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

Yuzhen Liu[1-3,6], Christopher D. McGann[1,6], Conor P. Herlihy[1,2,6], Mary Krebs[1,2], Thomas A. Perkins, Jr.[1], Rose Fields[1], Conor K. Camplisson[1,2], David Z. Nwizugbo[1,2], Chris Hsu[1,2], Shayan C. Avanessian[1-3], 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*]

[1]Department of Genome Sciences, University of Washington, Seattle, WA, USA

[2]Brotman Bay Institute for Precision Medicine, Seattle, WA, USA

[3]Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA

[4]Department of Pharmacology, University of Washington, Seattle, WA, USA

[5]Institute of Stem Cell and Regenerative Medicine, University of Washington, USA

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

*Correspondence e-mails: beliveau@uw.edu; dkschwep@uw.edu

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 sample multiplexed quantitative proteomics, targeted chemical perturbations, and next-generation sequencing to quantify DNA-protein 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 (kilobases to megabases), 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 functions[1]. Therefore, developing methods to decipher DNA-protein interactions in cells has been a focus of technology development efforts for decades[2]. For instance, chromatin immunoprecipitation followed by sequencing (ChIP-seq[3]), which has emerged as a core technology for epigenomics[4], surveys the genome-wide binding profile of a target DNA-associated protein. ChIP-seq and related technologies (e.g., DamID[5], CUT&Tag[6]) have produced an abundance of high-quality datasets that enabled the establishment of database consortia such as ENCODE[7,8] and IHEC[9], 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 residues[17], 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 produce visible signal[18]. As noted in follow-up work, PICh was effective for repeat sequences but would require significant additional work to extend to more complex genomic sequences[19]. 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 washes[19], limiting detection sensitivity. As a consequence, an input of one trillion cells was required for one purification and successful identification of proteins interacting with telomeres[16].

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-BERST[12] and GLoPro[13], 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 (MS). 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. Nevertheless, a notable limitation of CRISPR-guided proximity labeling is requiring the introduction of the fusion dCas9-APEX enzyme and sgRNAs into 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 cells[24]. 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 proximity labeling to cell lines amenable to genetic engineering. Owing to the large numbers of cells required and the need to maximize sensitivity, previous methods often compared only 1–2 biological replicates[12,13]. In some cases, this was limited by the use of stable-isotope-based quantification methods that can only multiplex up to three samples per analysis[12]. Thus, an unmet need exists for extensible methods capable of scaling and profiling multiple genomic loci. Moreover, these methods would ideally be capable of scaling and multiplexing comparisons between multiple local proteomes or one local proteome in response to multiple stimuli or perturbations.

We address these pressing 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-MAP[25] and pSABER[26] 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 the recovered material is compatible with sample multiplexed proteomics[27] and drug perturbation. We benchmark the specificity of our approach by recovering telomere-specific DNA binding proteins after targeting telomeric DNA. We further showcase the scalability and sample multiplexing capacity 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 tags[27]. We also establish that DNA O-MAP can be used to capture functionally relevant DNA-DNA interactions, read out by DNA sequencing, from intervals as small as 20 kilobases. Additionally, we show that DNA OMAP can obtain the local proteome from non-repetitive elements such as the HOXA and HOXB gene clusters as well as differential proteomes before and after drug directed enzyme inhibition. 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-specific chromatin interactions.

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 specific genomic sites then enables the pull-down and purification of chromatin associated proteins and DNA from *trans*-interacting genomic loci. As in RNA O-MAP, the specificity of ISH and/or biotinylation can be assessed by microscopy using a small sample of cells immobilized on solid support before the cell pellets enter affinity purification downstream. 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.

DNA O-MAP deploys a scalable in-solution hybridization-biotinylation workflow.

During the development of DNA O-MAP, we refined a suspension-based hybridization workflow for cost-efficient genomic labeling (**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 in parallel by one experimentalist. Critically, this approach reduces reagent costs by ~1,000-fold relative to conventional ISH protocols performed on solid substrates, making the labeling of millions or more cells with oligo-based ISH probe sets, including those targeting non-repetitive DNA, cost-feasible.

DNA O-MAP reveals the organization of the telomeric proteome.

To demonstrate that O-MAP can successfully purify proteins from small genomic viewpoints, we selected human telomeres for initial testing (**Figure 1B**). Mammalian telomeres are several kilobases of tandemly repeated arrays of 5'-TTAGGG-3' hexamers with terminal 3' single-stranded overhangs at the ends of chromosomes[28]. Telomeric DNA is specifically bound by a proteinaceous cap that protects the natural chromosome ends from being recognized as damaged DNA—the shelterin complex[29,30]. 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 exclusively at the ends of the chromosomes. Accordingly, this well-defined set of proteins has been widely accepted as goalposts for a successful locus-specific 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% (~500kb) of the human genome[31]. 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 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 1C**). 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 1D-E**). 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 1F**). We then overlaid each called interactor on direct protein interaction data and found that DNA O-MAP enabled greater coverage of known protein interactors, even those not previously identified as enriched at telomeres by other methods. In addition to shelterins, 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]. In addition, HNRNPU belongs to the telomerase-associated proteome[40] where it binds the telomeric G-quadruplex to prevent RPA from recognizing chromosome ends[41]. Taken together, this data supports the effectiveness of DNA O-MAP for sensitively and selectively isolating loci-specific proteomes.

Figure 1

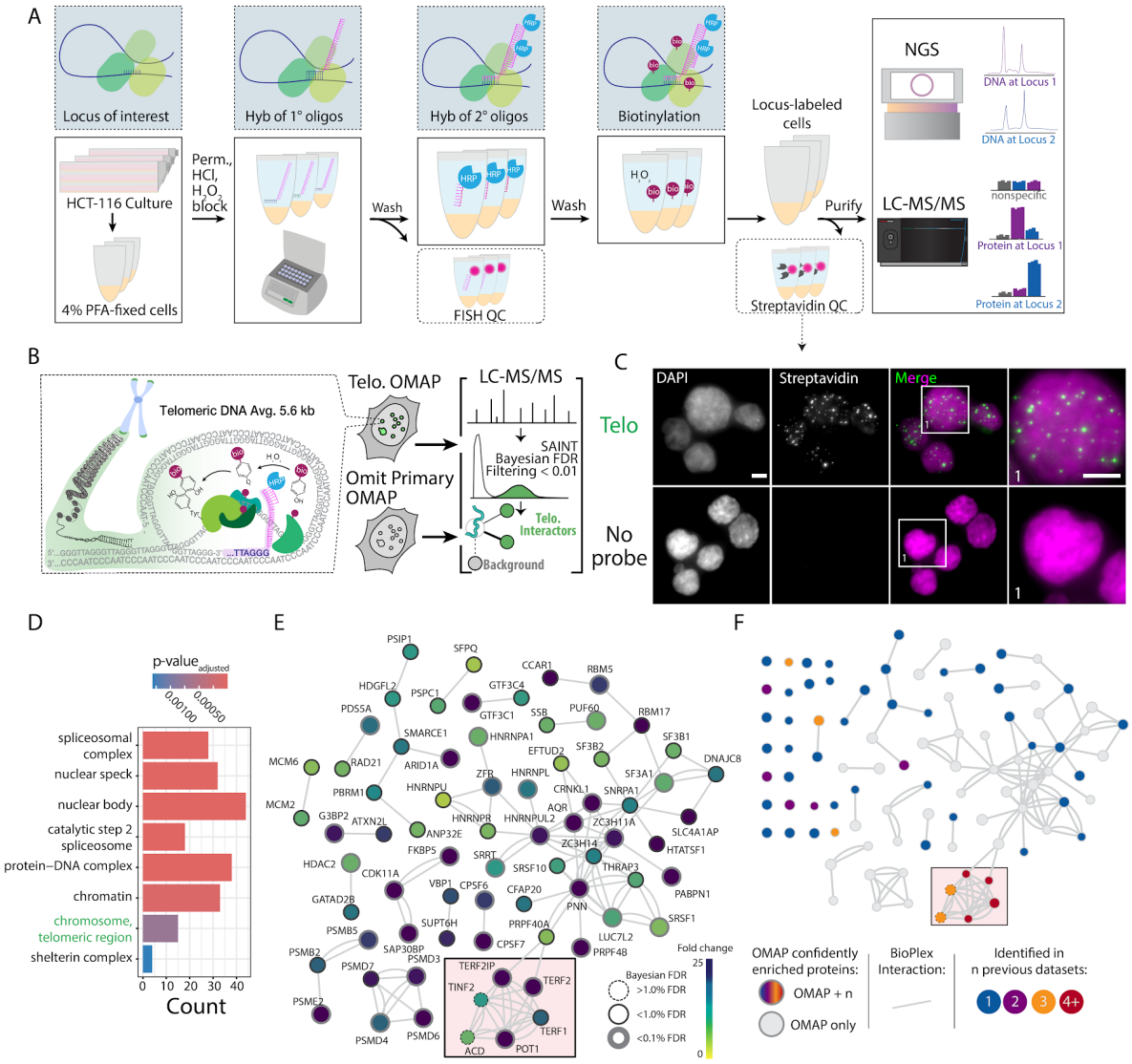


Figure 1. Overview of DNA O-MAP workflow and label-free quantitative proteomics analysis of telomeres. A) Schematic of DNA O-MAP. B) Overview of telomere targeted DNA O-MAP experiment. C) Fluorescent microscopy data showing the observed patterns of DNA (DAPI, left) and *in situ* biotinylation detected by staining with fluorescent streptavidin conjugates (middle, left). D) Significant gene sets identified by the Gene Set Enrichment Analysis of the proteins enriched by the telomere probe. E) 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 C, excluding unconnected nodes. F) 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 µm.

DNA O-MAP enables multiplexed detection of locus proteomes.

We next evaluated the utility of DNA O-MAP to quantitatively delineate locus-specific proteomes. We integrated sample multiplexing quantitative[27,42,43] proteomics downstream of DNA O-MAP to enable spectral quantification of all samples simultaneously (**Figure 2A**). In our experimental design, we selected three well-characterized DNA loci with distinct protein occupants in the human genome: 1) telomeres, 2) peri-centromeric alpha satellite repeats; 3) the mitochondrial genome (**Figure 2B**). Centromeres are epigenetically defined chromosomal loci where kinetochore proteins assemble for spindle microtubule attachment to ensure equal chromosome segregation during cell division[44,45]. Human centromeres are located within the AT-rich alpha satellite repeats, which are higher-order repeats composed of 171-base-pair monomeric units[46,47]. Due to the sequence independence of centromeres, we utilized a previously described probe[26,48] that targets a subset of alpha satellite repeats to represent centromeres, hereafter denoted as the 'Pan Alpha Sat.' probe. The predicted genome-wide binding profile[49] of the pan-alpha probe closely overlaps with centromeres and covers an estimated 35 Mb (**Figure S1**). 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[50,51]. In HCT-116 cells, mtDNA copy number has been reported to range from 310–677[52]. To demonstrate the locus-specificity of biotinylation using the new oligo/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. After *in situ* HRP-mediated labeling, we performed thermal reversal of fixation of cells prior to lysis, enrichment of biotinylated proteins[53], tryptic digestion, and labeling with isobaric TMTpro barcodes[27]. We note that artificial lysine alkylation due to cellular fixation with PFA may affect TMTpro labeling of protein, thus we tracked artificial lysine modifications during mass spectrometric analysis to ensure minimal effects of alkylation on protein quantification (1.38% of lysines were alkylated).

Figure 2

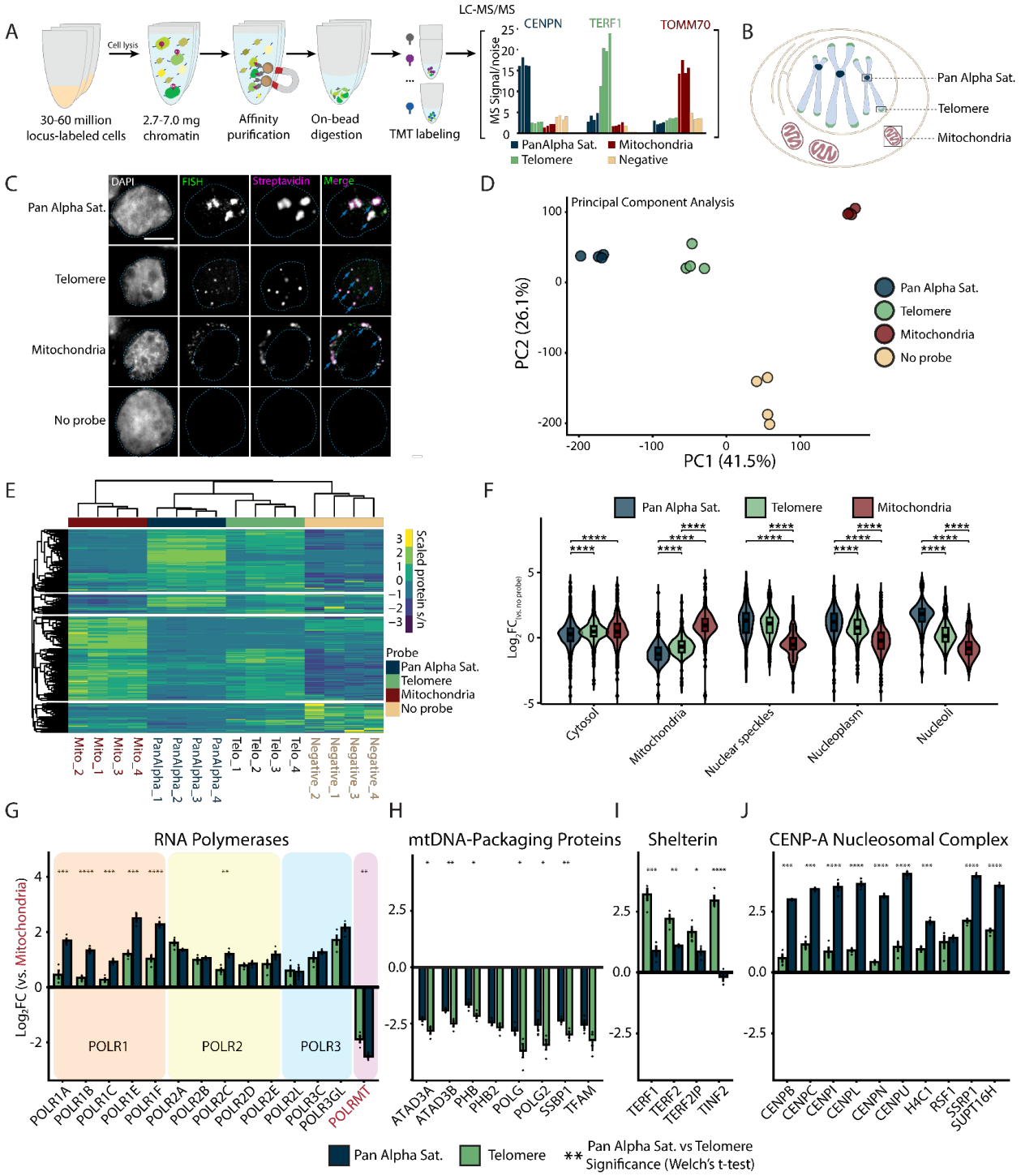


Figure 2. DNA O-MAP reveals distinct features of the sub-proteomes at peri-centromeric alpha satellites, telomeres, and the mitochondrial genome. A) Workflow of DNA O-MAP integrated with sample multiplexing quantitative proteomics B) Schematic of the three DNA loci examined in the TMT16plex experiment: 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 µm. 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

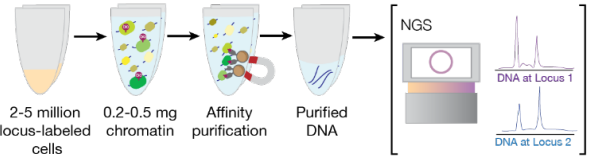
clustering of scaled intensities of proteins enriched by the pan-alpha probe, telomere probe, mitochondrial genome oligo pool, and no-primary-probe control. F) Log_2 fold change of proteins compared to no-primary-probe control, grouped by HPA subcellular location. Significance calculated based on Welch's t-test for pairwise comparisons (****: p-value <0.0001). G–J) Log_2 fold change of proteins compared to mitochondrial probe enriched proteins for the RNA Polymerases (G), mtDNA nucleoid packaging proteins[54] (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).

In total we quantified 3,055 proteins across all four conditions (**Figure 2D–E**). We observed consistent proteome enrichment by principal component analysis and correlation analyses, with tight clustering of replicates (**Figure 2D–E, S2**). Based on Human Protein Atlas annotations[55], 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, S3**). Notably, the pan-alpha probe enriched proteins from the nucleoli, consistent with the known nucleoli-centromere associations[56]; chromosomal passenger complex member AURKB, consistent with the centromeric localization of AURKB in early mitosis to ensure faithful chromosome segregation[57,58] and the localization of chromosomal passenger complex members to pericentromeric heterochromatin[59,60]. We also observed pericentromeric enrichment of spindle and chromosomal segregation associated proteins TPX2[61] and KIF20A[62] (**Figure S3, S4**).

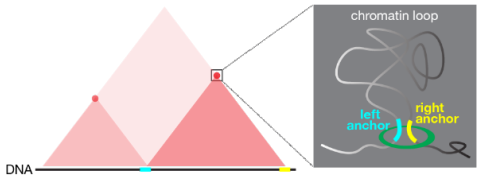
Next, we explored the enrichment of several multi-unit protein complexes across the examined loci. To dissect the differences between enriched proteomes for each probe, we chose a subset of proteins of interest and measured the fold change of the two nuclear targets compared to mitochondria. RNA Polymerase 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[63,64], 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 ($log_{2\ \text{Pan-Alpha Sat./Mito.}}$= -2.51; $log_{2\ \text{Telomere/Mito.}}$= -1.88). Similarly, we observed enrichment of mtDNA-packaging nucleoid components[54] 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 S4**). We also observed enrichment of catenins CTNNB1 and CTNND1 at telomeres (**Figure S3**). The transcription factor CTNNB1 has been observed at the transcriptional start site of h*TERT* where it regulates h*TERT* expression[65]. The h*TERT* gene is located in the subtelomeric region of chromosome 5 (chr5:1,253,167-1,295,068) and expressed in HCT-116 cells[66]. Collectively, these results demonstrate the sensitivity and subcompartment specificity of DNA O-MAP and highlight how coupling quantitative proteomics with DNA O-MAP can distinguish differential compartment components even for ubiquitous chromatin constituents like histones.

Figure 3

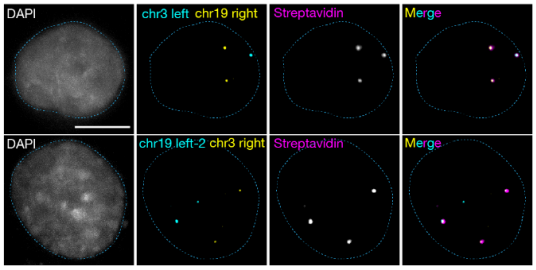
A



B

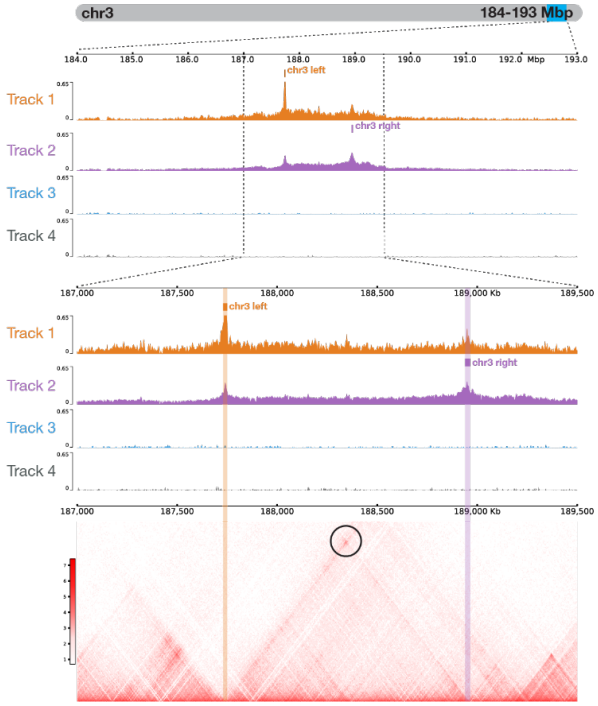


C



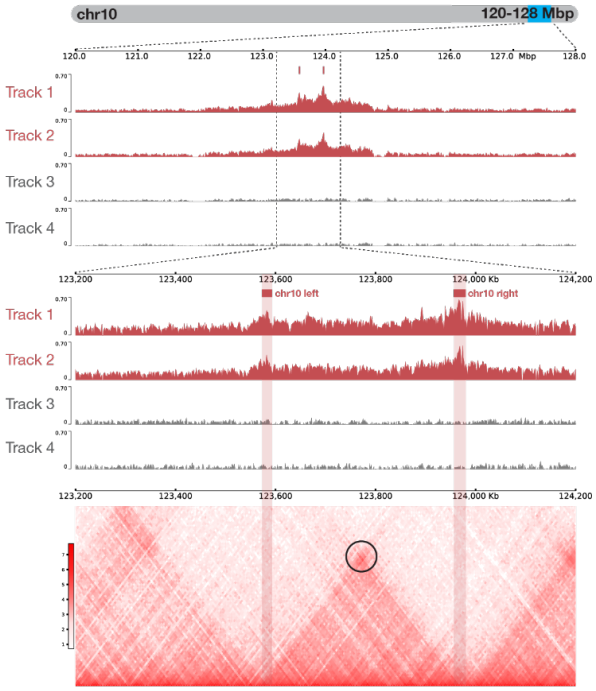
D

	Probed anchors	Expected contact anchors
Track 1	chr3 left	chr3 right
Track 2	chr3 right	chr3 left
Track 3	chr10 non-loop anchor	None
Track 4	No primary probe	None



E

	Probed anchors	Expected contact anchors
Track 1 Track 2	chr3 left, chr10 right, chr19 right	chr3 right, chr10 left, chr19 left-1, chr19 left-2
Track 3 Track 4	No primary probe	None



F

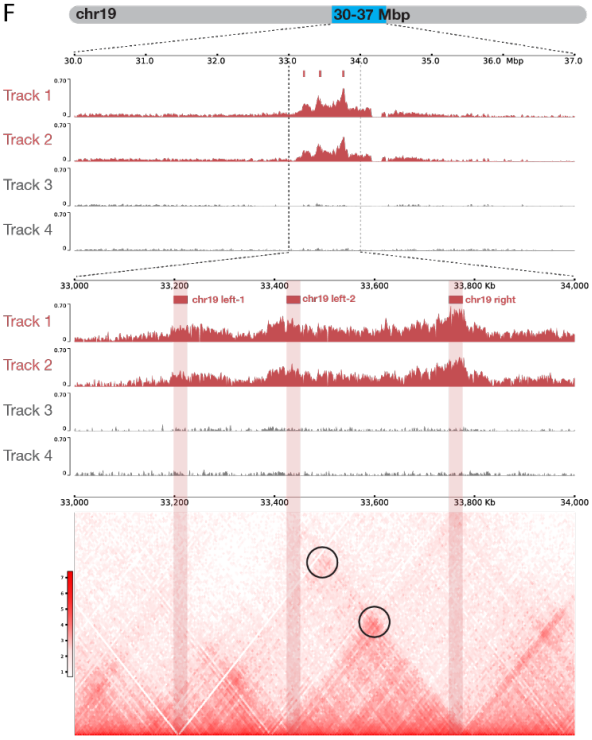


Figure 3. DNA O-MAP efficiently labels single-copy chromatin loop anchors. A) Workflow of DNA O-MAP integrated with biotin purification sequencing B) Schematic of a pair of chromatin loop anchors on a hypothetical Hi-C map and 3-dimensional space C) DNA FISH and the streptavidin staining of the proteins 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).

DNA O-MAP can uncover 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 small, 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 to often stall at CTCF-bound sites while dynamically moving along the genome, creating contact domains of preferential DNA-DNA interaction[67]. In HCT-116 cells, these contacts between chromatin loop anchors have been captured genome-wide with *in situ* Hi-C[68]. 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 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, S5A**). 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 except itself (**Figure S5B**). Lastly, in the cells that received no primary oligos, no pronounced enrichment was observed genome wide (**Figure S5B**).

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, S6A**), whereas no pronounced enrichment was observed in the no-primary-probe controls genome-wide (**Figure S6B**). 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**). 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 DNA-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 small, single-copy DNA intervals as we did earlier with repetitive intervals targeting the telomeres, alpha satellite repeats, and the mitochondrial genome. To investigate this, we utilized the Hox A and Hox B 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. Here we designed probes to target 83kb and 81kb, respectively, for the two loci. We performed DNA O-MAP targeting either HOXA, HOXB, or no primary hybridization probes from about 50 million human K562 cells in three replicates each (**Figure 4A**). Identification of biotinylated proteins was carried out using label-free quantitative proteomics. . We also performed DNA O-MAP in human K562 cells with co-hybridization of both fluorescent oligos and HRP oligos to demonstrate locus-specific biotinylation and FISH signal within the same cell. Biotinylation patterns for both HOXA and HOXB showed strong concordance with FISH signal (**Figure 4B**). From our label-free proteomics analysis, we identified 42 proteins that were significantly differentially enriched between the HOXA and HOXB loci (**Figure 4C**). Of the differentially enriched proteins, we see increased abundance of HDAC3 from our Hox B labeled sample compared to HOXA. This agrees with ENCODE ChIP-seq data in K562 cells, which shows more peaks being called at the HOXB loci than within HOXA (**Figure 4D**). Other enriched proteins, such as SWI/SNF factor SMARCB1 at Hox A, have been described to interact with the Hox genes[69], 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 another experiment 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 (PRC). The main catalytic component of the PRC is EZH2, which is responsible for trimethylating histone 3 at lysine 27 (H3K27me3). Deposition of H3K27me3 has been well characterized as a mark of transcriptional repression associated with PRC activity. Additionally, EZH2 inhibition has become a prominent target in cancer therapies[70,71]. Here we took advantage of sample multiplexing quantitative[27,42,43] proteomics downstream of DNA O-MAP to enable spectral quantification of all samples simultaneously (**Figure 4E**). To quantify the local proteomes in response to drug

inhibition, we prepared replicate (n=3) samples for each probe following treatment with either the EZH2 inhibitor GSK126 or DMSO control. After *in situ* HRP-mediated labeling, we performed thermal reversal of fixation of cells prior to lysis, enrichment of biotinylated proteins[53], tryptic digestion, and labeling with isobaric TMTpro barcodes[27]. In total, we identified 11 proteins at *HOXA* and 8 proteins from HOXB with significantly altered enrichment following treatment with GSK126. (**Figure 4F, G**). Of the differentially enriched factors identified, we see a high proportion of RNA-interacting proteins, such as ELAV1, RMB14, ILF3, HNRNPK, HNRNPA1, PRPF40A, and PRPF8. Along these lines, recent evidence has suggested that proteins that bind long non-coding RNAs are critical for the establishment of PRC repressed loci[72,73]. Several proteins we identified here also have established connections to PRC repression of the *HOX* genes, such as GATA1 which has been suggested to mediate the switch to non-canonical PRC function[74]. Taken together, these data demonstrate that DNA O-MAP can elucidate loci-specific proteomes from non-repetitive genomic loci. Furthermore, we establish that DNA O-MAP has sufficient sensitivity to detect inhibitor induced changes to the local interactome at unique loci as small as 80kb when coupled with quantitative proteomics.

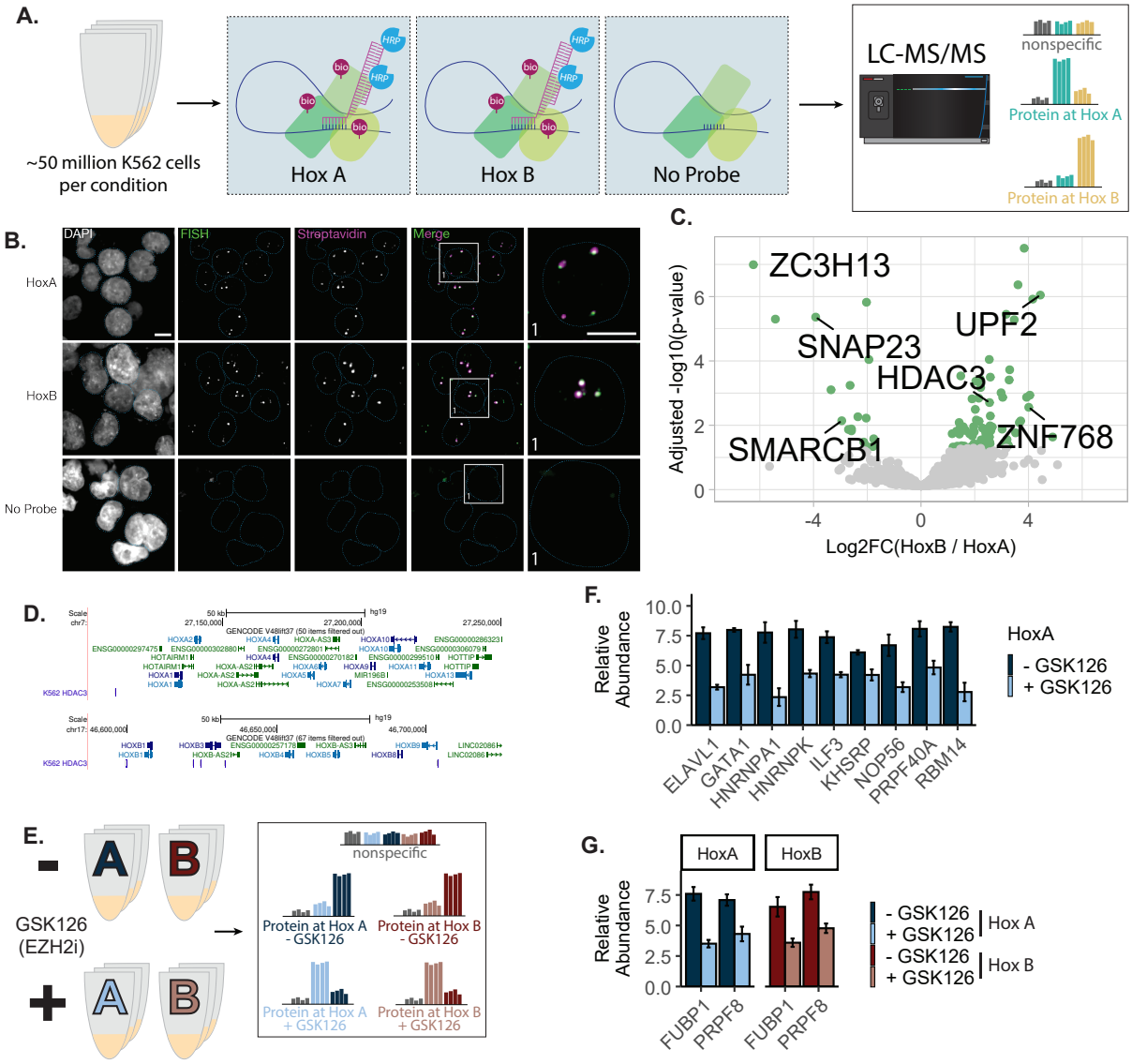


Figure 4. DNA O-MAP efficiently identifies the local proteome of the Hox A & B gene clusters. A) Schematic of DNA O-MAP being applied to the Hox A and Hox B gene clusters for identification of differentially enriched proteins. B) Representative images depicting overlap of FISH and Streptavidin labeling at Hox A and Hox B loci. C) Volcano plot of proteins identified at Hox A and Hox B loci. Each dot represents a single protein with proteins of interest called out in black. Green dots indicate proteins that passed significant enrichment thresholds of greater than 2-fold change and corrected p-value < 0.05. D) ENCODE ChIP-seq data in K562 cells at Hox A and Hox B gene cluster loci for HDAC3. 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 Hox A. G. Bar chart showing proteins with significantly altered abundance at HoxA and HoxB following treatment with GSK126. Corrected p-value < 0.05.

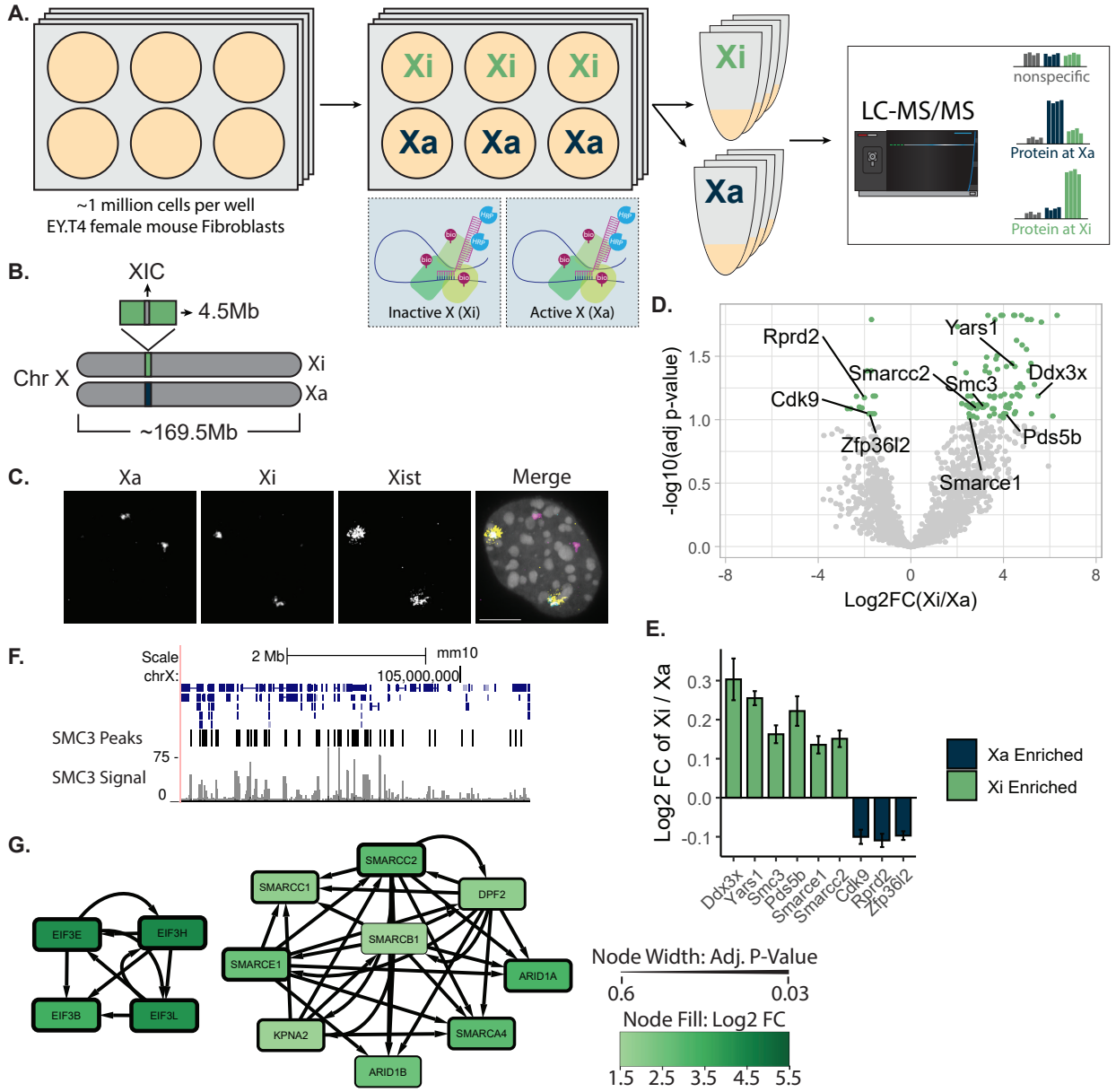


Figure 5. DNA O-MAP elucidates the homolog-resolved chromosome X proteome. A) Schematic of DNA O-MAP being applied to Xi and Xa for identification of differentially enriched proteins. B) Schematic showing the region of the X chromosome targeted by our primary hybridization probes. C) Representative images depicting overlap of Xist FISH and Xi Streptavidin labeling while spatially differentiated from Xa. Scale bar is 10 uM. D) Volcano plot of proteins identified at Xa and Xi. Each dot represents a single protein with proteins of interest called out in black. Green dots indicate proteins that passed significant enrichment thresholds of greater than 2-fold change and corrected p-value < 0.05. E) Bar chart showing example proteins with significant enrichment at Xi. Corrected p-value < 0.1. F) ENCODE ChIP-seq data in mouse fibroblast cells at chromosome X for SMC3. G) Protein interaction network of selected proteins enriched at Xi. Node width is a function of corrected p-value, and node color is a function of enrichment (Log2 Fold Change).

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. This silencing is carried out through cooperative activity of the non-coding RNA Xist and the PRC, which leads to subsequent H3K27me3, along with other factors[72,73,75–77]. As a result, Xi is inundated with a number of chromatin silencing factors. Thus, these two homologs present a vastly different chromatin environment. In order to target the two homologs separately, we took advantage of 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. In doing so we were able to design two primary hybridization probe sets to target the same 4.5Mb region on either Xi or Xa. The approach of leveraging SNPs for homolog specific probe binding has been previously validated for use with other ISH-based methods[78].

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 roughly one million cells (**Figure 5A**). 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. We also performed DNA O-MAP with co-hybridization of both fluorescent oligos and HRP oligos to demonstrate locus-specific biotinylation and FISH signal within the same cell. Biotinylation patterns for both Xi and Xa showed strong concordance with FISH signal (**Figure 5C**). From our label-free proteomics analysis, we identified 96 proteins that were significantly differentially enriched at either Xi or Xa (**Figure 5D**). Of those significantly enriched proteins we identified Xi linked factors, such as SWI/SNF proteins Samrce1, Smarcc2 and members of the cohesin complex Smc3 and Pds5b, where both complexes have been previously shown to interact with the inactive X chromosome[79]. Interestingly, we also see enrichment of the X-linked gene Ddx3x which has been previously implicated as a mediator of sex-based differences in neurodevelopment and disease[80]. Additionally, we see enrichment of proteins associated with a more active transcriptional pattern at Xa, such as RPRD2, ZFP36l2, and CDK9 (**Figures 5D & E**). Furthermore, protein-protein interaction network mapping of Xi enriched proteins shows enrichment of the EIF3 family and SWI/SNF factors (**Figure 5G**). Together, these data demonstrate that our DNA O-MAP approach is able to obtain homolog-specific differentially enriched local proteomes from a 4.5Mb, non-repetitive, region on the X chromosome. Furthermore, these results establish that our method has the sensitivity to obtain, to our knowledge, the first ever loci-specific homolog resolved proteome from as few as 3 million cells per replicate.

Discussion

By combining the versatility of hybridization-based genome targeting with robustness of proximity biotinylation, DNA O-MAP offers a scalable approach to study DNA-associated proteomes through a locus specific lens. 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 peri-centromeres. DNA O-MAP's ability to target single-copy loci, as evidenced by the chromatin loop anchor, HOXA and HOXB gene cluster, and X-chromosome targeting experiments, demonstrate the ability to study protein-mediated DNA interactions at a finer resolution than previously possible. This is further exemplified by the application of DNA O-MAP to the HOXA and HOXB loci following EXH2 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 RNAs[25] and DNA loci. Using oligos to target the DNA locus, DNA O-MAP can be theoretically adapted for use in any sample types amenable to *in situ* hybridization, including cultured cells, tissue sections, and primary tissue samples[26,49,81]. 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 cells[36], spindle-associated proteome dynamics at the pericentromere[82], and molecular drivers of hetero- or euchromatin formation[83] at nearly any locus in the human genome (O-MAP probes can feasibly cover >99% of the human genome)[49,81].

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 likely need 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 locus specific 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% CO_2. 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 Hox A (chr7:27092311 - 27175959), Hox B (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 PaintSHOP[81] 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 Hox A, Hox B, Xi, and Xa targeting primary oligo libraries was performed similarly to as previously described[79]. 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× 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× 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 2uL (4 units) of DNase I and 50uL of Ultra Pure Water (UPW) (10977-015) to each reaction. The reverse transcription reaction contained 45 µl of synthesized RNA, 15 µl of 5× 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 20ul 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 ul of 0.5M EDTA and 37.5 ul 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 described[84] in 100 ul-volume containing 10 mM MgSO4, 300 uM dATP/dCTP/dTTP mix, 100 nM Clean.G hairpin, 80 U/ml Bst DNA Polymerase, Large Fragment (NEB M0275L), 1 uM hairpin, and 1 uM 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 uM 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 of cell pellets

EY.T4 oligo hybridizations were performed on cells fixed to glass, with all washes being performed in 1mL. Cells were rinsed once with fresh phosphate buffer 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 500uL 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 followed by two 2X SSC-T buffer exchanges to remove residual Hyb1. Cell pellets were then washed in 2X SSC-T at 60°C, followed by two two-minute washes in 2X SSC-T at room temperature. 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 locus 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 um 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 um, 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 µg/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.

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 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. Assuming 20 µg of eluate protein, 200 ng LysC (Wako) was added to the beads in a 50-ul 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. 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)[85] method and dried in a vacuum concentrator.

For label free telomere-enriched samples, one sample consisted of HCT-116-Rad21-mAID cells[86] . 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. MS1 scans were conducted 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, 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, and 0.5 m/z 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 enabled[42]. Using a human fasta downloaded from Uniprot, fixed modifications for the TMTpro mass (+304.207146) were added to n-terminal residues and lysines. Carbamidomethly (+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 was analyzed with MSFragger[87] 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. Fixed modifications included Carbamidomethyl (+57.021464) on cysteines. Variable modifications included were 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 Percolator[88].

Multiplexed raw mass spectrometry data was analyzed using the Comet[89] 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 were 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 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). Log_2 fold changes of the mean of the biological replicates were also calculated for each biological condition. Human Protein Atlas[55] 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[90,91]. Significant hits were those with a SAINT calculated FDR less than 1%[92]. BioPlex interaction networks were accessed through the online BioPlex Explorer[93] (https://bioplex.hms.harvard.edu/). Networks were imaged using Cytoscape 3.10.02[94]. Protein complex members were accessed through CORUM[95]. Gene set enrichment analysis was performed with clusterProfiler[96] and fgsea[97] 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 ul 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 ul 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 Cutadapt[98]. 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>=30[99]. Duplicate reads were removed using Picard 3.1.1[100]. Eluate reads were normalized to input reads using DeepTools[101] bamCompare with the following parameters: –binSize 20 –normalizeUsing BPM –smoothLength 60 –extendReads 150. Normalized data were visualized using Coolbox 0.3.9[102].

Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange[103] 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.

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., 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.

Competing Interest Statement

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.

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 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.

References

1. Bickmore, W. A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270–1284 (2013).
2. Jerkovic, I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. *Nat. Rev. Mol. Cell Biol.* **22**, 511–528 (2021).
3. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
4. Ho, J. W. K., Alekseyenko, A. A., Kuroda, M. I. & Park, P. J. Genome-wide mapping of protein-DNA interactions by ChIP-seq. in *Tag-Based Next Generation Sequencing* 139–151 (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2012).
5. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins

using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424–428 (2000).
6. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
7. Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889 (2020).
8. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
9. Bujold, D. *et al.* The International Human Epigenome Consortium Data Portal. *Cell Syst* **3**, 496-499.e2 (2016).
10. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
11. Hoffman, M. M., Buske, O., Bilmes, J. A. & Noble, W. S. Segway: a dynamic Bayesian network method for segmenting genomic data. *Invert. Neurosci.* (2009).
12. Gao, X. D. *et al.* C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9–APEX2. *Nat. Methods* **15**, 433–436 (2018).
13. Myers, S. A. *et al.* Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nat. Methods* **15**, 437–439 (2018).
14. Qiu, W. *et al.* Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system. *Nucleic Acids Res.* **47**, e52 (2019).
15. Ugur, E., Bartoschek, M. D. & Leonhardt, H. Locus-Specific Chromatin Proteome Revealed by Mass Spectrometry-Based CasID. *Methods Mol. Biol.* **2175**, 109–121 (2020).
16. Déjardin, J. & Kingston, R. E. Purification of proteins associated with specific genomic Loci. *Cell* **136**, 175–186 (2009).
17. Silahtaroglu, A. N., Tommerup, N. & Vissing, H. FISHing with locked nucleic acids (LNA): evaluation of different LNA/DNA mixmers. *Mol. Cell. Probes* **17**, 165–169 (2003).
18. Beliveau, B. J. *et al.* Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21301–21306 (2012).
19. Ide, S. & Dejardin, J. End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. *Nat. Commun.* **6**, 6674 (2015).
20. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).
21. Cho, K. F. *et al.* Proximity labeling in mammalian cells with TurboID and split-TurboID. *Nat. Protoc.* **15**, 3971–3999 (2020).
22. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2015).
23. Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol.* **30**, 1143–1148 (2012).
24. Mangeot, P. E. *et al.* Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat. Commun.* **10**, 45 (2019).
25. Tsue, A. F. *et al.* Oligonucleotide-directed proximity-interactome mapping (O-MAP): A unified method for discovering RNA-interacting proteins, transcripts and genomic loci in situ. *bioRxiv* (2023) doi:10.1101/2023.01.19.524825.
26. Attar, S. *et al.* Programmable peroxidase-assisted signal amplification enables flexible

detection of nucleic acid targets in cellular and histopathological specimens. *bioRxiv* (2023) doi:10.1101/2023.01.30.526264.
27. Li, J. *et al.* TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nat. Methods* **17**, 399–404 (2020).
28. Chakravarti, D., LaBella, K. A. & DePinho, R. A. Telomeres: history, health, and hallmarks of aging. *Cell* **184**, 306–322 (2021).
29. Sfeir, A. & de Lange, T. Removal of shelterin reveals the telomere end-protection problem. *Science* **336**, 593–597 (2012).
30. de Lange, T. Shelterin-Mediated Telomere Protection. *Annu. Rev. Genet.* **52**, 223–247 (2018).
31. Myung, K. *et al.* Regulation of telomere length and suppression of genomic instability in human somatic cells by Ku86. *Mol. Cell. Biol.* **24**, 5050–5059 (2004).
32. Huttlin, E. L. *et al.* Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**, 3022-3040.e28 (2021).
33. Schweppe, D. K., Huttlin, E. L., Harper, J. W. & Gygi, S. P. BioPlex Display: An Interactive Suite for Large-Scale AP-MS Protein-Protein Interaction Data. *J. Proteome Res.* **17**, 722–726 (2018).
34. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
35. Liu, X. *et al.* In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* **170**, 1028-1043.e19 (2017).
36. Garcia-Exposito, L. *et al.* Proteomic Profiling Reveals a Specific Role for Translesion DNA Polymerase η in the Alternative Lengthening of Telomeres. *Cell Rep.* **17**, 1858–1871 (2016).
37. LaBranche, H. *et al.* Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat. Genet.* **19**, 199–202 (1998).
38. Zhang, Q.-S., Manche, L., Xu, R.-M. & Krainer, A. R. hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA* **12**, 1116–1128 (2006).
39. Flynn, R. L. *et al.* TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature* **471**, 532–536 (2011).
40. Fu, D. & Collins, K. Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation. *Mol. Cell* **28**, 773–785 (2007).
41. Izumi, H. & Funa, K. Telomere Function and the G-Quadruplex Formation are Regulated by hnRNP U. *Cells* **8**, (2019).
42. Schweppe, D. K. *et al.* Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *J. Proteome Res.* **19**, 2026–2034 (2020).
43. Navarrete-Perea, J., Yu, Q., Gygi, S. P. & Paulo, J. A. 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 (2018).
44. McKinley, K. L. & Cheeseman, I. M. The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* **17**, 16–29 (2016).

45. Talbert, P. B. & Henikoff, S. The genetics and epigenetics of satellite centromeres. *Genome Res.* **32**, 608–615 (2022).
46. McNulty, S. M. & Sullivan, B. A. Alpha satellite DNA biology: finding function in the recesses of the genome. *Chromosome Res.* **26**, 115–138 (2018).
47. Altemose, N. *et al.* Complete genomic and epigenetic maps of human centromeres. *Science* **376**, eabl4178 (2022).
48. Deng, Z. & Beliveau, B. J. An open source 16-channel fluidics system for automating sequential fluorescent in situ hybridization (FISH)-based imaging. *HardwareX* **12**, e00343 (2022).
49. Aguilar, R. *et al.* Tigerfish designs oligonucleotide-based in situ hybridization probes targeting intervals of highly repetitive DNA at the scale of genomes. *Nat. Commun.* **15**, 1027 (2024).
50. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465 (1981).
51. Rackham, O. & Filipovska, A. Organization and expression of the mammalian mitochondrial genome. *Nat. Rev. Genet.* **23**, 606–623 (2022).
52. Zhou, K. *et al.* A novel next-generation sequencing-based approach for concurrent detection of mitochondrial DNA copy number and mutation. *J. Mol. Diagn.* **22**, 1408–1418 (2020).
53. Paek, J. *et al.* Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell* **169**, 338-349.e11 (2017).
54. Matilainen, O., Quirós, P. M. & Auwerx, J. Mitochondria and Epigenetics - Crosstalk in Homeostasis and Stress. *Trends Cell Biol.* **27**, 453–463 (2017).
55. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, (2017).
56. Bersaglieri, C. *et al.* Genome-wide maps of nucleolus interactions reveal distinct layers of repressive chromatin domains. *Nat. Commun.* **13**, 1483 (2022).
57. Liang, C. *et al.* Centromere-localized Aurora B kinase is required for the fidelity of chromosome segregation. *J. Cell Biol.* **219**, (2020).
58. Broad, A. J., DeLuca, K. F. & DeLuca, J. G. Aurora B kinase is recruited to multiple discrete kinetochore and centromere regions in human cells. *J. Cell Biol.* **219**, (2020).
59. Rangasamy, D., Berven, L., Ridgway, P. & Tremethick, D. J. Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. *EMBO J.* **22**, 1599–1607 (2003).
60. Ono, T., Fang, Y., Spector, D. L. & Hirano, T. Spatial and temporal regulation of Condensins I and II in mitotic chromosome assembly in human cells. *Mol. Biol. Cell* **15**, 3296–3308 (2004).
61. Kufer, T. A. *et al.* Human TPX2 is required for targeting Aurora-A kinase to the spindle. *J. Cell Biol.* **158**, 617–623 (2002).
62. Khongkow, P. *et al.* Paclitaxel targets FOXM1 to regulate KIF20A in mitotic catastrophe and breast cancer paclitaxel resistance. *Oncogene* **35**, 990–1002 (2016).
63. Politz, J. C. R., Scalzo, D. & Groudine, M. Something silent this way forms: the functional organization of the repressive nuclear compartment. *Annu. Rev. Cell Dev. Biol.* **29**, 241–270 (2013).
64. Rodrigues, A. *et al.* Nucleoli and the nucleoli–centromere association are dynamic during

normal development and in cancer. *MBoC* **34**, br5 (2023).
65. Hoffmeyer, K. *et al.* Wnt/β-catenin signaling regulates telomerase in stem cells and cancer cells. *Science* **336**, 1549–1554 (2012).
66. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e16 (2017).
67. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
68. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
69. Weber, C. M. *et al.* mSWI/SNF promotes Polycomb repression both directly and through genome-wide redistribution. *Nat. Struct. Mol. Biol.* **28**, 501–511 (2021).
70. Zeng, J. *et al.* Targeting EZH2 for cancer therapy: From current progress to novel strategies. *Eur. J. Med. Chem.* **238**, 114419 (2022).
71. Guo, Y. *et al.* Regulation of EZH2 protein stability: new mechanisms, roles in tumorigenesis, and roads to the clinic. *EBioMedicine* **100**, 104972 (2024).
72. Jansz, N. *et al.* Smchd1 targeting to the inactive X is dependent on the Xist-HnrnpK-PRC1 pathway. *Cell Rep.* **25**, 1912-1923.e9 (2018).
73. Pintacuda, G. *et al.* HnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish Polycomb-mediated chromosomal silencing. *Mol. Cell* **68**, 955-969.e10 (2017).
74. Xu, J. *et al.* Developmental control of polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. *Mol. Cell* **57**, 304–316 (2015).
75. Bousard, A. *et al.* The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO Rep.* **20**, e48019 (2019).
76. Markaki, Y. *et al.* Xist nucleates local protein gradients to propagate silencing across the X chromosome. *Cell* **184**, 6174-6192.e32 (2021).
77. Wang, C.-Y., Colognori, D., Sunwoo, H., Wang, D. & Lee, J. T. PRC1 collaborates with SMCHD1 to fold the X-chromosome and spread Xist RNA between chromosome compartments. *Nat. Commun.* **10**, 2950 (2019).
78. Nir, G. *et al.* Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* **14**, e1007872 (2018).
79. Minajigi, A. *et al.* Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276 (2015).
80. Mossa, A. *et al.* Sex-specific perturbations of neuronal development caused by mutations in the autism risk gene DDX3X. *Nat. Commun.* **16**, 4512 (2025).
81. Hershberg, E. A. *et al.* PaintSHOP enables the interactive design of transcriptome- and genome-scale oligonucleotide FISH experiments. *Nat. Methods* **18**, 937–944 (2021).
82. Santos-Barriopedro, I., van Mierlo, G. & Vermeulen, M. Off-the-shelf proximity biotinylation for interaction proteomics. *Nat. Commun.* **12**, 5015 (2021).
83. Iglesias, N. *et al.* Native Chromatin Proteomics Reveals a Role for Specific Nucleoporins in Heterochromatin Organization and Maintenance. *Mol. Cell* **77**, 51-66.e8 (2020).
84. Kishi, J. Y., Schaus, T. E., Gopalkrishnan, N., Xuan, F. & Yin, P. Programmable autonomous synthesis of single-stranded DNA. *Nat. Chem.* **10**, 155–164 (2018).
85. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).

86. Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep.* **15**, 210–218 (2016).
87. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520 (2017).
88. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **4**, 923–925 (2007).
89. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24 (2013).
90. Choi, H. *et al.* SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat. Methods* **8**, 70–73 (2011).
91. Teo, G. *et al.* SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *J. Proteomics* **100**, 37–43 (2014).
92. Choi, H. *et al.* Analyzing protein-protein interactions from affinity purification-mass spectrometry data with SAINT. *Curr. Protoc. Bioinformatics* **Chapter 8**, 8.15.1-8.15.23 (2012).
93. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425–440 (2015).
94. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
95. Ruepp, A. *et al.* CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.* **36**, D646-50 (2008).
96. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
97. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* 060012 (2021) doi:10.1101/060012.
98. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
99. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
100. Picard. https://broadinstitute.github.io/picard/.
101. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-5 (2016).
102. Xu, W. *et al.* CoolBox: a flexible toolkit for visual analysis of genomics data. *BMC Bioinformatics* **22**, 489 (2021).
103. Vizcaíno, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226 (2014).