Published in final edited form as:

Methods Enzymol. 2019 ; 616: 365–383. doi:10.1016/bs.mie.2018.10.030.

Adapting dCas9-APEX2 for subnuclear proteomic profiling

Xin D. Gao, Tomás C. Rodríguez, and Erik J. Sontheimer*

RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA, United States

Abstract

Genome organization and subnuclear protein localization are essential for normal cellular function and have been implicated in the control of gene expression, DNA replication, and genomic stability. The coupling of chromatin conformation capture (3C), chromatin immunoprecipitation and sequencing, and related techniques have continuously improved our understanding of genome architecture. To profile site-specifically DNA-associated proteins in a high-throughput and unbiased manner, the RNA-programmable CRISPR–Cas9 platform has recently been combined with an enzymatic labeling system to allow proteomic landscapes at repetitive and nonrepetitive loci to be defined with unprecedented ease and resolution. In this chapter, we describe the dCas9-APEX2 experimental approach for specifically targeting a DNA sequence, enzymatically labeling local proteins with biotin, and quantitatively analyzing the labeled proteome. We also discuss the optimization and extension of this pipeline to facilitate its use in understanding nuclear and chromosome biology.

1. Introduction

The 3D organization of the genome can be defined using several methods with continuously improving resolution and temporal specificity. This has been achieved primarily through the coupling of chromosome conformation capture (3C)-derived methods, microscopy techniques, and chromatin Immuno-precipitation and sequencing (ChIP-seq) (Bonev & Cavalli, 2016). However, while 3C-derived approaches allow genome-wide mapping of DNA–DNA interactions, ChIP-seq does not offer an unbiased survey of site-specifically DNA-bound factors. The applications of ChIP-seq are further limited to proteins with known nuclear localization and to those for which high-affinity antibodies or epitope tags are available.

To provide a more complete depiction of the protein microenvironment surrounding genomic loci, several “reverse-ChIP” methods have been developed (Rusk, 2009). The common purpose of these increasingly diverse methods is to identify protein factors spatially proximal to a target locus. Proteomics of isolated chromatin segments (PICH), for example, isolates cross-linked chromatin using an oligonucleotide probe hybridized to a genomic region of interest (Déjardin & Kingston, 2009). The use of nuclease-dead CRISPR–Cas9 (dCas9) probes in place of modified oligonucleotides provides an additional, streamlined

*Corresponding author: erik.sontheimer@umassmed.edu.

approach to isolating cross-linked chromatin (Fujita & Fujii, 2013; Liu et al., 2017; Tsui et al., 2018). However, a substantial hurdle posed by formaldehyde cross-linking is the nonspecific enrichment of spatially distant proteins.

Cross-linking-independent “proximity-labeling” approaches have also been employed to capture the protein landscape within specified nuclear subcompartments. BioID, for example, is designed to biotinylate proteins adjacent to a promiscuous BirA* biotin ligase-bait fusion (Roux, Kim, Raida, & Burke, 2012). dCas9-BirA* (CasID) fusion extends this method with sequence-specific localization and biotin labeling by programmable single-guide RNA (sgRNA) (Schmidtman, Anton, Rombaut, Herzog, & Leonhardt, 2016). While the radius of biotin ligation using BirA* is better defined (~20–30nm) than enrichment by cross-linking, the long reaction time usually necessary for this proximity-labeling method is unlikely to capture proteins only transiently associated to the target locus, though recent advances are ameliorating this limitation (Branon et al., 2018).

dCas9-APEX2 biotinylation at genomic elements by restricted spatial tagging (C-BERST) and similar methods addresses, in part, the limitations of BirA*- and cross-linking-based reverse-ChIP (Gao et al., 2018; Myers et al., 2018). C-BERST exploits APEX2, an engineered soybean ascorbate peroxidase, as a means for proximity labeling that (i) does not depend on cross-linking, (ii) tags proteins within a ~20-nm radius, and (iii) requires only a ~1-min reaction time (Hung et al., 2014; Rhee et al., 2013). Specific DNA sequence targeting can be programmed by sgRNAs. These features allow for the site-specific characterization of dynamic genomic regulatory events.

APEX2-mediated labeling is achievable in biotin-phenol (BP)-treated live cells with the addition of H₂O₂. A short, easily quenchable reaction generates biotin-phenoxyl radicals that covalently bind to electron-rich amino acid side chains within a confined space (Rhee et al., 2013). APEX2 fusion with dCas9 and mCherry allows for construct localization to a genomic region of interest and microscopic validation of genomic sequence targeting, respectively. Stably transduced, mCherry-positive cells are sorted prior to biotinylation. Furthermore, dCas9-mCherry-APEX2 contains an N-terminal ligand-tunable degron domain (Banaszynski, Chen, Maynard-Smith, Ooi, & Wandless, 2006) and drug-inducible promoter (Das, Tenenbaum, & Berkhout, 2016) to limit leaky construct expression and resulting reactive oxygen generation.

Biotinylated nuclear proteins extracted from treated cells are isolated by streptavidin affinity purification and identified with liquid chromatography-tandem mass spectrometry (LC-MS/MS).

Stable isotope labeling with amino acids in cell culture (SILAC) improves the quantitative measure of proteins identified by MS (Ong & Mann, 2007) (Fig. 1).

2. Generation of cell lines stably expressing dCas9-APEX2 and sgRNAs

2.1 Construction and selection of dCas9-APEX2 cells

We reasoned that tightly controlled dCas9-APEX2 expression will increase the specificity of labeling and reduce the noise generated by background nucleoplasmic dCas9-APEX2. Therefore, we constructed an inducible dCas9-mCherry-APEX2 under a CMV promoter with two embedded tetO sites. Additionally, a destabilization domain was appended to the N-terminus of dCas9 to further attenuate leaky expression in a manner that is tunable by the Shield1 ligand that can suppress degradation (Banaszynski et al., 2006). The sgRNA plasmid also includes a Tet repressor followed by a self-cleaving P2A peptide and a blue fluorescent protein (BFP)-coding sequence, driven by the human PGK promoter (Fig. 2). Therefore, in the stable cell line expressing both dCas9-mCherry-APEX2 and sgRNAs, dCas9-mCherry-APEX2 is maximally expressed in the presence of doxycycline and Shield1. Through viral transduction, the two components are introduced into the cell line of interest. Stable cell lines or clonal lines can be FACS-sorted (mCherry and BFP double-positive) or drug-selected (Mortensen, Chestnut, Hoeffler, & Kingston, 1997). Here, we describe the procedure for constructing stable cell lines by lentiviral infection.

1. Seed HEK293T cells into 1-well of 6-well plate 1 day before transfection. Cells should be about 80% confluent on the second day.
2. Transfect 293T cells with 0.5µg packaging plasmid dR8.2 (Addgene ID: 8455), 0.25µg VSV-G envelop plasmid (Addgene ID: 8454), and 1.5µg dCas9-mCherry-APEX2 plasmids (Addgene ID: 108570) following the Mirus-recommended protocol. U2OS cells should be maintained for later transduction.
3. 48h after transfection, collect and concentrate the lentiviral supernatant using Takara lentiviral concentrator following the recommended protocol. Meanwhile, 1.8×10^5 U2OS cells are plated onto a well (6-well plate). Add 1mL concentrated virus supernatant to plated U2OS cells. Spin the plate at $1200 \times g$ for 30min at room temperature. Carefully remove the plate from the centrifuge and place it into an incubator. Maintain the cells at 37°C and 5% CO₂ for 48h.
4. Cells are passaged and plated onto a 10-cm Petri dish. Once cells are expanded, FACS is performed to sort out mCherry-positive cells under sterile conditions. Sorted cells are expanded and maintained for sgRNA-encoding lentivirus transduction.

2.2 Design and cloning of sgRNAs

Previous reports have indicated that sgRNA expression levels limit Cas9 function in human cells (Jinek et al., 2013). Live-cell imaging of genomic loci via CRISPR-Cas9 system also emphasizes the importance of sgRNA expression efficiency and assembly with dCas9-GFP (Chen et al., 2013). It is crucial to carefully engineer and validate sgRNAs. Here, we describe the general procedure of sgRNA design.

1. Search for 5'- (N)₁₉-NGG-3' sequences within the DNA region of interest. NGG is the PAM sequence recognized by the *Streptococcus pyogenes* Cas9

(SpyCas9) protein. The base-pairing region (spacer) of the sgRNA is represented as (N)₁₉.

2. To clone the spacer sequence into the *Bsp*MI-digested backbone (Addgene ID: 117687), phosphorylate and anneal forward oligo 5'-accg (N)₁₉-3' and reverse oligo 5'-aaac (N)₁₉-3' [the (N)₁₉ in the reverse oligo is reverse complementary to the (N)₁₉ in the forward oligo]. The PCR program is as follows: 5'-phosphorylation at 37°C for 30min, denature at 98°C for 5min, and reanneal at 97°C decreasing by 1°C/min until 12°C is reached.
3. Clone the annealed sgRNA fragment into the digested plasmid according to manufacturer's instructions using Quick T4 DNA ligase (NEB, #M2200S).

2.3 Expression of sgRNAs using lentiviral vectors and selection of stable cell line

Similar to the protocol described for dCas9-mCherry-APEX2 viral transduction, sgRNA lentiviruses were packaged in HEK293T cells. The amount of sgRNA viral particles used for transduction should be at least threefold (Chen et al., 2013) that of dCas9-mCherry-APEX2.

1. Once cells are expanded and recovered from transduction, select BFP and mCherry double-positive cells using FACS (doxycycline and Shield1 are added 21h before FACS).
2. Select the population that represents both high BFP and low mCherry expression (P1), reflecting high sgRNA to dCas9-mCherry-APEX2 ratio (Fig. 3). Using live-cell imaging, P1 should show the lowest nucleolar background signal compared to the other three subpopulations. Generation of clonal cell lines is strongly recommended for selection and validation when targeting nonrepetitive regions.

3. Assessing dCas9-APEX2-specific DNA targeting and biotin-labeling signals at repetitive loci

3.1 Live-cell imaging of a repetitive locus for assessing targeting efficiency

CRISPR-based imaging platforms (dCas9 tethered to a fluorescent protein) are well established for visualizing genomic elements, especially at repetitive elements. We used live-cell imaging to assess the on-target localization of dCas9-mCherry-APEX2 molecules. Imaging procedures are briefly described here.

1. Seed dCas9-APEX/sgRNA stable cell lines onto 35 10mm glass-bottom dishes supplemented with doxycycline and Shield1 21h before imaging.
2. Image live cells (70%–80% confluent) with a deconvolution microscope capable of 3D-imaging. Images can be processed with MetaMorph or similar software for the generation of z-stacks and processing into maximum projection composite images.
3. Sorted subpopulations or clonal cell lines are assessed by CRISPR-based imaging. Cells demonstrating minimal nucleolar localization, high signal-to-

background ratio, and robust puncta are optimal for downstream biotin-labeling steps (Chen et al., 2016; Gu et al., 2018). DNA FISH or immunofluorescence experiments can be carried out to confirm the specific DNA sequence targeting by dCas9-mCherry-APEX2 using specific DNA probes or marker proteins (Chen & Huang, 2014) (Fig. 4).

3.2 Assessing biotin labeling

Neutravidin-conjugated fluorescent antibodies can be used to visualize biotin-tagged proteins in background-labeling and target-labeling samples. Optimal stable cells and clonal populations will show a high degree of overlap between mCherry foci and neutravidin-fluorophore signals, suggesting efficient biotinylation that is mostly restricted to the target site.

1. Cells are seeded and grown on glass cover slips. Dox and Shield1 are added to induce dCas9-mCherry-APEX2 expression. After 21h, live cells are incubated with BP and H₂O₂ for 30 and 1min, respectively.
2. Cells are fixed in PHEM [0.05 M PIPES/0.05 M HEPES (pH 7.4), 0.01 M EGTA, 0.01 M MgCl₂] supplemented with 2% paraformaldehyde for 15min and then subjected to a 2-min extraction with 0.1% Triton X-100 in PHEM.
3. After two brief PBS washes, the cells are blocked with 1% BSA in TBST at 4°C overnight.
4. Cells are incubated with neutravidin conjugated with Oregon Green 488 for 2h at room temperature and washed three times with blocking solution (10min/wash). Cells are then incubated with secondary antibodies for 1h at room temperature and subjected to another three washes in blocking solution followed by two washes in PBS.
5. Cells are mounted with ProLong antifade (without DAPI) and visualized by fluorescence microscopy (Fig. 5).

3.3 Repetitive locus target optimization

PCR amplification of repetitive genomic regions often yields nonspecific, undesired products (Hommelsheim, Frantzeskakis, Huang, & Ülker, 2014). Validation of sgRNA binding to repetitive target elements by chromatin immunoprecipitation followed by quantitative polymerase chain reaction (ChIP-qPCR) is therefore not recommended. While ChIP-seq may serve as an alternative, traditional sequencing alignment pipelines discard repetitive elements due to nonunique mapping on reference genomes. Raw reads must, therefore, be manually filtered and assessed for enrichment of target repetitive sequences. Though alternative strategies have been proposed to map “multireads” to custom reference genomes (Schwartz & Längst, 2016), they are not commonly employed and will not be discussed here. We demonstrate read quantification using R, a statistical computing interface available on Linux, Mac, and Windows operating systems.

1. Block two aliquots of 50 μ L protein G Dynabeads by washing twice with 400 μ L PBS/BSA (PBS with 5mg/mL BSA) and incubating beads with 400 μ L PBS/BSA on a rotating wheel at 4°C for at least 3h.
2. Culture 1×10^7 U2OS cells stably expressing mCherry-APEX2-dCas9 transduced with sequence targeting or nonspecific sgRNAs and treat with doxycycline (2ng/mL) for 21h.
3. Wash with PBS. Treat with 1% formaldehyde for 10min and quench immediately after with 0.125 M glycine.
4. Harvest cells using a plate scraper and lyse in 1mL RIPA buffer [50mM Tris-HCl (pH 7.5), 150mM NaCl, 0.125% SDS, 0.125% sodium deoxycholate, and 1% Triton X-100 in Millipore water] with 1% $\times$ freshly supplemented Halt Protease Inhibitor for 10min on ice.
5. Centrifuge lysate at $2300 \times g$ for 5min at 4°C to isolate the nuclei. Lyse in 0.5% SDS-containing RIPA lysis buffer.
6. Shear chromatin to 200–700bp fragments by sonication at 4°C.
Optional: To measure chromatin fragment size, add 1 μ L RNase A to 5 μ L lysate and incubate overnight at 65°C. Add 7.5 μ L of 20mg/mL proteinase K and incubate at 50°C for 2h. Purify DNA using a spin column and visualize on a 1% agarose gel.
7. Add 4 μ g anti-mCherry antibody (Thermo; PA5–34974) to 500 μ L nuclear lysate. Add 4 μ g Rabbit IgG Isotype Control (Thermo 31235) to remaining 500 μ L nuclear lysate (mock sample). Rotate samples at 4°C for 3h.
8. Wash blocked beads twice with 400 μ L PBS/BSA. Remove PBS/BSA on magnetic stand. Add nuclear lysate and antibody mixtures to beads. Rotate samples at 4°C overnight.
9. Wash each sample twice with RIPA lysis buffer, once with 1 M KCl, once with 0.1 M Na₂CO₃, once with 2 M urea in 10mM Tris-HCl (pH 8.0), and twice with RIPA lysis buffer.
10. Centrifuge the samples at 18,000rpm for 1s, place on magnetic stand, and remove the supernatant completely.
11. Add 40 μ L of elution buffer and incubate at 65°C for 15min of shaking. Centrifuge at maximum speed for 1s, place on a magnetic rack, and transfer the supernatant into a new 1.5mL tube. Add 160 μ L TE+1% SDS to beads, vortex, centrifuge 18,000rpm for 1s, place on a magnetic rack, and combine the supernatant with the previous eluate. Optional: Take 5 μ L of eluate for Western blotting (IP) and keep at –20°C.
12. Add 1 μ L RNase A and incubate all samples overnight at 65°C to digest RNA and reverse the cross-links.

13. To each of the 200 μ L aliquots, add 7.5 μ L of 20mg/mL proteinase K and incubate at 50°C for 2h.
14. Add 1mL Buffer PB (QIAGEN) and 10 μ L of 3 M NaOAc (pH 5.2) and incubate at 37°C for 30min. Purify DNA using QIAquick column and elute with 50 μ L TE (pH 8.0).
15. Using 15 μ L ChIP-DNA, proceed with end repair using NEBNext End Repair module (NEB Cat. E6050), purify with 1.8 $\times$ AMPure XP beads (Beckman-Coulter Cat. A63880) according to the manufacturer's protocol, and elute into 44 μ L 0.1 $\times$ TE.
16. Process end-repaired DNA with a dA-tailing reaction using NEBNext dA-Tailing module (NEB Cat. E6053), purify with 1.8 $\times$ AMPure XP beads, and elute DNA into 19 μ L 0.1 $\times$ TE.
17. Prepare Y-adaptor mix according to Zhang, Theurkauf, Weng, and Zamore (2012). Add 1 μ L Y-adaptor (~1.6 μ M) mix, 6 μ L quick ligation reaction buffer (5 $\times$), and 4 μ L Quick T4 DNA ligase to eluted DNA (total volume 30 μ L). Incubate in a thermal cycler for 15min at 20°C. Purify with 1.5 $\times$ AMPure XP beads and elute DNA into 20 μ L 0.1 $\times$ TE.
18. Incubate ligated DNA with 25 μ L 2 $\times$ Q5 PCR mix and 2.5 μ L of 10 μ M Illumina barcode PCR primer (Zhang et al., 2012) 2–1 (anti-mCherry ChIP sample) or 2–2 (mock IgG sample) in a thermal cycler (98°C for 40s, 65°C for 30s, and 72°C for 30s).
19. Add 2.5 μ L of 10 μ M Universal PCR Primer 1 (Zhang et al., 2012) to the above PCR mix. Incubate for 10 cycles (98°C for 10s, 65°C for 30s, 72°C for 30s), followed by incubation at 72°C for 3min. Purify with 1.0 $\times$ AMPure XP beads and elute DNA into 25 μ L 0.1 $\times$ TE.

Note: We recommend TOPO PCR cloning of ChIP library DNA followed by Sanger sequencing of 5–10 individual clones to show enrichment of desired repetitive regions. However, this method is not quantitative and should not replace Illumina sequencing for validation of gRNA binding.
20. Trim .fastq files containing raw Illumina sequencing reads of 150 nucleotide length using the Bioconductor ShortRead R. By default, this package trims sequences at positions in which two nucleotides in a five-nucleotide bin have a quality encoding phred score<20.
21. Discard reads that do not satisfy a minimum specified length requirement (e.g., the length of one PAM+sgRNA-complementary sequence).
22. Quantify the number of reads in the anti-mCherry and mock IgG ChIP samples that contain at least one sgRNA's worth of target repetitive DNA.
23. The proportion of anti-mCherry ChIP reads reflecting the desired repetitive sequence should be significantly higher than that of the IgG mock sample.

Additional analysis may be required to suit different experiments and biological questions.

3.4 Nonrepetitive locus target optimization

In cells stably expressing dCas9-mCherry-APEX, ChIP-qPCR is necessary to confirm biotinylation construct localization to a specified genomic site targeted by single or multiplexed guides.

1. Follow Section 3.3, steps 1–14.
2. Design qPCR primers targeting the sgRNA-binding site, regions ~5kb upstream and downstream of the sgRNA-binding site, and a reference gene (e.g., *GAPDH*). RTPrimerDB and RefGenes offer a large variety of validated primers and reference genes, respectively, for use in qPCR experiments (Hruz et al., 2011; Lefever, Vandesompele, Speleman, & Pattyn, 2009).

Note: ChIP-seq may offer sgRNA-binding validation of higher resolution to that of qPCR. Library preparation can be performed according to Gao et al. (2018) after step 1 above. Genome indexing, alignment, peak-calling, and statistical analysis of enriched sequences should be performed according to the ENCODE Project Transcription Factor ChIP-seq pipeline (ENCODE Project Consortium, 2012).

3. sgRNA-binding sites that are significantly enriched compared to ± 5 kb negative-control sites are acceptable for use in C-BERST experiments.

4. SILAC labeling, C-BERST biotinylation, and enrichment of biotinylated proteins

The APEX-based enzymatic labeling system has been widely used in mammalian cells for subcellular compartment proteomic studies. Mammalian cells are incubated with BP for 30min to achieve efficient delivery and subsequently treated for 1min with H_2O_2 to initiate biotin labeling. This enzymatic tagging system offers a unique advantage for capturing proteome changes in dynamic and transient processes. Additionally, quantitative proteomic methods such as SILAC and tandem mass tagging (TMT) can be integrated into the C-BERST pipeline for more robust data analysis. A quantitative approach can also facilitate de novo protein factor discovery. Here, we describe the general procedure of amino acid isotope cell labeling and biotinylation.

4.1 SILAC labeling and biotinylation

SILAC is used to barcode experimental and control samples and improve enriched-protein detection sensitivity. The general steps are described later:

1. Day 0:
In preparation for the proximity-labeling experiment, early passage, sorted, stably transduced U2OS cells are grown in SILAC media. Samples expressing target-locus sgRNA are cultured in “heavy” SILAC media containing L-

arginine- $^{13}\text{C}_6$, $^{15}\text{N}_4$ (Arg10), and L-lysine- $^{13}\text{C}_6$, $^{15}\text{N}_2$ (Lys8) (Sigma). Control samples expressing nonspecific sgRNA are cultured in “medium” SILAC media which contained L-arginine- $^{13}\text{C}_6$ (Arg6) and L-lysine-4,4,5,5- d_4 (Lys4) (Sigma). Untransduced U2OS cells are cultured in “light” SILAC media containing L-arginine (Arg0) and L-lysine (Lys0) (Sigma).

2. Days 1–10:
Passage cells >5 times to allow for sufficient incorporation of SILAC isotopes.
3. Day 11:
Doxycycline (2 $\mu\text{g}/\text{mL}$) and Shield1 (250nM) are added to each isotope culture 21h before treatment with BP and H_2O_2 . When targeting repetitive genetic elements (e.g., centromere or telomere repeats), 6×10^7 cells provide sufficient amounts of biotinylated protein.
4. Incubate cells with 500 μM BP for 30min at 37°C. 1mM H_2O_2 is then added to initiate biotinylation for 1min on a horizontal shaker at room temperature. Add quencher solution (5mM trolox, 10mM sodium ascorbate, and 10mM sodium azide) to stop the reaction and wash cells five times (two quencher washes, two DPBS washes, and one quencher wash) to remove excess BP. Each wash should last >1min.

4.2 Streptavidin bead enrichment of biotinylated protein

Nuclear isolation reduces nonspecific capture of background proteins and minimizes contamination by endogenous cytoplasmic biotin. This step improves the efficiency of C-BERST-tagged nuclear factor enrichment by streptavidin affinity pulldown.

5. Day 11 (cont.):
Remove quenched cells from the dish using a cell scraper and collect in 10mL DPBS. Centrifuge at $300 \times g$ for 5min. Aspirate the supernatant and resuspend the cells in 10mL DPBS. Repeat centrifugation.
6. Aspirate the supernatant and resuspend in 7.5mL nuclei buffer [10mM PIPES (pH 7.4), 0.1% NP-40, 10mM KCl, 2mM MgCl_2 , 1mM DTT (add immediately before use), and 1 $\times$ Halt Protease Inhibitor (add immediately before use)]. Incubate cells on ice for 10min. Transfer cells to homogenizer and dounce the cells with 20 strokes. Incubate cells on ice for 20min. Dounce the cells with another 20 strokes.
7. Layer the cells on a sucrose cushion made of 20mL 30% sucrose [10mM PIPES (pH 7.4), 10mM KCl, 2mM MgCl_2 , 30% sucrose, 1mM DTT (add immediately before use)] at the bottom and 3.5mL 10% sucrose (the same buffer components but with 10% sucrose prechilled at 4°C in a 50-mL Falcon tube) at the top. Spin the tube at $1000 \times g$ for 15min.
8. Remove supernatant and interphase carefully. Nuclei are found in the pellet. Resuspend the pellet in 800 μL DPBS and centrifuge cells in a 1.5-mL Eppendorf tube at 4000rpm for 5min. Use 500 μL lysis buffer [50mM Tris-HCl (pH 7.5),

150mM NaCl, 0.125% SDS, 0.125% sodium deoxycholate, 1% Triton X-100 in Millipore water] to lyse each sample for 10min on ice.

9. Sonicate cells (30s on/off cycles) at high intensity using a Diagenode Bioruptor for 15min. Cell lysates are clarified by centrifugation at 13,000rpm for 10min. Mix clarified protein (~2.5mg for each state) from heavy, medium, and light samples and begin affinity pulldown using 800µL Dynabeads MyOne Streptavidin T1. Incubate overnight at 4°C.
10. Day 12:

Wash each sample with a series of buffers to remove nonspecifically bound proteins: twice with RIPA lysis buffer, once with 1 M KCl, once with 0.1 M Na₂CO₃, once with 2 M urea in 10mM Tris-HCl (pH 8.0), and twice with RIPA lysis buffer.
11. Elute protein in 50µL 3× protein loading buffer supplemented with 2mM biotin and 20mM DTT for 10min at 65°C. Load 50µL eluate into a 4%–12% SDS-PAGE precast gel (Bio-Rad). Run samples halfway down the gel and stain with Coomassie. Excise each lane and cut into five slices for in-gel digestion and LC-MS/MS analysis.

5. Mass spectrometry analysis

Quantitative proteomics can improve the identification of locus-specific bound factors by dCas9-APEX2. Although label-free quantification is effective in profiling some subcellular proteomes, SILAC or TMT methods are recommended for analyzing locus-specific subnuclear space. Here, we describe the general procedure for SILAC mass spectrometry analysis.

1. Place excised protein gel slices into a 1.5-mL tube and incubate in 1mL water for 30min. Remove water. Add 200µL of 250mM ammonium bicarbonate and 20µL of a 45mM solution of DTT and incubate the samples at 50°C for 30min.
2. Cool the samples to room temperature and add 20µL of a 100-mM iodoacetamide solution. Allow the alkylation reaction to proceed for 30min.
3. Wash the gel slices twice with 1mL water. Remove the water and add 1mL of 1:1 (50mM ammonium bicarbonate:acetonitrile) to each tube. Incubate samples at room temperature for 1h. Aspirate the solution and add 200µL acetonitrile to each tube. The gel slices turn opaque white.
4. Aspirate the acetonitrile and dry the gel slices further in a Speedvac. Rehydrate gel slices in 100µL of 4ng/µL sequencing-grade trypsin in 0.01% ProteaseMAX Surfactant:50mM ammonium bicarbonate. Add bicarbonate buffer as needed to ensure complete submersion of the gel slices. Incubate samples at 37°C for 18h.
5. Transfer the supernatant from each sample into a separate 1.5-mL tube. Process gel slices further with 200µL of 80:20 acetonitrile:1% formic acid. Combine

slices and corresponding supernatants. Dry the samples completely in a Speedvac.

6. Reconstitute tryptic peptide digests in 25 μ L 5% acetonitrile containing 0.1% (v/v) trifluoroacetic acid and separate on a NanoAcquity (Waters) UPLC. In brief, a 3.0- μ L injection is loaded in 5% acetonitrile containing 0.1% formic acid at 4.0 μ L/min onto a 100- μ m I.D. fused-silica precolumn packed with 2cm of 5 μ m (200Å) Magic C18AQ (Bruker-Michrom) and eluted using a gradient at 300nL/min onto a 75- μ m I.D. analytical column packed with 25cm of 3 μ m (100Å) Magic C18AQ particles to a gravity-pulled tip. The solvents are A, water (0.1% formic acid); and B, acetonitrile (0.1% formic acid). A linear gradient is developed from 5% solvent A to 35% solvent B in 60min. Ions are introduced by positive electrospray ionization via liquid junction into a Q Exactive hybrid mass spectrometer. Mass spectra are acquired over m/z 300–1750 at 70,000 resolution (m/z 200) and data-dependent acquisition selected the top 10 most abundant precursor ions for tandem mass spectrometry by HCD fragmentation using an isolation width of 1.6Da, collision energy of 27, and a resolution of 17,500.
7. Process the raw data files with Mascot Distiller (Matrix Science, version 2.6) prior to database searching with Mascot Server (version 2.6) against the Uniprot_Human database. Search parameters include trypsin specificity with two missed cleavages. The variable modifications of oxidized methionine, pyroglutamic acid for N-terminal glutamine, N-terminal acetylation of the protein, BP on tyrosine, and a fixed modification for carbamidomethyl cysteine are considered. For SILAC labels, the medium samples are labeled with Lys4 and Arg6 and the heavy samples are labeled with Lys8 and Arg10. The mass tolerances are 10ppm for the precursor and 0.05Da for the fragments. SILAC ratio quantitation is accomplished using Mascot Distiller, and the results from Mascot Distiller are loaded into the Scaffold Viewer for peptide/protein validation and SILAC label quantitation. For SILAC experiments, protein identification is subject to a two-peptide cut-off. For proteins detectable in the heavy sample but that lack an empirical heavy/light ratio value (due to low background detection in the light sample), peak areas of all the identified peptides in the Distiller file are used to calculate heavy/light ratios.

6. Further configuration of C-BERST

C-BERST can be successfully applied to identify protein factors associated with repetitive DNA elements using relatively few cells (~40 million) with high specificity compared to similar methods (Déjardin & Kingston, 2009; Garcia-Exposito et al., 2016; Liu et al., 2017; Schmidtmann et al., 2016; Tsui et al., 2018). The rapid kinetics of the APEX2 rapid system allow C-BERST to characterize a broad range of dynamic nuclear processes (e.g., cellular differentiation, extracellular stimulus response, and cell-cycle progression). Early attempts to use dCas9-APEX2-based proteomic mapping at single-copy loci (hTERT and c-Myc promoters) (Myers et al., 2018) required pooling of enriched factors from multiple 293T cell lines, each containing a unique sgRNA. To further improve the specificity of the proteomic

mapping, especially at nonrepetitive genomic elements, C-BERST will be explored with an sgRNA multiplexing strategy for improving method specificity by increasing the signal-to-noise ratio. Moreover, C-BERST and other labeling methods (such as BioID) can be used in tandem to improve specificities by diminishing the number of false negatives resulting from inefficient labeling due to differences in surface-accessible amino acid distribution or the suitability of certain peptides for MS analysis. Another potential path forward is to target loci with two orthogonal dCas9s, each tethered to one-half of a split labeling enzyme (Martell et al., 2016; Schopp et al., 2017). This would enable background labeling to be minimized.

7. Summary

Here, we have described a procedure to define the proteomic space surrounding distinct genomic loci using proximity biotinylation. Like other reverse-ChIP methods, C-BERST addresses some important pitfalls of traditional ChIP. Namely, no prior knowledge of nuclear localization or regulatory function is required to identify proteins associated with a target DNA element. C-BERST combines the flexible binding capacity of dCas9 with the rapid kinetics of APEX2 biotinylation for unbiased enrichment of a discrete subnuclear protein environment. A suite of nonspecific sgRNAs are readily available for definition of nuclear background proteins to which site-specific data are compared. While C-BERST has been demonstrated to provide high-resolution proteomic mapping of repetitive sequences using a single sgRNA, multiplexing of guides makes proximity biotinylation possible at single-copy loci as well. Using alternative guide expression vectors and CRISPR systems more amenable to multiplexing may further streamline C-BERST at these regions.

References

- Banaszynski LA, Chen L-C, Maynard-Smith LA, Ooi AGL, & Wandless TJ (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell*, 126(5), 995–1004. [PubMed: 16959577]
- Bonev B, & Cavalli G (2016). Organization and function of the 3D genome. *Nature Reviews Genetics*, 17(11), 661.
- Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, et al. (2018). Efficient proximity labeling in living cells and organisms with turboid. *Nature Biotechnology*, 36(9), 880–887.
- Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155(7), 1479–1491. [PubMed: 24360272]
- Chen B, Hu J, Almeida R, Liu H, Balakrishnan S, Covill-Cooke C, et al. (2016). Expanding the CRISPR imaging toolset with *Staphylococcus aureus* Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Research*, 44(8), e75. [PubMed: 26740581]
- Chen B, & Huang B (2014). Imaging genomic elements in living cells using CRISPR/Cas9. *Methods in Enzymology*, 546, 337–354. [PubMed: 25398348]
- Das AT, Tenenbaum L, & Berkhout B (2016). Tet-on systems for doxycycline-inducible gene expression. *Current Gene Therapy*, 16(3), 156–167. [PubMed: 27216914]
- Déjardin J, & Kingston RE (2009). Purification of proteins associated with specific genomic loci. *Cell*, 136(1), 175–186. [PubMed: 19135898]
- ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. [PubMed: 22955616]

- Fujita T, & Fujii H (2013). Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochemical and Biophysical Research Communications*, 439(1), 132–136. [PubMed: 23942116]
- Gao XD, Tu L-C, Mir A, Rodriguez T, Ding Y, Leszyk J, et al. (2018). C-BERST: Defining subnuclear proteomic landscapes at genomic elements with dCas9–APEX2. *Nature Methods*, 15(6), 433–436. [PubMed: 29735996]
- Garcia-Exposito L, Bournique E, Bergoglio V, Bose A, Barroso-Gonzalez J, Zhang S, et al. (2016). Proteomic profiling reveals a specific role for translesion DNA polymerase η in the alternative lengthening of telomeres. *Cell Reports*, 17(7), 1858–1871. [PubMed: 27829156]
- Gu B, Swigut T, Spencley A, Bauer MR, Chung M, Meyer T, et al. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science*, 359(6379), 1050–1055. [PubMed: 29371426]
- Hommelsheim CM, Frantzeskakis L, Huang M, & Ülker B (2014). PCR amplification of repetitive DNA: A limitation to genome editing technologies and many other applications. *Scientific Reports*, 4, 5052. [PubMed: 24852006]
- Hruz T, Wyss M, Docquier M, Pfaffl MW, Masanetz S, Borghi L, et al. (2011). RefGenes: Identification of reliable and condition specific reference genes for RT-qPCR data normalization. *BMC Genomics*, 12, 156. [PubMed: 21418615]
- Hung V, Zou P, Rhee H-W, Udeshi ND, Cracan V, Svinkina T, et al. (2014). Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Molecular Cell*, 55(2), 332–341. [PubMed: 25002142]
- Jinek M, East A, Cheng A, Lin S, Ma E, & Doudna J (2013). RNA-programmed genome editing in human cells. *eLife*, 2, e00471. [PubMed: 23386978]
- Lefever S, Vandesompele J, Speleman F, & Pattyn F (2009). RTPrimerDB: The portal for real-time PCR primers and probes. *Nucleic Acids Research*, 37(Database issue), D942–D945. [PubMed: 18948285]
- Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, et al. (2017). In situ capture of chromatin interactions by biotinylated dCas9. *Cell*, 170(5), 1028–1043.e19. [PubMed: 28841410]
- Martell JD, Yamagata M, Deerinck TJ, Phan S, Kwa CG, Ellisman MH, et al. (2016). A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nature Biotechnology*, 34(7), 774–780.
- Mortensen R, Chestnut JD, Hoeffler JP, & Kingston RE (1997). Selection of transfected mammalian cells. *Current Protocols in Neuroscience*, (1), 4–6.
- Myers SA, Wright J, Peckner R, Kalish BT, Zhang F, & Carr SA (2018). Discovery of proteins associated with a predefined genomic locus via dCas9–APEX-mediated proximity labeling. *Nature Methods*, 15(6), 437. 10.1038/s41592-018-0007-1. [PubMed: 29735997]
- Ong S-E, & Mann M (2007). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols*, 1, 2650.
- Rhee H-W, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, et al. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*, 339(6125), 1328–1331. [PubMed: 23371551]
- Roux KJ, Kim DI, Raida M, & Burke B (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of Cell Biology*, 196(6), 801–810. [PubMed: 22412018]
- Rusk N (2009). Reverse ChIP. *Nature Methods*, 6, 187.
- Schmidtman E, Anton T, Rombaut P, Herzog F, & Leonhardt H (2016). Determination of local chromatin composition by CasID. *Nucleus*, 7(5), 476–484. [PubMed: 27676121]
- Schopp IM, Amaya Ramirez CC, Debeljak J, Kreibich E, Skribbe M, Wild K, et al. (2017). Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nature Communications*, 8, 15690.
- Schwartz U, & Längst G (2016). Bioinformatic analysis of ChIP-seq data on the repetitive ribosomal RNA gene. In Németh A (Ed.), *The nucleolus: Methods and protocols* (pp. 225–230). New York, NY: Springer New York.

- Tsui C, Inouye C, Levy M, Lu A, Florens L, Washburn MP, et al. (2018). dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. *Proceedings of the National Academy of Sciences*, 115(12), E2734–E2741. 10.1073/pnas.1718844115.
- Zhang Z, Theurkauf WE, Weng Z, & Zamore PD (2012). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly (A) selection. *Silence*, 3(1), 9. [PubMed: 23273270]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1.

Diagram of the C-BERST workflow. Cells stably expressing sgRNA and inducible dSpyCas9-APEX2 are generated by lentiviral transduction. Following dox and Shield1 induction (21h), cells are incubated with biotin-phenol and then H₂O₂ to activate a burst of biotin-phenoxyl radical generation by dSpyCas9-APEX2, leading to proximity labeling of nearby proteins. Following quenching, nuclei isolation, and protein extraction, biotinylated proteins are enriched by streptavidin selection and analyzed by LC-MS/MS.

Fig. 2.

Inducible dSpyCas9-mCherry-APEX2 expression. The dSpyCas9-mCherry-APEX2 and sgRNA lentiviral expression constructs are shown. *Top*: dSpyCas9-mCherry-APEX2 under the control of the pCMV_TetO inducible promoter. The mCherry fusion is included to enable quantification of dSpyCas9 expression level as well as its subcellular localization. *DD*, Shield1-repressible degradation domain; *LTR*, long terminal repeat; *NLS*, nuclear localization signal. *Bottom*: sgRNA/TetR/BFP expression construct. *BFP*, blue fluorescent protein; *P2A*, 2A self-cleaving peptide; *pPGK*, PGK promoter; *pU6*, U6 promoter; *sgNS*, nonspecific sgRNA; *sgTelo*, telomere-targeting sgRNA; *tetR*, tet repressor.

Fig. 3.

Selection of stable cells. The mCherry- and BFP-positive cells are selected by FACS. The P1 population corresponds to high BFP (as a surrogate for sgRNA and TetR) and low mCherry expression, providing optimal signal-to-noise ratio to maximize the fraction of telomere-localized dSpyCas9-mCherry-APEX2. Three independent experiments were performed.

Fig. 4. Specific telomere targeting by dSpyCas9-mCherry-APEX2. Immunostaining of telomeric marker protein with primary anti-TERF2IP and secondary antibody conjugated with Alexa 488. Colocalization of dSpyCas9-mCherry-APEX2 foci with TERF2IP is observed ($n = 25$ cells examined). Scale bar, 5 μ m.

Fig. 5. dCas9-mCherry-APEX2 targets telomeres and enables restricted biotinylation of endogenous proteins. Fluorescence imaging of dSpyCas9-mCherry-APEX2 labeling in cells. Stable sgTelo and sgNS cells were labeled live or were only supplemented with H_2O_2 as a no-labeling control. Cells were then fixed and stained with neutravidin conjugated with OG488 to visualize biotinylated proteins. dCas9-mCherry-APEX2 localization is indicated by mCherry fluorescence. Two independent experiments were performed ($n = 25$ cells examined). Scale bar, $5\mu m$.