



Published in final edited form as:

Nat Methods. 2018 June ; 15(6): 433–436. doi:10.1038/s41592-018-0006-2.

C-BERST: Defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2

Xin D. Gao¹, Li-Chun Tu¹, Aamir Mir¹, Tomás Rodríguez¹, Yuehe Ding¹, John Leszyk^{2,4}, Job Dekker^{3,4,5}, Scott A. Shaffer^{2,4}, Lihua Julie Zhu^{6,7,8}, Scot A. Wolfe^{4,6}, and Erik J. Sontheimer^{1,8,*}

¹RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA 01605, U.S.A.

²Proteomics and Mass Spectrometry Facility, University of Massachusetts Medical School, Worcester, MA 01605, U.S.A.

³Program in Systems Biology, University of Massachusetts Medical School, Worcester, MA 01605, U.S.A.

⁴Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, U.S.A.

⁵Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, MA 01605, U.S.A.

⁶Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA 01605, U.S.A.

⁷Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, U.S.A.

⁸Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, U.S.A.

Abstract

Mapping proteomic composition at distinct genomic loci in living cells has been a long-standing challenge. Here we report that dCas9-APEX2 Biotinylation at genomic Elements by Restricted Spatial Tagging (C-BERST) allows the rapid, unbiased mapping of proteomes near defined

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>

*Correspondence: erik.sontheimer@umassmed.edu.

Note: Any Supplemental and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

X.D.G. and E.J.S. conceived the study. X.D.G., L.-C.T., J.D., S.A.W., and E.J.S. designed experiments. X.D.G. and T.R. performed C-BERST and ChIP-seq experiments, and J.L. conducted mass spectrometry procedures. X.D.G. and L.-C.T. processed fluorescence images, A.M. processed flow cytometry data, A.M. and T.R. processed ChIP-seq data, X.D.G., Y.D., and J.L. processed mass spectrometry data, and L.J.Z. conducted statistical analyses. All co-authors interpreted the data. X.D.G. and E.J.S. wrote the manuscript, and all authors revised and edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

genomic loci, as demonstrated for telomeres and centromeres. C-BERST enables the high-throughput identification of proteins associated with specific sequences, facilitating annotation of these factors and their roles in nuclear biology.

Chromosome organization is being defined at ever-increasing resolution through the use of Hi-C and related methods¹. Genome organization can also be analyzed in live cells by imaging, especially via fluorescent protein (FP) fusions to nuclease-dead *Streptococcus pyogenes* Cas9 (dSpyCas9), which can be directed to nearly any genomic region via single-guide RNAs (sgRNAs)². It has proven more difficult to map subnuclear proteomes onto 3-D genome landscapes in a comprehensive manner that avoids demanding fractionation protocols, specific DNA-associated protein fusions [e.g. in proximity-dependent biotin identification (BioID³)], or validated antibodies. dSpyCas9 has been combined with biotin ligases in approaches such as CasID⁴ and CAPTURE⁵ to allow the isolation of proteins associated with specific genomic regions in living cells. However, the efficiencies of these approaches usually necessitate long labeling times (hours), limiting the time resolution of dynamic processes.

Engineered ascorbate peroxidase (APEX2) has been used for an alternative live-cell biotinylation strategy called spatially restricted enzymatic tagging (SRET)^{6, 7}. In this approach, APEX2 is fused to a localized protein of interest, and cells are then treated with biotin-phenol (BP) and H₂O₂, generating a localized (~20nm radius) burst of diffusible but rapidly quenched biotin-phenoxyl radicals. These products react with electron-rich amino acid side chains (e.g. Tyr), leading to covalent biotinylation of nearby proteins and enabling identification by streptavidin selection and liquid chromatography/tandem mass spectrometry (LC-MS/MS). Notably, this method is extremely efficient (1 min H₂O₂ treatment). Based in part on the success of dSpyCas9-FP fusions in imaging, we reasoned that a dSpyCas9 derivative that emits radicals rather than photons could be used for rapid (1 min.) subnuclear proteomics. Here we use dSpyCas9-APEX2 fusions in the development of C-BERST (Fig. 1a) for genomic element-specific profiling of subnuclear proteomes in live cells.

To develop and validate this method, we used telomeres for benchmarking and proof-of-concept because they are associated with a well-defined suite of proteins and can be targeted by a well-established sgRNA (sgTelo)^{8, 9} in human U2OS cells. U2OS cells rely on alternative lengthening of telomeres (ALT) pathways to maintain telomere length without telomerase activation¹⁰. We transduced U2OS cells with a lentiviral vector expressing dSpyCas9 under the control of a tet-on CMV promoter and fused to nuclear localization signals (NLSs), a ligand-tunable degradation domain (DD)¹¹, mCherry, and APEX2 (Supplementary Fig. 1a). This combination allows control over dSpyCas9-mCherry-APEX2 protein levels (Supplementary Fig. 1b,c). mCherry-positive cells were then transduced with a lentiviral vector that encoded an sgRNA as well as a blue fluorescent protein (BFP) construct that also expresses the TetR repressor (Supplementary Fig. 1a). One sgRNA construct encodes sgTelo, and the other encodes a non-specific sgRNA (sgNS) whose target sequence is absent from the human genome¹². Sorted cells (Supplementary Fig. 1d and Supplementary Note 1) were confirmed to exhibit telomeric mCherry foci (Supplementary

Fig. 2), as well as telomeric biotinylation activity (Supplementary Fig. 3a and Supplementary Note 1), in the presence of sgTelo. ChIP-seq analysis of dCas9-mCherry-APEX2 confirmed sgTelo-directed localization to (TTAGGG)_n repeats (Supplementary Fig. 3b and Supplementary Note 1). These experiments indicate that sgTelo-guided dCas9-mCherry-APEX2 targets telomeres and enables restricted biotinylation of endogenous proteins.

For proteomic analysis, we induced APEX2 biotinylation with BP and H₂O₂ in the sgTelo and sgNS cells, and also included an sgTelo control from which H₂O₂ was omitted. Biotinylation of nucleoplasmic proteins in the sgNS sample serves as a reference, permitting an assessment of the telomere specificity of labeling with sgTelo. After nuclei isolation and streptavidin affinity purification, proteins were subjected to western blot and silver staining analyses to confirm dCas9-mCherry-APEX2 expression, biotinylation, and enrichment of biotinylated proteins (Supplementary Fig. 4a–c). Streptavidin-selected proteins were analyzed by LC-MS/MS. Using intensity-based absolute quantification (iBAQ) [a label-free quantification (LFQ) proteomic approach] values to measure enrichment in the sgTelo sample relative to sgNS [Benjamini-Hochberg (BH)-adjusted $p < 0.05$ and log₂ fold change (FC) ≥ 2.0], we found 30 out of 143 proteins that have been reported to be telomere-associated or otherwise linked to telomere function (Supplementary Fig. 4d, Supplementary Tables 1 and 2, and Supplementary Note 2). These results indicate that validated telomeric proteins can be identified rapidly and efficiently by C-BERST.

To further improve our assessments of differential C-BERST biotinylation, we used a more quantitative approach enabled by stable isotope labeling with amino acids in cell culture (SILAC). Telomere-targeted cells were cultured in heavy-isotope medium, sgNS cells in medium-isotope medium, and untransduced U2OS cells in light-isotope medium (Supplementary Fig. 5a,b). We performed biotinylation and purification as described above, except that equal amounts of protein lysates from heavy, medium, and light samples were mixed before streptavidin purification for three-state SILAC⁶. 913 proteins were identified in both the heavy and medium samples, and 885 of these were also detectable in the light sample. Using significance (BH-adjusted $p < 0.01$) and enrichment ($|\log_2$ fold change (FC) ≥ 2.5) cut-offs that were even more stringent than those used for LFQ, we identified 55 proteins that are strongly enriched in the sgTelo sample relative to sgNS (H/M) (Fig. 1b and Supplementary Table 3). Among these 55 hits, 34 are known telomere-associated factors, including all six shelterin components as well as subunits from 5 other complexes that contribute to ALT-associated pathways or processes (Supplementary Fig. 6a). Of the 55 H/M-enriched proteins, 54 were also strongly enriched (\log_2 FC ≥ 1) in H/L ratio, indicating that background detection in the absence of dCas9-mCherry-APEX2 biotinylation was minimal. Gene ontology (GO) analysis of the 55 H/M-enriched C-BERST hits reveals strong functional associations with terms such as telomere maintenance, homologous recombination, and DNA repair, all of which are important for ALT pathways¹⁰ (Supplementary Fig. 6b).

Telomere-associated proteomes from ALT+ cell lines have been defined previously by TERF1-BirA* BioID¹³ and proteomics of isolated chromatin segments (PICH)¹⁴. Over 50% of C-BERST IDs (Fig. 1c) were also detected by one or both of the other methods. The

remaining 23 proteins that were uniquely detected by C-BERST include seven known telomeric/ALT factors. Of the 18 proteins detected by all three approaches, 17 are known telomere-related factors. The remaining consensus hit [SLX4-interacting protein (SLX4IP)] was not previously validated as telomeric, but its identification by all three proteomic approaches strongly suggests that it has an unrecognized role in telomere function or maintenance (Fig. 1d). We used independent methods to confirm the telomere colocalization of SLX4IP, as well as a factor (RPA3) that was detected by C-BERST but missed by BioID and PICh (Fig. 1e, Supplementary Fig. 7, and Supplementary Note 3).

We extended C-BERST to centromeric alpha-satellite arrays in U2OS cells (Supplementary Fig. 8a) using a similar pipeline. The human alpha-satellite proteome from K562 cells has been analyzed by the PICh-related protocol HyCCAPP (hybridization capture of chromatin-associated proteins for proteomics)¹⁵, again enabling comparison. We first confirmed dCas9-mCherry-APEX2 inducible expression, specific centromere targeting¹⁶, and biotinylation (Supplementary Fig. 8b–d), and then used SILAC to identify 1,268 proteins (Supplementary Table 4) from each of two biological replicates. Among these 1,268 proteins, 460 were enriched to a statistically significant extent ($\log_2FC \geq 2.5$ and $p < 0.01$) in the sgAlpha vs. sgNS samples (H/M) (Fig. 2a). We identified subunits of the CENP-A nucleosome-associated complex¹⁷, CENP-A distal complex¹⁷, CENP-A loading factors¹⁸, chromosome passenger complex¹⁹, and other known centromere-associated proteins. 31 enriched proteins overlapped between C-BERST in U2OS cells and HyCCAPP in K562 cells (Supplementary Fig. 9a). C-BERST uniquely captured multiple centromeric factors including CENP-F and ATR, which were recently reported to engage RPA-coated centromeric R loops²⁰. GO analysis of the 460 C-BERST centromeric hits reveals strong functional associations with terms related to centromere maintenance or function¹⁰ (Supplementary Fig. 9b).

Our generation of both telomeric and centromeric C-BERST datasets affords the opportunity to compare protein enrichment at these two landmarks. 36 were identified in both (Fig. 2b and Supplementary Note 4). Significant GO terms for these 36 overlapping proteins include DNA replication and others that would be expected for both chromosomal elements. Significantly, all CENP factors were found among the 424 non-overlapping proteins from the sgAlpha dataset. Conversely, the 19 telomere-specific hits include five of the six shelterin subunits. These results provide strong evidence that C-BERST successfully profiles subnuclear proteomes enriched at distinct chromosomal elements.

By combining the flexibility of dSpyCas9 with the efficiency and rapid kinetics of APEX2 biotinylation, C-BERST promises to extend the unbiased definition of subnuclear proteomes to many other genomic elements, and to a range of dynamic processes (e.g. cellular differentiation, responses to extracellular stimuli, and cell cycle progression) that occur too rapidly to analyze via longer labeling procedures. C-BERST and BirA*-based methods favor biotinylation of distinct sets of proteins by virtue of their different labeling specificities; using these approaches in tandem would likely diminish the number of false negatives resulting from inefficient labeling due to differences in the surface-accessible amino acid distribution or the suitability of certain peptides for MS analysis. Importantly, C-BERST promises to augment and extend Hi-C and related methods by linking conformationally important cis-elements with their associated factors. Guide multiplexing should enable the

extension of C-BERST subnuclear proteomics to single-copy, non-repetitive loci. In the meantime, many types of repetitive elements within the genome play critically important roles in chromosome maintenance and function in ways that depend upon their associated proteins; C-BERST provides an unbiased method for defining subnuclear, locus-specific proteomes at such elements.

ONLINE METHODS

Construction of C-BERST plasmids

The Shield1- and doxycycline-inducible dSpyCas9-mCherry-APEX2 construct was made by subcloning Flag-APEX2 from Flag-APEX2-NES (Addgene 49386) into DD-dSpyCas9-mCherry²¹ using the pHAGE backbone. Two additional NLSs (SV40 and nucleoplasmin NLS) were inserted at each terminus to improve nuclear localization. The sgTelo-encoding construct was created by replacing the C3-guide RNA sequence (pCMV_C3-sgRNA_2XBroccoli/pPGK_TetR_P2A_BFP) with sgTelo sequences (using a plasmid provided by Hanhui Ma and Thoru Pederson). Non-specific sgRNA (sgNS)¹² and sgAlpha were constructed similarly. The sequences of the final constructs are provided in Supplementary Notes 5 and 6. SLX4IP-turboGFP plasmid was obtained from OriGene (catalog number: RG220896). RPA3-turboGFP was made by replacing the SLX4IP coding sequence with the human RPA3 coding sequence.

Cell culture and cell line construction

Human U2OS cells obtained from Thoru Pederson's lab (originally obtained from ATCC) were cultured in Dulbecco-modified Eagle's

Minimum Essential Medium (DMEM; Life Technologies) supplemented with 10% (vol/vol) FBS (Sigma). Lentiviral transduction was as described²¹. Six-fold higher titers of sgRNA-encoding lentiviruses were used for transduction relative to dSpyCas9-APEX2 lentivirus. Stably transduced cells were grown under the same conditions as the parental U2OS cells.

Flow cytometry

One day before performing FACS, dox (Sigma; 2 µg/ml) and Shield1 (Clontech; 250 nM) were added to the media. $\sim 2 \times 10^6$ cells expressing dSpyCas9-mCherry-APEX2 and BFP sgRNA were selected by FACS Aria cell sorter or analyzed with MacsQuant® VYB. Both instruments are equipped with 405- and 561-nm excitation lasers, and the emission signals were detected by using filters at 450/50 nm (wavelength/bandwidth) for BFP, and 610/20 nm (FACS Aria) or 615/20nm (MacsQuant) for mCherry. Bulk population and single cells (Supplementary Fig. 1b) were sorted into plates containing 1% GlutaMAX, 20% FBS, and 1% penicillin/streptomycin in DMEM medium.

Fluorescence microscopy

U2OS cells expressing sgRNA were seeded onto 170 µm, 35 × 10 mm glass-bottom dishes (Eppendorf) supplemented with dox and Shield1 21 hours before imaging. Live cells were imaged with a Leica DMi8 microscope equipped with a Hamamatsu camera (C11440-22CU), a 63× oil objective lens, and Microsystems software (LASX). Further

imaging processing was done with MetaMorph (Molecular Devices) or ImageJ (version: 2.0.0-rc-49/1.51d). Image contrast was set to ease visualization of cell, foci and nucleoplasmic background.

Immunofluorescence

Cells for immunofluorescence microscopy were grown on glass coverslips. The transfected cells or normal cells were fixed for 15 minutes in 2% paraformaldehyde in PHEM [0.05 M PIPES/0.05 M HEPES (pH 7.4), 0.01 M EGTA, 0.01 M MgCl₂], followed by a 2-minute extraction with 0.1% Triton X-100 in PHEM. After PBS washes, the cells were blocked by 1% BSA/1× TBST at 4°C overnight. Cells were first incubated with primary antibodies for two hours at room temperature and washed three times with blocking solution (10 minutes/wash). Cells were then incubated with secondary antibodies for one hour at room temperature, followed by another three blocking solution washes and two PBS washes²². Cells were mounted with ProLong antifade and visualized by fluorescence microscopy as described above. Neutravidin conjugated with OG488 experiment was described previously⁶. Image processing was done as described above.

C-BERST biotinylation

Six 15cm plates of U2OS cells ($\sim 6 \times 10^7$) expressing specific (sgTelo or sgAlpha) or nonspecific (sgNS) sgRNAs were used in this assay. Dox (2 µg/ml) and Shield1 (250 nM) were added 21 hours before biotinylation. Cells were then incubated with 500 µM biotin-phenol (BP) (Adipogen) for 30 minutes at 37°C. 1 mM H₂O₂ was then added to initiate of biotinylation for 1 minute on a horizontal shaker at room temperature. Six 15cm plates of sgTelo- or sgAlpha-expressing cells were treated in parallel, but without H₂O₂ addition, as a negative control. Quencher solution (5 mM trolox, 10 mM sodium ascorbate, and 10 mM sodium azide) was added to stop the reaction, and cells were washed five times (three quencher washes and two DPBS washes) to continue the quench and to remove excess BP.

Enrichment of biotinylated proteins

Cells were scraped off the plates and used for the preparation of isolated nuclei²³. Nuclei were washed with DPBS before lysis. RIPA lysis buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.125% SDS, 0.125% sodium deoxycholate and 1% Triton X-100 in Millipore water] with 1× freshly supplemented Halt Protease Inhibitor were used to lyse the cells for 10 minutes on ice. Cell lysates in 1.5 ml Eppendorf tubes were sonicated for 15 minutes with a Diagenode Bioruptor with 30 seconds on/off cycles at high intensity. Cell lysates were clarified by centrifugation at 13,000 rpm for 10 minutes. Clarified protein samples (~ 3.5 mg) were subjected to 400 µl Dynabeads MyOne Streptavidin T1 affinity purification overnight at 4°C. Each bead sample was washed with a series of buffers to remove non-specifically 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. Proteins were eluted in 70 µl 3× protein loading buffer supplemented with 2 mM biotin and 20 mM DTT with heating for 10 minutes at 95°C⁶. 50 µl eluents were loaded and run on a 4–12% SDS-PAGE gel (Bio-Rad) and run approximately 1cm off the loading well for in-gel digestion and LC-MS/MS analysis. All samples, including negative controls, contained ~ 75 kDa endogenously biotinylated proteins that are routinely detected in SRET-labeled

samples^{6,7}. The gel-fractionated sample used for LC-MS/MS (see below) corresponded to proteins from $\sim 4 \times 10^7$ cells.

Western blotting

Protein concentrations of the cell lysates were determined by BCA assay (Thermo). 50 μ g of each sample was mixed with protein loading buffer, boiled, and separated in SDS-PAGE gels. Proteins were transferred to PVDF membrane (Millipore), and blotted with Streptavidin-HRP (Thermo), or with anti-mCherry (Abcam) or anti-HDAC1 (Bethyl) antibodies. Additional details of the anti-SLX4IP and anti-RPA3 western analyses are described in the figure legends.

mCherry affinity purification of dSpyCas9-mCherry-APEX2 captured DNA and sequencing

1×10^7 U2OS cells stably expressing dCas9-mCherry-APEX2 transduced with sequence-targeting or non-specific sgRNAs were washed with PBS, fixed with 1% formaldehyde for 10 minutes and quenched with 0.125 M glycine for 5 minutes. Cells were harvested using a plate scraper and lysed in RIPA cell lysis buffer [50 mM Tris-HCl (pH 7.5), 150 mM 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 10 minutes on ice. Cell lysates were centrifuged at $2,300 \times g$ for 5 minutes at 4°C to isolate nuclei. Nuclei were suspended in 500 μ l of RIPA nuclear lysis buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5% SDS, 0.125% sodium deoxycholate and 1% Triton X-100 in Millipore water] with $1 \times$ freshly supplemented Halt Protease Inhibitor and subjected to sonication to shear chromatin fragments to an average size of 200–500 bp on a Diagenode Bioruptor with 30 seconds on/off cycles at high intensity for 15 minutes. Fragmented chromatin was centrifuged at $16,100 \times g$ for 10 minutes at 4°C . 450 μ l of supernatant was transferred to a new microcentrifuge tube. 4 μ g anti-mCherry antibody (Thermo PA5-34974) was added to each sample and incubated at 4°C for 3 hours. 50 μ l of blocked Protein G Dynabeads (Thermo 10003D) was added to each sample and rotated at 4°C overnight. After overnight incubation, Dynabeads were washed seven times as described above for selection of biotinylated proteins. Chromatin was eluted from Dynabeads in 200 μ l elution buffer [50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS] and transferred to a new microcentrifuge tube. Eluted chromatin was treated with 1 μ l RNase A and incubated overnight at 65°C to reverse crosslinks. 7.5 μ l of 20 mg/ml proteinase K was added to each sample followed by incubation for 2 hours at 50°C . ChIP DNA was then incubated with 1 ml Buffer PB (QIAGEN) and 10 μ l of 3M sodium acetate pH 5.2 at 37°C for 30 minutes. DNA was purified using QIAGEN quickspin column.

15 ng of ChIP DNA was processed for library preparation using the NEBNext ChIP-seq Library Prep Kit (New England Biolabs) according to the manufacturer's protocol.

15 ng of ChIP DNA was end-repaired using NEBNext End Repair module (NEB Cat. E6050) and purified with $1.8 \times$ AMPure XP beads (Beckman-Coulter Cat. A63880). End-repaired DNA was processed in a dA-tailing reaction using NEBNext dA-Tailing module (NEB Cat. E6053) and purified with $1.8 \times$ AMPure XP beads. Adaptor oligos 1 (5'-pGAT CCG AAG AGC ACA CGT CT-3') and 2 (5'-ACA CTC TTT CCC TAC ACG ACG CTC

TTC CGA TCT-3') used in Y-shaped adapter mix were ligated to dA-tailed DNA according to ref.²⁴ and purified with 1.5× AMPure XP beads. Ligated DNA was incubated in a thermal cycler (98°C for 40s, 65°C for 30s, and 72°C 30s) with one of the Illumina barcode primers and NEB Q5 Polymerase Master Mix. Primer 1 (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T-3') was added to mix for 10 cycles (98°C for 10s, 65°C for 30s, 72°C for 30s), followed by incubation at 72°C for 3 minutes. PCR-enriched DNA was purified with 1× AMPure XP beads.

Raw Illumina sequencing reads of 150 nucleotide length were processed as fastq files in R. Reads were trimmed using the Bioconductor ShortRead R package at positions which contained 2 nucleotides in a 5-nucleotide bin with a quality encoding less than phred score = 20. Reads with at least one (TTAGGG)₄ or (CCCTAA)₄ segment constituted a "hit", and were counted using the Bioconductor Biostrings R package. (number of hits / total trimmed reads) was calculated to assess the specificity of Cas9-mCherry-APEX2 for each sample.

SILAC labeling

On day 0, early-passage, sorted, stably transduced sgTelo or sgAlpha U2OS cells were grown in heavy SILAC media, which contained L-arginine-¹³C₆, ¹⁵N₄ (Arg10) and L-lysine-¹³C₆, ¹⁵N₂ (Lys8) (Sigma). Stable sgNS cells were grown in medium SILAC media, which contained L-arginine-¹³C₆ (Arg6) and L-lysine-4,4,5,5-*d*₄ (Lys4) (Sigma). Untransduced U2OS cells were grown in light SILAC media, which contained L-arginine (Arg0) and L-lysine (Lys0) (Sigma). Cells were grown for more than 10 days (>5 passages) to allow for sufficient incorporation of the isotopes. On day 11, dox and Shield1 were added to each isotope culture (4 15cm plates for each cell line) 21h before BP and H₂O₂ treatment. The biotinylation, nuclei isolation, and cell lysis followed the procedure described above. Before streptavidin affinity purification, equal amount of proteins measured by Pierce™ BCA Protein Assay Kit (~1mg from each isotope sample) were mixed in a 1:1:1 ratio (H:M:L). Streptavidin affinity purification and sample wash were described above. Proteins were eluted in 50 μl 3× protein loading buffer supplemented with 2 mM biotin and 20 mM DTT with heating for 10 minutes at 65°C. 50 μl eluents were loaded and run approximately to the center of the lane on a 4–12% SDS-PAGE gel (Bio-Rad). The coomassie-stained protein bands were excised and cut to five slices for in-gel digestion and LC-MS/MS analysis.

LC-MS/MS and proteomic analyses for LFQ

Unresolved protein bands from SDS-PAGE were cut into 1×1 mm pieces and placed in 1.5ml Eppendorf tubes with 1ml of water. After 30 minutes, water was removed and replaced with 70 μl of 250 mM ammonium bicarbonate. Proteins were then reduced by the addition of 20 μl of 45 mM 1,4-dithiothreitol (DTT), incubated at 50°C for 30 minutes, cooled to room temperature, alkylated with 20 μl of 100 mM iodoacetamide for 30 minutes, and washed twice with 1 ml water. The water was removed and replaced with 1 ml of 50 mM ammonium bicarbonate: acetonitrile (1:1) and incubated at room temperature for 1 hour. The solvent was then replaced with 200 μl acetonitrile, removed, and the pieces dried in a Speed Vac (Savant Instruments, Inc.). Gel pieces were then rehydrated in 75 μl of 4 ng/μl sequencing-grade trypsin (Promega) in 0.01% ProteaseMAX Surfactant (Promega) in

50 mM ammonium bicarbonate and incubated at 37°C for 21 hour. The supernatant was then removed to a 1.5 ml Eppendorf tube, the gel pieces further dehydrated with 100 µl of acetonitrile: 1% (v/v) formic acid (4:1), and the combined supernatants dried on a Speed Vac. Peptides were then reconstituted in 25 µl of 5% acetonitrile containing 0.1% (v/v) trifluoroacetic acid for LC-MS/MS.

Samples were analyzed on a NanoAcquity UPLC (Waters Corporation) coupled to a Q Exactive (Thermo Fisher Scientific) hybrid mass spectrometer. In brief, 1.0 µl aliquots were loaded at 4 µl/minute onto a custom-packed fused silica precolumn (100 µm ID) with Kasil frit containing 2 cm Magic C18AQ (5µm, 100Å) particles (Bruker Corporation). Peptides were then separated on a 75µm ID fused silica analytical column containing 25 cm Magic C18AQ (3µm, 100Å) particles (Bruker) packed in-house into a gravity-pulled tip. Peptides were eluted at 300 nl/minute with a linear gradient from 95% solvent A (0.1% (v/v) formic acid in water) to 35% solvent B (0.1% (v/v) formic acid in acetonitrile) in 60 minutes. Data was acquired by data-dependent acquisition according to a published method²⁵. Briefly, MS scans were acquired from m/z 300–1750 at a resolution of 70,000 (m/z 200) and followed by ten tandem mass spectrometry scans using HCD fragmentation using an isolation width of 1.6 Da, a collision energy of 27%, and a resolution of 17,500 (m/z 200). Raw data files were processed with Proteome Discoverer (Thermo, version 2.1.1.21) and searched with Mascot (Matrix Science, version 2.6) against the SwissProt *Homo sapiens* database. Search parameters used tryptic specificity considering up to 2 missed cleavages, a parent mass tolerance of 10 ppm, and a fragment mass tolerance of 0.05 Da. Fixed modification of carbamidomethyl cysteine was considered as were variable modifications of N-terminal acetylation, N-terminal conversion of Gln to pyroGlu, oxidation of methionine, and biotin-phenol conjugation of tyrosine. Results were loaded into Scaffold (Proteome Software Inc., version 4.8.4) for peptide and protein validation and quantitation using the Peptide Prophet and Protein Prophet algorithms^{26, 27}. The threshold for peptides was set to 80% (1.1% FDR) and 90% for proteins (3-peptide minimum). Contaminants such as human keratin were included in all statistical analyses and removed from the figures.

LC-MS/MS and proteomic analyses for SILAC

A fully resolved SDS-PAGE was cut into 5 fractions and each fraction was processed separately as described. Gel bands were cut into 1×1 mm pieces and placed in 1.5 mL Eppendorf tubes with 1mL of water for 30 minutes. The water was removed and 200µl of 250 mM ammonium bicarbonate was added. For reduction, 20 µl of a 45 mM solution of 1,4-dithiothreitol was added and the samples were incubated at 50°C for 30 minutes. The samples were cooled to room temperature and then, for alkylation, 20 µl of a 100 mM iodoacetamide solution was added and allowed to react for 30 minutes. The gel slices were washed twice with 1 mL water. The water was removed and 1mL of 50:50 (50 mM ammonium bicarbonate:acetonitrile) was placed in each tube and samples were incubated at room temperature for 1 hour. The solution was then removed and 200 µl of acetonitrile was added to each tube, at which point the gels slices turned opaque white. The acetonitrile was removed and gel slices were further dried in a Speed Vac. Gel slices were rehydrated in 100 µl of 4ng/µl of sequencing-grade trypsin in 0.01% ProteaseMAX Surfactant:50 mM ammonium bicarbonate. Additional bicarbonate buffer was added to ensure complete

submersion of the gel slices. Samples were incubated at 37°C for 18 hours. The supernatant of each sample was then removed and placed in a separate 1.5 mL Eppendorf tube. Gel slices were further extracted with 200 μ L of 80:20 (acetonitrile:1% formic acid). The extracts were combined with the supernatants of each sample. The samples were then completely dried down in a Speed Vac.

Tryptic peptide digests were reconstituted in 25 μ L 5% acetonitrile containing 0.1% (v/v) trifluoroacetic acid and separated on a NanoAcquity UPLC. In brief, a 3.0 μ L injection was loaded in 5% acetonitrile containing 0.1% formic acid at 4.0 μ L/min for 4.0 minute onto a 100 μ m I.D. fused-silica pre-column packed with 2 cm of 5 μ m (200Å) Magic C18AQ and eluted using a gradient at 300 nL/minute onto a 75 μ m I.D. analytical column packed with 25 cm of 3 μ m (100Å) Magic C18AQ particles to a gravity-pulled tip. The solvents were A, water (0.1% formic acid); and B, acetonitrile (0.1% formic acid). A linear gradient was developed from 5% solvent A to 35% solvent B in 60 minutes. Ions were introduced by positive electrospray ionization via liquid junction into a Q Exactive hybrid mass spectrometer (Thermo). Mass spectra were 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 1.6 Da, collision energy of 27, and a resolution of 17,500.

Raw data files were peak processed with Mascot Distiller prior to database searching with Mascot Server against the *Uniprot_Human* database. Search parameters included trypsin specificity with two missed cleavages. The variable modifications of oxidized methionine, pyroglutamic acid for N-terminal glutamine, N-terminal acetylation of the protein, biotin-phenol on tyrosine and a fixed modification for carbamidomethyl cysteine were considered. For SILAC labels, the medium samples were labeled with Lys4 and Arg6 and the heavy samples were labeled with Lys8 and Arg10. The mass tolerances were 10 ppm for the precursor and 0.05 Da for the fragments. SILAC ratio quantitation was accomplished using Mascot Distiller and the results from Mascot Distiller were loaded into the Scaffold Viewer for peptide/protein validation and SILAC label quantitation. For SILAC experiments, protein identification was subject to a two-peptide cut-off. For proteins detectable in the H sample but that lack an empirical H/L ratio value (due to low background detection in the L sample), peak areas of all the identified peptides in the Distiller file were used to calculate H/L ratios.

Data analysis

Data was first filtered to exclude proteins detected in only one of the dCas9-mCherry-APEX2/sgTelo (+BP, +H₂O₂) (“S1”) replicates, followed by log₂ transformation. Prior to the log₂ transformation, iBAQ values of 0 were replaced with the smallest iBAQ value from the corresponding sample in dCas9-mCherry-APEX2/sgTelo (+BP, -H₂O₂) (“S2”) or dCas9-mCherry-APEX2/sgNS (+BP, +H₂O₂) (“S3”) to avoid generation of infinite ratios. Moderated t-test with a paired design was used to compare the log₂ transformed iBAQ values between S1 and S3, S1 and S2, and S2 and S3 using limma package²⁸. To adjust for multiple comparisons, *p* values were adjusted using the Benjamini-Hochberg (BH) method²⁹. Proteins were selected for subsequent analysis if they were (*i*) significantly

enriched in both S1 vs. S3 and S1 vs. S2, (*ii*) not enriched in S2 vs. S3, and (*iii*) if S1/S3 and S1/S2 ratios were greater than 2.

Similarly, SILAC datasets were filtered to exclude proteins with H/M ratios detected in only one of the biological replicates. Detection in a biological replicate required identification in at least two of the three technical replicates that were done for each biological replicate; median values from the technical replicates were used for subsequent analyses. Proteins with BH-adjusted *p* values less than 0.05 (moderated t-test described above) are considered statistically significant. Proteins (with BH-adjusted *p* values < 0.01 and log₂ fold change 2.5) were selected for subsequent GO (David Bioinformatics) and overlap analysis. To determine whether the proteins identified in this experiment overlap significantly with three published datasets, hypergeometric test was used. Hypergeometric test was also used for testing the overlapping proteins between C-BERST telomere IDs and centromere IDs.

Protocols can also be accessed³⁰ at Protocol Exchange with digital object identifier (doi) 10.1038/protex.2018.036.

Life Sciences Reporting Summary

Further information on experimental design is available in the **Life Sciences Reporting Summary**.

Data availability

Mass spectrometry data that support the findings of this study have been deposited into the EMBO PRIDE archive with the dataset identifier 10.6019/PXD009216. Source data for all the graphical representations reported in the manuscript have been provided in Supplementary Table 6. All other data support the findings of this study are available from the corresponding authors on request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to all members of the Sontheimer, Wolfe and Dekker labs for advice and discussions, Tom Fazzio, Samyabrata Bhaduri and Michael Green for helpful feedback, Hanhui Ma, Tong Wu, David Grünwald, and Thoru Pederson for reagents, the Flow Cytometry Core Facility at UMass Medical School for cell sorting, and Lingji Zhu for assistance with figure preparation. This work was supported by 4D Nucleome grant U54 DK107980 from the National Institutes of Health to J.D., S.A.W. and E.J.S.

References

1. Davies JO, Oudelaar AM, Higgs DR, Hughes JR. How best to identify chromosomal interactions: a comparison of approaches. *Nat Methods*. 2017; 14:125–134. [PubMed: 28139673]
2. Dominguez AA, Lim WA, Qi LS. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol*. 2016; 17:5–15. [PubMed: 26670017]
3. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol*. 2012; 196:801–810. [PubMed: 22412018]

4. Schmidtman E, Anton T, Rombaut P, Herzog F, Leonhardt H. Determination of local chromatin composition by CasID. *Nucleus*. 2016; 7:476–484. [PubMed: 27676121]
5. Liu X, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell*. 2017; 170:1028–1043 e1019. [PubMed: 28841410]
6. Hung V, et al. Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol Cell*. 2014; 55:332–341. [PubMed: 25002142]
7. Rhee HW, et al. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*. 2013; 339:1328–1331. [PubMed: 23371551]
8. Chen B, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*. 2013; 155:1479–1491. [PubMed: 24360272]
9. Ma H, et al. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci U S A*. 2015; 112:3002–3007. [PubMed: 25713381]
10. Cesare AJ, Reddel RR. Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet*. 2010; 11:319–330. [PubMed: 20351727]
11. Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell*. 2006; 126:995–1004. [PubMed: 16959577]
12. Knight SC, et al. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*. 2015; 350:823–826. [PubMed: 26564855]
13. Garcia-Exposito L, et al. Proteomic Profiling Reveals a Specific Role for Translesion DNA Polymerase eta in the Alternative Lengthening of Telomeres. *Cell Rep*. 2016; 17:1858–1871. [PubMed: 27829156]
14. Dejardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. *Cell*. 2009; 136:175–186. [PubMed: 19135898]
15. Buxton KE, et al. Elucidating Protein-DNA Interactions in Human Alphoid Chromatin via Hybridization Capture and Mass Spectrometry. *J Proteome Res*. 2017; 16:3433–3442. [PubMed: 28704058]
16. Chen B, et al. Expanding the CRISPR imaging toolset with *Staphylococcus aureus* Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Res*. 2016; 44:e75. [PubMed: 26740581]
17. Foltz DR, et al. The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol*. 2006; 8:458–469. [PubMed: 16622419]
18. Verdaasdonk JS, Bloom K. Centromeres: unique chromatin structures that drive chromosome segregation. *Nat Rev Mol Cell Biol*. 2011; 12:320–332. [PubMed: 21508988]
19. Carmena M, Wheelock M, Funabiki H, Earnshaw WC. The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. *Nat Rev Mol Cell Biol*. 2012; 13:789–803. [PubMed: 23175282]
20. Kabeche L, Nguyen HD, Buisson R, Zou L. A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. *Science*. 2018; 359:108–114. [PubMed: 29170278]
21. Ma H, et al. CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *J Cell Biol*. 2016; 214:529–537. [PubMed: 27551060]
22. Follit JA, Tuft RA, Fogarty KE, Pazour GJ. The intraflagellar transport protein IFT20 is associated with the Golgi complex and is required for cilia assembly. *Mol Biol Cell*. 2006; 17:3781–3792. [PubMed: 16775004]
23. Nagano T, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. 2013; 502:59–64. [PubMed: 24067610]
24. Zhang Z, Theurkauf WE, Weng Z, Zamore PD. Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence*. 2012; 3:9. [PubMed: 23273270]
25. Kelstrup CD, et al. Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J Proteome Res*. 2014; 13:6187–6195. [PubMed: 25349961]

26. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem.* 2002; 74:5383–5392. [PubMed: 12403597]
27. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem.* 2003; 75:4646–4658. [PubMed: 14632076]
28. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Molec Biol.* 2004; 3:1–25.
29. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 1995; 57:289–300.
30. Gao XD, et al. C-BERST: Defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nature Protocols.* 2018; doi: 10.1038/protex.2018.036

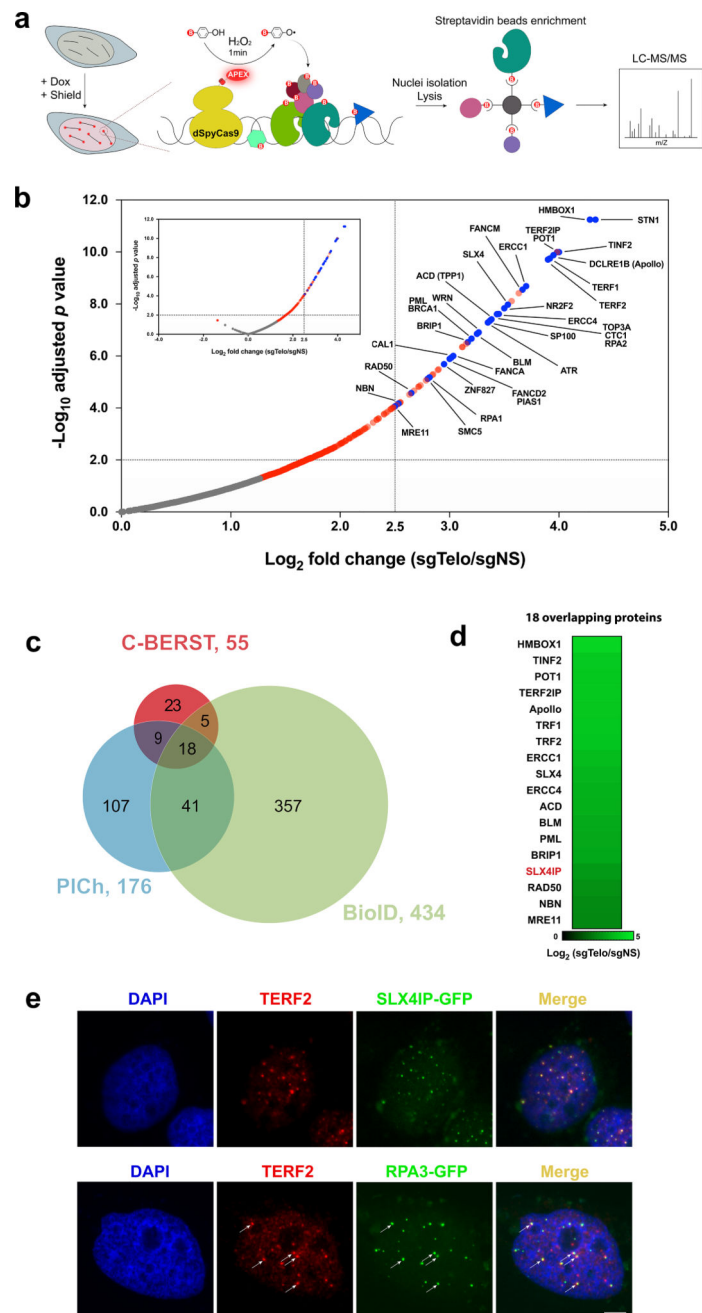


Figure 1. Using C-BERST to biotinylate telomere-associated proteins in living human cells. **(a)** Diagram of the C-BERST workflow. **(b)** Telomere-associated proteome identification by ratiometric C-BERST. A volcano plot is shown for C-BERST-labeled, telomere-associated proteins in U2OS cells. For each protein, the H/M SILAC ratio reflects the enrichment of identified proteins in sgTelo vs. sgNS cells. 359 proteins (indicated by blue and red dots) are statistically enriched [BH-adjusted p value < 0.05] with sgTelo, relative to sgNS controls. 55 proteins fall above a more stringent cut-off based on FDR and enrichment level (BH-adjusted p value < 0.01 and \log_2 FC ≥ 2.5). The 34 proteins indicated by blue dots (with

identities provided) are previously defined as either telomere-associated proteins or ALT pathway components. The volcano plot shows 97.7% of identified proteins (inset shows all proteins, including the few with SILAC H/M \log_2 ratios < 0). Two independent experiments were performed. (c) Venn diagram of statistically enriched (BH-adjusted p value < 0.01) telomeric proteins from ALT+ human cells, as detected by C-BERST (red), PICH (purple), and TERF1-BirA* BioID (green). 32 proteins from the C-BERST proteome were also detected by PICH, BioID, or both. (d) The 18 proteins found by telomeric C-BERST, BioID, and PICH are highly enriched in the C-BERST telomere proteome. To our knowledge, SLX4IP (denoted in red) has not been validated previously as telomere- or ALT-associated. (e) Colocalization of turboGFP-tagged SLX4IP and RPA3 with the telomeric marker protein TERF2. $\sim 0.6 \times 10^5$ U2OS cells were transiently transfected with 100ng SLX4IP-GFP expression plasmid or 50ng RPA3-GFP expression plasmid. Cells were then fixed and incubated with TERF2 primary antibody and secondary antibody conjugated with Alexa Fluor 647. DAPI stained cells were imaged (n = 20 cells examined). Representative images were from two independent experiments. Scale bar, 5 μ m.

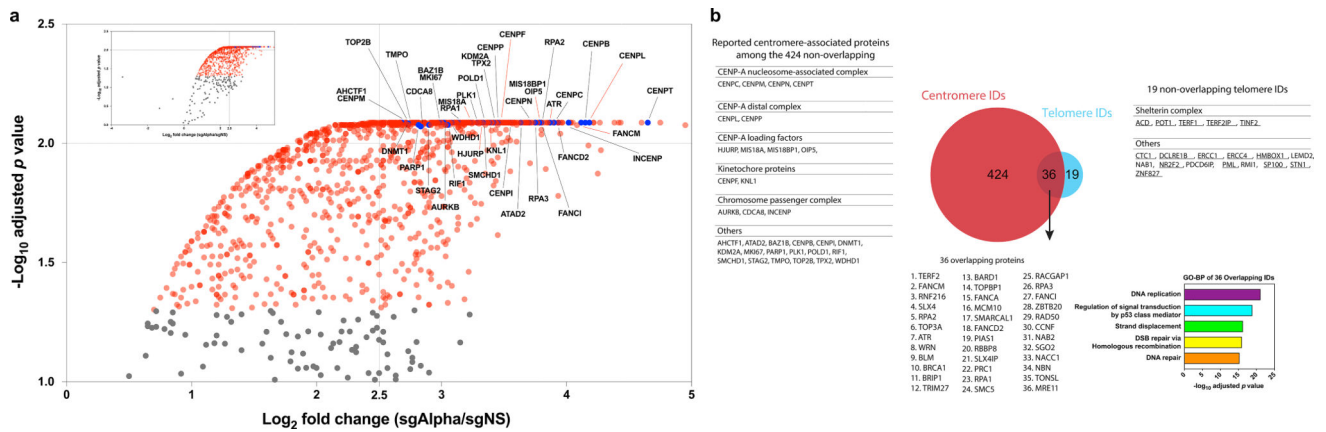


Figure 2. Successful capture of alpha-satellite-associated proteomes in live human cells by C-BERST. **(a)** Ratiometric C-BERST (using SILAC) was used to profile the alpha-satellite-associated proteome. A volcano plot of C-BERST-labeled, centromere-associated proteins in U2OS cells is shown. For each protein, the H/M SILAC ratio reflects the enrichment of identified proteins in sgAlpha vs. sgNS cells. 1,134 proteins (indicated by blue and red dots) are statistically enriched [BH-adjusted p value < 0.05] in the sgAlpha sample, relative to sgNS controls. 460 proteins fall above a more stringent cut-off based on FDR and enrichment level (BH-adjusted p value < 0.01 and \log_2 fold change ≥ 2.5). The 40 proteins indicated by blue dots (with identities provided) were previously defined as either centromere-associated (Supplementary Table 5) or were reported as components of the HyCCAPP centromere proteome (see text). The nine known centromere-associated proteins indicated by red lines are uniquely captured by C-BERST. The volcano plot shows 96.2% of identified proteins (inset shows all proteins, including the few with SILAC H/M ratio < 0 or $-\log_{10}$ adjusted p value < 1). Two independent experiments were performed. **(b)** Venn diagram of 55 C-BERST ALT/telomeric IDs and 460 centromeric IDs. 19 non-overlapping telomere IDs are listed on the right, with known ALT/telomeric proteins underlined. Among the 424 non-overlapping centromere IDs, 33 (listed on the left) are known or implicated as centromeric proteins. 36 overlapping proteins from both sets are listed below, as indicated. The five most significant GO-BP terms for the 36 overlapping ID are provided on the lower right.