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

Xin D. Gao[1], Li-Chun Tu[1], Aamir Mir[1], Tomás Rodriguez[1], Yuehe Ding[1], 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,*]

[1]RNA Therapeutics Institute
[2]Proteomics and Mass Spectrometry Facility
[3]Program in Systems Biology
[4]Department of Biochemistry and Molecular Pharmacology
[5]Howard Hughes Medical Institute
[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 rapid, unbiased mapping of proteomes near defined genomic loci, as demonstrated for telomeres and centromeres. By combining the spatially restricted enzymatic tagging enabled by APEX2 with programmable DNA targeting by dCas9, C-BERST has successfully identified nearly 50% of known telomere-associated factors and many known centromere-associated factors. We also identified and validated SLX4IP and RPA3 as telomeric factors, confirming C-BERST's utility as a discovery platform. C-BERST enables the rapid, 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 usually involves long (18-24h) labeling times, limiting the time resolution of dynamic processes. Similar considerations apply to the recently reported CAPTURE approach for subnuclear proteomics[5], in which dSpyCas9 itself is biotinylated by BirA* and used as an affinity handle.

72 Engineered ascorbate peroxidase (APEX2) has been used for an alternative live-cell
73 biotinylation strategy called spatially restricted enzymatic tagging (SRET)[6, 7]. In this approach,
74 APEX2 is fused to a localized protein of interest, and cells are then treated with biotin-phenol
75 (BP) and H_2O_2, generating a localized (within a ~20nm radius) burst of diffusible but rapidly
76 quenched biotin-phenoxyl radicals. These products react with electron-rich amino acid side
77 chains (e.g. Tyr, Trp, His and Cys), leading to covalent biotinylation of proteins in the vicinity of
78 the localized APEX2, thus allowing subsequent identification by streptavidin selection and LC-
79 MS/MS. Notably, this subcellular tagging method is extremely efficient (1 min H_2O_2 treatment),
80 allowing temporal control over the labeling process. Based in part on the success of dSpyCas9-FP
81 fusions in enabling subnuclear imaging in living cells, we reasoned that a dSpyCas9 derivative
82 that emits radicals rather than photons could be used for subnuclear proteomic analyses in a
83 manner that overcomes the unfavorable kinetics and high background of CasID. Here we use
84 dSpyCas9-APEX2 fusions in the development of C-BERST (**Fig. 1a**) for genomic element-
85 specific profiling of subnuclear proteomes in live cells.
86
87 **Results**
88
89 **C-BERST design and workflow**
90 To develop and validate this method, we sought genomic elements that are associated
91 with a well-defined suite of known protein factors, and that can be bound with dSpyCas9 with
92 high efficiency and specificity using an established sgRNA. For this purpose, we chose to target
93 telomeres in human U2OS cells. As with ~10-15% of cancer cell types, U2OS cells rely on
94 alternative lengthening of telomeres (ALT) pathways to maintain telomere length without
95 telomerase activation[8]. Cohorts of proteins associated with telomeres in ALT+ cells are well-
96 characterized, and they map to key pathways such as homologous recombination (HR) and
97 break-induced telomere synthesis[9]. Furthermore, an sgRNA (sgTelo) has already been established
98 for efficient telomere association of dSpyCas9[10, 11].
99 We transduced U2OS cells with a lentiviral vector expressing dSpyCas9 under the
100 control of a tet-on CMV promoter and fused to five nuclear localization signals (NLSs), a ligand-
101 tunable degradation domain (DD)[12], mCherry, and APEX2 (**Fig. 1b**). Maximal expression of
102 this dSpyCas9-mCherry-APEX2 fusion protein requires not only doxycycline (dox) but also the

103 Shield1 ligand to inactivate the DD in a dose-dependent fashion[13]. This combination allows
104 precise control over dSpyCas9-mCherry-APEX2 protein levels for optimal signal-to-noise levels.
105 mCherry-positive cells [collected by fluorescence-activated cell sorting (FACS)] were then
106 transduced with a separate lentiviral vector that included an sgRNA construct (driven by the U6
107 promoter) as well as a blue fluorescent protein (BFP) construct that also expresses the TetR
108 repressor (**Fig. 1b**). In one version of this construct the sgRNA cassette encodes sgTelo (for
109 labeling telomeres), and in the other it encodes a non-specific sgRNA (sgNS) that is
110 complementary to a bacteriophage-derived sequence that is absent from the human genome[14].
111 After 21h of dox and Shield1 induction, we again used fluorescence-activated sorting (FACS) to
112 sort four distinct BFP/mCherry double-positive cell populations (P1-P4) that correlate with
113 different expression levels of dSpyCas9-mCherry-APEX2 and BFP (as a surrogate for sgRNA
114 and TetR) (**Fig. 1c** and **Supplementary Fig. 1**). We reasoned that our signal-to-noise ratio of
115 telomeric vs. non-telomeric biotinylation would be maximized when sgTelo levels are saturating,
116 and when dSpyCas9-mCherry-APEX2 levels are limiting (relative to potential genomic binding
117 sites). Both conditions are expected to favor maximal partitioning of the sgRNA-programmed
118 dSpyCas9-mCherry-APEX2 into the desired telomere-associated state, with as little unlocalized
119 or mislocalized fusion protein as possible. Visual inspection by fluorescence microscopy
120 (**Supplementary Fig. 2**) confirmed that the sgTelo P1 cell population (with higher BFP
121 expression and lower mCherry expression, about 20% of the total sorted cells) exhibited the most
122 robust mCherry-labelled telomeric foci (as determined by colocalization with anti-TERF2IP
123 immunofluorescence) with the lowest amount of nucleolar or diffuse nucleoplasmic background[10].
124 We therefore used this population in our subsequent experiments.
125 To assess the distribution of biotinylated proteins in the nucleus, we used avidin-
126 conjugated Oregon Green 488 (OG 488) to probe sgTelo and sgNS cells after BP and H_2O_2
127 treatment. We found that biotinylated proteins were strongly enriched at telomeric foci in sgTelo
128 cells, whereas labeling was diffuse in sgNS cells (**Fig. 1d**). Efficient labeling was BP- and H_2O_2-
129 dependent. We also analyzed genome-wide dCas9-mCherry-APEX2 binding via anti-mCherry
130 chromatin immunoprecipitation and sequencing (ChIP-seq). Nearly 60% of total trimmed reads
131 from sgTelo cells contained at least one $(TTAGGG)_4$ sequence (the minimal length of telomeric
132 repeats with full complementarity to sgTelo). However, such $(TTAGGG)_4$-containing reads
133 comprised <0.5% of total trimmed reads from either sgNS or untransduced U2OS cells

134 (**Supplementary Fig. 3**). These experiments indicate that sgTelo-guided dCas9-mCherry-
135 APEX2 targets telomeres and enables restricted biotinylation of endogenous proteins near these
136 chromosomal elements.
137
138 **Label-free profiling of telomere-associated proteomes using C-BERST**
139 For proteomic analysis, we induced APEX2-catalyzed biotinylation with BP and H_2O_2 in
140 the sgTelo and sgNS P1-sorted cells (~6 x 10^7 cells for each guide), and also included an sgTelo
141 control in which the H_2O_2 was omitted. APEX2-catalyzed biotinylation of nucleoplasmic
142 proteins in the sgNS control sample serves as a reference, permitting an assessment of the
143 telomere specificity of labeling in the sgTelo sample. Nuclei were isolated from treated cells (to
144 reduce cytoplasmic background), and nuclear proteins were then extracted. Recovered proteins
145 (50 µg) were subjected to western blot analysis using streptavidin-conjugated horseradish
146 peroxidase (streptavidin-HRP) (**Fig. 2a**), as well as total protein visualization by Coomassie
147 staining (**Fig. 2b**). Samples were also probed with anti-mCherry antibodies (to detect dSpyCas9-
148 mCherry-APEX2) and HDAC1 (as a loading control) (**Fig. 2a**, bottom). Biotinylated proteins
149 were readily detected in both sgTelo and sgNS samples, but were largely absent in the -H_2O_2
150 control (**Fig. 2a**). In the sgNS sample, anti-mCherry and streptavidin-HRP signals were less
151 intense in comparison with the sgTelo sample, indicating that dSpyCas9-mCherry-APEX2
152 accumulation and activity are lower in the former. Biotinylated proteins were then isolated using
153 streptavidin beads and analyzed by SDS-PAGE and silver staining (**Fig. 2c**). Aside from the
154 ~75kDa endogenously biotinylated proteins routinely detected in SRET-labeled samples[6, 7], only
155 background levels of proteins were detected in the no-H_2O_2 control sample, indicating successful
156 purification. All three samples were subjected to in-gel trypsin digestion followed by LC-MS/MS
157 to identify the biotin-labeled proteins. These analyses were done with two biological replicates
158 prepared on different days.
159 The two sgTelo replicates yielded at least three peptides from 930 and 851 proteins, >85%
160 of which (792) were detected in both (**Supplementary Table 1**). For these 792 proteins, we
161 used intensity-based absolute quantification (iBAQ) values [a label-free quantification (LFQ)
162 proteomic approach] to determine the degree of enrichment in the sgTelo sample relative to the
163 sgNS sample. Some of these 792 proteins (104 in the first replicate, and 340 in the second)
164 yielded no spectra whatsoever in the corresponding sgNS sample, consistent with sgTelo

specificity. In those cases, to avoid infinitely large sgTelo/sgNS enrichment scores, we assigned those proteins the smallest non-zero iBAQ value from the proteins positively identified in that sgNS dataset. The sgTelo/sgNS iBAQ ratios were then analyzed by moderated t-test, yielding 143 proteins whose enrichment in sgTelo was statistically significant [Benjamini-Hochberg (BH)-adjusted $p < 0.05$] [**Fig. 2d** (red and blue dots) and **Supplementary Table 1**]. Strikingly, the six subunits of the shelterin complex (a telomere-binding complex that protects ends from chromosome fusion[15]) were among the seven most significantly enriched proteins (**Fig. 2d** and **Supplementary Table 1**). Another highly enriched protein was Apollo, a 5'→3' exonuclease that interacts with the shelterin component TRF2 and functions in the ALT pathway[16]. Overall, among the 143 most significantly sgTelo-enriched proteins (**Supplementary Table 1**), 30 have been reported previously to be associated with telomeres or linked to telomere function (**Supplementary Table 2**). These results indicate that validated telomeric proteins can be identified rapidly and efficiently by C-BERST.

Ratiometric proteomics enhances the sensitivity and specificity of C-BERST

To further improve our assessments of differential C-BERST biotinylation with specific vs. non-specific sgRNAs, we used a more quantitative proteomic approach enabled by stable isotope labeling with amino acids in cell culture (SILAC) to analyze telomere-associated proteomes. sgTelo/dCas9-mCherry-APEX2 cells were cultured in heavy-isotope medium, sgNS/dCas9-mCherry-APEX2 cells were cultured in medium-isotope medium, and untransduced U2OS cells were cultured in light-isotope medium, each for at least 5 passages to allow sufficient incorporation of isotope-labeled arginine and lysine (**Supplementary Fig. 4a**). We then induced dCas9-mCherry-APEX2 expression by dox and Shield1 for 21 hours, with comparable accumulation with either sgTelo or sgNS (**Supplementary Fig. 4b**). Biotinylation and cell lysis were then performed as described above, except that equal amounts (~1 mg, measured by Pierce BCA Protein Assay Kit) of protein lysates from heavy, medium, and light samples (H:M:L = 1:1:1) were mixed before streptavidin affinity purification for three-state SILAC[6]. 913 proteins were identified in both the heavy and medium samples, and 885 of these were also detectable in the light (no-APEX2 background) sample. Using significance (BH-adjusted $p < 0.01$) and enrichment ([$\log_2$ fold change (FC) $\geqslant 2.5$]) cut-offs that were even more stringent than those used for the label-free analysis (**Fig. 2d**), we identified 55 proteins that are strongly enriched in

196 the sgTelo sample relative to sgNS (H/M) (**Fig. 3** and **Supplementary Table 3**). Among these
197 55 proteomic hits, 34 are known telomere-associated factors, including all six shelterin
198 components as well as subunits from 5 other complexes that are known to contribute to ALT-
199 associated pathways or processes (**Supplementary Fig. 5**). All but one of the 55 H/M-enriched
200 proteins (BARD1) were also strongly enriched ($\log_2$ FC $\geqslant$ 1) in H/L ratio, indicating that
201 background detection in the absence of dCas9-mCherry-APEX2 biotinylation was minimal.
202 Gene ontology (GO) analysis of the 55 H/M-enriched C-BERST hits reveals strong functional
203 associations with terms such as telomere maintenance, DNA replication, DNA repair, and
204 homologous recombination, all of which are important for ALT pathways[8] (**Fig. 4a**).
205 Telomere-associated proteomes from ALT+ cell lines have been defined previously by
206 TRF1-BirA* BioID in U2OS cells[17], and by biochemical purification [proteomics of isolated
207 chromatin segments (PICh)] from WI38-VA13 cells[18]. Protein identifications from these analyses,
208 as well as from our C-BERST SILAC dataset, were examined for overlap as depicted in the
209 Venn diagram shown in **Fig. 4b**. Of the 55 proteins identified by C-BERST, 32 (~58%) were
210 also detected by one or both of the other methods [23 by BioID (p = 7.29 x 10^{-32}), 27 by PICh (p
211 = 2.04 x 10^{-50}), and 18 by both]. The remaining 23 proteins that were uniquely detected by C-
212 BERST include seven known telomeric/ALT factors (ATR, CTC1, FANCA, FANCD2,
213 FANCM, SMC5, and WRN). Of the 18 proteins detected by all three approaches, 17 are known
214 telomere-related factors. The remaining consensus hit [SLX4-interacting protein (SLX4IP)] was
215 not previously validated as telomeric in the BioID and PICh studies; nonetheless its identification
216 by all three proteomic approaches strongly suggests that it has a previously unappreciated role in
217 telomere function or maintenance (**Fig. 4c**). Such a role could be related to that of its binding
218 partner SLX4 in the resolution of telomere recombination intermediates[19].
219 To validate the ability of C-BERST to identify novel or provisional telomeric or ALT-
220 related proteins, we used independent methods to assess telomere colocalization of SLX4IP, as
221 well as a factor (RPA3) that was detected by C-BERST [ranked 44th ($\log_2$FC) among the 55
222 enriched proteins] but missed by BioID and PICh. We transiently transfected U2OS cells with
223 turboGFP-tagged SLX4IP, and then analyzed cells for its appearance in endogenous TRF2-
224 colocalizing foci, as indicated by anti-TRF2 immunofluorescence **(Fig 4d)**. SLX4IP-GFP signal
225 was evident in some but not all TRF2 foci, confirming that it localizes to a subset of telomeres.
226 This conclusion was further corroborated by immunofluorescence detection of endogenous

227 SLX4IP, which again revealed TRF2 colocalization (**Supplementary Fig. 6**). We also
228 transfected U2OS cells with RPA3-GFP, and again we detected it in a subset of TRF2 foci (**Fig**
229 **4d**). Due to high background staining by the commercial anti-RPA3 antibody, we were unable
230 to analyze TRF2 colocalization by endogenous RPA3, but western analyses confirmed that the
231 RPA3-GFP fusion protein (like the anti-SLX4IP fusion protein) was expressed at or below the
232 levels of the endogenous protein under the conditions used for fluorescence microscopy
233 (**Supplementary Fig. 7**). RPA3 is a subunit of the RPA complex, which has known functions
234 in ALT pathways[20]; in addition, other RPA subunits (RPA1 and RPA2) were enriched in PICh[18]
235 as well as C-BERST. Therefore the partial telomeric localization of RPA3 is not altogether
236 surprising, despite the fact that it was not detected by either PICh or BioID. The incomplete
237 overlap for SLX4IP and RPA3 with the telomeric marker is consistent with previously defined
238 non-telomeric functions for both SLX4IP[21] and RPA3[22], and indicate that C-BERST is
239 sufficiently sensitive and specific to detect telomere-associated factors even with proteins that are
240 only partially (or perhaps transiently) telomeric.
241
242 **C-BERST subnuclear proteomics at centromeres**
243 To extend our subnuclear proteomic approach to other genomic elements, we targeted
244 dCas9-mCherry-APEX2 to centromeric alpha-satellite arrays in U2OS cells, and then used C-
245 BERST to profile the protein components of alphoid chromatin (**Figure 5a**) using a similar
246 pipeline as that described above for telomeres. The human alpha satellite proteome from K562
247 cells has been analyzed previously using the PICh-related protocol known as HyCCAPP
248 (hybridization capture of chromatin-associated proteins for proteomics), again providing a basis
249 for comparison. Via live-cell imaging (**Figure 5b**) and western blotting (**Supplementary Fig.**
250 **8**), we have confirmed dCas9-mCherry-APEX2 inducible expression, specific centromere
251 targeting[23], and biotinylation. We used SILAC proteomic analysis (heavy, sgAlpha; medium,
252 sgNS; light, untransduced U2OS cells lacking dCas9-mCherry-APEX2) and identified 1268
253 proteins (**Supplementary Table 4**) from each of two biological replicates (based upon
254 detection in at least two of three technical replicates of each biological replicate). Among these
255 1268 proteins, 460 were enriched to a statistically significant extent ($\log_2 FC \geq 2.5$ and $p < 0.01$)
256 in the sgAlpha vs. sgNS samples (H/M). We have identified four highly enriched subunits of the
257 CENP-A nucleosome-associated complex[24] (CENP-C, -M, -N, and –T), and one subunit (CENP-

258 P) of the CENP-A distal complex[24]. We also identified nearly all CENP-A loading factors[25],
259 including HJURP, Mis18α, Mis18β, and MIS18BP1, among the enriched proteins. Many other
260 known centromere-associated proteins were also identified as enriched in sgAlpha including
261 CENP-B, -F, -I, and -L, as well as KNL1. Additionally, we identified three subunits of the
262 chromosome passenger complex [INCENP, CDCA8, and Aurora kinase B (AURKB)], which
263 has been suggested to localize to the centromere during mitosis[26]. We also found Fanconi anemia
264 pathway proteins such as FANCM, which has been implicated in promoting centromere
265 stability[27]. There are 31 overlapping enriched proteins ($p = 3.84 \times 10^{-38}$) between C-BERST and
266 HyCCAPP, despite the use of different cancer cell lines in the respective studies
267 (**Supplementary Fig. 9a**). C-BERST uniquely captured multiple known centromeric factors
268 including CENP-F and ATR, which were recently reported to localize to centromeres and
269 engage RPA-coated R loops[28]. GO analysis of the 460 C-BERST centromeric hits reveals strong
270 functional associations with terms such as DNA repair, DNA replication, sister chromatid
271 cohesion, double-strand break repair by homologous recombination, mitotic nuclear division,
272 and cell division, all of which are related to centromere maintenance or function[8]
273 (**Supplementary Fig. 9b**).
274 Our generation of both telomeric and centromeric C-BERST datasets affords the
275 opportunity to compare SILAC-based protein enrichment at these two chromosomal landmarks.
276 Of the 55 and 460 C-BERST enriched proteins at ALT+ telomeres and centromeres,
277 respectively, 36 were identified in both ($p = 1.31 \times 10^{-57}$) (**Supplementary Fig. 10**). Significant
278 GO terms for these 36 overlapping proteins include DNA replication, regulation of signal
279 transduction by p53 class mediator, strand displacement, double strand break repair via
280 homologous recombination, and DNA repair, each of which would be expected for both
281 categories of chromosomal elements. Significantly, all CENP factors were found among the 424
282 non-overlapping proteins from the sgAlpha centromeric dataset. Conversely, the 19 telomere-
283 specific hits (enriched only with sgTelo and not sgAlpha) include five of the six shelterin subunits
284 (TRF1, TERF2IP, TIN2, POT1, TPP1); intriguingly, the sixth (TRF2) has previously been
285 reported to associate with CENP-F[29]. These results provide strong evidence that C-BERST
286 successfully measures subnuclear protein enrichment at distinct chromosomal elements.
287
288

Discussion

We demonstrate that C-BERST successfully maps subnuclear proteomes associated with genomic landmarks. Using the extensively investigated ALT telomeric proteome as an initial benchmark for our ratiometric implementation of C-BERST, we recover approximately 44% of known ALT-associated proteins (34 of 78, **Supplementary Tables 2 and 3**) as strongly enriched at telomeres, in addition to factors involved in all reported biological processes that contribute to ALT. Combining C-BERST with SILAC made it possible to set very high enrichment cut-offs (BH-adjusted $p < 0.01$, $\log_2 FC \geqslant 2.5$]) while still retaining excellent representation of the known telomeric or centromeric factors that we employed as benchmarks; lower thresholds can be set where appropriate to cast a wider net for previously unknown factors. We used fluorescence microscopy to validate a strongly enriched C-BERST hit (SLX4IP) whose telomeric localization had not been previously confirmed, and another (RPA3) that was missed by previous subnuclear proteomic approaches such as BioID[17] and PICh[18]. Importantly, we extended C-BERST to a second category of genomic elements (alpha satellite centromeric sequences) and identified a distinct set of enriched factors that included many known centromeric factors, and that largely excluded known telomere-specific factors (including those enriched by telomeric C-BERST). These results provide strong indications that C-BERST can successfully profile subnuclear proteomes based upon proximity to specific classes of genomic sequence. C-BERST is also compatible with other ratiometric approaches such as tandem mass tagging (TMT)[30]. Although ratiometric approaches increase the sensitivity and quantitative rigor of C-BERST enrichment, more economical label-free quantitation can also be used successfully with C-BERST, as we showed for telomeres.

By combining the flexibility of RNA-guided dSpyCas9 genome binding with the efficiency and rapid kinetics of APEX2-catalyzed 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 the longer labeling procedures often necessary for CasID[4], CAPTURE[5], and related BirA*-based approaches. Furthermore, we found 40-60 million cells to provide a sufficient sample size for telomeric and centromeric C-BERST, in contrast to the $>10^9$ cells reported as inputs for CasID, CAPTURE, and PICh[4, 5, 18]. The ability to apply C-BERST to smaller populations of cells provides an obvious cost savings, and may also

320 be important when available cell numbers are limiting due to their proliferation properties or
321 challenges in their manipulation. Finally, C-BERST and BirA*-based methods favor
322 biotinylation of distinct sets of proteins by virtue of their different labeling specificities (lysines for
323 BirA*, and predominantly tyrosines for C-BERST); using these approaches in tandem would
324 likely diminish the number of false negatives resulting from inefficient labeling due to differences
325 in the surface-accessible amino acid distribution or the suitability of certain peptides for MS
326 analysis.
327 Importantly, C-BERST promises to augment and extend Hi-C and related methods by
328 linking conformationally important cis-elements with the factors that associate with them. Guide
329 RNA multiplexing should enable the extension of C-BERST subnuclear proteomics to single-
330 copy, non-repetitive loci. In the meantime, many types of repetitive elements within the genome,
331 like telomeres and centromeres, play critically important roles in chromosome maintenance and
332 function in ways that depend upon their associated proteins; C-BERST provides an unbiased
333 method for sampling subnuclear, locus-specific proteomics at these elements to define protein
334 factors that are critical to their functions.
335
336 **METHODS**
337 Methods, data files, and any associated references are available in the online version of the paper.
338 *Note: Any Supplemental and Source Data files are available in the online version of the paper.*
339
340 **ACKNOWLEDGEMENTS**
341 We are grateful to all members of the Sontheimer, Wolfe and Dekker labs for advice and
342 discussions, Tom Fazzio, Samyabrata Bhaduri and Michael Green for helpful feedback, Hanhui
343 Ma, Tong Wu, David Grünwald, and Thoru Pederson for reagents, the Flow Cytometry Core
344 Facility at UMass Medical School for cell sorting, and Lingji Zhu for assistance with figure
345 preparation. This work was supported by 4D Nucleome grant U54 DK107980 from the
346 National Institutes of Health to J.D., S.A.W. and E.J.S.
347
348 **AUTHOR CONTRIBUTIONS**
349 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
350 experiments. X.D.G. and T.R. performed C-BERST and ChIP-seq experiments, and J.L.

351 conducted mass spectrometry procedures. X.D.G and L.-C.T. processed the fluorescence images,
352 A.M. processed flow cytometry data, X.D.G., Y.D., and J.L. processed mass spectrometry data,
353 and L.J.Z. conducted statistical analyses. All co-authors interpreted the data. X.D.G. and E.J.S
354 wrote the manuscript, and all authors revised and edited the manuscript.
355

356 **COMPETING FINANCIAL INTERESTS**

357 The authors declare no competing financial interests.
358

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

361 **REFERENCES**
362

363 1. Davies, J.O., Oudelaar, A.M., Higgs, D.R. & Hughes, J.R. How best to identify
364 chromosomal interactions: a comparison of approaches. *Nat Methods* **14**, 125-134 (2017).
365 2. Dominguez, A.A., Lim, W.A. & Qi, L.S. Beyond editing: repurposing CRISPR-Cas9 for
366 precision genome regulation and interrogation. *Nat Rev Mol Cell Biol* **17**, 5-15 (2016).
367 3. Roux, K.J., Kim, D.I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein
368 identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **196**, 801-810
369 (2012).
370 4. Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F. & Leonhardt, H. Determination of
371 local chromatin composition by CasID. *Nucleus* **7**, 476-484 (2016).
372 5. Liu, X. et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* **170**,
373 1028-1043 e1019 (2017).
374 6. Hung, V. et al. Proteomic mapping of the human mitochondrial intermembrane space in live
375 cells via ratiometric APEX tagging. *Mol Cell* **55**, 332-341 (2014).
376 7. Rhee, H.W. et al. Proteomic mapping of mitochondria in living cells via spatially restricted
377 enzymatic tagging. *Science* **339**, 1328-1331 (2013).
378 8. Cesare, A.J. & Reddel, R.R. Alternative lengthening of telomeres: models, mechanisms and
379 implications. *Nat Rev Genet* **11**, 319-330 (2010).
380 9. Dilley, R.L. et al. Break-induced telomere synthesis underlies alternative telomere
381 maintenance. *Nature* **539**, 54-58 (2016).
382 10. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized
383 CRISPR/Cas system. *Cell* **155**, 1479-1491 (2013).
384 11. Ma, H. et al. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad*
385 *Sci U S A* **112**, 3002-3007 (2015).
386 12. Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G. & Wandless, T.J. A rapid,
387 reversible, and tunable method to regulate protein function in living cells using synthetic
388 small molecules. *Cell* **126**, 995-1004 (2006).
389 13. Ma, H. et al. CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *J Cell Biol*
390 **214**, 529-537 (2016).
391 14. Knight, S.C. et al. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*
392 **350**, 823-826 (2015).

393 15. de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres.
394 *Genes Dev* **19**, 2100-2110 (2005).
395 16. Lenain, C. et al. The Apollo 5' exonuclease functions together with TRF2 to protect
396 telomeres from DNA repair. *Curr Biol* **16**, 1303-1310 (2006).
397 17. Garcia-Exposito, L. et al. Proteomic Profiling Reveals a Specific Role for Translesion DNA
398 Polymerase eta in the Alternative Lengthening of Telomeres. *Cell Rep* **17**, 1858-1871 (2016).
399 18. Dejardin, J. & Kingston, R.E. Purification of proteins associated with specific genomic Loci.
400 *Cell* **136**, 175-186 (2009).
401 19. Sobinoff, A.P. & Pickett, H.A. Alternative Lengthening of Telomeres: DNA Repair Pathways
402 Converge. *Trends Genet* **33**, 921-932 (2017).
403 20. Grudic, A. et al. Replication protein A prevents accumulation of single-stranded telomeric
404 DNA in cells that use alternative lengthening of telomeres. *Nucleic Acids Res* **35**, 7267-7278
405 (2007).
406 21. Meissner, B. et al. Frequent and sex-biased deletion of SLX4IP by illegitimate V(D)J-
407 mediated recombination in childhood acute lymphoblastic leukemia. *Hum Mol Genet* **23**, 590-
408 601 (2014).
409 22. Lin, Y.L. et al. The evolutionarily conserved zinc finger motif in the largest subunit of human
410 replication protein A is required for DNA replication and mismatch repair but not for
411 nucleotide excision repair. *J Biol Chem* **273**, 1453-1461 (1998).
412 23. Chen, B. et al. Expanding the CRISPR imaging toolset with Staphylococcus aureus Cas9 for
413 simultaneous imaging of multiple genomic loci. *Nucleic Acids Res* **44**, e75 (2016).
414 24. Foltz, D.R. et al. The human CENP-A centromeric nucleosome-associated complex. *Nat Cell*
415 *Biol* **8**, 458-469 (2006).
416 25. Verdaasdonk, J.S. & Bloom, K. Centromeres: unique chromatin structures that drive
417 chromosome segregation. *Nat Rev Mol Cell Biol* **12**, 320-332 (2011).
418 26. Carmena, M., Wheelock, M., Funabiki, H. & Earnshaw, W.C. The chromosomal passenger
419 complex (CPC): from easy rider to the godfather of mitosis. *Nat Rev Mol Cell Biol* **13**, 789-803
420 (2012).
421 27. Yan, Z. et al. A histone-fold complex and FANCM form a conserved DNA-remodeling
422 complex to maintain genome stability. *Mol Cell* **37**, 865-878 (2010).

423 28. Kabeche, L., Nguyen, H.D., Buisson, R. & Zou, L. A mitosis-specific and R loop-driven
424 ATR pathway promotes faithful chromosome segregation. *Science* **359**, 108-114 (2018).
425 29. Giannone, R.J. et al. The protein network surrounding the human telomere repeat binding
426 factors TRF1, TRF2, and POT1. *PLoS One* **5**, e12407 (2010).
427 30. Zhang, L. & Elias, J.E. Relative Protein Quantification Using Tandem Mass Tag Mass
428 Spectrometry. *Methods Mol Biol* **1550**, 185-198 (2017).
429

430 **ONLINE METHODS**

431

432 **Construction of C-BERST plasmids.** The Shield1- and doxycycline-inducible dSpyCas9-
433 mCherry-APEX2 construct was made by subcloning Flag-APEX2 from Flag-APEX2-NES
434 (Addgene 49386) into DD-dSpyCas9-mCherry[13] using the pHAGE backbone. Two additional
435 NLSs (SV40 and nucleoplasmin NLS) were inserted at each terminus to improve nuclear
436 localization. The sequence of the final plasmid is provided in the **Supplementary Note**. The
437 sgTelo-encoding construct was created by replacing the C3-guide RNA sequence (pCMV_C3-
438 sgRNA_2XBroccoli/pPGK_TetR_P2A_BFP) with sgTelo sequences (using a plasmid provided
439 by Hanhui Ma and Thoru Pederson). Non-specific sgRNA (sgNS)[14] and sgAlpha were
440 constructed similarly. SLX4IP-turboGFP plasmid was obtained from OriGene (catalog number:
441 RG220896). RPA3-turboGFP was made by replacing the SLX4IP coding sequence with the
442 human RPA3 coding sequence.

443

444

445 **Cell culture and cell line construction.** Human U2OS cells obtained from Thoru
446 Pederson's lab (originally obtained from ATCC) were cultured in Dulbecco-modified Eagle's
447 Minimum Essential Medium (DMEM; Life Technologies) supplemented with 10% (vol/vol) FBS
448 (Sigma). Lentiviral transduction was as described[13]. Six-fold higher titers of sgRNA-encoding
449 lentiviruses were used for transduction relative to dSpyCas9-APEX2 lentivirus. Stably transduced
450 cells were grown under the same conditions as the parental U2OS cells.

451

452

453 **Flow cytometry.** One day before performing FACS, dox (Sigma; 2 μg/ml) and Shield1
454 (Clontech; 250 nM) were added to the media. ~2 x 10^6 cells expressing dSpyCas9-mCherry-
455 APEX2 and BFP sgRNA were selected by FACSAria cell sorter or analyzed with MacsQuant®
456 VYB. Both instruments are equipped with 405- and 561-nm excitation lasers, and the emission
457 signals were detected by using filters at 450/50 nm (wavelength/bandwidth) for BFP, and
458 610/20 nm (FACSAria) or 615/20nm (MacsQuant) for mCherry. Bulk population and single
459 cells (**Supplementary Fig. 1b**) were sorted into plates containing 1% GlutaMAX, 20% FBS,
460 and 1% penicillin/streptomycin in DMEM medium.

461

462 **Fluorescence microscopy.** U2OS cells expressing sgRNA were seeded onto 170 µm, 35 × 10
463 mm glass-bottom dishes (Eppendorf) supplemented with dox and Shield1 21 hours before
464 imaging. Live cells were imaged with a Leica DMi8 microscope equipped with a Hamamatsu
465 camera (C11440-22CU), a 63x oil objective lens, and Microsystems software (LASX). Further
466 imaging processing was done with MetaMorph (Molecular Devices). Image contrast was set to
467 ease visualization of cell, foci and nucleoplasmic background.
468

469 **Immunofluorescence microscopy.** Cells for immunofluorescence microscopy were grown
470 on glass coverslips. The transfected cells or normal cells were fixed for 15 minutes in 2%
471 paraformaldehyde in PHEM [0.05 M PIPES/0.05 M HEPES (pH 7.4), 0.01 M EGTA, 0.01 M
472 $MgCl_2$], followed by a 2-min. extraction with 0.1% Triton X-100 in PHEM. After PBS washes,
473 the cells were blocked by 1% BSA/1X TBST at 4°C overnight. Cells were first incubated with
474 primary antibodies for two hours at room temperature and washed three times with blocking
475 solution (10 minutes/wash). Cells were then incubated with secondary antibodies for one hour at
476 room temperature, followed by another three blocking solution washes and two PBS washes[31].
477 Cells were mounted with ProLong antifade and visualized by fluorescence microscopy as
478 described above. Neutravidin conjugated with OG488 experiment was described previously[6].
479

480 **C-BERST biotinylation protocol.** Six 15cm plates of U2OS cells (~6 x 10^7) expressing
481 specific (sgTelo or sgAlpha) or nonspecific (sgNS) sgRNAs were used in this assay. Dox (2 µg/ml)
482 and Shield1 (250 nM) were added 21 hours before biotinylation. Cells were then incubated with
483 500 µM biotin-phenol (BP) (Adipogen) for 30 minutes at 37°C. 1 mM H_2O_2 was then added to
484 initiate of biotinylation for 1 minute on a horizontal shaker at room temperature. Six 15cm plates
485 of sgTelo- or sgAlpha-expressing cells were treated in parallel, but without H_2O_2 addition, as a
486 negative control. Quencher solution (5 mM trolox, 10 mM sodium ascorbate, and 10 mM
487 sodium azide) was added to stop the reaction, and cells were washed five times (three quencher
488 washes and two DPBS washes) to continue the quench and to remove excess BP.
489

490 **Enrichment of biotinylated proteins.** Cells were scraped off the plates and used for the
491 preparation of isolated nuclei[32]. Nuclei were washed with DPBS before lysis. RIPA lysis buffer

492 [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.125% SDS, 0.125% sodium deoxycholate and 1%
493 Triton X-100 in Millipore water) with 1x freshly supplemented Halt Protease Inhibitor were used
494 to lyse the cells for 10 minutes on ice. Cell lysates in 1.5 ml Eppendorf tubes were sonicated for
495 15 minutes with a Diagenode Bioruptor with 30s on/off cycles at high intensity. Cell lysates were
496 clarified by centrifugation at 13,000 rpm for 10 minutes. Clarified protein samples (~3.5 mg)
497 were subjected to 400 µl Dynabeads MyOne Streptavidin T1 affinity purification overnight at
498 4°C. Each bead sample was washed with a series of buffers to remove non-specifically bound
499 proteins: twice with RIPA lysis buffer, once with 1 M KCl, once with 0.1 M Na_2CO_3, once with
500 2 M urea in 10mM Tris-HCl, pH 8.0, and twice with RIPA lysis buffer. Proteins were eluted in
501 70 µl 3x protein loading buffer supplemented with 2 mM biotin and 20 mM DTT with heating
502 for 10min at 95˚C[6]. 50 µl eluents were loaded and run on a 4-12% SDS-PAGE gel (Bio-Rad) and
503 run approximately 1cm off the loading well for in-gel digestion and LC-MS/MS analysis. The
504 gel-fractionated sample used for LC-MS/MS (see below) corresponded to proteins from ~4 x 10^7
505 cells.
506
507 **Western blotting.** Protein concentrations of the cell lysates were determined by BCA assay
508 (Thermo). 50 ug of each sample was mixed with protein loading buffer, boiled, and separated in
509 SDS-PAGE gels. Proteins were transferred to PVDF membrane (Millipore), and blotted with
510 Streptavidin-HRP (Thermo), or with anti-mCherry (Abcam) or anti-HDAC1 (Bethyl) antibodies.
511 Additional details of the anti-SLX4IP and anti-RPA3 western analyses are described in the figure
512 legends.
513
514 **mCherry affinity purification of dSpyCas9-mCherry-APEX2 captured DNA and**
515 **sequencing.** 1 × 10^7 U2OS cells stably expressing dCas9-mCherry-APEX2 transduced with
516 sequence-targeting or non-specific sgRNAs were washed with PBS, fixed with 1% formaldehyde
517 for 10 min and quenched with 0.125 M glycine for 5 min. Cells were harvested using a plate
518 scraper and lysed in RIPA cell lysis buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.125%
519 SDS, 0.125% sodium deoxycholate and 1% Triton X-100 in Millipore water] with 1x freshly
520 supplemented Halt Protease Inhibitor for 10 minutes on ice. Cell lysates were centrifuged at
521 2,300 x g for 5 min at 4°C to isolate nuclei. Nuclei were suspended in 500 µl of RIPA nuclear
522 lysis buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5% SDS, 0.125% sodium deoxycholate

523 and 1% Triton X-100 in Millipore water] with 1x freshly supplemented Halt Protease Inhibitor
524 and subjected to sonication to shear chromatin fragments to an average size of 200-500 bp on a
525 Diagenode Bioruptor with 30s on/off cycles at high intensity for 15 minutes. Fragmented
526 chromatin was centrifuged at 16,100 x g for 10 min at 4°C. 450 µl of supernatant was transferred
527 to a new microcentrifuge tube. 4 µg anti-mCherry antibody (Thermo PA5-34974) was added to
528 each sample and incubated at 4°C for 3h. 50µl of blocked Protein G Dynabeads (Thermo
529 10003D) was added to each sample and rotated at 4°C overnight. After overnight incubation,
530 Dynabeads were washed seven times as described above for selection of biotinylated proteins.
531 Chromatin was eluted from Dynabeads in 200µl elution buffer [50 mM Tris-HCl (pH 8.0), 10
532 mM EDTA, 1% SDS] and transferred to a new microcentrifuge tube. Eluted chromatin was
533 treated with 1 µl RNase A and incubated overnight at 65°C to reverse crosslinks. 7.5 µl of 20
534 mg/ml proteinase K was added to each sample followed by incubation for 2h at 50°C. ChIP
535 DNA was then incubated with 1ml Buffer PB (QIAGEN) and 10 µl of 3M sodium acetate pH 5.2
536 at 37°C for 30 minutes. DNA was purified using QIAGEN quickspin column.
537 15 ng of ChIP DNA was processed for library preparation using the NEBNext ChIP-seq Library
538 Prep Kit (New England Biolabs) according to the manufacturer's protocol.
539 15 ng of ChIP DNA was end-repaired using NEBNext End Repair module (NEB Cat. E6050)
540 and purified with 1.8x AMPure XP beads (Beckman-Coulter Cat. A63880). End-repaired DNA
541 was processed in a dA-tailing reaction using NEBNext dA-Tailing module (NEB Cat. E6053) and
542 purified with 1.8x AMPure XP beads. Adaptor oligos 1 (5'-pGAT CGG AAG AGC ACA CGT
543 CT-3') and 2 (5'-ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT-3') used in Y-
544 shaped adapter mix were ligated to dA-tailed DNA according to ref. [33] and purified with 1.5x
545 AMPure XP beads. Ligated DNA was incubated in a thermal cycler (98°C for 40s, 65°C for 30s,
546 and 72°C 30s) with Illumina barcode primers 2-1 (5'-CAA GCA GAA GAC GGC ATA CGA
547 GAT CGT GATGTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3'), 2-2 (5'-
548 CAA GCA GAA GAC GGC ATA CGA GAT ACA TCGGTG ACT GGA GTT CAG ACG
549 TGT GCT CTT CCG ATC T-3'), 2-3 (5'-CAA GCA GAA GAC GGC ATA CGA GAT GCC
550 TAAGTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3') and NEB Q5
551 Polymerase Master Mix. Primer 1 (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT
552 CTT TCC CTA CAC GAC GCT CTT CCG ATC T-3') was added to mix for 10 cycles (98°C

553 for 10s, 65°C for 30s, 72°C for 30s), followed by incubation at 72°C for 3 minutes. PCR-
554 enriched DNA was purified with 1x AMPure XP beads.
555 Raw Illumina sequencing reads of 150 nucleotide length were processed as fastq files in R. Reads
556 were trimmed using the Bioconductor ShortRead R package at positions which contained 2
557 nucleotides in a 5-nucleotide bin with a quality encoding less than phred score = 20. Reads with
558 at least one $(TTAGGG)_4$ or $(CCCTAA)_4$ segment constituted a "hit", and were counted using
559 the Bioconductor Biostrings R package. (number of hits / total trimmed reads) was calculated to
560 assess the specificity of Cas9-mCherry-APEX2 for each sample.
561
562 **SILAC labeling.** On day 0, early-passage, sorted, stably transduced sgTelo or sgAlpha U2OS
563 cells were grown in heavy SILAC media, which contained L-arginine-$^{13}C_6$, $^{15}N_4$ (Arg10) and L-
564 lysine-$^{13}C_6$, $^{15}N_2$ (Lys8) (Sigma). Stable sgNS cells were grown in medium SILAC media, which
565 contained L-arginine-$^{13}C_6$ (Arg6) and L-lysine-4,4,5,5-d_4 (Lys4) (Sigma). Untransduced U2OS
566 cells were grown in light SILAC media, which contained L-arginine (Arg0) and L-lysine (Lys0)
567 (Sigma). Cells were grown for more than 10 days (>5 passages) to allow for sufficient
568 incorporation of the isotopes. On day 11, dox and Shield1 were added to each isotope culture (4
569 plates for each cell line) 21h before BP and H_2O_2 treatment. The biotinylation, nuclei isolation,
570 and cell lysis followed the procedure described above. Before streptavidin affinity purification,
571 equal amount of proteins measured by Pierce™ BCA Protein Assay Kit (~1mg from each
572 isotope sample) were mixed in a 1:1:1 ratio (H:M:L). Streptavidin affinity purification and sample
573 wash were described above. Proteins were eluted in 50 µl 3x protein loading buffer supplemented
574 with 2 mM biotin and 20 mM DTT with heating for 10min at 65˚C. 50 µl eluents were loaded
575 and run approximately to the center of the lane on a 4-12% SDS-PAGE gel (Bio-Rad). The
576 coomassie-stained protein bands were excised and cut to five slices for in-gel digestion and LC-
577 MS/MS analysis.
578
579 **LC-MS/MS and proteomic analyses for LFQ.** Unresolved protein bands from SDS-
580 PAGE were cut into 1x1 mm pieces and placed in 1.5ml Eppendorf tubes with 1ml of water.
581 After 30 min, water was removed and replaced with 70 µl of 250 mM ammonium bicarbonate.
582 Proteins were then reduced by the addition of 20 µl of 45 mM 1,4-dithiothreitol, incubated at
583 50˚C for 30 min, cooled to room temperature, alkylated with 20 µl of 100 mM iodoacetamide for

584 30 min, and washed twice with 1 ml water. The water was removed and replaced with 1 ml of 50
585 mM ammonium bicarbonate: acetonitrile (1:1) and incubated at room temperature for 1 hr. The
586 solvent was then replaced with 200 µl acetonitrile, removed, and the pieces dried in a Speed Vac.
587 Gel pieces were then rehydrated in 75 µl of 4 ng/µl sequencing-grade trypsin (Promega) in 0.01%
588 ProteaseMAX Surfactant (Promega) in 50 mM ammonium bicarbonate and incubated at 37˚C
589 for 21 hr. The supernatant was then removed to a 1.5 ml Eppendorf tube, the gel pieces further
590 dehydrated with 100 µl of acetonitrile: 1% (v/v) formic acid (4:1), and the combined supernatants
591 dried on a Speed Vac. Peptides were then reconstituted in 25 µl of 5% acetonitrile containing 0.1%
592 (v/v) trifluoroacetic acid for LC-MS/MS.

593 Samples were analyzed on a NanoAcquity UPLC (Waters Corporation) coupled to a Q Exactive
594 (Thermo Fisher Scientific) hybrid mass spectrometer. In brief, 1.0 µl aliquots were loaded at 4
595 µl/min onto a custom-packed fused silica precolumn (100 µm ID) with Kasil frit containing 2 cm
596 Magic C18AQ (5µm, 100Å) particles (Bruker Corporation). Peptides were then separated on a
597 75µm ID fused silica analytical column containing 25 cm Magic C18AQ (3µm, 100Å) particles
598 (Bruker) packed in-house into a gravity-pulled tip. Peptides were eluted at 300 nl/min with a
599 linear gradient from 95% solvent A (0.1% (v/v) formic acid in water) to 35% solvent B (0.1%
600 (v/v) formic acid in acetonitrile) in 60 min. Data was acquired by data-dependent acquisition
601 according to a published method[34]. Briefly, MS scans were acquired from *m/z* 300-1750 at a
602 resolution of 70,000 (*m/z* 200) and followed by ten tandem mass spectrometry scans using HCD
603 fragmentation using an isolation width of 1.6 Da, a collision energy of 27%, and a resolution of
604 17,500 (*m/z* 200). Raw data files were processed with Proteome Discoverer (Thermo, version
605 2.1.1.21) and searched with Mascot (Matrix Science, version 2.6) against the SwissProt *Homo*
606 *sapiens* database. Search parameters used tryptic specificity considering up to 2 missed cleavages,
607 a parent mass tolerance of 10 ppm, and a fragment mass tolerance of 0.05 Da. Fixed
608 modification of carbamidomethyl cysteine was considered as were variable modifications of N-
609 terminal acetylation, N-terminal conversion of Gln to pyroGlu, oxidation of methionine, and
610 biotin-phenol conjugation of tyrosine. Results were loaded into Scaffold (Proteome Software Inc.,
611 version 4.8.4) for peptide and protein validation and quantitation using the Peptide Prophet and
612 Protein Prophet algorithms[35, 36]. The threshold for peptides was set to 80% (1.1% FDR) and 90%
613 for proteins (3-peptide minimum). Contaminants such as human keratin were included in all
614 statistical analyses and removed from the figures.

615

616 **LC-MS/MS and proteomic analyses for SILAC.** A fully resolved SDS-PAGE was cut into
617 5 fractions and each fraction was processed separately as described. Gel bands were cut into 1x1
618 mm pieces and placed in 1.5 mL eppendorf tubes with 1mL of water for 30 min. The water was
619 removed and 200µl of 250 mM ammonium bicarbonate was added. For reduction, 20 µl of a 45
620 mM solution of 1,4-dithiothreitol (DTT) was added and the samples were incubated at 50˚C for
621 30 min. The samples were cooled to room temperature and then, for alkylation, 20 µl of a 100
622 mM iodoacetamide solution was added and allowed to react for 30 min. The gel slices were
623 washed twice with 1 mL water. The water was removed and 1mL of 50:50 (50 mM ammonium
624 bicarbonate:acetonitrile) was placed in each tube and samples were incubated at room
625 temperature for 1h. The solution was then removed and 200 µl of acetonitrile was added to each
626 tube, at which point the gels slices turned opaque white. The acetonitrile was removed and gel
627 slices were further dried in a Speed Vac (Savant Instruments, Inc.). Gel slices were rehydrated in
628 100 µl of 4ng/µl of sequencing-grade trypsin (Sigma) in 0.01% ProteaseMAX Surfactant
629 (Promega):50 mM ammonium bicarbonate. Additional bicarbonate buffer was added to ensure
630 complete submersion of the gel slices. Samples were incubated at 37˚C for 18 hrs. The
631 supernatant of each sample was then removed and placed in a separate 1.5 mL eppendorf tube.
632 Gel slices were further extracted with 200 µl of 80:20 (acetonitrile:1% formic acid). The extracts
633 were combined with the supernatants of each sample. The samples were then completely dried
634 down in a Speed Vac.
635 Tryptic peptide digests were reconstituted in 25 µL 5% acetonitrile containing 0.1% (v/v)
636 trifluoroacetic acid and separated on a NanoAcquity (Waters) UPLC. In brief, a 3.0 µL injection
637 was loaded in 5% acetonitrile containing 0.1% formic acid at 4.0 µL/min for 4.0 min onto a 100
638 µm I.D. fused-silica pre-column packed with 2 cm of 5 µm (200Å) Magic C18AQ (Bruker-
639 Michrom) and eluted using a gradient at 300 nL/min onto a 75 µm I.D. analytical column
640 packed with 25 cm of 3 µm (100Å) Magic C18AQ particles to a gravity-pulled tip. The solvents
641 were A, water (0.1% formic acid); and B, acetonitrile (0.1% formic acid). A linear gradient was
642 developed from 5% solvent A to 35% solvent B in 60 minutes. Ions were introduced by positive
643 electrospray ionization via liquid junction into a Q Exactive hybrid mass spectrometer (Thermo).
644 Mass spectra were acquired over *m/z* 300-1750 at 70,000 resolution (*m/z* 200) and data-
645 dependent acquisition selected the top 10 most abundant precursor ions for tandem mass

646 spectrometry by HCD fragmentation using an isolation width of 1.6 Da, collision energy of 27,
647 and a resolution of 17,500.
648 Raw data files were peak processed with Mascot Distiller (Matrix Science, version 2.6) prior to
649 database searching with Mascot Server (version 2.6) against the *Uniprot_Human* database. Search
650 parameters included trypsin specificity with two missed cleavages. The variable modifications of
651 oxidized methionine, pyroglutamic acid for N-terminal glutamine, N-terminal acetylation of the
652 protein, biotin-phenol on tyrosine and a fixed modification for carbamidomethyl cysteine were
653 considered. For SILAC labels, the medium samples were labeled with Lys4 and Arg6 and the
654 heavy samples were labeled with Lys8 and Arg10. The mass tolerances were 10 ppm for the
655 precursor and 0.05 Da for the fragments. SILAC ratio quantitation was accomplished using
656 Mascot Distiller and the results from Mascot Distiller were loaded into the Scaffold Viewer
657 (Proteome Software, Inc., version 4.8.4) for peptide/protein validation and SILAC label
658 quantitation. For SILAC experiments, protein identification was subject to a two-peptide cut-off.
659 For proteins detectable in the H sample but that lack an empirical H/L ratio value (due to low
660 background detection in the L sample), peak areas of all the identified peptides in the Distiller file
661 were used to calculate H/L ratios.
662
663 **Data analysis.** Data was first filtered to exclude proteins detected in only one of the dCas9-
664 mCherry-APEX2/sgTelo (+BP, +H_2O_2) ("S1") replicates, followed by log2 transformation. Prior
665 to the $\log_2$ transformation, iBAQ values of 0 were replaced with the smallest iBAQ value from
666 the corresponding sample in dCas9-mCherry-APEX2/sgTelo (+BP, -H_2O_2) ("S2") or dCas9-
667 mCherry-APEX2/sgNS (+BP, +H_2O_2) ("S3") to avoid generation of infinite ratios. Moderated t-
668 test with a paired design was used to compare the log2 transformed iBAQ values between S1 and
669 S3, S1 and S2, and S2 and S3 using limma package[37]. To adjust for multiple comparisons, *p*
670 values were adjusted using the Benjamini-Hochberg (BH) method[38]. Proteins were selected for
671 subsequent analysis if they were (*i*) significantly enriched in both S1 vs. S3 and S1 vs. S2, (*ii*) not
672 enriched in S2 vs. S3, and (*iii*) if S1/S3 and S1/S2 ratios were greater than 2.
673 Similarly, SILAC datasets were filtered to exclude proteins with H/M ratios detected in only one
674 of the biological replicates. Detection in a biological replicate required identification in at least
675 two of the three technical replicates that were done for each biological replicate; median values
676 from the technical replicates were used for subsequent analyses. Proteins with BH-adjusted *p*

677 values less than 0.01 (moderated t-test described above) are considered statistically significant.
678 Proteins (with BH-adjusted p values < 0.01 and $\log_2$ fold change ≥ 2.5) were selected for
679 subsequent GO (David Bioinformatics) and overlap analysis. To determine whether the proteins
680 identified in this experiment overlap significantly with three published datasets, hypergeometric
681 test was used. Hypergeometric test was also used for testing the overlapping proteins between C-
682 BERST telomere IDs and centromere IDs.
683
684

685 31. Follit, J.A., Tuft, R.A., Fogarty, K.E. & Pazour, G.J. The intraflagellar transport protein
686 IFT20 is associated with the Golgi complex and is required for cilia assembly. *Mol Biol Cell* **17**,
687 3781-3792 (2006).
688 32. Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure.
689 *Nature* **502**, 59-64 (2013).
690 33. Zhang, Z., Theurkauf, W.E., Weng, Z. & Zamore, P.D. Strand-specific libraries for high
691 throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence* **3**, 9
692 (2012).
693 34. Kelstrup, C.D. et al. Rapid and deep proteomes by faster sequencing on a benchtop
694 quadrupole ultra-high-field Orbitrap mass spectrometer. *J Proteome Res* **13**, 6187-6195 (2014).
695 35. Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. Empirical statistical model to
696 estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal*
697 *Chem* **74**, 5383-5392 (2002).
698 36. Nesvizhskii, A.I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying
699 proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646-4658 (2003).
700 37. Smyth, G.K. Linear models and empirical Bayes methods for assessing differential expression
701 in microarray experiments. *Stat Appl Genet Molec Biol* **3**, 1-25 (2004).
702 38. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful
703 approach to multiple testing. *J R Stat Soc B* **57**, 289-300 (1995).

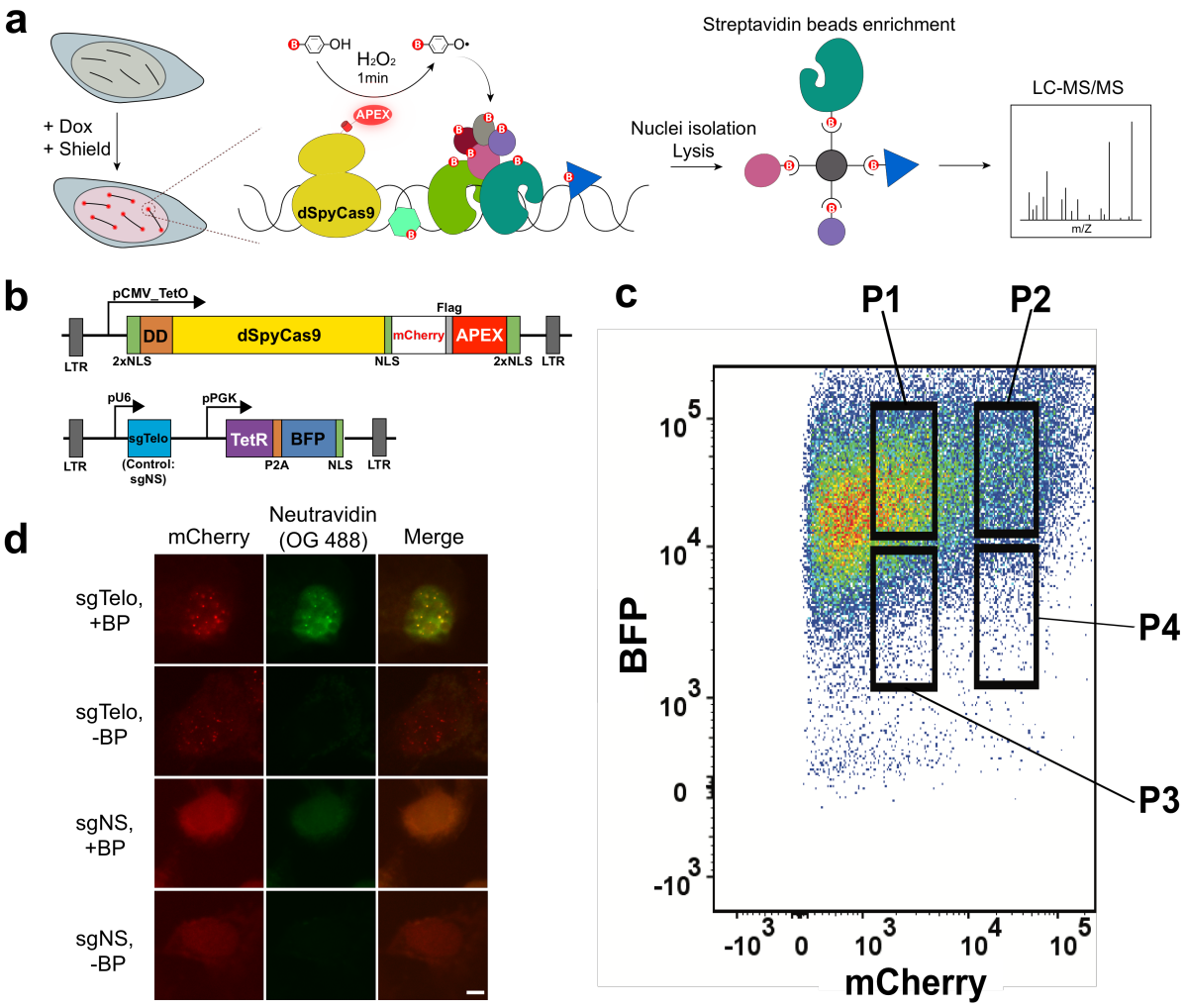

a
+ Dox
+ Shield
H2O2
1min
APEX
dSpyCas9
Streptavidin beads enrichment
Nuclei isolation
Lysis
LC-MS/MS
m/Z
b
pCMV_TetO
Flag
DD
dSpyCas9
mCherry
APEX
LTR
2xNLS
NLS
2xNLS
LTR
pU6
pPGK
sgTelo
(Control: sgNS)
TetR
BFP
LTR
P2A
NLS
LTR
c
P1
P2
P3
P4
BFP
mCherry
d
mCherry
Neutravidin (OG 488)
Merge
sgTelo, +BP
sgTelo, -BP
sgNS, +BP
sgNS, -BP

Figure 1

705 **Figure 1** Using C-BERST to biotinylate telomere-associated proteins in living human cells. (**a**)
706 Diagram of the C-BERST workflow. U2OS cells stably expressing sgRNA and inducible
707 dSpyCas9-APEX2 are generated by lentiviral transduction. Following dox and Shield1 induction
708 (21 h), cells are incubated with biotin-phenol (BP, 30 min) and then H_2O_2 (1 min) to activate a
709 burst of biotin-phenoxyl radical generation by dSpyCas9-APEX2, leading to proximity-labeling
710 of nearby proteins. Following quenching, nuclei isolation and protein extraction, biotinylated
711 proteins are enriched by streptavidin selection and analyzed by LC-MS/MS. (**b**) The dSpyCas9-
712 mCherry-APEX2 and sgRNA lentiviral expression constructs. Top: dSpyCas9-mCherry-APEX2
713 under the control of the pCMV_TetO inducible promoter. The mCherry fusion is included to
714 enable quantification of dSpyCas9 expression level as well as its subcellular localization. NLS,
715 nuclear localization signal; LTR, long terminal repeat; DD, Shield1-repressible degradation
716 domain. Bottom: sgRNA/TetR/BFP expression construct. pU6, U6 promoter; pPGK, PGK
717 promoter; sgTelo, telomere-targeting sgRNA; sgNS, non-specific sgRNA; tetR, tet repressor;
718 P2A, 2A self-cleaving peptide; BFP, blue fluorescent protein. (**c**) FACS sorting of untransduced
719 (blue) and mCherry- and BFP-positive cells (red). The P1 population corresponds to high BFP (as
720 a surrogate for sgRNA and TetR) and low mCherry expression, providing optimal signal-to-noise
721 ratio to maximize the fraction of telomere-localized dSpyCas9-mCherry-APEX2. (**d**)
722 Fluorescence imaging of dSpyCas9-mCherry-APEX2 labeling in cells. Stable sgTelo and sgNS
723 cells were labeled live as described in (**a**) or were only supplemented with H_2O_2 as a no-labeling
724 control. Cells were then fixed and stained with neutravidin conjugated with OG488 to visualize
725 biotinylated proteins. dCas9-mCherry-APEX2 localization are indicated by mCherry
726 fluorescence. Scale bar, 5 μm.

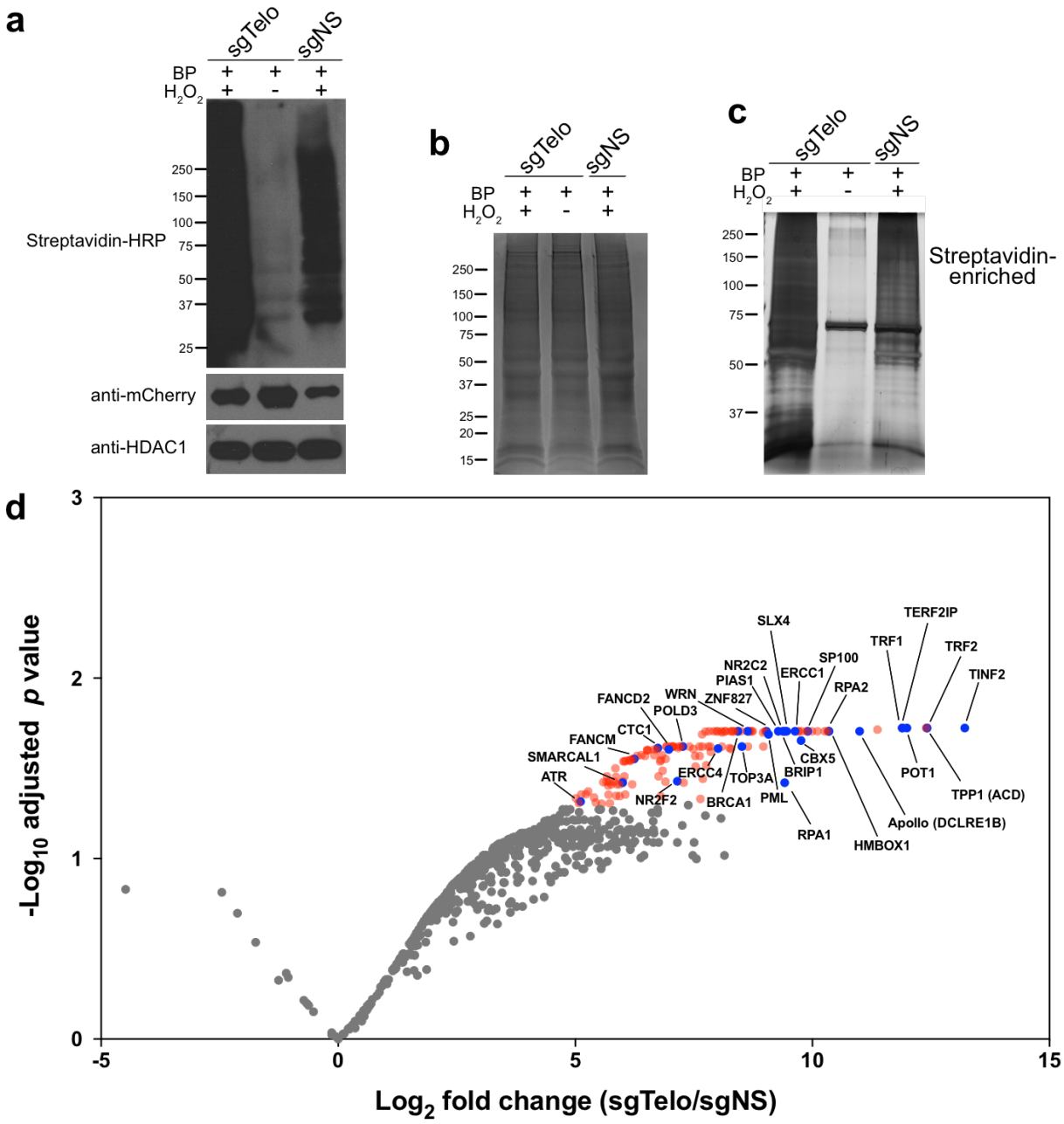

a
sgTelo
sgNS
BP
H2O2
Streptavidin-HRP
anti-mCherry
anti-HDAC1
b
c
Streptavidin-enriched
d
-Log10 adjusted p value
Log2 fold change (sgTelo/sgNS)
SLX4
TERF2IP
TRF1
TRF2
TINF2
SP100
NR2C2
PIAS1
ERCC1
RPA2
FANCD2
WRN
ZNF827
POLD3
CTC1
FANCM
SMARCAL1
ATR
CBX5
BRIP1
ERCC4
TOP3A
NR2F2
BRCA1
PML
POT1
TPP1 (ACD)
Apollo (DCLRE1B)
RPA1
HMBOX1

Figure 2

731 **Figure 2** Successful capture of telomere-associated proteins in living human cells by C-BERST
732 using Label Free Quantification (LFQ). (**a**) Top: Western blot analysis of dSpyCas9-mCherry-
733 APEX2 biotinylation, as detected by streptavidin-HRP. sgRNAs, BP treatment, and H_2O_2
734 treatment are indicated at the top of each lane. Anti-mCherry was used to detect dSpyCas9-
735 mCherry-APEX2 (middle), and anti-HDAC1 was used as a loading control (bottom). (**b**)
736 Coomassie-stained SDS-PAGE of total protein from isolated nuclei following biotin labeling. (**c**)
737 Silver-stained SDS-PAGE of biotin-labeled proteins enriched with streptavidin-coated beads. In
738 **a-c**, the mobilities of protein markers (in kDa) are indicated on the left of each panel. (**d**) Volcano
739 plot of C-BERST-labeled, telomere-associated proteins in U2OS cells. Intensity-based absolute
740 quantification (iBAQ) values from the MS analyses were calculated for each identified protein for
741 all three samples (sgTelo + H_2O_2, sgTelo - H_2O_2, and sgNS + H_2O_2). 143 proteins (indicated by
742 blue and red) are statistically enriched [Benjamini-Hochberg (BH)-adjusted p value < 0.05] in the
743 sgTelo + H_2O_2 sample, relative to both control samples. The 30 proteins indicated by blue dots
744 (with identities provided) are previously defined as either telomere-associated proteins or ALT
745 pathway components. These include all six shelterin components.
746
747
748
749
750
751

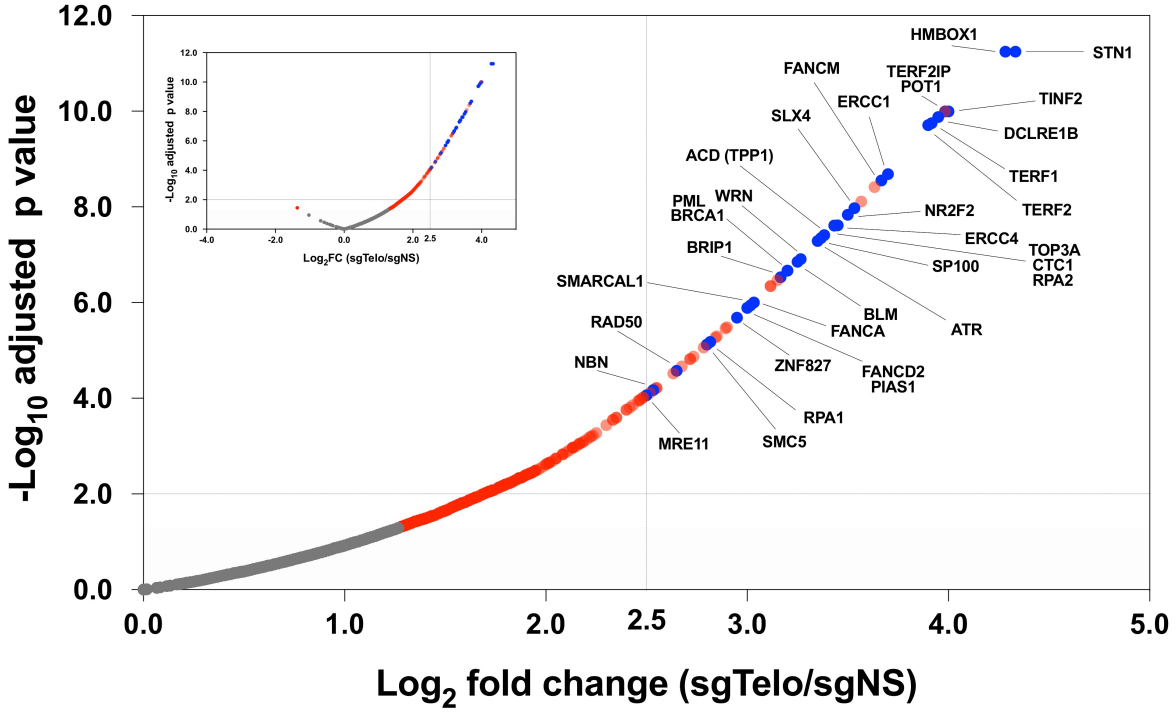


Figure 3

752

753 **Figure 3** Ratiometric C-BERST tagging strategy improves telomere-associated proteome
754 identification. Volcano plot of C-BERST-labeled, telomere-associated proteins in U2OS cells.
755 For each protein, the H/M SILAC ratio reflects the enrichment of identified proteins in sgTelo
756 vs. sgNS cells. 359 proteins (indicated by blue and red) are statistically enriched [Benjamini-
757 Hochberg (BH)-adjusted p value < 0.05] in the sgTelo, relative to sgNS controls (indicated by red
758 dots). 55 proteins fall within the cut-off based on FDR and enrichment level (BH-adjusted p value
759 < 0.01 and $\log_2$ fold change ≥ 2.5). The 34 proteins indicated by blue dots (with identities
760 provided) are previously defined as either telomere-associated proteins or ALT pathway
761 components. The volcano plot shows 97.1% of identified proteins (inset shows all proteins,
762 including the few with SILAC H/M $\log_2$ ratios < 0).

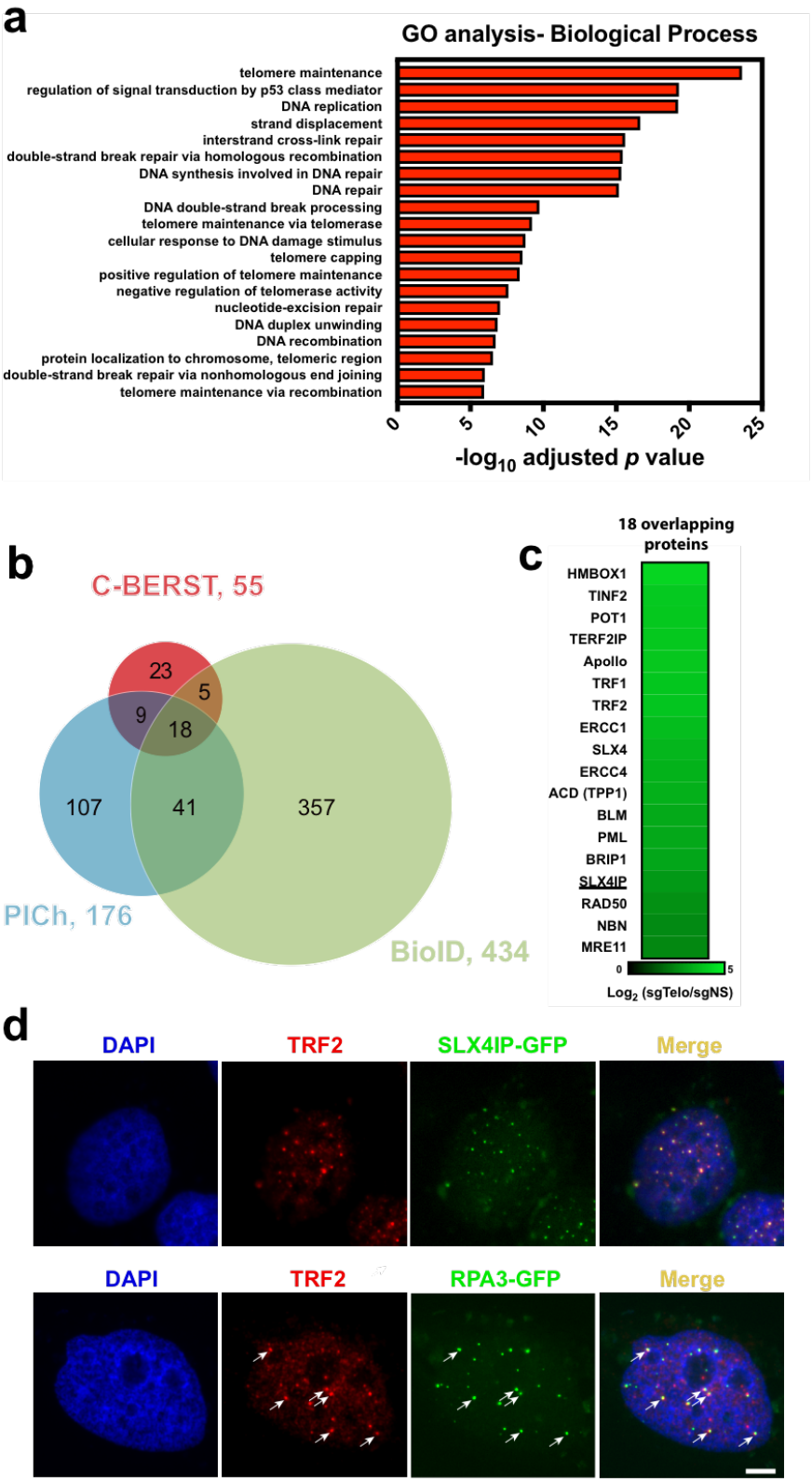
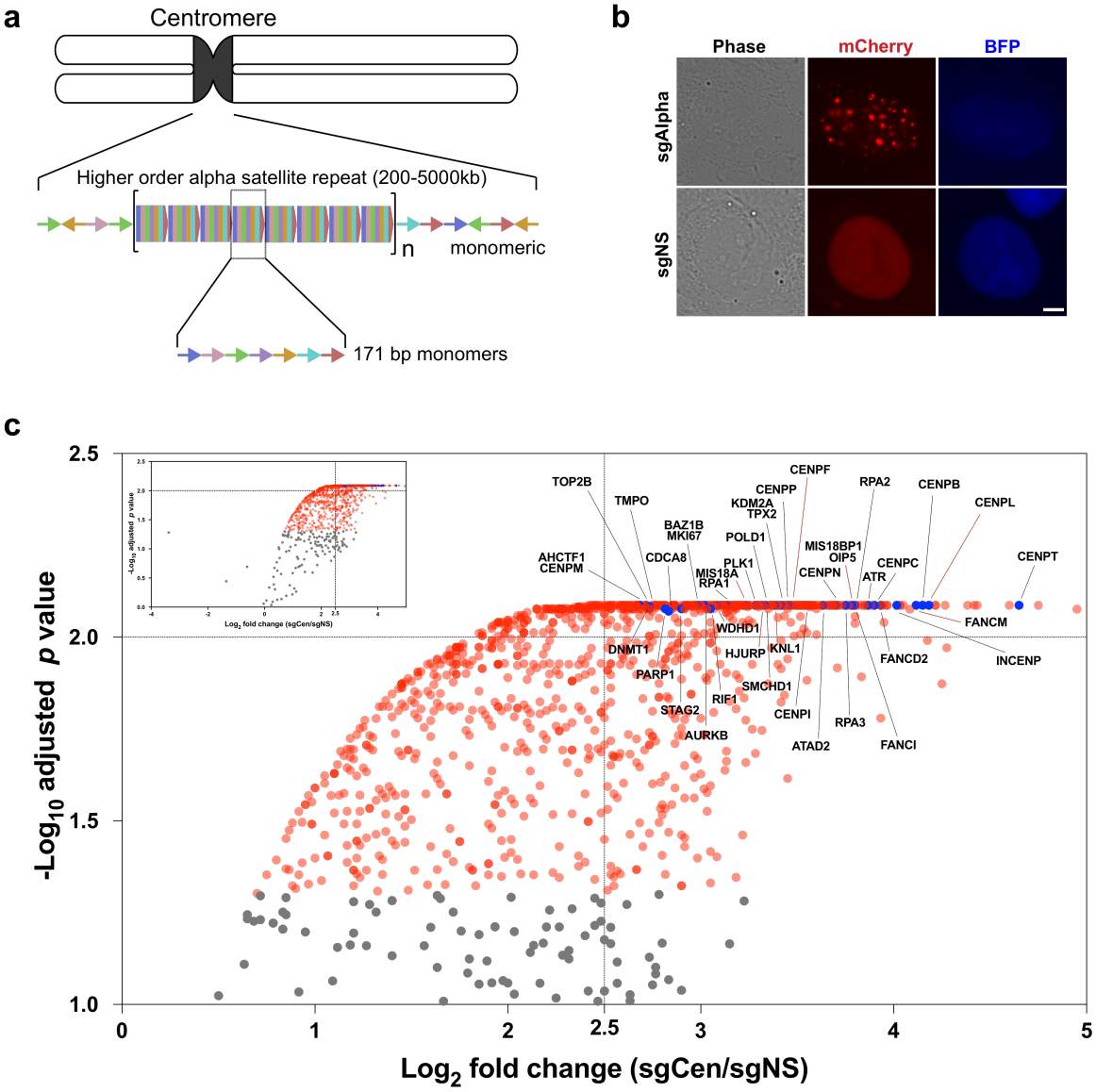

a
GO analysis- Biological Process
telomere maintenance
regulation of signal transduction by p53 class mediator
DNA replication
strand displacement
interstrand cross-link repair
double-strand break repair via homologous recombination
DNA synthesis involved in DNA repair
DNA repair
DNA double-strand break processing
telomere maintenance via telomerase
cellular response to DNA damage stimulus
telomere capping
positive regulation of telomere maintenance
negative regulation of telomerase activity
nucleotide-excision repair
DNA duplex unwinding
DNA recombination
protein localization to chromosome, telomeric region
double-strand break repair via nonhomologous end joining
telomere maintenance via recombination
0
5
10
15
20
25
-log10 adjusted p value
b
C-BERST, 55
23
5
9
18
107
41
357
PICh, 176
BioID, 434
c
18 overlapping proteins
HMBOX1
TINF2
POT1
TERF2IP
Apollo
TRF1
TRF2
ERCC1
SLX4
ERCC4
ACD (TPP1)
BLM
PML
BRIP1
SLX4IP
RAD50
NBN
MRE11
0
5
Log2 (sgTelo/sgNS)
d
DAPI
TRF2
SLX4IP-GFP
Merge
DAPI
TRF2
RPA3-GFP
Merge

Figure 4

763

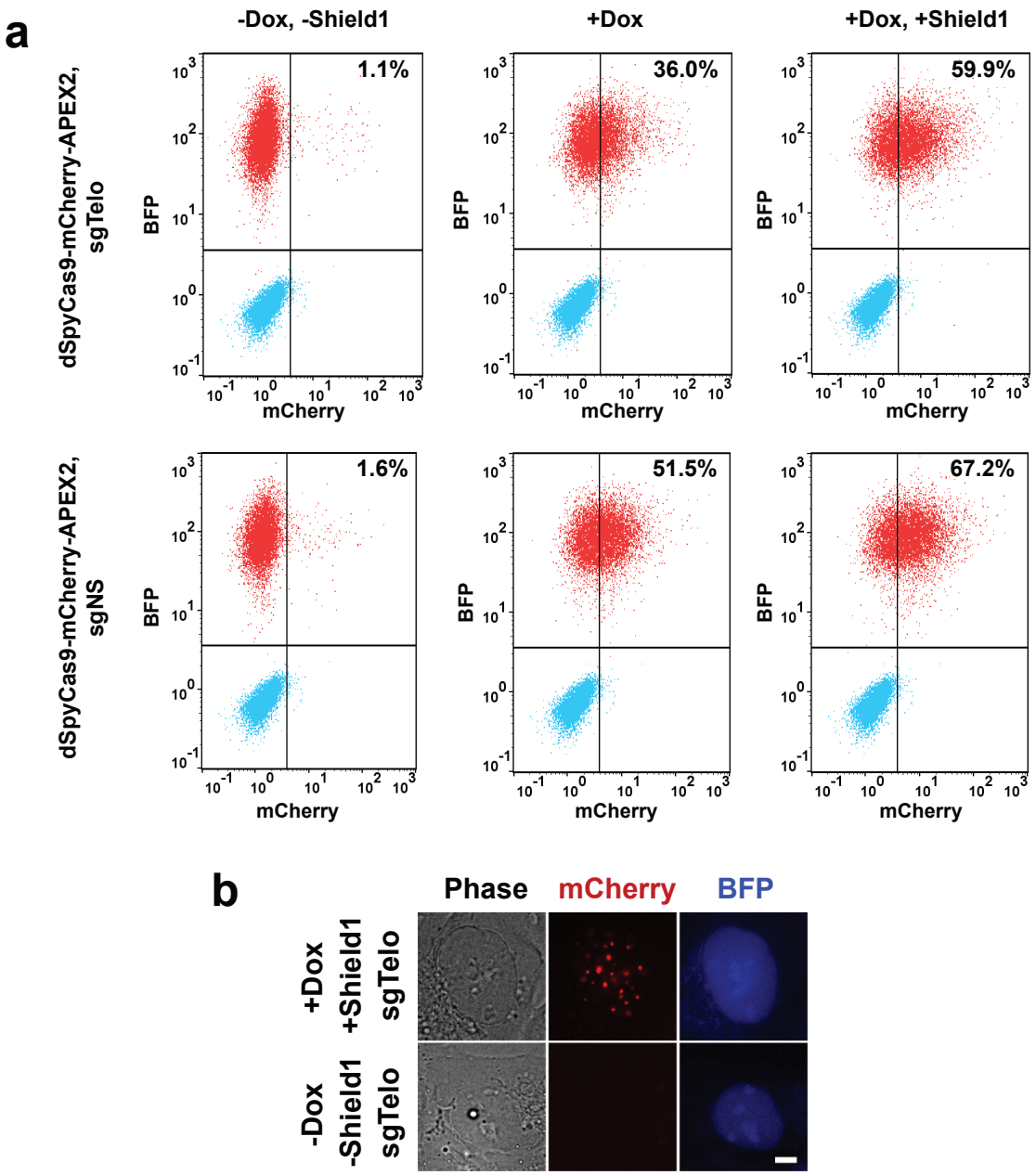
764 **Figure 4** Comparison of the C-BERST telomere-associated proteome (based on the SILAC
765 dataset) with other approaches, and validation of novel telomeric factors. (**a**) Gene Ontology-
766 Biological Process (GO-BP) analysis on the 55 telomeric/ALT proteins identified by C-BERST.
767 The x-axis is the $-\log_{10} p$ value (BH-adjusted) for the C-BERST-detected proteins associated with
768 each GO term given on the left. The 20 most statistically significant GO terms are displayed. (**b**)
769 Venn diagram of statistically enriched (BH-adjusted p value < 0.01) telomeric proteins from ALT
770 human cells, as detected by C-BERST (red), PICh (purple), and TRF1-BirA* BioID (green). 32
771 proteins from the C-BERST proteome were also detected by PICh, BioID, or both. (**c**) A heat
772 map of the C-BERST $\log_2$ fold-change enrichment scores for the 18 telomeric proteins identified
773 by all three proteomic approaches from (**b**). All 18 proteins are highly enriched in the C-BERST
774 telomere proteome. The one underlined protein (SLX4IP) has not been reported previously (to
775 our knowledge) as telomere- or ALT-associated. (**d**) Colocalization of turboGFP-tagged SLX4IP
776 and RPA3 with telomeric marker protein TRF2. 0.3×10^5 U2OS cells were transiently
777 transfected with 100ng SLX4IP-GFP expression plasmid or 50ng RPA3-GFP expression plasmid.
778 Cells were then fixed and incubated with TRF2 primary antibody and secondary antibody
779 conjugated with Alexa Fluor 647. DAPI stained cells were imaged ($n \geq 20$ cells examined). Scale
780 bar, 5 μm.



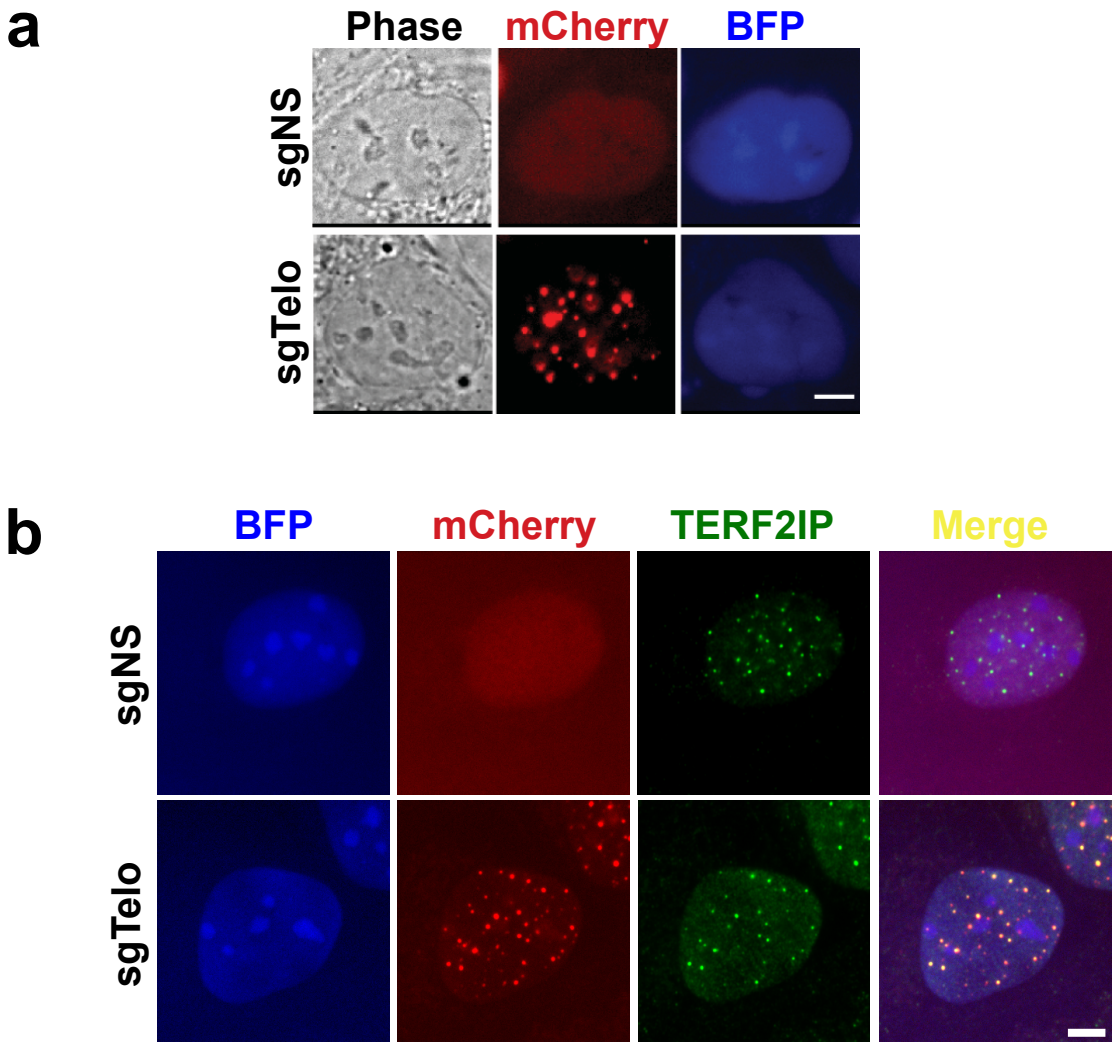
781

782 **Figure 5** Successful capture of alpha satellite associated proteomes in live human cells by C-
783 BERST. (**a**) Schematic diagram of alpha satellite repeat position and arrangement at a
784 centromere. (**b**) Live-cell imaging of centromere localization by dSpyCas9-mCherry-APEX2 in
785 U2OS cells, using the P1-sorted population defined by the FACS workflow in **Fig. 1c**.
786 dSpyCas9-mCherry-APEX2 exhibited centromeric foci with sgAlpha but not with sgNS. Scale
787 bar, 5 µm. (**c**) Ratiometric C-BERST (using SILAC) was used to profile the alpha satellite
788 associated proteome. A volcano plot of C-BERST-labeled, centromere-associated proteins in
789 U2OS cells is shown. For each protein, the H/M SILAC ratio reflects the enrichment of

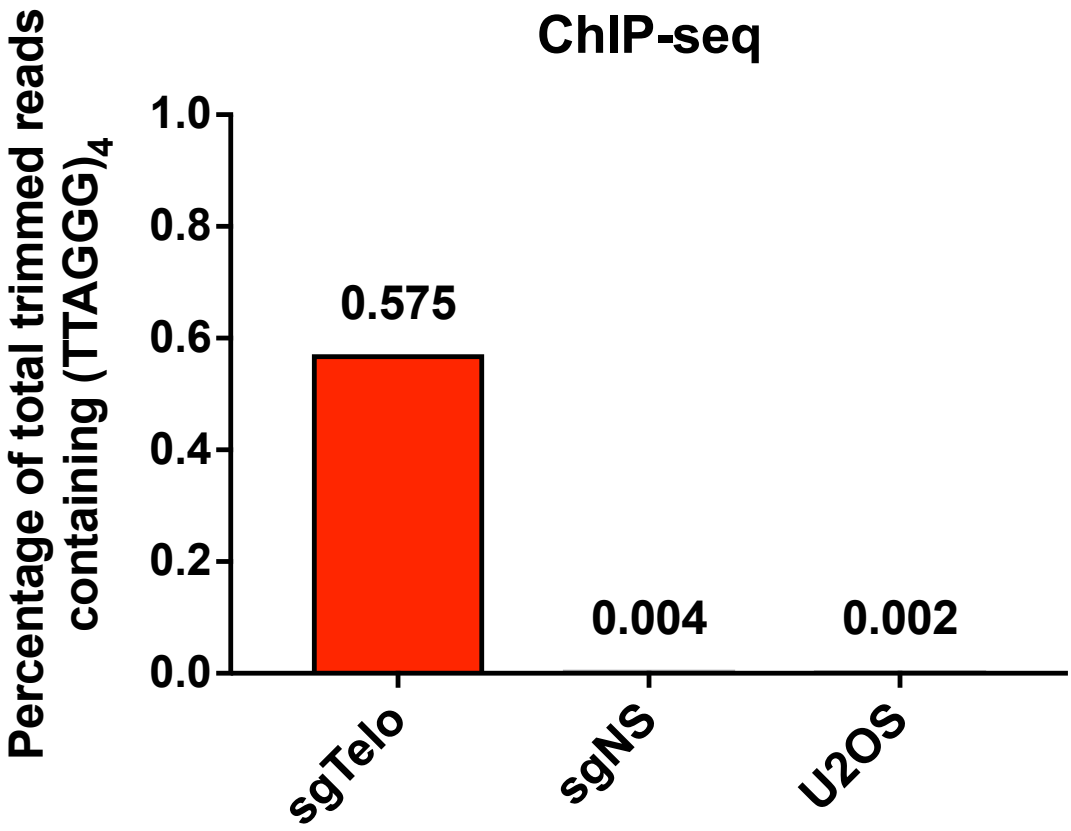
790 identified proteins in sgAlpha vs. sgNS cells. 1134 proteins (indicated by blue and red dots) are
791 statistically enriched [Benjamini-Hochberg (BH)-adjusted p value < 0.05] in the sgAlpha sample,
792 relative to sgNS controls. 460 proteins fall within the cut-off based on FDR and enrichment level
793 (BH-adjusted p value < 0.01 and $\log_2$ fold change ≥ 2.5). The 40 proteins indicated by blue dots
794 (with identities provided) were previously defined as either centromere-associated proteins
795 (**Supplementary Table 5**) or were reported as components of the HyCCAPP centromere
796 proteome (see text). The nine known centromere-associated proteins indicated by red lines are
797 uniquely captured by C-BERST. The volcano plot shows 96.2% of identified proteins (inset
798 shows all proteins, including the few with SILAC H/M ratio < 0 and -log10 adjusted p value < 1).



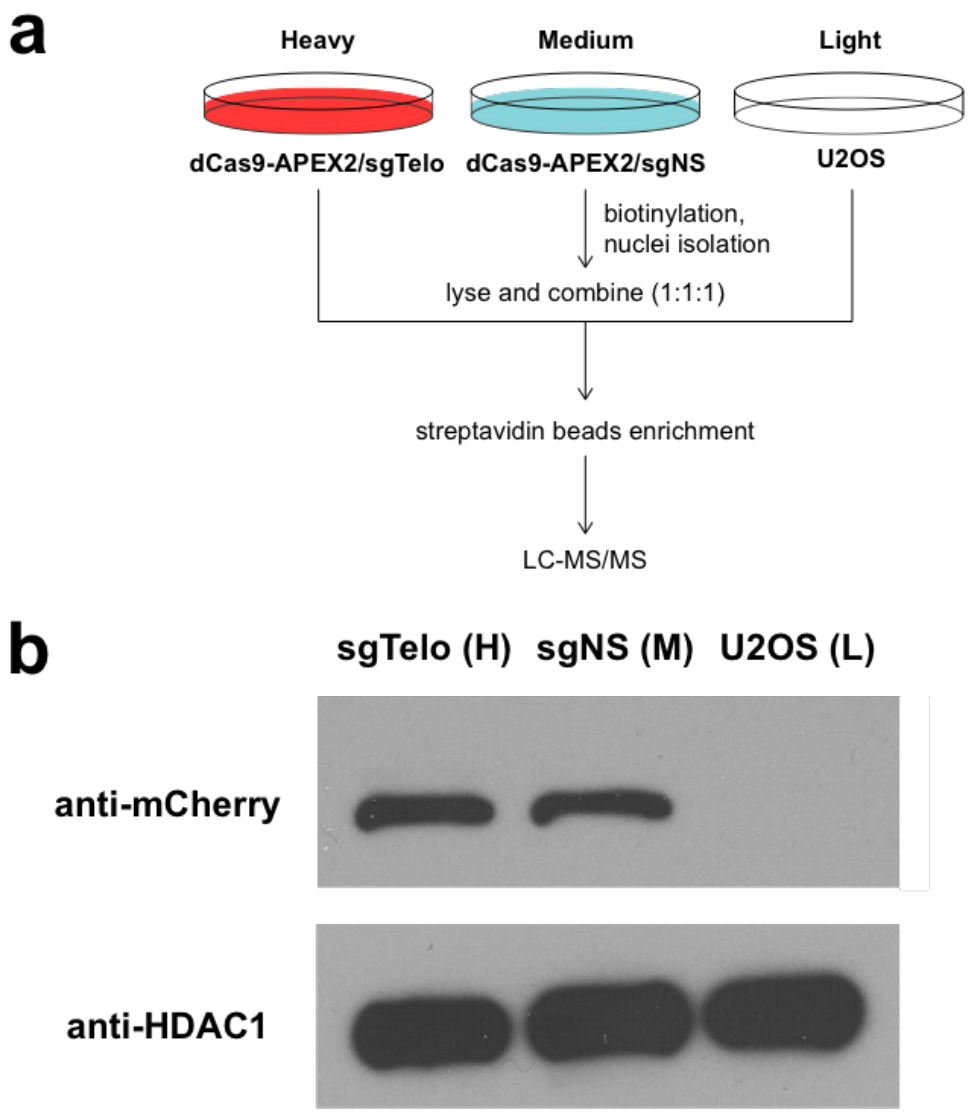
Supplementary Fig. 1 Inducible dSpyCas9-mCherry-APEX2 expression. (**a**) Flow cytometry was used to measure the percentage of mCherry+ and BFP+ double-positive cells under different induction conditions. Stable U2OS cells expressing sgTelo (top row) or sgNS (bottom row) were exposed to three conditions 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.



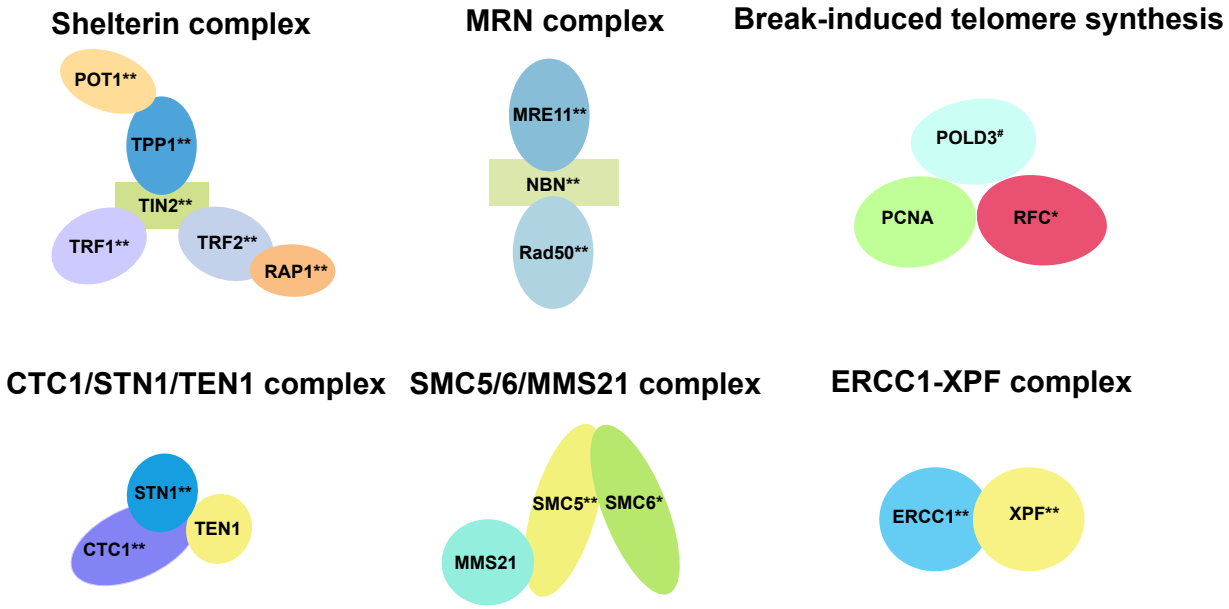
Supplementary Fig. 2 Specific telomere targeting by dSpyCas9-mCherry-APEX2. (**a**) Live-cell imaging of telomere localization by dSpyCas9-mCherry-APEX2 in U2OS cells, using the P1-sorted population defined by the FACS workflow in **Fig. 1c**. dSpyCas9-mCherry-APEX2 exhibited telomeric foci with sgTelo but not with sgNS. (**b**) 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 μm.



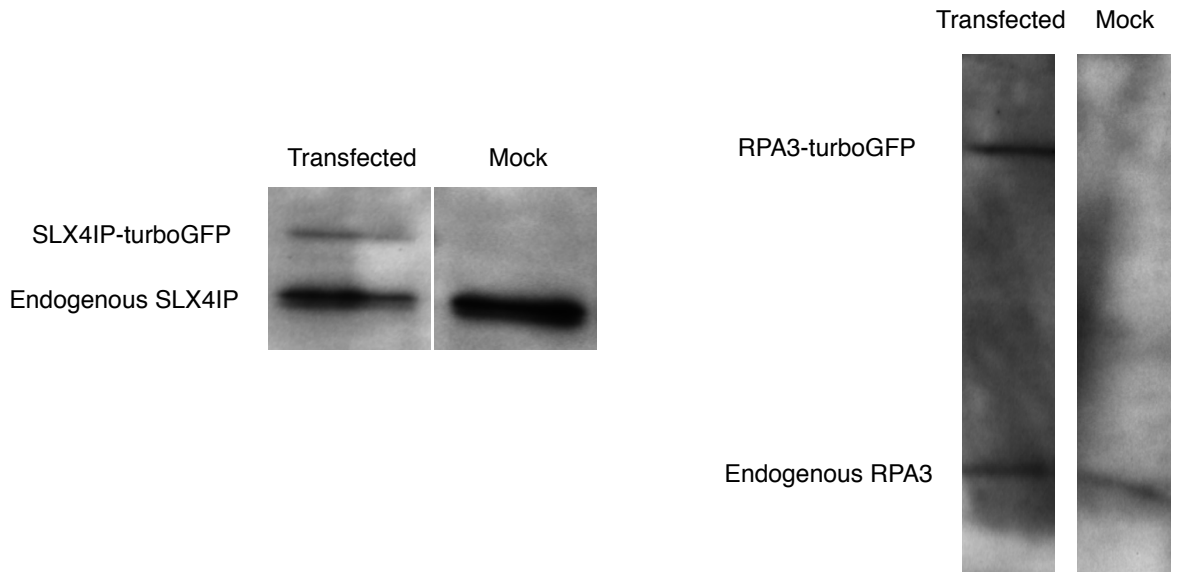
818
819 **Supplementary Fig. 3** anti-mCherry chromatin immunoprecipitation shows genome-wide
820 binding of sgTelo-programmed dSpyCas9-mCherry-APEX2. The reads were trimmed by
821 adaptor removal and filtering. The percentage of total trimmed reads that include at least one
822 $(TTAGGG)_4$ telomeric sequence (the minimum length required for complete sgTelo
823 complementarity) is shown. Values were averaged from two independent biological replicates,
824 except for the U2OS ChIP-seq, which was only performed once.



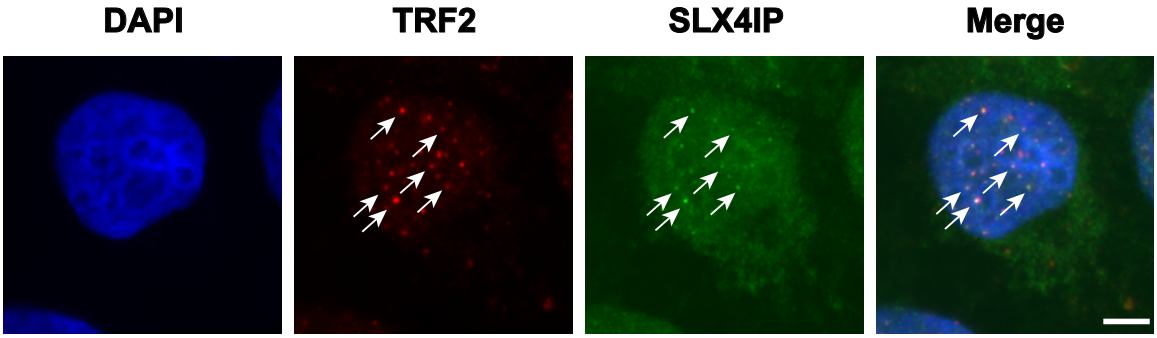
825
826 **Supplementary Fig. 4** (**a**) Schematic diagram of SILAC workflow. Cells were grown in
827 different isotope culture media for at least five passages. dSpyCas9-mCherry-APEX2 proteins
828 were induced by dox and Shield1 21 hours before biotinylation. Following biotinylation and
829 nuclei isolation, cell lysates were sonicated and mixed in a 1:1:1 ratio. (**b**) Anti-mCherry was used
830 to detect dSpyCas9-mCherry-APEX2 (top), and anti-HDAC1 was used as a loading control
831 (bottom) in the western blot analysis.



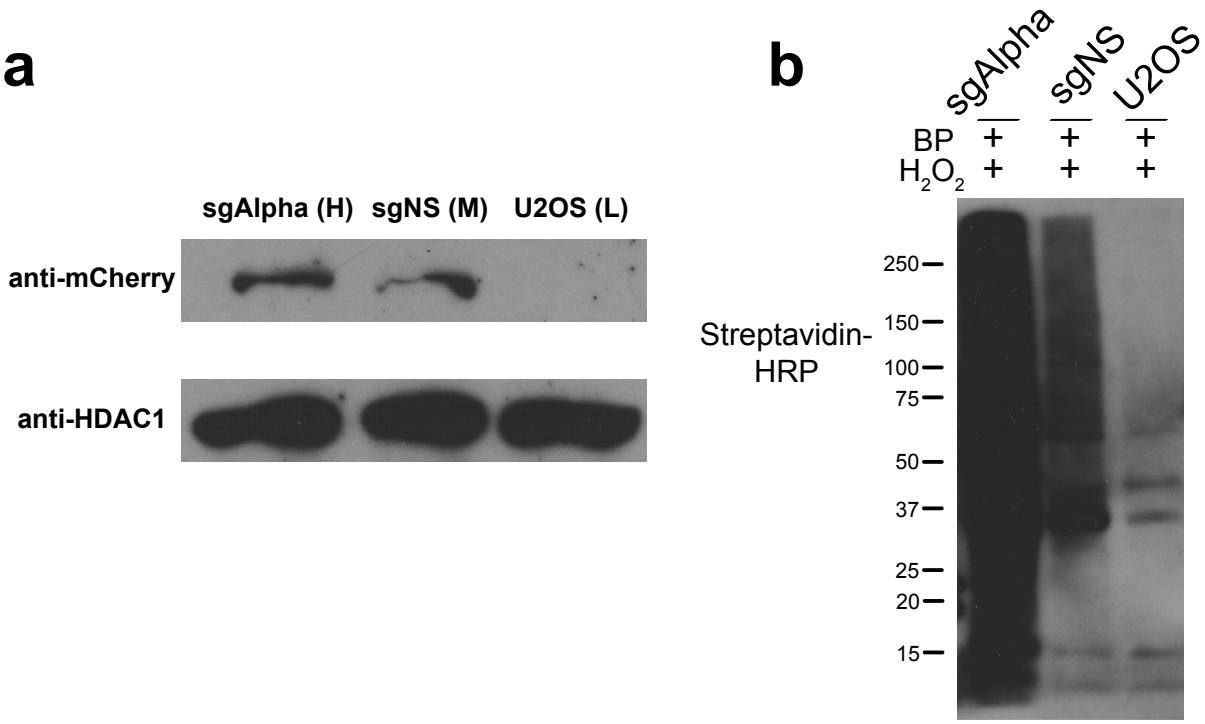
Supplementary Fig. 5 C-BERST specifically detects components of multiple complexes and factors implicated in ALT pathways. Proteins denoted by double asterisks were significantly enriched (p value < 0.01) and meet the SILAC cut-off $\log_2$ fold change ≥ 2.5 in the sgTelo labeling sample, relative to the sgNS labeling samples (H/M) ratio (see **Supplementary Table 3**). Proteins denoted by single asterisks were also detected but with lower degrees of significance (BH-adjusted p value < 0.05) in the sgTelo labeling sample (**Supplementary Table 3**). Proteins denoted by a hashtag were enriched and statistically significant in LFQ. Components of the RAD9/RAD1/HUS1 complex (see **Supplementary Table 3**) were not detected.



Supplementary Fig. 6 Western blot analysis of exogenous, turboGFP-tagged SLX4IP or RPA3 expression, in comparison with the corresponding endogenous protein. ~0.1 x 10^5 U2OS cells transfected with SLX4IP-turboGFP or RPA3-turboGFP expression plasmid (100 ng and 50 ng, respectively) were lysed in 1x RIPA lysis buffer, and proteins were resolved by SDS-PAGE. Western blots were probed by primary SLX4IP or RPA3 antibody and anti-rabbit secondary antibody conjugated with HRP. The gel lanes shown in each panel were cropped from identical exposures of the same western blot membranes.

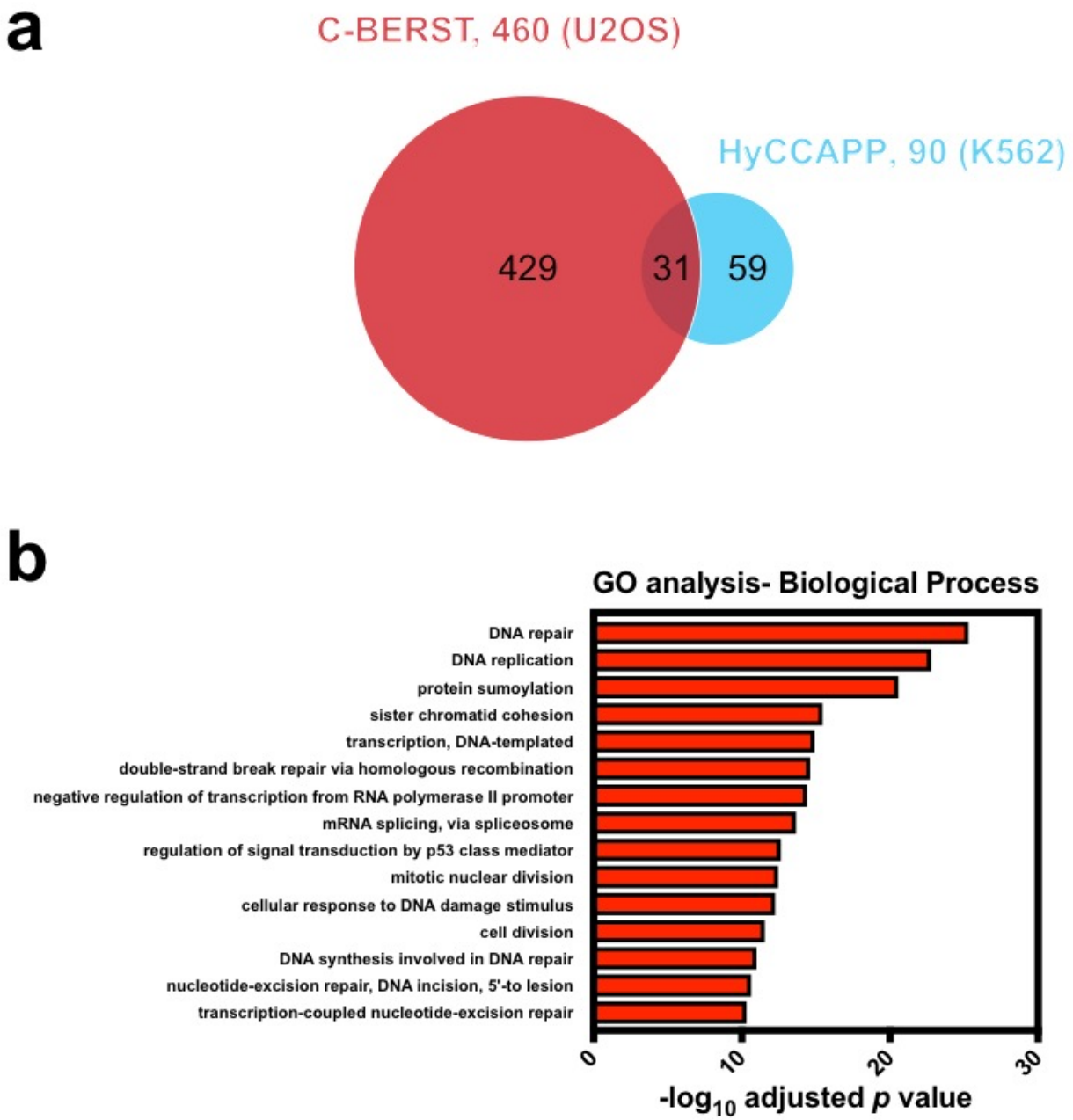


Supplementary Fig. 7 Coimmunostaining of TRF2 (telomeric marker protein) and SLX4IP protein. Primary goat anti-TRF2 and rabbit anti-SLX4IP were used to detect endogenous TRF2 and SLX4IP in the fixed U2OS cells. Secondary donkey anti-goat conjugated with Alexa 647 and mouse anti-rabbit conjugated with CruzFluor™ 488 were then incubated with cells. Scale bar, 5 μm.

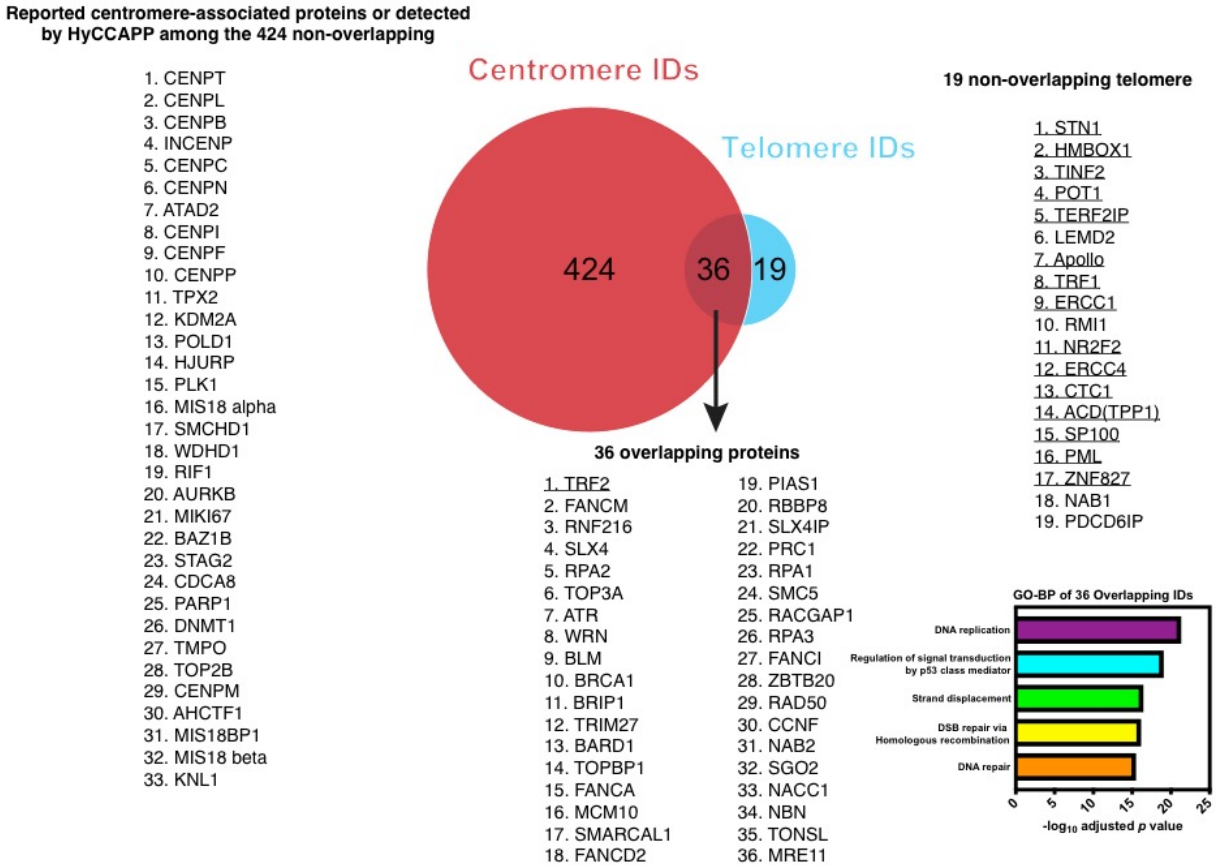


Sup fig 8

Supplementary Fig. 8 (**a**) Western blot analysis of dSpyCas9-mCherry-APEX2 expression using anti-mCherry to detect dSpyCas9-mCherry-APEX2 (top). Anti-HDAC1 was used as a loading control (bottom). (**b**) Streptavidin-HRP blotting analysis of biotinylated proteins in sgAlpha- and sgNS-expressing cells. Untransduced U2OS cells were used as a control.



Supplementary Fig. 9 (**a**) Venn diagram of statistically enriched (BH-adjusted p value < 0.01) 460 centromeric proteins from U2OS cells, as detected by C-BERST (red), and 90 centromeric proteins from K562 cells, as detected by HyCCAPP (cyan) (see text). 31 proteins from the C-BERST proteome were detected by both. (**b**) Gene Ontology-Biological Process (GO-BP) analysis on 460 centromeric proteins identified by C-BERST. The x-axis is the $-\log_{10} p$ value (BH-adjusted) for the C-BERST-detected proteins associated with each GO term given on the left. The 15 most statistically significant GO terms are displayed.



Supplementary Fig. 10 Venn diagram of 55 C-BERST ALT/telomeric IDs and 460 centromeric IDs. 19 non-overlapping telomere IDs are listed on the right. Known ALT/telomeric proteins are 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 the Venn diagram, as indicated. The five most significant GO-BP terms for the 36 overlapping ID are provided on the lower right.

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