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

Xin D. Gao[1], Li-Chun Tu[1], Aamir Mir[1], Job Dekker[2,3,4], Scott A. Shaffer[3,5], Lihua Julie Zhu[6,7,8], Scot A. Wolfe[3,6], and Erik J. Sontheimer[1,8,*]

[1]RNA Therapeutics Institute
[2]Program in Systems Biology
[3]Department of Biochemistry and Molecular Pharmacology
[4]Howard Hughes Medical Institute
[5]Proteomics and Mass Spectrometry Facility
[6]Department of Molecular, Cell and Cancer Biology
[7]Program in Bioinformatics and Integrative Biology
[8]Program in Molecular Medicine

University of Massachusetts Medical School
Worcester, MA 01605
U.S.A.

*Correspondence: erik.sontheimer@umassmed.edu

Mapping proteomic composition at distinct genomic loci and subnuclear landmarks 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 unbiased mapping of proteomes near defined genomic loci, as demonstrated for telomeres. C-BERST enables the high-throughput identification of proteins associated with specific sequences, facilitating annotation of these factors and their roles in nuclear and chromosome biology.

Three-dimensional organization of chromosomes is being defined at ever-increasing resolution through the use of Hi-C and related high-throughput methods[1]. Genome organization can also be analyzed in live cells by fluorescence 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 (BioID[3])], or validated antibodies. dSpyCas9 has enabled a BioID-derived subnuclear proteomic technique called CasID[4], in which biotin ligase (BirA*) fusion to dSpyCas9 allows proteins associated with specific genomic regions to be biotinylated on neighboring, exposed lysine residues in live cells. Streptavidin affinity selection and liquid chromatography/tandem mass spectrometry (LC-MS/MS) is then used to identify the tagged proteins. However, this approach is relatively inefficient and requires long (18-24h) labeling times, leading to increased background levels and limited 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)[5, 6]. In this approach, APEX2 is fused to a localized protein of interest, and cells are then treated with biotin-phenol and H_2O_2, generating a localized (within a ~20nm radius) burst of diffusible but rapidly quenched biotin-phenoxyl radicals. These products react with electron-rich amino acid side chains (e.g. Tyr, Trp, His and Cys), leading to covalent biotinylation of proteins in the vicinity of the localized APEX2, thus allowing subsequent identification by streptavidin selection and LC-MS/MS. Notably, this subcellular tagging method is extremely efficient (1 min H_2O_2 treatment),

allowing temporal control over the labeling process. Based in part on the success of dSpyCas9-FP fusions in enabling subnuclear imaging in living cells, we reasoned that a dSpyCas9 derivative that emits radicals rather than photons could be used for subnuclear proteomic analyses in a manner that overcomes the unfavorable kinetics and high background of CasID. 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 sought genomic elements that are associated with a well-defined suite of known protein factors, and that can be bound with dSpyCas9 with high efficiency and specificity using an established sgRNA. For this purpose, we chose to target telomeres in human U2OS cells. As with ~10-15% of cancer cell types, U2OS cells rely on alternative lengthening of telomeres (ALT) pathways to maintain telomere length without telomerase activation[7]. Cohorts of proteins associated with telomeres in ALT+ cells are well-characterized and map to key pathways such as homologous recombination (HR) and break-induced telomere synthesis[8]. Furthermore, an sgRNA (sgTelo) has already been established for efficient telomere association of dSpyCas9[9, 10].

We transduced U2OS cells with a lentiviral vector expressing dSpyCas9 under the control of a tet-on CMV promoter and fused to five nuclear localization signals (NLSs), a ligand-tunable degradation domain (DD)[11], mCherry, and APEX2 (**Fig. 1b**). Maximal expression of this dSpyCas9-mCherry-APEX2 fusion protein requires not only doxycycline (dox) but also the Shield1 ligand to inactivate the DD in a dose-dependent fashion[12]. This combination allows precise control over dSpyCas9-mCherry-APEX2 protein levels for optimal signal-to-noise levels. mCherry-positive cells [collected by fluorescence-activated cell sorting (FACS)] were then transduced with a separate lentiviral vector that included an sgRNA construct (driven by the U6 promoter) as well as a blue fluorescent protein (BFP) construct that also expresses the TetR repressor (**Fig. 1b**). In one version of this construct the sgRNA cassette encodes sgTelo (for labeling telomeres), and in the other it encodes a non-specific sgRNA (sgNS) that is complementary to a bacteriophage-derived sequence that is absent from the human genome[13]. After 21h of dox and Shield1 induction, we again used fluorescence-activated sorting (FACS) to sort four distinct BFP/mCherry double-positive cell populations (P1-P4) that correlate with different expression levels of dSpyCas9-mCherry-APEX and BFP (as a surrogate for sgRNA and TetR) (**Fig. 1c** and **Supplementary Fig. 1**). We reasoned that our signal-to-noise ratio of

102 telomeric vs. non-telomeric biotinylation would be maximized when sgTelo levels are saturating,
103 and when dSpyCas9-mCherry-APEX2 levels are limiting (relative to potential genomic binding
104 sites). Both conditions are expected to favor maximal partitioning of the sgRNA-programmed
105 dSpyCas9-mCherry-APEX into the desired telomere-associated state, with as little unlocalized or
106 mislocalized fusion protein as possible. Visual inspection by fluorescence microscopy (**Fig. 1d**)
107 confirmed that the sgTelo P1 cell population (with higher BFP expression and lower mCherry
108 expression) exhibited the most robust mCherry-labelled telomeric foci with the lowest amount of
109 nucleolar or diffuse nucleoplasmic background[9]. We therefore used this population in our
110 subsequent experiments. Diffuse nucleoplasmic labeling, as opposed to distinct foci, was observed
111 in the sgNS P1 population (**Fig. 1d**). APEX2-catalyzed biotinylation of nucleoplasmic proteins in
112 the sgNS control sample serves as a reference, permitting an assessment of the telomere
113 specificity of labeling in the sgTelo sample.

114 We induced APEX2-catalyzed biotinylation with biotin-phenol and H_2O_2 in the sgTelo
115 and sgNS P1 cells, and also included an sgTelo control in which the H_2O_2 was omitted. Nuclei
116 were isolated from these cells (to reduce cytoplasmic background), and nuclear proteins were
117 then extracted. Recovered proteins (50 µg) were subjected to western analysis using streptavidin-
118 conjugated horseradish peroxidase (streptavidin-HRP) (**Fig. 1e**), as well as total protein
119 visualization by Coomassie staining (**Fig. 1f**). Samples were also probed with anti-mCherry
120 antibodies (to detect dSpyCas9-mCherry-APEX2) and HDAC1 (as a loading control) (**Fig. 1e**,
121 bottom). Biotinylated proteins were readily detected in both sgTelo and sgNS samples, but were
122 largely absent in the -H_2O_2 control (**Fig. 1e**). In the sgNS sample, anti-mCherry and
123 streptavidin-HRP signals were less intense in comparison with the sgTelo sample, indicating that
124 dSpyCas9-mCherry-APEX2 accumulation and activity are lower in the former. Biotinylated
125 proteins were then isolated using streptavidin beads and analyzed by SDS-PAGE and silver
126 staining (**Fig. 1g**). Aside from the ~75kDa endogenously biotinylated proteins routinely detected
127 in SRET-labeled samples[5, 6], only background levels of proteins were detected in the no-H_2O_2
128 control sample, indicating successful purification. All three samples were subjected to in-gel
129 trypsin digestion followed by LC-MS/MS to identify the biotin-labeled proteins. These analyses
130 were done with two biological replicates prepared on different days.

131 The two sgTelo replicates yielded at least two unique peptides from 1,106 and 958
132 proteins, >80% of which (876) were detected in both. For these 876 proteins, we used intensity-

133 based absolute quantification (iBAQ) values to determine the degree of enrichment in the sgTelo
134 sample relative to the sgNS sample. Some of these 876 proteins (140 in the first replicate, and
135 403 in the second) yielded no spectra whatsoever in the corresponding sgNS sample, consistent
136 with sgTelo specificity. In those cases, to avoid infinitely large sgTelo/sgNS enrichment scores,
137 we assigned those proteins the smallest non-zero iBAQ value from the proteins positively
138 identified in that sgNS dataset. The sgTelo/sgNS iBAQ ratios were then analyzed by moderated
139 t-test, yielding 192 proteins whose enrichment in sgTelo was statistically significant ($p < 0.05$)
140 [**Fig. 2a** (red and blue dots) and **Supplementary Table 1**]. Strikingly, the six subunits of the
141 shelterin complex (a telomere-binding complex that protects ends from chromosome fusion[14])
142 were the six most significantly enriched proteins (**Fig. 2a** and **Supplementary Table 1**).
143 Another highly enriched protein was Apollo, a 5'→3' exonuclease that interacts with the
144 shelterin component TRF2 and functions in the ALT pathway[15]. Overall, among the 192 most
145 significantly sgTelo-enriched proteins (**Supplementary Table 1**), 32 have been reported
146 previously to be associated with telomeres or linked to telomere function (**Supplementary**
147 **Table 1**). These include proteins from complexes known to contribute to ALT-associated
148 pathways or processes (**Supplementary Fig. 2**). Gene ontology (GO) analysis of the 192 C-
149 BERST proteomic hits reveals strong functional associations with terms such as telomere
150 maintenance, DNA replication, DNA repair, and homologous recombination, all of which are
151 important for ALT pathways[7] (**Fig. 2b**).
152 Telomere-associated proteomes from ALT+ cell lines have been defined previously by
153 TRF1-BirA* BioID in U2OS cells[16], and by biochemical purification [proteomics of isolated
154 chromatin segments (PICh)] from WI38-VA13 cells[17]. Protein identifications from these analyses,
155 as well as from our C-BERST dataset, were examined for overlap as depicted in the Venn
156 diagram shown in **Fig. 2c**. Of the 192 proteins identified by C-BERST, 74 (~38%) were also
157 detected by one or both of the other methods [62 by BioID ($p = 7.02 \times 10^{-56}$), 31 by PICh ($p =$
158 1.58×10^{-30}), and 19 by both]. Of the 19 proteins detected by all three approaches (**Fig. 2d**), 17
159 are among the top third of the 192 most significant C-BERST hits based on sgTelo/sgNS
160 enrichment, and 16 are known telomere-related factors. The remaining three of the nineteen
161 proteins identified by PICh, BioID and now C-BERST [nitric oxide synthase-interacting protein
162 (NOSIP), SLX4-interacting protein (SLX4IP), and core-binding factor subunit beta (CBFB) (**Fig.**
163 **2d**)] appear likely to have previously unappreciated roles as telomere-associated factors.

164 In conclusion, we demonstrate that C-BERST successfully maps subnuclear proteomes
165 associated with genomic landmarks. Using the extensively investigated ALT telomeric proteome
166 as a benchmark, we recover approximately 41% of known ALT-associated proteins (32 of 78,
167 **Supplementary Tables 1 and 2**), as well as factors involved in all reported biological
168 processes that contribute to ALT. By combining the flexibility of RNA-guided dSpyCas9 genome
169 binding with the efficiency and rapid kinetics of APEX2-catalyzed biotinylation, C-BERST
170 promises to extend the unbiased definition of subnuclear proteomes to many genomic elements,
171 and to a range of dynamic processes (e.g. cellular differentiation, responses to extracellular stimuli,
172 and cell cycle progression) that occur too rapidly to analyze via the 18-24h labeling procedures
173 inherent to CasID. Importantly, C-BERST promises to augment and extend Hi-C and related
174 methods by linking conformationally important cis-elements with the factors that associate with
175 them. Our initial implementation of C-BERST focused on minimizing the background labeling
176 catalyzed by un- or mis-localized dSpyCas9-mCherry-APEX2, mostly by optimizing its
177 expression level relative to that of its sgRNA. This method is amenable to further improvements
178 including isotope labeling[5] or tandem mass tagging[18] that will enable ratiometric assessments of
179 differential biotinylation when dSpyCas9-mCherry-APEX2 is guided by specific vs. non-specific
180 sgRNAs. Such improvements, in conjunction with sgRNA multiplexing, may prove to be
181 important for applying C-BERST to the definition of subnuclear proteomes associated with non-
182 repetitive loci. In the meantime, many types of repetitive elements within the genome, like
183 telomeres, play critically important roles in chromosome maintenance and function in ways that
184 depend upon associated proteins; C-BERST provides an unbiased method for sampling
185 subnuclear, locus-specific proteomics at these elements to define protein factors critical to the
186 functions of these elements.
187

188 **METHODS**

189 Methods, data files, and any associated references are available in the online version of the paper.
190 *Note: Any Supplemental and Source Data files are available in the online version of the paper.*
191

192 **ACKNOWLEDGEMENTS**

193 We are grateful to all members of the Sontheimer, Wolfe and Dekker labs for advice and
194 discussions, Tom Fazzio, Samya Bhaduri and Michael Green for helpful feedback, Hanhui Ma,

David Grünwald, and Thoru Pederson for reagents, Yuehe Ding (Mello lab) for help with mass spectrometry data analysis, John Leszyk for mass spectrometry, 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.

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. performed the experiments. L.-C.T. processed the fluorescence images, A.M. processed flow cytometry data, and L.J.Z. conducted statistical analyses. X.D.G, L.-C.T., A.M., S.A.S., and L.J.Z. analyzed data, and 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.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html

212 **REFERENCES**
213

214 1. Davies, J.O., Oudelaar, A.M., Higgs, D.R. & Hughes, J.R. *Nat Methods* **14**, 125-134 (2017).
215 2. Dominguez, A.A., Lim, W.A. & Qi, L.S. *Nat Rev Mol Cell Biol* **17**, 5-15 (2016).
216 3. Roux, K.J., Kim, D.I., Raida, M. & Burke, B. *J Cell Biol* **196**, 801-810 (2012).
217 4. Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F. & Leonhardt, H. *Nucleus* **7**, 476-484
218 (2016).
219 5. Hung, V. et al. *Mol Cell* **55**, 332-341 (2014).
220 6. Rhee, H.W. et al. *Science* **339**, 1328-1331 (2013).
221 7. Cesare, A.J. & Reddel, R.R. *Nat Rev Genet* **11**, 319-330 (2010).
222 8. Dilley, R.L. et al. *Nature* **539**, 54-58 (2016).
223 9. Chen, B. et al. *Cell* **155**, 1479-1491 (2013).
224 10. Ma, H. et al. *Proc Natl Acad Sci U S A* **112**, 3002-3007 (2015).
225 11. Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G. & Wandless, T.J. *Cell* **126**,
226 995-1004 (2006).
227 12. Ma, H. et al. *J Cell Biol* **214**, 529-537 (2016).
228 13. Knight, S.C. et al. *Science* **350**, 823-826 (2015).
229 14. de Lange, T. *Genes Dev* **19**, 2100-2110 (2005).
230 15. Lenain, C. et al. *Curr Biol* **16**, 1303-1310 (2006).
231 16. Garcia-Exposito, L. et al. *Cell Rep* **17**, 1858-1871 (2016).
232 17. Dejardin, J. & Kingston, R.E. *Cell* **136**, 175-186 (2009).
233 18. Loh, K.H. et al. *Cell* **166**, 1295-1307 e1221 (2016).

234

235 **ONLINE METHODS**

236

237 **Construction of C-BERST plasmids.** The Shield1- and doxycycline-inducible dSpyCas9-
238 mCherry-APEX2 construct was made by subcloning Flag-APEX2 from Flag-APEX2-NES
239 (Addgene 49386) into DD-dSpyCas9-mCherry[12] using the pHAGE backbone. Two NLSs (SV40
240 and nucleoplasmin NLS) were inserted at each terminus to improve nuclear localization. The
241 sequence of the final plasmid is provided in the **Supplementary Note**. The sgTelo-encoding
242 construct was created by replacing the C3-guide RNA sequence (pCMV_C3-
243 sgRNA_2XBroccoli/pPGK_TetR_P2A_BFP) with sgTelo sequences (using a plasmid provided
244 by Hanhui Ma and Thoru Pederson). Non-specific sgRNA (sgNS)[13] was used similarly to create
245 the analogous sgNS construct.

246

247 **Cell culture and cell line construction.** Human U2OS cells obtained from Thoru
248 Pederson's lab (originally obtained from ATCC) were cultured in Dulbecco-modified Eagle's
249 Minimum Essential Medium (DMEM; Life Technologies) supplemented with 10% (vol/vol) FBS
250 (Sigma). Lentiviral transduction was as described[12]. Six-fold higher titers of sgRNA-encoding
251 lentiviruses were used for transduction relative to dSpyCas9-APEX2 lentivirus.

252

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

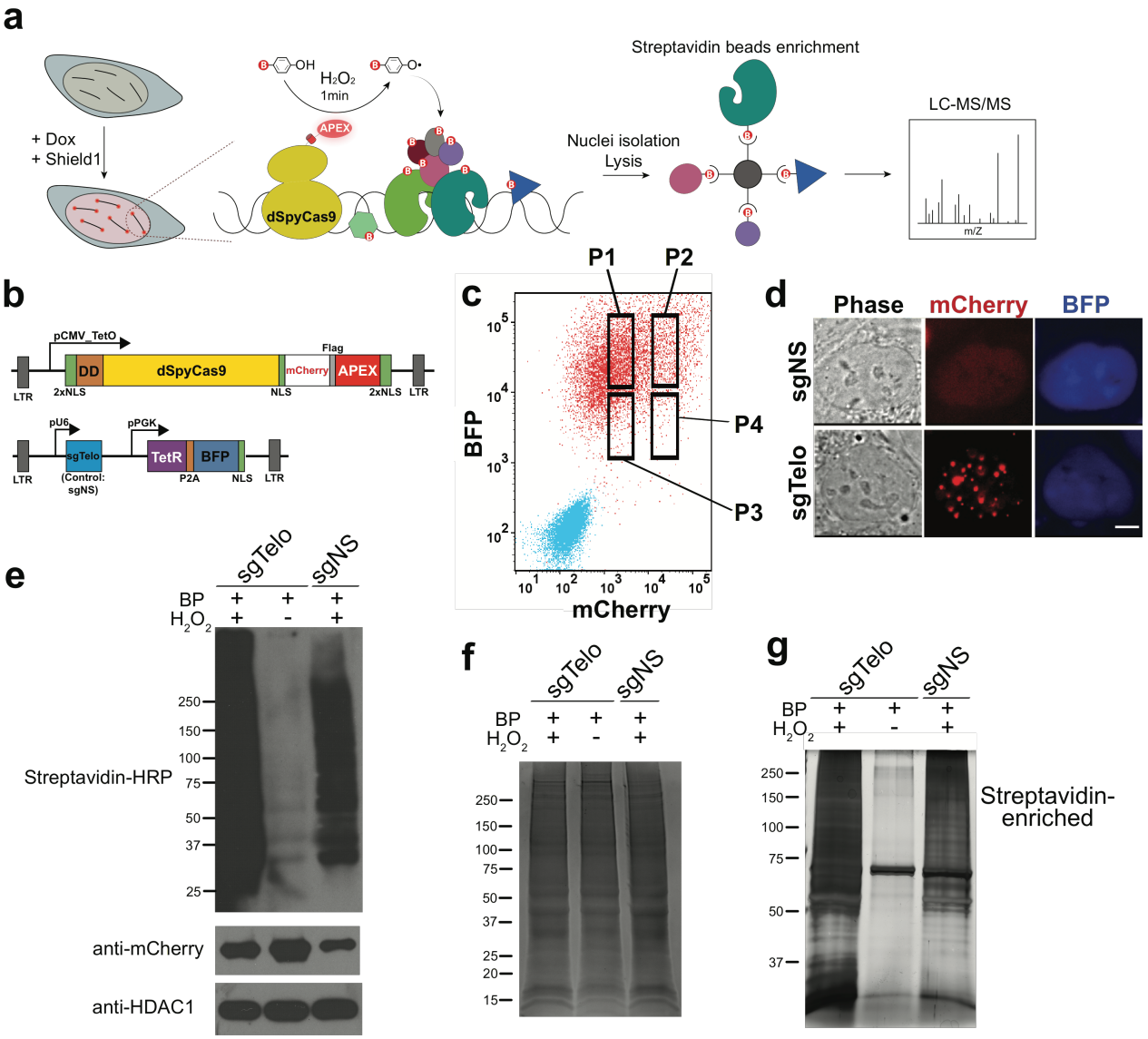
261

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

266 imaging processing was done with MetaMorph (Molecular Devices). Image contrast was set to
267 ease visualization of cell, foci and nucleoplasmic background.
268
269 **C-BERST biotinylation protocol.** Six 15cm plates of U2OS cells expressing sgTelo or sgNS
270 were used in this assay. Dox (2 µg/ml) and Shield1 (250 nM) were added 21 hours before
271 biotinylation. Cells were then incubated with 500 µM biotin-phenol (BP) (Adipogen) for 30
272 minutes at 37°C. 1 mM H_2O_2 was then added to initiate of biotinylation for 1 minute on a
273 horizontal shaker at room temperature. Six 15cm plates of sgTelo-expressing cells were treated in
274 parallel, but without H_2O_2 addition, as a negative control. Quencher solution (5 mM trolox, 10
275 mM sodium ascorbate, and 10 mM sodium azide) was added to stop the reaction, and cells were
276 washed five times (three quencher washes and two DPBS washes) to continue the quench and to
277 remove excess BP.
278
279 **Enrichment of biotinylated proteins.** Cells were scraped off the plates and used for the
280 preparation of isolated nuclei[19]. Nuclei were washed with DPBS before lysis. RIPA lysis buffer
281 [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.125% SDS, 0.125% sodium deoxycholate and 1%
282 Triton X-100 in Millipore water) with 1x freshly supplemented Halt Protease Inhibitor were used
283 to lyse the cells for 10 minutes on ice. Cell lysates in 1.5 ml Eppendorf tubes were sonicated for
284 15 minutes with a Diagenode Bioruptor with 30s on/off cycles at high intensity. Cell lysates were
285 clarified by centrifugation at 13,000 rpm for 10 minutes. Clarified protein samples (~3.5 mg)
286 were subjected to 400 µl Dynabeads MyOne Streptavidin T1 affinity purification overnight at
287 4°C. Each bead sample was washed with a series of buffers to remove non-specifically bound
288 proteins: twice with RIPA lysis buffer, once with 1 M KCl, once with 0.1 M Na_2CO_3, once with
289 2 M urea in 10mM Tris-HCl, pH 8.0, and twice with RIPA lysis buffer. Proteins were eluted in
290 70 µl 3x protein loading buffer supplemented with 2 mM biotin and 20 mM DTT with heating
291 for 10min at 95˚C[5]. 50 µl eluents were loaded and run on a 4-12% SDS-PAGE gel (Bio-Rad) and
292 run approximately 1cm off the loading well for in-gel digestion and LC-MS/MS analysis.
293
294 **Western blotting.** Protein concentrations of the cell lysates were determined by BCA assay
295 (Thermo). 50 ug of each sample was mixed with protein loading buffer, boiled, and separated in

296 SDS-PAGE gels. Proteins were transferred to PVDF membrane (Millipore), and blotted with
297 Streptavidin-HRP (Thermo), or with anti-mCherry (Abcam) or anti-HDAC1 (Bethyl) antibodies.
298
299 **LC-MS/MS and proteomic analyses.** Unresolved protein bands from SDS-PAGE were cut
300 into 1x1 mm pieces and placed in 1.5ml Eppendorf tubes with 1ml of water. After 30 min, water
301 was removed and replaced with 70 µl of 250 mM ammonium bicarbonate. Proteins were then
302 reduced by the addition of 20 µl of 45 mM 1,4-dithiothreitol, incubated at 50˚C for 30 min,
303 cooled to room temperature, alkylated with 20 µl of 100 mM iodoacetamide for 30 min, and
304 washed twice with 1 ml water. The water was removed and replaced with 1 ml of 50 mM
305 ammonium bicarbonate: acetonitrile (1:1) and incubated at room temperature for 1 hr. The
306 solvent was then replaced with 200 µl acetonitrile, removed, and the pieces dried in a Speed Vac.
307 Gel pieces were then rehydrated in 75 µl of 4 ng/µl sequencing-grade trypsin (Promega) in
308 0.01% ProteaseMAX Surfactant (Promega) in 50 mM ammonium bicarbonate and incubated at
309 37˚C for 21 hr. The supernatant was then removed to a 1.5 ml Eppendorf tube, the gel pieces
310 further dehydrated with 100 µl of acetonitrile: 1% (v/v) formic acid (4:1), and the combined
311 supernatants dried on a Speed Vac. Peptides were then reconstituted in 25 µl of 5% acetonitrile
312 containing 0.1% (v/v) trifluoroacetic acid for LC-MS/MS.
313
314 Samples were analyzed on a NanoAcquity UPLC (Waters Corporation) coupled to a Q Exactive
315 (Thermo Fisher Scientific) hybrid mass spectrometer. In brief, 1.0 µl aliquots were loaded at 4
316 µl/min onto a custom-packed fused silica precolumn (100 µm ID) with Kasil frit containing 2 cm
317 Magic C18AQ (5µ, 100Å) particles (Bruker Corporation). Peptides were then separated on a
318 75µm ID fused silica analytical column containing 25 cm Magic C18AQ (3µm, 100Å) particles
319 (Bruker) packed in-house into a gravity-pulled tip. Peptides were eluted at 300 nl/min with a
320 linear gradient from 95% solvent A (0.1% (v/v) formic acid in water) to 35% solvent B (0.1%
321 (v/v) formic acid in acetonitrile) in 60 min. Data was acquired by data-dependent acquisition
322 according to a published method[20]. Briefly, MS scans were acquired from *m/z* 300-1750 at a
323 resolution of 70,000 (*m/z* 200) and followed by ten tandem mass spectrometry scans using HCD
324 fragmentation using an isolation width of 1.6 Da, a collision energy of 27%, and a resolution of
325 17,500 (*m/z* 200). Raw data files were processed with Proteome Discoverer (Thermo, version
326 2.1.1.21) and searched with Mascot (Matrix Science, version 2.6) against the SwissProt *Homo*

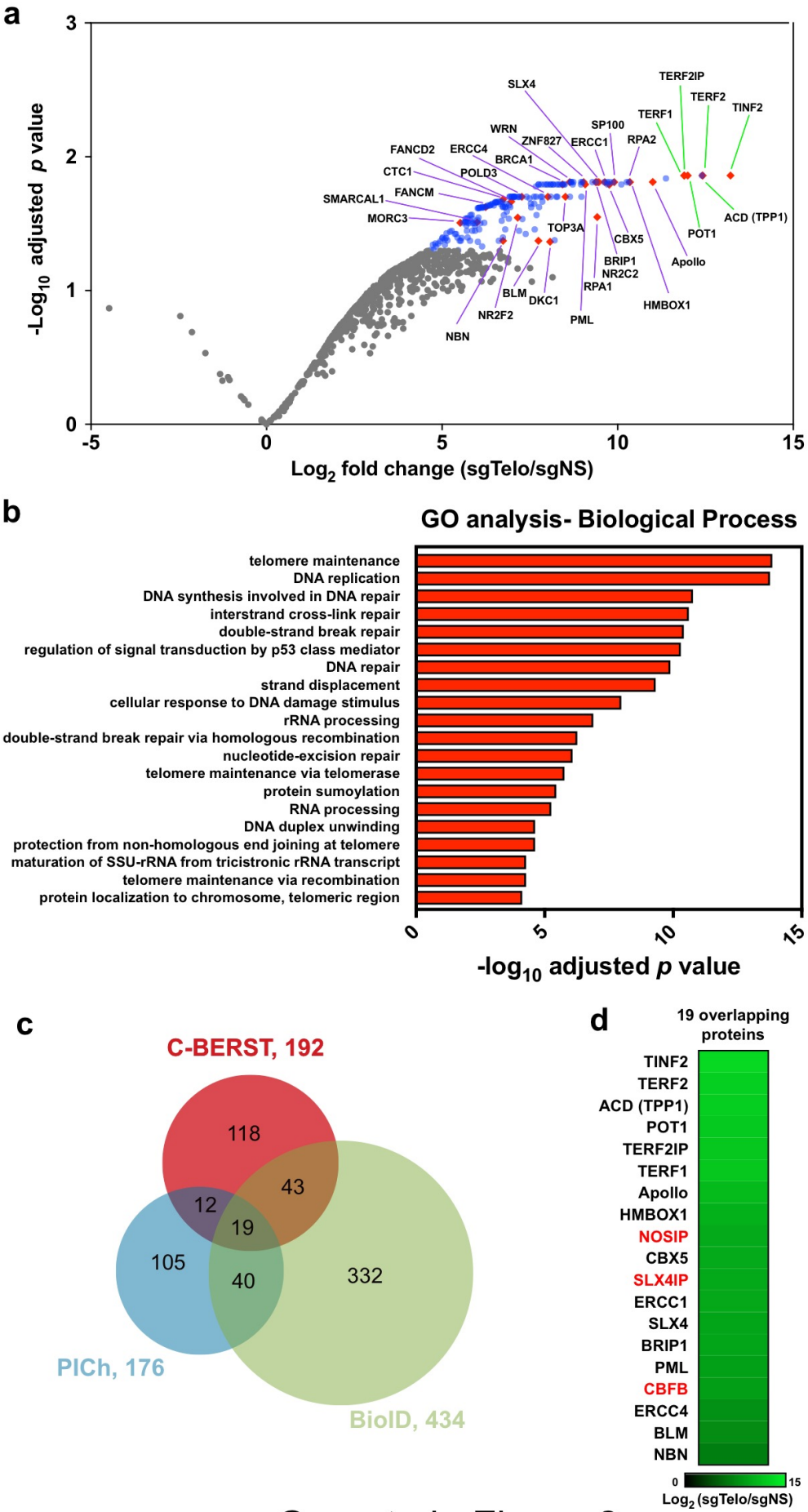
327 *sapiens* database. Search parameters used tryptic specificity considering up to 2 missed cleavages,
328 a parent mass tolerance of 10 ppm, and a fragment mass tolerance of 0.05 Da. Fixed
329 modification of carbamidomethyl cysteine was considered as were variable modifications of N-
330 terminal acetylation, N-terminal conversion of Gln to pyroGlu, oxidation of methionine, and
331 biotin-phenol conjugation of tyrosine. Results were loaded into Scaffold (Proteome Software Inc.,
332 version 4.8.2) for peptide and protein validation and quantitation using the Peptide Prophet and
333 Protein Prophet algorithms[21, 22]. The threshold for peptides was set to 70% (1.1% FDR) and 90%
334 for proteins (2-peptide minimum). Contaminants such as human keratin were included in all
335 statistical analyses and removed from the figures.
336
337 **Data analysis.** Data was first filtered to exclude proteins detected in only one of the S1
338 replicates, followed by log2 transformation. Prior to the $\log_2$ transformation, iBAQ values of 0
339 were replaced with the smallest iBAQ value from the corresponding sample in S2 or S3 to avoid
340 generation of infinite ratios. Moderated t-test with a paired design was used to compare the log2
341 transformed iBAQ values between S1 and S3, S1 and S2, and S2 and S3 using limma package[23].
342 To adjust for multiple comparisons, p values were adjusted using the Benjamini-Hochberg (BH)
343 method[24]. Proteins with BH-adjusted p values less than 0.05 are considered statistically significant.
344 Proteins were selected for subsequent GO (David Bioinformatics) and overlap analysis if they
345 were (*i*) significantly enriched in both S1 vs. S3 and S1 vs. S2, (*ii*) not enriched in S2 vs. S3, and
346 (*iii*) if S1/S3 and S1/S2 ratios were greater than 2. To determine whether the proteins identified
347 in this experiment overlap significantly with two published datasets, a hypergeometric test was
348 used.
349
350
351 19. Nagano, T. et al. *Nature* **502**, 59-64 (2013).
352 20. Kelstrup, C.D. et al. *J Proteome Res* **13**, 6187-6195 (2014).
353 21. Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. *Anal Chem* **74**, 5383-5392 (2002).
354 22. Nesvizhskii, A.I., Keller, A., Kolker, E. & Aebersold, R. *Anal Chem* **75**, 4646-4658 (2003).
355 23. Smyth, G.K. *Stat Appl Genet Molec Biol* **3**, 1-25 (2004).
356 24. Benjamini, Y. & Hochberg, Y. *J R Stat Soc B* **57**, 289-300 (1995).



Gao et al., Figure 1

357

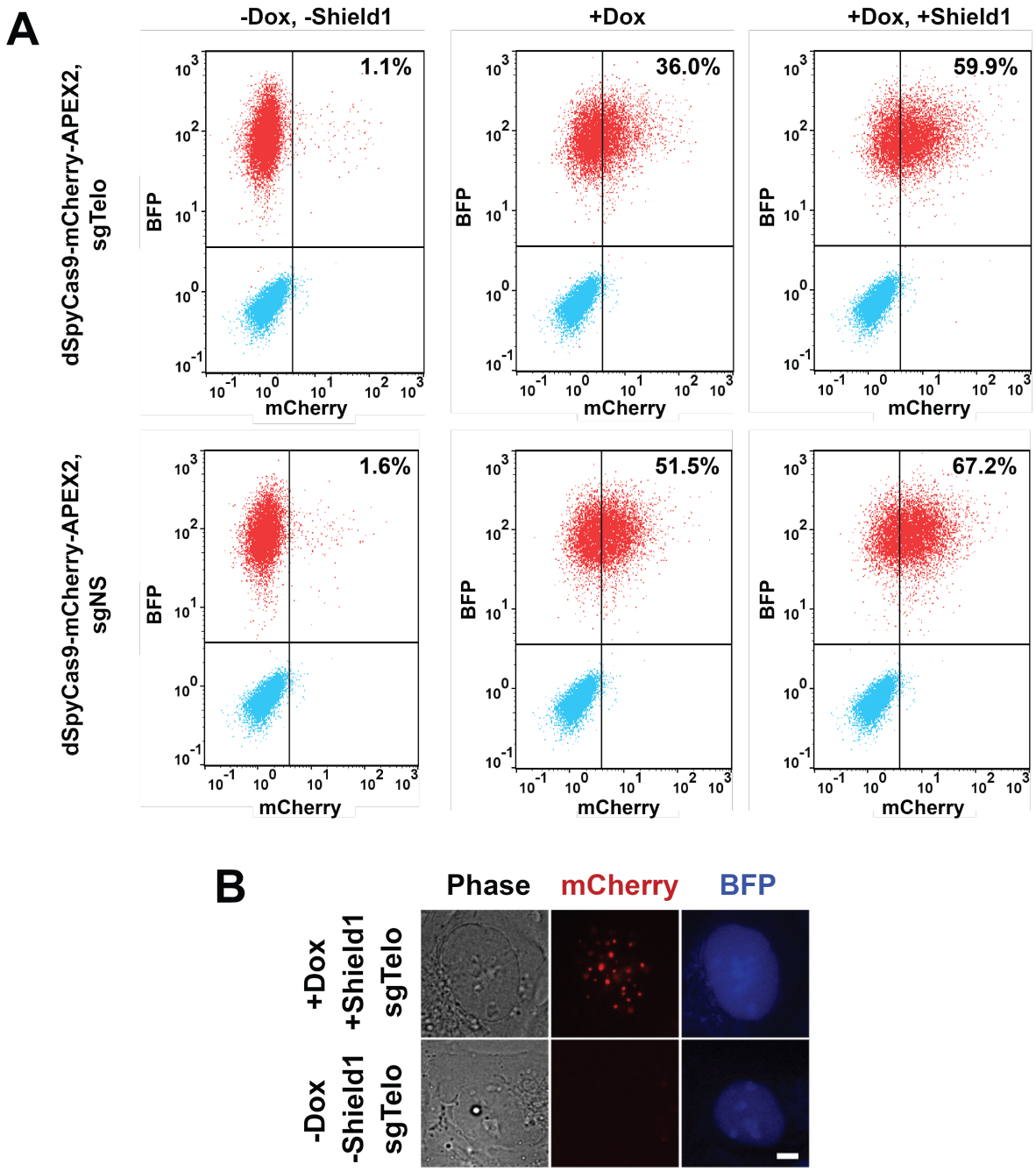
358 **Figure 1** Using C-BERST to biotinylate telomere-associated proteins in living human cells. (**a**)
359 Diagram of the C-BERST workflow. U2OS cells stably expressing sgRNA and inducible
360 dSpyCas9-APEX2 are generated by lentiviral transduction. Following dox and Shield1 induction
361 (21 h), cells are incubated with biotin-phenol (BP, 30 min) and then H_2O_2 (1 min) to activate a
362 burst of biotin-phenoxyl radical generation by dSpyCas9-APEX2, leading to proximity-labeling
363 of nearby proteins. Following quenching, nuclei isolation and protein extraction, biotinylated
364 proteins are enriched by streptavidin selection and analyzed by LC-MS/MS. (**b**) The dSpyCas9-
365 mCherry-APEX2 and sgRNA lentiviral expression constructs. Top: dSpyCas9-mCherry-APEX2
366 under the control of the pCMV_TetO inducible promoter. The mCherry fusion is included to
367 enable quantification of dSpyCas9 expression level as well as its subcellular localization. NLS,
368 nuclear localization signal; LTR, long terminal repeat; DD, Shield1-repressible degradation
369 domain. Bottom: sgRNA/TetR/BFP expression construct. pU6, U6 promoter; pPGK, PGK
370 promoter; sgTelo, telomere-targeting sgRNA; sgNS, non-specific sgRNA; tetR, tet repressor;
371 P2A, 2A self-cleaving peptide; BFP, blue fluorescent protein. (**c**) FACS sorting of untransduced
372 (blue) and mCherry- and BFP-positive cells (red). The P1 population corresponds to high BFP (as
373 a surrogate for sgRNA and TetR) and low mCherry expression, providing optimal signal-to-noise
374 ratio to maximize the fraction of telomere-localized dSpyCas9-mCherry-APEX2. (**d**) Live-cell
375 imaging of telomere localization by dSpyCas9-mCherry-APEX2 in U2OS cells, using the P1-
376 sorted population defined in (**c**). dSpyCas9-mCherry-APEX2 exhibited telomeric foci with
377 sgTelo but not with sgNS. Scale bar, 5 μm. (**e**) Top: Western blot analysis of dSpyCas9-
378 mCherry-APEX2 biotinylation, as detected by streptavidin-HRP. sgRNAs, BP treatment, and
379 H_2O_2 treatment are indicated at the top of each lane. Anti-mCherry was used to detect
380 dSpyCas9-mCherry-APEX2 (middle), and anti-HDAC1 was used as a loading control (bottom).
381 (**f**) Coomassie-stained SDS-PAGE of total protein from isolated nuclei following biotin labeling.
382 (**g**) Silver-stained SDS-PAGE of biotin-labeled proteins enriched with streptavidin-coated beads.
383 In **e-g**, the mobilities of protein markers (in kDa) are indicated on the left of each panel.



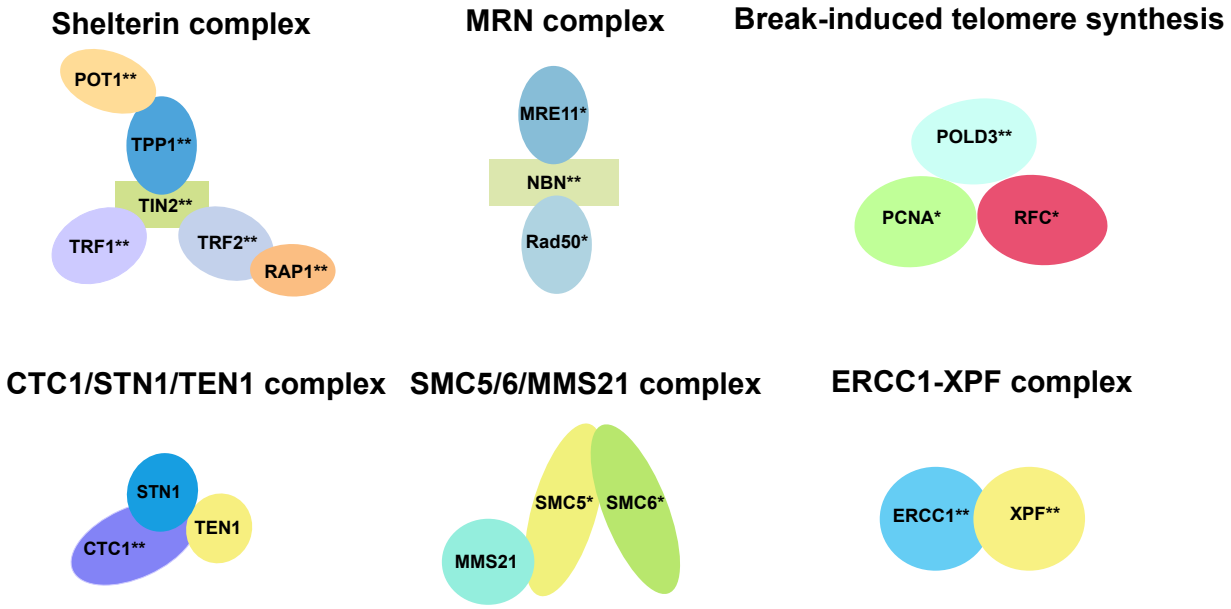
384

Gao et al., Figure 2

385 **Figure 2** Successful capture of telomere-associated proteins in living human cells by C-BERST.
386 (**a**) Volcano plot of C-BERST-labeled, telomere-associated proteins in U2OS cells. Intensity-
387 based absolute quantification (iBAQ) values from the MS analyses were calculated for each
388 identified protein for all three samples (sgTelo + H_2O_2, sgTelo - H_2O_2, and sgNS + H_2O_2). 192
389 proteins (indicated by blue and red) are statistically enriched [Benjamini-Hochberg (BH)-adjusted
390 p value < 0.05] in the sgTelo + H_2O_2 sample, relative to both control samples. The 32 proteins
391 indicated by red diamonds (with identities provided) are previously defined as either telomere-
392 associated proteins or ALT pathway components. These include all six shelterin components
393 (indicated by green lines). (**b**) Gene Ontology-Biological Process (GO-BP) analysis on the 192
394 telomeric/ALT proteins identified by C-BERST. The x-axis is the $-\log_{10} p$ value (BH-adjusted)
395 for the C-BERST-detected proteins associated with each GO term given on the left. The 20 most
396 statistically significant GO terms are displayed. (**c**) Venn diagram of statistically enriched (BH-
397 adjusted p value < 0.05) telomeric proteins from ALT human cells, as detected by C-BERST
398 (red), PICh (purple), and TRF1-BirA* BioID (green). 74 proteins from the C-BERST proteome
399 were also detected by PICh, BioID, or both. (**d**) A heat map of the C-BERST $\log_2$ fold-change
400 enrichment scores for the 19 telomeric proteins identified by all three proteomic approaches from
401 (**c**). All 19 proteins are highly enriched in the C-BERST telomere proteome. The three proteins
402 indicated in red have not been reported previously as telomere- or ALT-associated.



Supplementary Fig. 1 Inducible dSpyCas9-mCherry-APEX2 expression. (**a**) Flow cytometry was used to measure the percentage of mCherry+ and BFP+ double-positive cells under different induction conditions. Stable U2OS cells expressing sgTelo (top row) or sgNS (bottom row) were exposed to three conditions for 21h before flow cytometry: no inducers (left), dox only (2 µg/ml, middle), or a combination of dox (2 µg/ml) and Shield1 (250 nM) (right). Cyan: untransduced cells; red: transduced cells. With both sgRNAs, dox and Shield1 in combination yield the highest percentages of double-positive cells. Specific percentages of mCherry+, BFP+ cells are indicated in each plot. (**b**) Live-cell imaging of clonal cells derived from the sgTelo P1 population (see **Fig. 1c**). When inducers are omitted, dSpyCas9-mCherry-APEX2 expression and telomeric accumulation are not observed. Scale bar, 5 µm.



Supplementary Fig. 2 C-BERST specifically detects components of multiple complexes and factors implicated in ALT pathways. Proteins denoted by double asterisks were significantly enriched (BH-adjusted p value < 0.05) in the sgTelo + H_2O_2 sample, relative to the sgTelo - H_2O_2 and sgNS + H_2O_2 samples (see **Supplementary Table 1**). Proteins denoted by single asterisks were also detected but with lower degrees of enrichment (BH-adjusted p value > 0.05) in the sgTelo + H_2O_2 sample (**Supplementary Table 1**). Components of the RAD9/RAD1/HUS1 complex (see **Supplementary Table 2**) were not detected.

Supplementary Note dSpyCas9-mCherry-APEX2, sgTelo, and sgNS sequences.

Amino acid sequence of dSpyCas9-mCherry-APEX2

Legend: NLS DD dSpyCas9 *mCherry* FLAG APEX2

MAPKKKRKVEDKRPAATKKAGQAKKKKEDACGVQVETISPGDGRTFPKRGQTCVVHYTGMLE
DGKKVDSSRDRNKPFKFMLGKQEVIRGWEEGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPH
ATLVFDVELLKPEGSEFGSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKK
NLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREK
IEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR
ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
HQSITGLYETRIDLSQLGGDGSTSGSPKKKRKVGSGS*MVSKGEEDNMAIIKEFMRFKVHMEGS*
VNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLK
LSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEA
SSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYT
*IVEQYERAEGRHSTGGMDELYK*SGGRGGGGSDYKDDDDKGKSYPTVSADYQDAVEKAKKKL
RGFIAEKRCAPLMLRLAFHSAGTFDKGTKTGGPFGTIKHPAELAHSANNGLDIAVRLLEPLKAEF
PILSYADFYQLAGVVAVEVTGGPKVPFHPGREDKPEPPPEGRLPDPTKGSDHLRDVFGKAMGL
TDQDIVALSGGHTIGAAHKERSGFEGPWTSNPLIFDNSYFTELLSGEKEGLLQLPSDKALLSDP
VFRPLVDKYAADEDAFFADYAEAHQKLSELGFADALEPKKKRKVEDKRPAATKKAGQAKKKKG
S*

Nucleotide sequence of sgRNA/TetR_P2P_BFP

Legend: *U6 promoter* Guide sequence sgRNA scaffold *hPGK promoter* TetR P2A BFP NLS

ccccttcaccgagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgta
aacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactat
*catatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc*gNNNNNNNNNNNNNN
NNNNNNNNttaagagctatgctggaaacagcatagcaagtttaaataaggctagtccgttatcaacttgaaaaagtggcaccgag
tcggtgctttttttgaattctcgacctcgagacaaatggcagtattcatccacaattttaaaagaaaagggggattgggggtacagtgc
aggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaacaaattacaaaaattcaaaattttcgg
gtttattacgggacagcagagatccactttggccgcggctcgaggggg*ttggggttgcgcctttccaaggcagccctgggtttgcgc*
agggacgcggctgctctgggcgtggttccgggaaacgcagcggcgccgaccctgggtctcgcacattcttcacgtccgttcgcagcg
tcacccggatcttcgccgctacccttgtgggcccccggcgacgcttcctgctccgcccctaagtcgggaaggttccttgcggttcgcgg
cgtgccggacgtgacaaacggaagccgcacgtctcactagtaccctcgcagacggacagcgccagggagcaatggcagcgcgc
cgaccgcgatgggctgtggccaatagcggctgctcagcagggcgcgccgagagcagcggccgggaaggggcggtgcgggagg
cggggtgtggggcggtagtgtgggccctgttcctgcccgcgcggtgttccgcattctgcaagcctccggagcgcacgtcggcagtcgg
*ctccctcgttgaccgaatcaccgacctctctccccag*ggggatccgtgagtttggggacccttgattgttctttctttttcgctattgtaaaatt
catgttatatggagggggcaaagttttcagggtgttgtttagaatgggaagatgtcccttgtatcaccatggaccctcatgataattttgtttc
tttcactttctactctgttgacaaccattgtctcctcttattttcttttcattttctgtaacttttttcgttaaactttagcttgcatttgtaacgaatttttaaa
ttcacttttgtttatttgtcagattgtaagtactttctctaatcactttttttcaaggcaatcagggtatattatattgtacttcagcacagttttagag
aacaattgttataattaaatgataaggtagaatatttctgcatataaattctggctggcgtggaaatattcttattggtagaaacaactacat
cctggtcatcatcctgcctttctctttatggttacaatgatatacactgtttgagatgaggataaaatactctgagtccaaaccgggcccctc
tgctaaccatgttcatgccttcttctttttcctacagctcctgggcaacgtgctggttattgtgctgtctcatcattttggcaaagaattgtaatac
gactcactatagggcgaattgatatgtctagattagataaaagtaaagtgattaacagcgcattagagctgcttaatgaggtcggaatc
gaaggtttaacaacccgtaaactcgcccagaagctaggtgtagagcagcctacattgtattggcatgtaaaaaataagcgggctttgc
tcgacgccttagccattgagatgttagataggcaccatactcacttttgccctttagaaggggaaagctggcaagatttttacgtaataa
cgctaaaagttttagatgtgctttactaagtcatcgcgatggagcaaaagtacatttaggtacacggcctacagaaaaacagtatgaaa
ctctcgaaaatcaattagccttttatgccaacaaggtttttcactagagaatgcattatatgcactcagcgctgtggggcattttactttagg
ttgcgtattggaagatcaagagcatcaagtcgctaaagaagaaagggaaacacctactactgatagtatgccgccattattacgaca
agctatcgaattatttgatcaccaaggtgcagagccagccttcttattcggccttgaattgatcatatgcggattagaaaaacaacttaaa
tgtgaaagtgggtccgcgtacagcggctcccgggagttcgctagcggtgctactaacttcagcctgctgaagcaggctggagacgtg
gaggagaaccctggacctggtagtggaacgcgtatggtgtctaagggcgaagagctgattaaggagaacatgcacatgaagctgt
acatggagggcaccgtggacaaccatcacttcaagtgcacatccgagggcgaaggcaagccctacgagggcacccagaccatg
agaatcaaggtggtcgagggcggccctctccccttcgccttcgacatcctggctactagcttcctctacggcagcaagaccttcatcaa
ccacacccagggcatccccgacttcttcaagcagtccttccctgagggcttcacatgggagagagtcaccacatacgaagatgggg
gcgtgctgaccgctacccaggacaccagcctccaggacggctgcctcatctacaacgtcaagatcagaggggtgaacttcacatcc
aacggccctgtgatgcagaagaaaacactcggctgggaggccttcaccgagacgctgtaccccgctgacggcggcctggaaggc
agaaacgacatggccctgaagctcgtgggcgggagccatctgatcgcaaacgccaagaccacatatagatccaagaaacccgct
aagaacctcaagatgcctggcatctactatgtggactacagactggaaagaatcaaggaggccaacaacgaaacctacgtcgagc
agcacgaggtggcagtggccagatactgcgacctccctagcaaactggggcacaagcttaatccaaagaaaaagcggaaagtg

Guide sequences:

sgTelo: gTTAGGGTTAGGGTTAGGGTT

sgNS: gAATCTCGCTTATATAACGAG