association are dynamic during normal development and in cancer. *Mol. Biol. Cell* **34**:br5 <https://doi.org/10.1091/mbc.e22-06-0237> | PubMed
69. Hoffmeyer K., et al. (2012) Wnt/ β -catenin signaling regulates telomerase in stem cells and cancer cells. *Science* **336**:1549-1554 <https://doi.org/10.1126/science.1218370> | PubMed
70. Tsherniak A., et al. (2017) Defining a cancer dependency map. *Cell* **170**:564-576.e16 <https://doi.org/10.1016/j.cell.2017.06.010> | PubMed
71. Rowley M. J., Corces V. G (2018) Organizational principles of 3D genome architecture. *Nat. Rev. Genet* **19**:789-800 <https://doi.org/10.1038/s41576-018-0060-8> | PubMed
72. Rao S. S. P., et al. (2015) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **162**:687-688 <https://doi.org/10.1016/j.cell.2015.07.024>
73. Chen Y., et al. (2018) Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol* **217**:4025-4048 <https://doi.org/10.1083/jcb.201807108> | PubMed
74. Steens J., Klein D (2022) HOX genes in stem cells: Maintaining cellular identity and regulation of differentiation. *Front. Cell Dev. Biol* **10**:1002909 <https://doi.org/10.3389/fcell.2022.1002909> | PubMed

75. Pinto P. B., Domsch K., Lohmann I (2024) Hox function and specificity - A tissue centric view. *Semin. Cell Dev. Biol* 35-43 <https://doi.org/10.1016/j.semcdb.2022.11.011> | PubMed
76. Hubert K. A., Wellik D. M (2024) Correction: Hox genes in development and beyond. *Development* **151** <https://doi.org/10.1242/dev.202770> | PubMed
77. Shah N., Sukumar S (2010) The Hox genes and their roles in oncogenesis. *Nat. Rev. Cancer* **10**:361-371 <https://doi.org/10.1038/nrc2826> | PubMed
78. Khan I., Amin M. A., Eklund E. A., Gartel A. L (2024) Regulation of HOX gene expression in AML. *Blood Cancer J* **14** <https://doi.org/10.1038/s41408-024-01004-y> | PubMed
79. Zhang J., et al. (2020) An integrative ENCODE resource for cancer genomics. *Nat. Commun* **11**:3696 <https://doi.org/10.1038/s41467-020-14743-w> | PubMed
80. Weber C. M., et al. (2021) mSWI/SNF promotes Polycomb repression both directly and through genome-wide redistribution. *Nat. Struct. Mol. Biol* **28**:501-511 <https://doi.org/10.1038/s41594-021-00604-7> | PubMed
81. Cao R., et al. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**:1039-1043 <https://doi.org/10.1126/science.1076997> | PubMed
82. O'Meara M. M., Simon J. A (2012) Inner workings and regulatory inputs that control Polycomb repressive complex 2. *Chromosoma* **121**:221-234 <https://doi.org/10.1007/s00412-012-0361-1> | PubMed
83. Zeng J., et al. (2022) Targeting EZH2 for cancer therapy: From current progress to novel strategies. *Eur. J. Med. Chem* **238** <https://doi.org/10.1016/j.ejmech.2022.114419> | PubMed
84. Guo Y., et al. (2024) Regulation of EZH2 protein stability: new mechanisms, roles in tumorigenesis, and roads to the clinic. *EBioMedicine* **100**:104972 <https://doi.org/10.1016/j.ebiom.2024.104972> | PubMed
85. Jansz N., et al. (2018) Smchd1 targeting to the inactive X is dependent on the Xist-HnrnpK-PRC1 pathway. *Cell Rep* **25**:1912-1923.e9 <https://doi.org/10.1016/j.celrep.2018.10.044> | PubMed
86. Pintacuda G., et al. (2017) HnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish Polycomb-mediated chromosomal silencing. *Mol. Cell* **68**:955-969.e10 <https://doi.org/10.1016/j.molcel.2017.11.013> | PubMed
87. Xu J., et al. (2015) Developmental control of polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. *Mol. Cell* **57**:304-316 <https://doi.org/10.1016/j.molcel.2014.12.009> | PubMed
88. Bousard A., et al. (2019) The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO Rep* **20**:e48019 <https://doi.org/10.15252/embr.201948019> | PubMed
89. Markaki Y., et al. (2021) Xist nucleates local protein gradients to propagate silencing across the X chromosome. *Cell* **184**:6174-6192.e32 <https://doi.org/10.1016/j.cell.2021.10.022> | PubMed
90. Wang C.-Y., Colognori D., Sunwoo H., Wang D., Lee J. T (2019) PRC1 collaborates with SMCHD1 to fold the X-chromosome and spread Xist RNA between chromosome compartments. *Nat. Commun* **10**:2950 <https://doi.org/10.1038/s41467-019-10755-3> | PubMed
91. Yildirim E., Sadreyev R. I., Pinter S. F., Lee J. T (2011) X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription. *Nat. Struct. Mol. Biol* **19**:56-61 <https://doi.org/10.1038/nsmb.2195> | PubMed
92. Beliveau B. J., et al. (2015) Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun* **6**:7147 <https://doi.org/10.1038/ncomms8147> | PubMed
93. Nir G., et al. (2018) Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet* **14**:e1007872 <https://doi.org/10.1371/journal.pgen.1007872> | PubMed

94. Minajigi A., et al. (2015) A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**:aab2276 <https://doi.org/10.1126/science.aab2276> | PubMed
95. Mossa A., et al. (2025) Sex-specific perturbations of neuronal development caused by mutations in the autism risk gene DDX3X. *Nat. Commun* **16**:4512 <https://doi.org/10.1038/s41467-025-59680-8> | PubMed
96. Hershberg E. A., et al. (2021) PaintSHOP enables the interactive design of transcriptome- and genome-scale oligonucleotide FISH experiments. *Nat. Methods* **18**:937-944 <https://doi.org/10.1038/s41592-021-01187-3> | PubMed
97. Santos-Barriopedro I., van Mierlo G., Vermeulen M (2021) Off-the-shelf proximity biotinylation for interaction proteomics. *Nat. Commun* **12**:5015 <https://doi.org/10.1038/s41467-021-25338-4> | PubMed
98. Iglesias N., et al. (2020) Native chromatin proteomics reveals a role for specific nucleoporins in heterochromatin organization and maintenance. *Mol. Cell* **77**:51-66.e8 <https://doi.org/10.1016/j.molcel.2019.10.018> | PubMed
99. Simonis M., et al. (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet* **38**:1348-1354 <https://doi.org/10.1038/ng1896> | PubMed
100. Mathew B., Bathla S., Williams K. R., Nairn A. C (2022) Deciphering spatial protein-protein interactions in brain using proximity labeling. *Mol. Cell. Proteomics* **21**:100422 <https://doi.org/10.1016/j.mcpro.2022.100422> | PubMed
101. Kishi J. Y., Schaus T. E., Gopalkrishnan N., Xuan F., Yin P (2018) Programmable autonomous synthesis of single-stranded DNA. *Nat. Chem* **10**:155-164 <https://doi.org/10.1038/nchem.2872> | PubMed
102. Rappsilber J., Ishihama Y., Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem* **75**:663-670 <https://doi.org/10.1021/ac026117i> | PubMed
103. Natsume T., Kiyomitsu T., Saga Y., Kanemaki M. T (2016) Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Rep* **15**:210-218 <https://doi.org/10.1016/j.celrep.2016.03.001> | PubMed
104. Kong A. T., Leprevost F. V., Avtonomov D. M., Mellacheruvu D., Nesvizhskii A. I (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**:513-520 <https://doi.org/10.1038/nmeth.4256> | PubMed
105. Käll L., Canterbury J. D., Weston J., Noble W. S., MacCoss M. J (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **4**:923-925 <https://doi.org/10.1038/nmeth1113> | PubMed
106. Eng J. K., Jahan T. A., Hoopmann M. R (2013) Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13**:22-24 <https://doi.org/10.1002/pmic.201200439> | PubMed
107. Choi H., et al. (2011) SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat. Methods* **8**:70-73 <https://doi.org/10.1038/nmeth.1541> | PubMed
108. Teo G., et al. (2014) SAINTexpress: improvements and additional features in Significance Analysis of INteractome software. *J. Proteomics* **100**:37-43 <https://doi.org/10.1016/j.jprot.2013.10.023> | PubMed
109. Choi H., et al. (2012) Analyzing protein-protein interactions from affinity purification-mass spectrometry data with SAINT. *Curr. Protoc. Bioinformatics* **8**:8.15.1-8.15.23 <https://doi.org/10.1002/0471250953.bi0815s39> | PubMed
110. Huttlin E. L., et al. (2015) The BioPlex network: A systematic exploration of the human interactome. *Cell* **162**:425-440 <https://doi.org/10.1016/j.cell.2015.06.043> | PubMed
111. Shannon P., et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**:2498-2504 <https://doi.org/10.1101/gr.1239303> | PubMed

112. Ruepp A., et al. (2008) CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res* **36**:D646-50 <https://doi.org/10.1093/nar/gkm936> | PubMed
113. Yu G., Wang L.-G., Han Y., He Q.-Y (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**:284-287 <https://doi.org/10.1089/omi.2011.0118> | PubMed
114. Korotkevich G., et al. (2016) Fast gene set enrichment analysis. *bioRxiv* <https://doi.org/10.1101/060012>
115. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17** <https://doi.org/10.14806/ej.17.1.200>
116. Langmead B., Salzberg S. L (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**:357-359 <https://doi.org/10.1038/nmeth.1923> | PubMed
117. Broad Institute (2019) Picard. <https://broadinstitute.github.io/picard/>
118. Ramírez F., et al. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**:W160-5 <https://doi.org/10.1093/nar/gkw257> | PubMed
119. Xu W., Zhong Q., Lin D., Li G., Cao G. (2021) CoolBox: A flexible toolkit for visual analysis of genomics data. *bioRxiv* <https://doi.org/10.1101/2021.04.15.439923>
120. Vizcaíno J. A., et al. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol* **32**:223-226 <https://doi.org/10.1038/nbt.2839> | PubMed

Peer reviews

Reviewer #1 (Public review):

The new experiments on the HOX and XIC look strong. A limited (conservative) number of proteins are determined to be enriched at the respective loci. And the number of cells used is a good advancement for these kinds of methods.

Unfortunately, the warnings about mitochondrial to nuclear comparisons and validations do not appear to be taken seriously. It's not that "...there could be non-specific nuclear comparison." There are definitely non-specific enriched proteins. Minimizing false positives is the responsibility of those developing the method and generating the hit lists. I think you saying our probes go to where they are supposed to and label the proteins in that compartment is fine. But that is as far as that should go. Any non-validated protein hits in those comparisons need to be removed. It will contaminate the literature by having all the proteins in 1E, S4D-F, and S5 reported (even though it appears there is no tables reporting the new proteins claimed to be associated with that locus. Why is that?).

I think the line "...we have not made any claims about new proteins at specific loci." is the heart of the issue. What is the point of this method then? Isn't it to identify unknown proteins at a locus of interest? Without that, it's just generating a long list of proteins, where an unknown number of which are likely erroneous, and highlighting the ones you already knew to be there. Along those lines, it is not validation to show proteins that we already knew were at a locus are at the locus. Validation is developing a method to help find new things, then testing those new things to confirm the new method's fidelity.

The comparison of OMAP identified proteins to the several other methods that look at similar regions is not there. A Figure 1F is referred to in the rebuttal but is not in the manuscript. If you mean the Bioplex comparison, that is not the goal. The goal of this analysis to see how much overlap, if any, is being identified across methods. OMAP has so many proteins claimed to be associated with telomeres that are not tested or validated, it would be nice if other methods see similar ones.

Minor points: You have now done label free proteomics. A) Methodological details are needed. It is not clear if you mean MS1 or DIA based quant. B) Do you need all the language about how multiplexed proteomics is enabling this methods?

Labeling the all the enriched proteins in the volcano plots would be nice. I don't want to see just the "relevant" ones that support your claims. I want to see all the "new" ones your discovery method is claiming to discover.

<https://doi.org/10.7554/eLife.102489.2.sa3>

Reviewer #2 (Public review):

Summary

The authors introduce DNA O-MAP, a method that combines oligo-based in situ hybridization with peroxidase-mediated proximity biotinylation to profile proteins and DNA-DNA interactions linked to targeted genomic regions. In the revised manuscript, they expand the method beyond repetitive elements by profiling non-repetitive gene clusters (HOXA and HOXB), studying inhibitor-induced chromatin remodeling, and differentiating homolog-specific proteomes on both the active and inactive X chromosome. These additions considerably broaden the scope of the work and indicate that DNA O-MAP is currently most effective for analyzing gene-cluster size or domain-level chromatin environments, rather than focusing on individual promoters or cis-regulatory elements.

Strengths

The study demonstrates that DNA O-MAP can be applied to both repetitive domains and non-repetitive genomic regions, including gene clusters spanning 80 kilobases and larger single-copy chromosomal intervals, rather than isolated cis-regulatory elements.

Orthogonal validation using ENCODE ChIP-seq data supports several differentially enriched proteins observed between the HOXA and HOXB gene clusters proteomes.

The ability to detect quantitative changes in local protein environments after chemical perturbation demonstrates the method's sensitivity at the level of extended genomic domains.

Homolog-resolved analysis of the active and inactive X chromosome provides an additional demonstration of biological specificity and technical flexibility at the megabase scale.

The revised manuscript appropriately frames DNA O-MAP as a method for interrogating local domain-level genomic environments, rather than exhaustively defining the protein composition of individual regulatory elements.

Weaknesses

As with all proximity labeling approaches, the effective resolution of DNA O-MAP is constrained by the spatial distance of peroxidase-mediated labeling rather than by genomic distance. Consequently, for gene-cluster-scale targets, enrichment extends beyond the targeted interval into surrounding chromosomal regions, potentially limiting the method's specificity at the level of individual promoters, enhancers, or gene bodies.

Specificity is demonstrated through comparative and internally controlled analyses rather than through a quantitative estimate of false discovery rate for locus specificity. Readers should therefore interpret individual protein enrichments as indicative of local chromatin environments rather than definitive evidence of direct binding to a specific regulatory element.

Orthogonal validation is necessarily selective and hypothesis-driven. A broader validation would be required before newly enriched proteins can be interpreted as bona fide region-resident factors.

Comparisons to prior locus-proteomics methods remain indirect and should be interpreted primarily in terms of demonstrated feasibility, scalability, and reduced cell-number requirements rather than absolute performance or resolution.

<https://doi.org/10.7554/eLife.102489.2.sa2>

Reviewer #3 (Public review):

Significance of the Findings:

The study by Liu et al. presents a novel method, DNA-O-MAP, which combines locus-specific hybridisation with proximity biotinylation to isolate specific genomic regions and their associated proteins. The potential significance of this approach lies in its purported ability to target genomic loci with heightened specificity by enabling extensive washing prior to the biotinylation reaction, theoretically improving the signal-to-noise ratio when compared with other methods such as dCas9-based techniques. Should the method prove successful, it could represent a notable advancement in the field of chromatin biology, particularly in establishing the proteomes of individual chromatin regions-an extremely challenging objective that has not yet been comprehensively addressed by existing methodologies.

Strength of the Evidence:

The evidence presented by the authors is somewhat mixed, and the robustness of the findings appears to be preliminary at this stage. While certain data indicate that DNA-O-MAP may function effectively for repetitive DNA regions, a number of the claims made in the manuscript are either unsupported or require further substantiation. There are significant concerns about the resolution of the method, with substantial biotinylation signals extending well beyond the intended target regions (megabases around the target), suggesting a lack of specificity and poor resolution, particularly for smaller loci. Furthermore, comparisons with previous techniques are unfounded since the authors have not provided direct comparisons with the same mass spectrometry (MS) equipment and protocols. Additionally, although the authors assert an advantage in multiplexing, this claim appears overstated, as previous methods could achieve similar outcomes through TMT multiplexing. Therefore, while the method has potential, the evidence requires more rigorous support, comprehensive benchmarking, and further experimental validation to demonstrate the claimed improvements in specificity and practical applicability.

<https://doi.org/10.7554/eLife.102489.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Public Reviews:

Reviewer #1 (Public review):

Summary:

The authors describe a method to probe both the proteins associated with genomic elements in cells, as well as 3D contacts between sites in chromatin. The approach is interesting and promising, and it is great to see a proximity labeling method like this

that can make both proteins and 3D contacts. It utilizes DNA oligomers, which will likely make it a widely adopted method. However, the manuscript over-interprets its successes, which are likely due to the limited appropriate controls, and of any validation experiments. I think the study requires better proteomic controls, and some validation experiments of the "new" proteins and 3D contacts described. In addition, toning down the claims made in the paper would assist those looking to implement one of the various available proximity labeling methods and would make this manuscript more reliable to non-experts.

Strengths:

- (1) The mapping of 3D contacts for 20 kb regions using proximity labeling is beautiful.
- (2) The use of in situ hybridization will probably improve background and specificity.
- (3) The use of fixed cells should prove enabling and is a strong alternative to similar, living cell methods.

Weaknesses:

- (1) A major drawback to the experimental approach of this study is the "multiplexed comparisons". Using the mtDNA as a comparator is not a great comparison - there is no reason to think the telomeres/centrosomes would look like mtDNA as a whole. The mito proteome is much less complex. It is going to provide a large number of false positives. The centromere/telomere comparison is ok, if one is interested in what's different between those two repetitive elements.

We appreciate the reviewers' point here. In fact we selected the mitochondrial DNA as a target for just the reason that the reviewer notes. mtDNA should be spatially distinct from the nuclear targets and allow us to determine if we were in fact seeing spatially distinct proteins at the interorganelle (mtDNA vs. telomeres/centrosomes) and intraorganelle (telomeres vs centromeres) levels.

But the more realistic use case of this method would be "what is at a specific genomic element"? A purely nuclear-localized control would be needed for that. Or a genomic element that has nothing interesting at it (I do not know of one).

We have now added two studies in Figure 4 and Figure 5 detailing the use of OMAP to investigate specific genomic elements. In this case the Hox clusters (*HOXA* and *HOXB*) and haplotype-specific analysis of X-chromosome inactivation centers in female murine (EY.T4) cells. The controls in these cases are more specific, in line with those suggested by the reviewer as we (1) compare *HOXA* and *HOXB* with or without EZH2 inhibition using the same sets of probes and (2) specifically compare the region surrounding the XIC in female cells for the inactive and active X chromosomes.

You can see this in the label-free work: non-specific, nuclear GO terms are enriched likely due to the random plus non-random labeling in the nucleus. What would a Telo vs general nucleus GSEA look like? (GSEA should be used for quantitative data, no GO). That would provide some specificity. Figures 2G and S4A are encouraging, but a) these proteins are largely sequestered in their respective locations, and b) no validation by an orthogonal method like ChIP or Cut and Run/Tag is used.

We performed GSEA on the enrichment scores for the label-free proteomics data from the SAINT output in Figure 1D and that several of these proteins (e.g., those highlighted in Figure 2A: TERF1, CENPN, TOM70) have already been extensively validated to co-localize to these locations.

To the reviewers request for additional validation, we analyzed ChIP-seq data for several proteins to determine if they were enriched surrounding specific loci. In the case of the HoxA/B analysis, we found that HDAC3 and TCF12 were enriched at *HOXB* compared to *HOXA*, and SMARCB1 and ZC3H13 were enriched at *HOXA* compared to *HOXB* (Figure 4C). HDAC3 and TCF12 ChIP data confirmed increased peak calls at *HOXB* and SMARCB1 and ZC3H13 ChIP data confirmed increased peak calls at *HOXA* for these four selected proteins (Figure 4D).

You can also see this in the enormous number of "enriched" proteins in the supplemental volcano plots. The hypothesis-supporting ones are labeled, but do the authors really believe all of those proteins are specific to the loci being looked at? Maybe compared to mitochondria, but it's hard to believe there are not a lot of false positives in those blue clouds. I believe the authors are more seeing mito vs nucleus + Telo than the stated comparison. For example, if you have no labeling in the nucleus in the control (Figures 1C and 2C) you cannot separate background labeling from specific labeling. Same with mito vs. nuc+Telo. It is not the proper control to say what is specifically at the Telo.

We agree with the reviewer that compared to mitochondrial targeting, there could be non-specific nuclear comparisons. We note again though that we purposefully stayed away from using the word “specifically” when describing the proteomics work developed here. The reason being that we are not atlas a large number of targets to define specificity. Instead, we highlight in Figure 2 that we did observe differences in proteins associating with telomeres and mitochondrial DNA. That may be non-specific, and in fact, this is also why we decided to include two nuclear targets to determine what might be specifically enriched. Thus, we compared centromeric and telomeric protein enrichment as determined by OMAP and observed consistent differential enrichment of shelterin proteins at telomeres (Figure 2I) and CENP-A complex members at centromeres (Figure 2J). We could have done the relative comparisons to no-oligo controls, analogous to how CASPEX compared targeted analyses to no-sgRNA controls (PMID: 29735997). However, we found that the mitochondrial targeted samples were generally better as a comparator because (1) we have clear means to validate differences and (2) the local environment around DNA is being labeled.

I would like to see a Telo vs nuclear control and a Centromere vs nuc control. One could then subtract the background from both experiments, then contrast Telo vs Cent for a proper, rigorous comparison. However, I realize that is a lot of work, so rewriting the manuscript to better and more accurately reflect what was accomplished here, and its limitations, would suffice.

Assuming the nuclear control was the same, It is unclear how this ratio-of-ratios ($[(\text{Telo}/\text{Ctrl})/(\text{Cent}/\text{ctrl})]$) experiment would be inherently different from the direct comparison between Telo and Centromere. Again, assuming the backgrounds are derived from the same cellular samples. More than likely adding the extra ratios could increase the artifactual variance in the estimates, reducing the power of the comparisons as has been seen in proteomics data using ratio-of-ratio comparisons in the past (Super-SILAC).

(2) A second major drawback is the lack of validation experiments. References to literature are helpful but do not make up for the lack of validation of a new method claiming new protein-DNA or DNA-DNA interactions. At least a handful of newly described proximal proteins need to be validated by an orthogonal method, like ChIP qPCR, other genomic methods, or gel shifts if they are likely to directly bind DNA. It is ok to have false positives in a challenging assay like this. But it needs to be well and clearly estimated and communicated.

We appreciate the reviewers' point here. To be clear, we have not made any claims about new proteins at specific loci. Instead we validated that known telomeric and centromeric

associating proteins were consistently enriched by DNA OMAP (Figure 2). We also want to emphasize that while valuable, the current paper is not an atlasing paper to define the full and specific proteomes of two genomic loci. We instead show how this method can be used to observe quantitative differences in proteins enriched at certain loci (*HOXA/B* work, Figure 4) and even between haplotypes (Xi/Xa work, Figure 5).

(3) The mapping of 3D contacts for 20 kb regions is beautiful. Some added discussion on this method's benefits over HiC-variants would be welcomed.

We appreciate the reviewers' point here and have added the following text to the discussion: "Additionally, we show that this method is also able to detect DNA-DNA contacts through biotinylation of loop anchors. Our approach functions similarly to 4C[86]. However, our approach of biotin labeling of contacts does not rely on pairwise ligation events. Thus, detection of contacts through DNA O-MAP will vary in the sampling of DNA-DNA contacts in comparison."

(4) The study claims this method circumvents the need for transfectable cells. However, the authors go on to describe how they needed tons of cells, now in solution, to get it to work. The intro should be more in line with what was actually accomplished.

We took the reviewers point and have worked to scale down the DNA OMAP experiments while revising this manuscript. As noted in Figure 5, we have been able to scale this work down to work on plates with ~10x fewer cells than with our initial experiments. This is on top of the initial DNA OMAP work in Figure 1 and 2, as well as our additional work in Figure 4, where we are using 30-60 million cells in solutions which is still 10x less material than previous work (PMID: 29735997). Thus, the newest DNA OMAP platform uses ~100x fewer cells than previous work.

(5) Comments like "Compared to other repetitive elements in the human genome...." appear to circumvent the fact that this method is still (apparently) largely limited to repetitive elements. Other than Glopro, which did analyze non-repetitive promoter elements, most comparable methods looked at telomeres. So, this isn't quite the advancement you are implying. Plus, the overlap with telomeric proteins and other studies should be addressed. However, that will be challenging due to the controls used here, discussed above.

As noted above, we have added Figures 4 and 5 to address the reviewer concerns by targeting multiple non-repetitive loci (*HOXA* and *HOXB* clusters and a 4.5Mb region straddling X-inactivation center on both the active and inactive X homolog). Targeting the regions around the X-inactivation center shows the potential to perform haplotype-resolved proteome analysis of chromatin interactors.

For the telomeric protein overlap, we tried to do this specifically in Figure 1F, we agree with the reviewer that the controls used dramatically change the proteins considered enriched. The goal of the network analysis was to show (1) that we identify proteins previously observed in telomere proteomic datasets and (2) that we gain a more complete view of proteins based on capturing more known interacting proteins than many previous methods as was noted for the RNA OMAP platform (PMID: 39468212). For example, we observed enrichment of PRPF40A in the telomeric DNA OMAP data. From the Bioplex interactome, PRPF40A was observed to interact with TERF2IP and TERF2, suggesting that through these interactions PRPF40A may colocalize at telomeres. Similarly, we observed enrichment of SF3A1, SF3B1, and SF3B2. The SF3 proteins are known regulators of telomere maintenance (PMID: 27818134), but have not previously been observed in telomeric proteomics datasets, except now in DNA OMAP.

We have added the following text to the Results to clarify these points:

“To benchmark DNA O-MAP, we compared the full set of telomeric proteins to proteins observed in five established telomeric datasets (PICH, C-BERST, CAPLOCUS, CAPTURE, BioID)12,14,16,35,36 (Figure 1F). DNA O-MAP captured both previously observed telomeric interacting proteins (shelterins) as well as telomere associated proteins (ribonucleoproteins). We identified multiple heterogeneous nuclear ribonucleoproteins (hnRNPs) previously annotated as telomere-associated, including HNRNPA1 and HNRNPU. HNRNPA1 has been demonstrated to displace replication protein A (RPA) and directly interact with single-stranded telomeric DNA to regulate telomerase activity37–39. HNRNPU belongs to the telomerase-associated proteome40 where it binds the telomeric G-quadruplex to prevent RPA from recognizing chromosome ends41. We mapped DNA O-MAP enriched telomeric proteins to the BioPlex protein interactome and observed that in addition to capturing proteins from previously observed telomeric datasets (Figure 1F), DNA O-MAP enriched for interactors of previously observed telomeric proteins. Previous data found RBM17 and SNRPA1 at telomeres, and in BioPlex these proteins interact with three SF3 proteins (SF3A1, SF3B1, SF3B2). Though they were not identified in previous telomeric proteome datasets, all three of these SF3 proteins were enriched in the DNA O-MAP telomeric data. Furthermore, through interactions with G-quadruplex binding factors, these SF3 proteins are regulators of telomere maintenance (PMID: 27818134). Taken together, this data supports the effectiveness of DNA O-MAP for sensitively and selectively isolating loci-specific proteomes.”

Reviewer #2 (Public review):

Summary

Liu and MacGann et al. introduce the method DNA O-MAP that uses oligo-based ISH probes to recruit horseradish peroxidase for targeted proximity biotinylation at specific DNA loci. The method's specificity was tested by profiling the proteomic composition at repetitive DNA loci such as telomeres and pericentromeric alpha satellite repeats. In addition, the authors provide proof-of-principle for the capture and mapping of contact frequencies between individual DNA loop anchors.

Strengths

Identifying locus-specific proteomes still represents a major technical challenge and remains an outstanding issue (1). Theoretically, this method could benefit from the specificity of ISH probes and be applied to identify proteomes at non-repetitive DNA loci. This method also requires significantly fewer cells than other ISH- or dCas9-based locus-enrichment methods. Another potential advantage to be tested is the lack of cell line engineering that allows its application to primary cell lines or tissue.

We thank the reviewers for their comments and note that we have followed up on the idea of targeting non-repetitive DNA loci (*HOXA* and *HOXB* clusters and a 4.5Mb section of the X chromosome on each homolog) in the revised manuscript (Figures 4 and 5).

Weaknesses

The authors indicate that DNA O-MAP is superior to other methods for identifying locus-specific proteomes. Still, no proof exists that this method could uncover proteomes at non-repetitive DNA loci. Also, there is very little validation of novel factors to confirm the superiority of the technique regarding specificity.

Our primary claim for DNA OMAP is that it requires orders of magnitude fewer cells than previous studies. Based on comments along these lines from both reviewers, we performed DNA OMAP targeting non-repetitive DNA loci (*HOXA* and *HOXB* clusters and a 4.5Mb section of the X chromosome on each homolog) in the revised manuscript (Figure 4 and 5). For the X chromosome targeting, we used ~3 million cells per condition with methods that we

optimized during revision. When targeting *HOXA* and *HOXB*, we were able to identify HDAC3 and TCF12 enrichment at *HOXB* compared to *HOXA* as well as ZC3H13 and SMARB1 enrichment at *HOXA* compared to *HOXB*, which is consistent with ChIP-seq reads from ENCODE for these proteins (Figure 4C, D). Both the *HOXA* and X chromosome work help to address limitations noted in the Gauchier et al. paper the reviewer notes as both show progress towards overcoming “the major signal-to-noise ratio problem will need to be addressed before they can fully describe the specific composition of single-copy loci”.

The authors first tested their method's specificity at repetitive telomeric regions, and like other approaches, expected low-abundant telomere-specific proteins were absent (for example, all subunits of the telomerase holoenzyme complex). Detecting known proteins while identifying noncanonical and unexpected protein factors with high confidence could indicate that DNA O-MAP does not fully capture biologically crucial proteins due to insufficient enrichment of locus-specific factors. The newly identified proteins in Figure 1E might still be relevant, but independent validation is missing entirely. In my opinion, the current data cannot be interpreted as successfully describing local protein composition.

We analyzed ChIP-seq reads for our *HOXA* and *HOXB* (Figure 4C,D) which recapitulate our findings for four of our differentially enriched proteins. We also note that with the addition of the nonrepetitive loci (Figures 4 and 5), we have performed DNA OMAP on seven different targets (telomeres, pericentromeres, mitoDNA, *HOXA*, *HOXB*, Xi, and Xa) and identified expected targets at each of these. The consistency of these data, which mirrors the consistency of the RNA implementation of OMAP (PMID: 39468212), reinforces that we can successfully enrich local proteomes at genomic loci.

Finally, the authors could have discussed the limitations of DNA O-MAP and made a fair comparison to other existing methods (2-5). Unlike targeted proximity biotinylation methods, DNA O-MAP requires paraformaldehyde crosslinking, which has several disadvantages. For instance, transient protein-protein interactions may not be efficiently retained on crosslinked chromatin. Similarly, some proteins may not be crosslinked by formaldehyde and thus will be lost during preparation (6).

Based on this critique we have gone back through the manuscript to improve the fairness of our comparisons and expanded the limitations in our discussion section.

To the point about fixation, Schmiedeberg et al., which the reviewer references, does describe crosslinking requiring longer interactions (~5 s). Yet, as featured in reviews, many additional studies have found that “it has been possible to perform ChIP on transcription factors whose interactions with chromatin are known from imaging studies to be highly transient” (Review PMID: 26354429). We note similar results in proteomics analysis in Subbotin and Chait that state that the linkage of lysine-based fixatives like formaldehyde and “glutaraldehyde to reactive amines within the cellular milieu were sufficient to preserve even labile and transient interactions (PMID: 25172955).

(1) Gauchier M, van Mierlo G, Vermeulen M, Dejardin J. Purification and enrichment of specific chromatin loci. *Nat Methods*. 2020;17(4):380-9.

(2) Dejardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. *Cell*. 2009;136(1):175-86.

(3) Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell*. 2017;170(5):1028-43 e19.

(4) Villaseñor R, Pfaendler R, Ambrosi C, Butz S, Giuliani S, Bryan E, et al. ChromID identifies the protein interactome at chromatin marks. *Nat Biotechnol*. 2020;38(6):728-36.

(5) Santos-Barriopedro I, van Mierlo G, Vermeulen M. Off-the-shelf proximity biotinylation for interaction proteomics. *Nat Commun.* 2021;12(1):5015.

(6) Schmiedeberg L, Skene P, Deaton A, Bird A. A temporal threshold for formaldehyde crosslinking and fixation. *PLoS One.* 2009;4(2):e4636.

Reviewer #3 (Public review):

Significance of the Findings:

The study by Liu et al. presents a novel method, DNA-O-MAP, which combines locus-specific hybridisation with proximity biotinylation to isolate specific genomic regions and their associated proteins. The potential significance of this approach lies in its purported ability to target genomic loci with heightened specificity by enabling extensive washing prior to the biotinylation reaction, theoretically improving the signal-to-noise ratio when compared with other methods such as dCas9-based techniques. Should the method prove successful, it could represent a notable advancement in the field of chromatin biology, particularly in establishing the proteomes of individual chromatin regions - an extremely challenging objective that has not yet been comprehensively addressed by existing methodologies.

Strength of the Evidence:

The evidence presented by the authors is somewhat mixed, and the robustness of the findings appears to be preliminary at this stage. While certain data indicate that DNA-O-MAP may function effectively for repetitive DNA regions, a number of the claims made in the manuscript are either unsupported or require further substantiation. There are significant concerns about the resolution of the method, with substantial biotinylation signals extending well beyond the intended target regions (megabases around the target), suggesting a lack of specificity and poor resolution, particularly for smaller loci.

We thank the reviewers for their comments and note that we have followed up on the idea of targeting non-repetitive DNA loci (HOX clusters and part of the X chromosome) in the revised manuscript (**Figures 4 and 5**).

Furthermore, comparisons with previous techniques are unfounded since the authors have not provided direct comparisons with the same mass spectrometry (MS) equipment and protocols. Additionally, although the authors assert an advantage in multiplexing, this claim appears overstated, as previous methods could achieve similar outcomes through TMT multiplexing. Therefore, while the method has potential, the evidence requires more rigorous support, comprehensive benchmarking, and further experimental validation to demonstrate the claimed improvements in specificity and practical applicability.

We have made the comparisons as best as possible. In fact, we found it difficult to find examples of recent implementations of many of these methods. Purchasing the exact mass spectrometers or performing every version of chromatin proteomics would be well beyond the scope of this work. On the other hand, OMAP has already generated data for three manuscripts. We are making the claim that using the instrumentation and methods available to us, we were able to reduce the number of cells required to analyze a given genomic loci. We then applied TMT multiplexing to further improve the throughput and perform replicate analyses. To fully validate that one protein exists at one loci and no other would require exhaustive atlasing of protein-genomic interactions which would be well beyond the scope of this single paper. Similarly, ChIP for every target identified to assess an empirical FDR would be well beyond the scope of this work.

Recommendations for the authors:

Reviewing Editor Comments:

In summary, all three reviewers raised major concerns about the limitations of the method, many of which could be resolved by more precise and transparent language about these limitations. If you choose to resubmit a revised version, you should address questions like: What scale does "individual locus" refer to? At what scale can the method map protein-DNA interactions at individual targeted loci, rather than large repetitive domains? What is the estimated false discovery rate for a set of enriched proteins? The eLife assessment for this version of the manuscript is based on reviewer concerns. Note that this assessment can be updated after receiving a response to reviewer comments.

Reviewer #1 (Recommendations for the authors):

(1) The first couple of paragraphs make it sound like your method would exclusively benefit from sample multiplexing with MS-based proteomics. That is a bit misleading. The other stated methods use TMT. They don't use it to compare very different genomic (or compartmental) regions, but there is no reason cberst, glopro or CasID could not.

A good point and we have updated the manuscript to reflect this. While previous methods generally did not use TMT, they could be adapted to do so and, similar to OMAP, improved by the use of more replicates in their analyses.

(2) Please make the colors in 1F for the dataset overlap easier to read. 2 and 4+ are too similar.

We appreciate the comment on making the colors easier to discern. Along these lines we've changed the color of "2" to make it easier to distinguish from "4+".

(3) Label as many dots as legible in your volcano plots.

We've labeled a number of proteins that are relevant to the discussion in this paper as well as some additional proteins. We feel that additional labeling would detract from the points that we are trying to make in individual figure panels about groups of proteins, rather than general remodeling of all proteins.

(4) Figure 2E needs a divergent color scheme since it crosses 0. And is it scaled, log-transformed, or both? And compared to what then?

Figure 2E (heatmap) is z-scaled relative protein abundance measurements based on TMTpro reporter ion signal to noise ("s/n"). We have added additional information to the legend to highlight the information that the reviewer points out here. For the color, we are unsure of what is being asked for, as above 0 is red and below 0 is blue.

(5) Unclear what you are implying with "...only 1-2 biological replicates." I would omit or clarify.

Fair point, we have updated the manuscript to omit this section to simplify the introduction.

(6) H2O2 and biotin phenols might be toxic to living organisms. But so is 4% PFA and ISH. I realize you are trying to justify your new approach but you don't need to do it with exaggerated contrasts. This O-MAP is a great approach and probably more likely for people to adopt it because it's DNA ISH based. Plus, with the clicking, you are likely not displacing proteins via Cas9 landing.

We appreciate the reviewer's comments about adoption and lack of protein displacement. We've scaled back on the claims and added more about limitations owing to crosslinking and

ISH.

(7) How much genome does the Cent regions take up? You state 500 kb for Telos.

In the text we delineate how large of a region the PanAlpha probes target “The genome-wide binding profile of the pan-alpha probe closely overlaps with centromeres (Figure S1) and covers approximately 35 Mb of the genome according to *in silico* predictions.” Additionally, we’ve added Table S4 to summarize target locus sizes for all of the included targets.

(8) You seem to be underestimating the lysine labeling. Is that after TMT labeling and analysis? If so, you're already ignoring what couldn't be seen. I don't think it's that important but you included it, so please describe clearly why it's an issue and how much of an issue it is. How does that relate to lit values? And it's not just TMTpro, it's any lysine labeler.

We appreciate the reviewers point about specifying the reasoning and the lack of clarity around overall lysine labeling. That 1.38% is the number of peptides with remainder modifications due to formaldehyde crosslinking. For overall acylation of lysines with TMT labels, we generally expect (and achieve) >97% labeling of lysines with TMT reagents as the Kuster and Carr labs nicely demonstrated across a range of labeling conditions (PMID: 30967486).

Decrosslinking is a critical step generally for proteomics workflows on fixed or FFPE tissues and thus we sought to explore whether we could achieve sufficiently low residual lysine alkylation to enable protein quantitation by TMTpro reagents (or any lysine labeler, as the reviewer notes). For TMTpro-based methods on peptides, this is less of a concern generally as protease cleavage frees new primary amines at the N-termini of peptides which can be labeled for quantitation. But in part since we are describing a proteomics method on fixed tissues we wanted to share these data and the potential inclusion of residual fixation modifications for readers to potentially take into consideration when performing this method.

Reviewer #3 (Recommendations for the authors):

Liu et al. describe an original locus labelling approach that enables the isolation of specific genomic regions and their associated proteins. I have mixed views on this work, which, in my opinion, remains preliminary at this stage. Establishing the proteome of a single chromatin region is one of the most complex challenges in chromatin biology, as extensively discussed in Gauchier et al. (2020). Any breakthrough towards this goal is of significant interest to the community, making this manuscript potentially compelling. Indeed, some data suggest that the method works for repetitive DNA to some extent. However, much of the data is not very convincing, and in the case of small DNA targets, it argues against the use of DNA-O-MAP.

In contrast to existing methods, DNA-O-MAP combines locus-specific hybridisation in situ (using affordable oligonucleotides) with proximity biotinylation. A major advantage of this strategy over other locus-specific biotinylation methods is the possibility of extensively washing excess or non-specifically hybridised probes before the biotinylation reaction, theoretically limiting biotinylation to the target region and thus significantly enhancing the signal-to-noise ratio. Other methods involving proximity biotinylation, such as targeted dCas9, do not have this capacity, meaning biotinylation occurs not only at the locus where a small fraction of dCas9 molecules is targeted but also around non-bound dCas9 molecules (representing the vast majority of dCas9 expressed in a given cell). This aspect potentially represents an interesting advance.

We thank the reviewer for their thoughts and critiques, which we hope have in part relieved concerns pertaining to limitation on repetitive elements. To the latter points, we confirmed

this with new specificity analysis that showed labeling to be highly specific to a given probe locus (Figure S3).

Below, I outline the significant issues:

The manuscript implies that DNA-O-MAP has better sensitivity than earlier techniques like CAPTURE, GLOPRO, or PICH. The authors state that PICH uses one trillion cells (which I doubt is accurate), and other methods require 300 million cells, whereas DNA-O-MAP uses only 60 million cells, suggesting the latter is more feasible. However, these earlier experiments were conducted almost 15 and 6 years ago, when mass spectrometry (MS) sensitivity was considerably lower than that of current instruments. The authors cannot know whether the proteome obtained by previous methods using 60 million cells, but analysed with current MS technology, would yield results inferior to those of DNA-O-MAP. Unless the authors directly compare these methods using the same number of cells and identical MS setups, I find their argument unjustified and misleading.

Based on the instrumentation listed, we actually do have a good idea of how sensitivity changes may have affected identifications and overall sensitivity. For example, the CASPEX data was collected on an Orbitrap Fusion Lumos, while our data was collected on an Orbitrap Fusion Eclipse. From our work characterizing these two instruments during the Eclipse development (PMID: 32250601), we do actually know that the ion optics improvements boosted sensitivity of the Eclipse used in our work compared to the Lumos by ~50%, meaning if GLOPRO was run on an Eclipse it would still require >200 million cells per replicate for input.

It is suggested that DNA-O-MAP is capable of 'multiplexing', whereas previous methods are not. This statement is also misleading. As I understand it, the targeted regions do not originate from a common pool of cells. Instead, TMT multiplexing only occurs after each group of cells has been independently labelled (Telo, Centro, Mito, control). Therefore, previous methods could also perform multiplexing with TMT. Moreover, it is unclear how each proteome was compared: one would expect many more proteins from centromeres than from telomeres (I am unsure about the number of mitochondria in these cells) since these regions are significantly larger than telomeres (possibly 10 to 100 times larger?). Have the authors attempted to normalise their proteomics data to the size (concatenated) of each target? This is particularly relevant when comparing histone enrichment at chromatin regions of differing sizes.

We agree with the reviewers that this was overstated. In fact the GLOPRO paper notes that they performed a MYC analysis with a previous generation of TMT that could multiplex 10 samples. We have amended the manuscript to be more specific in those contexts. As stated in the methods section, "Samples were column normalized for total protein concentration", to account for the amount of protein and size of the different targets.

Figure 1C shows streptavidin dots resembling telomeres. To substantiate this claim, simultaneous immunofluorescence with a telomere-specific protein (e.g., TRF1 or TRF2) is required. It is currently unknown whether all or only a subset of telomeres are targeted by DNA-O-MAP, and it is also unclear if some streptavidin foci are non-telomeric. Quantification is needed to indicate the reproducibility of the labelling (the same comment applies to the centromere probes later in the manuscript; an immunofluorescence assay with CENPB would be informative, alongside quantifications).

We understand the reviewer's concern about specificity and reproducibility of DNA-O-MAP. To address this we have added analysis showing the efficiency and specificity of our FISH and biotin labeling for Telomere, PanAlpha, and Mitochondria targeting oligos (Figure S3). We found that biotin deposition was highly specific to the intended targets with an average across the three probes of 98% specificity.

Perhaps more importantly, the authors suggest that it may be possible to enrich proteins that are not necessarily present at the target locus but are instead in spatial proximity (e.g., RNA polymerase I subunits enriched upon centromere targeting). Does this not undermine the purpose of retrieving locus-specific proteomes?

The goal of DNA OMAP is to identify a local neighborhood of proteins around a specific genomic loci, similar to GLOPRO. As we note in the work presented in Figure 4 and 5 now, these neighborhoods are inherently interesting for comparison of quantitative changes that occur around a genomic locus.

Possibly related to the previous issue, when DNA-O-MAP is used to assess DNA-DNA interactions, probes covering regions of 20-25 kb are employed. Therefore, one would expect these regions to be significantly biotinylated compared to flanking regions. However, Genome Browser screenshots indicate extensive biotinylation signals spanning several megabases around the 20-25 kb targets. If the method were highly resolutive, the target region would be primarily enriched, with possibly discrete lower enrichment at distant interacting regions. The lack of discrete enrichment suggests poor resolution, likely due to the likely large scale of proximity biotinylation. This compromises the effectiveness of DNA-O-MAP, especially if it is intended to target small loci with complex sequences. Could the authors quantify the absolute number of reads from the target region compared to those from elsewhere in the genome (both megabases around the locus and other chromosomes, where many co-enriched regions seem to exist)? This would provide insights into both enrichment and specificity.

Thanks for this suggestion, we have included a new Figure S8 to look at normalized read depth as a function of distance from the genomic target. The resolution of DNA OMAP, like all peroxidase mediated proximity labeling methods, is not dependent on the sequence length of the DNA region, but the 30-40nm of physical space around the HRP molecule that is targeted to the genomic loci.

Minor Issues:

(1) Page 3, second paragraph: It is unclear why probes producing a visible signal in situ necessarily translates to their ability to retrieve a specific proteome.

We have revised the manuscript to de-emphasize the visible signal aspect of probe targeting and re-emphasize our initial point that the number of probes needed to properly target unique regions makes the use of locked nucleic acid probes cost-prohibitive. The basic point though, we and others previously showed with RNA OMAP (PMID: 39468212) and Apex/proximity labeling strategies, the ability to deposit biotin and visualize generally directly translates to recovery of proximally labeled proteins (PMID: 26866790).

(2) Page 3, last paragraph: "to reach a higher degree of enrichment...": Has it been demonstrated that direct protein biotinylation provides higher enrichment of relevant proteins? Certainly, there is higher enrichment of proteins, but whether they are relevant is another matter.

Our point here was that the methods using direct protein biotinylation have higher levels of enrichment and thus require less cells than the previously mentioned PICH method, which is why we wrote the following: "In the case of GLoPro, APEX-based proximity labeling enhanced protein detection sensitivity, reducing the input required for each replicate analysis to ~300 million cells—a 10-fold reduction in cell input compared to PICH which used 3 billion cells."

Regarding if these proteins are relevant or not, we show enrichment of known proteins that are critical to the function of their occupied genomic region at telomeres and centromeres. Additionally, we've made added quantitative comparisons to assess relevance in our analysis

of Hox and our targeted region of the X chromosome through comparisons to ChIP data at these regions. The improved enrichment that we've established in our initial submission as well as in the updated version also means that we can further scale down the number of cells required.

(3) Figure 2B is misleading; it appears as though all three regions are targeted in the same cell, suggesting true multiplexing, which, I believe, is not the case.

To avoid any potential confusion about how the samples were derived we've updated this figure panel to show three separate cells, each with a different region being targeted.

(3) If I understand correctly, the 'no probe' control should primarily retrieve endogenously biotinylated proteins (carboxylases), which are mainly found in mitochondria. Why does the Pearson clustering in Supplementary Figure 2 not place this control proteome closer to the mitochondrial proteome?

Under the assumption that the ~10 carboxylases are biotinylated at the same levels in all cells, yet the proportion of these carboxylases compared to all enriched proteins for a given target is markedly reduced. Thus, as a proportion of the enriched proteome we note in Figure S4 that mitochondrial DNA OMAP enriches proteins besides the carboxylases. We believe this explains why the 'no probe' sample can be clearly separated along PC2 in Figure 2D.

(4) Was CENPA enriched in the centromere DNA-O-MAP? If not, have the authors scaled up (e.g., with ten times more cells) to see if the local proteome becomes deeper and detects relevant low-abundance proteins like CENPA or HJURP? This would be very informative.

We did not observe CENPA, and we had originally contemplated the experiment the reviewer suggested, but noted that CENPA has only two tryptic peptides (>7 AA, <35AA), and they are both in the commonly phosphorylated region of the protein. Rather than scale up these experiments, we decided to attempt DNA OMAP on the non-repetitive locus experiments.

(5) Using a few million cells, I do not see how the starting chromatin amount could range from 0.5 to 7 mg, as shown in Figures 2 and 3. How were these figures calculated? One diploid cell contains approximately 6 pg of DNA/chromatin, which means one billion cells represent about 6 mg of DNA/chromatin (a typical measurement for these methods).

Thanks to the reviewer for catching this, that should have been the total lysate amount, not chromatin mass. We have corrected Figures 2 and 3.

(6) Figure S1: There is no indication of the metrics used for the shades of red.

We have added a gradient legend to depict this.

(7) What is the purpose of HCl in the experiment?

HCl treatment was done to reduce autofluorescence for imaging (PMID: 39548245).

(8) I could not find the MS dataset on the server using the provided accession number (PDX054080).

Thank you for pointing this out, we have confirmed the dataset is public now and added the new datasets for the Xi/Xa and Hox studies. We also note that the accession should be "PXD054080"

(9) Why desthiobiotin instead of biotin?

We have tested both; desthiobiotin was helpful to reduce adsorption to surfaces. Either biotin or desthiobiotin can be used, though, for OMAP.

<https://doi.org/10.7554/eLife.102489.2.sa0>