A well-controlled BioID design for endogenous bait proteins

Giel Vandemoortele[1,2], Delphine De Sutter[1,2], Aline Moliere[1,2], Jarne Pauwels[1,2], Kris Gevaert[1,2] and Sven Eyckerman[1,2,*]

[1]VIB Center for Medical Biotechnology, VIB, B-9000 Ghent, Belgium

[2]Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium

Keywords: protein protein interaction; BioID proximity labeling; (T2A) autocleavage; genome engineering; p53; base editor

*Correspondence to Prof. Dr. Sven Eyckerman

VIB Center for Medical Biotechnology

A. Baertsoenkaai 3, B-9000 Ghent, Belgium

Tel: +32-9-264.92.18

Fax: +32-9-264.94.90

Email: sven.eyckerman@vib-ugent.be

ABSTRACT

The CRISPR/Cas9 revolution is profoundly changing the way life sciences technologies are used. Many assays now rely on engineered clonal cell lines to eliminate overexpression of bait proteins. Control cell lines are typically non-engineered cells or engineered clones implying a considerable risk for artefacts because of clonal variation. Genome engineering can also transform BioID, a proximity labelling method that relies on fusing a bait protein to a promiscuous biotin ligase, BirA*, resulting in the tagging of vicinal proteins. We here propose an innovative design to enable BioID for endogenous proteins wherein we introduce a T2A-BirA* module at the C-terminus of endogenous p53 by genome engineering, leading to bi-cistronic expression of both p53 and BirA* under control of the endogenous promoter. By targeting a Cas9-cytidine deaminase base editor to the T2A auto-cleavage site, we can efficiently derive an isogenic population expressing a functional p53-BirA* fusion protein. Using quantitative proteomics we show significant benefits over classical ectopic expression of p53-BirA*, and we provide a first well-controlled view on the proximal proteins of endogenous p53 in colon carcinoma cells. This novel application for base editors expands the CRISPR/Cas9 toolbox and can be a valuable addition for synthetic biology.

36 **ABBREVIATIONS**

37	AP-MS	Affinity Purification – Mass Spectrometry
38	dCas9	Catalytically inactive 'dead' Cas9
39	DDA	Data-Dependent Acquisition
40	DSB	DNA double Strand Break
41	GO	Gene Ontology
42	Indel	Insertion or deletion
43	IRES	Internal Ribosome Entry Site
44	LFQ	Label-Free Quantitation
45	nCas9	Nickase Cas9
46	PEI	Polyethylenime
47	rAAV	recombinant Adeno-Associated virus
48	UGI	Uracil DNA Glycosylase Inhibitor

49

50 **INTRODUCTION**

51 Elucidating the composition and function of protein complexes is crucial for advancing our
52 understanding of key cellular processes, both in health as well as disease [1]. Fuelled by an ever-
53 growing arsenal of so-called interactomics technologies [2], vast amounts of data are now available
54 for further exploration [3, 4]. With expanding lists of interactors for every protein challenged, there is
55 a growing consensus in the field that many reported interactions are highly contextual or spatially
56 restricted [5]. In addition, it became apparent that approaches in which a bait protein is overexpressed
57 lack the subtlety that is needed to further dissect protein-protein interactions and assess physiological
58 functions of individual interactions [6]. Indeed, protein overexpression can have a detrimental effect
59 on protein localization and folding, and distort downstream signalling [5, 7]. Protein overexpression
60 also tends to mask complex feedback control mechanisms affecting a protein at physiological levels
61 and generally increases the overall stress status of a cell [8, 9]. Several efforts have been made to
62 avoid bait protein overexpression, with the use of primary antibodies to pull-down endogenous
63 protein complexes as prime example [10, 11]. In reality, the specificity and the sensitivity of antibodies
64 are often inadequate for studying endogenous protein complexes in a reliable manner and can
65 therefore lead to problems with reproducibility [12]. Despite improved methodologies it remains
66 challenging to generate specific antibodies with favourable affinities for each protein [13, 14]
67 especially when one wants to specifically study mutant proteins bearing amino acid substitutions or
68 truncated proteins originating from, for instance, alternative translation start sites (i.e. proteoforms
69 [15]). For such situations, Flp-In^{TM} cell lines allowing tunable expression provide a valuable alternative
70 [16]. However, this requires thoughtful and precise fine-tuning of expression levels to mimic as closely
71 as possible physiological protein levels [17]. Artificial chromosomes have also been used to express
72 transgenes at physiological levels [18].

73 A technology that is gaining popularity to map protein interactomes is BioID. This method relies on a
74 promiscuous biotin ligase (BirA*) fused to a bait protein of interest [19]. BirA* is a mutagenized *E. coli*

75 biotin ligase (carrying the R118G mutation) that no longer displays substrate specificity. Upon
76 expressing a bait-BirA* fusion protein, proximity-dependent biotinylation occurs of proteins that are
77 near-neighbours of this bait protein in the cell. As endogenous biotinylation is a rare and highly specific
78 modification in mammalian cells [20, 21], biotinylated prey proteins are easily discriminated after
79 purification by streptavidin matrices and mass spectrometry (MS) analysis. Moreover, the need for
80 lysis is removed since the labelling of prey proteins occurs in intact cells before the homogenization
81 step. Since its initial publication in 2012, BioID has been used to study proximal and interacting
82 proteins (i.e. the proxisome) of bait proteins in several organisms and cell types producing datasets
83 complementary to affinity purification – MS (AP-MS) approaches [22]. In bait overexpression
84 experiments, as controls, BirA* is often expressed alone or fused to irrelevant baits (e.g. eDHFR, eGFP),
85 although cellular localizations and expression levels might differ between BirA*-fusion proteins and
86 BirA* alone, which therefore confound results [19].
87 Bi-cistronic expression systems can provide a way to ensure similar expression levels for a bait protein
88 and a detached BirA* module. 2A peptides cause ribosomal skipping during translation and
89 recommencement of translation immediately afterwards [23]. 2A peptides display increased
90 efficiency when compared to Internal Ribosome Entry Site (IRES) sequences and are only 18 amino
91 acids in length, which is an additional advantage when considering genome engineering approaches
92 [24].
93 Since the emergence of CRISPR/Cas9, many applications have been published with several of them
94 outgrowing classical genome engineering [25]. A recent application is on base editing in which Cas9 is
95 fused to a cytidine deaminase. This enzyme converts cytosine into uracil which has the base-pairing
96 properties of thymine [26] and thus introduces mutations. The base editing activity can be targeted
97 to confined sequence windows of approximately five nucleotides by loading Cas9 with a specific guide
98 RNA (gRNA). As base editing is typically performed by nickase Cas9 (nCas9) or catalytically inactive
99 (dead, dCas9) fusion constructs, no DNA double strand breaks (DSBs) are introduced, implying that
100 minimal numbers of insertion or deletion (indel) mutations are generated in this process. To avoid

101 base excision repair which would revert the introduced mutation, extra moieties were added to these
102 base editors, such as an uracil DNA glycosylase inhibitor (UGI) [27]. Up to now, base editing
103 applications mainly focussed on correcting disease mutations, on genetic diversification using tiled
104 gRNA libraries [27-29], or on introducing premature stop codons as alternative to classic knock-out
105 strategies [30].
106 p53 is arguably the most studied protein in health and disease [31, 32]. Historically dubbed the
107 guardian of the genome [33], the encoding gene is mutated in 50% of all reported human cancer cases
108 [34], making it an attractive starting point for the development of anticancer strategies [35]. Because
109 of the long-lasting research focus on p53, the current interactome of p53 consists of over 1000 protein
110 interaction partners (BioGRID, version 3.4 [3]). p53 acts as a homotetrameric, sequence-specific
111 transcription factor capable of inducing large gene networks comprising hundreds of target genes [36,
112 37]. Known pathways regulated by p53, either by transactivation or protein interactions, include DNA
113 damage repair, cell cycle arrest, senescence and apoptosis following exposure to a plethora of stress
114 conditions such as DNA damage, telomere erosion, hypoxia, replication stress and oncogene
115 activation. More recently, additional processes under p53 control such as autophagy, metabolism and
116 ferroptosis were revealed illustrating the need for continued research on this protein [38].
117 In this report, we use a novel knock-in strategy to enable BioID experiments for endogenous p53 in
118 HCT116 colon carcinoma cells as classic overexpression experiments in this cell model gave suboptimal
119 results. In our knock in strategy, a control cell line was first generated by introduction of a T2A-BirA*
120 cassette at the C-terminus of endogenous p53 using CRISPR/Cas9 genome engineering. These cells
121 express the BirA* protein at equal levels to p53 and thus provide a good system to assess biotinylation
122 background. Targeted inactivation of the T2A peptide sequence by a Cas9 cytidine deaminase fusion
123 (BE3) results in the generation of a cell line that expresses a functional p53-BirA* fusion protein from
124 the endogenous promoter allowing proximal protein mapping for endogenous p53. This reduces
125 genome engineering efforts considerably and eliminates artefacts due to clonal variation when a

126 selection step is added in the protocol. Furthermore, the concept of auto-cleavage switching in fusion

127 proteins can find applications in synthetic biology.

128 ## EXPERIMENTAL PROCEDURES

129 **Cell culture**

130 HCT116 colon carcinoma cells were purchased from ATCC (ATCC CCL-247) and cultured in McCoy's 5A
131 modified medium (Gibco, 22330070). Genomically characterized HEK293T embryonic kidney cells [39]
132 were cultured in DMEM (Gibco, 61965026) supplemented with fetal bovine serum (FBS) (Gibco,
133 10270106) to a final concentration of 10%, 12500 units penicillin-streptomycin (Gibco, 15070063) and
134 HEPES (Gibco, 15630056) to a final concentration of 10 mM. Growth medium was switched to McCoy's
135 5A modified medium or DMEM supplemented with 10% dialyzed FBS (Gibco, 26400044) for HCT116 2
136 weeks prior to seeding cells for a BioID experiment. Cells were passaged when cultures reached 70%
137 confluence. All cell lines were deemed clear of mycoplasma by using a mycoplasma PCR detection kit
138 (Minerva Biolabs, MIN-11-1100).

139 **Plasmids and cloning procedures**

140 The BioID biotin ligase BirA* was transferred from the pcDNA3.1 mycBioID vector (Addgene #35700)
141 by PCR to generate the pAav-TP53-T2A-BirA* targeting construct (available via Addgene, #115652).
142 All primers used in this manuscript can be found in table S1. Homology regions were generated by PCR
143 on genomic HCT116 DNA as described before [40, 41]. The pMet7-FLAG-PQS1-TP53-T2A-BirA*-Myc
144 construct was generated by standard restriction enzyme cloning procedures using the FLAG-PQS1-p53
145 expression vector published earlier [40]. This construct was further mutagenized by PCR to obtain a
146 construct bearing an inactivated T2A sequence for a BioID transient transfection experiment. A
147 puromycin selection cassette was incorporated in the BE3 (Addgene, plasmid #73021) vector for
148 endogenous mutagenesis purposes ($BE3_{puro}$) using a linear dsDNA fragment (IDT DNA Technologies,
149 gBlock). gRNA sequences incorporated in a U6-driven expression cassette were ordered as gBlocks
150 (Table S2) prior to blunt-ended ligation in a pCR-Blunt vector (Thermo Fisher Scientific, K2750).

151 **CRISPR/Cas9 genome engineering**

152 HCT116 *TP53*$^{+/T2A\text{-}BirA*}$ cells were generated using a combination of CRISPR/Cas9 and recombinant

153 Adeno-associated virus (rAAV)-mediated template delivery as described before [40, 41]. Briefly, a
154 *TP53* C-terminal targeted gRNA (5' ACGCACACCUAUUGCAAGCA 3') was cloned in a Cas9 expression
155 construct (Addgene #48139) as described by Ran *et al.* [42]. Cells were seeded in a 24-well plate 24 h
156 before transfection with the Cas9 plasmid using FugeneHD (Promega, E2311). 48 h after transfection,
157 cells were infected with targeting construct packaged in rAAV. Neomycin selection (Gibco, 11811031)
158 was performed for two weeks starting 24 h after infection. Afterwards, cells were manually diluted
159 and seeded as single cells in 96-wells for further clonal expansion until on average, clones reached
160 70% confluence at which point they were split into a lysis plate and culture plate. Screening of clonal
161 populations in the lysis plate was based on junction PCR screening using GoTaq G2 Hot Start
162 polymerase (Promega, M7405). Input lysates for PCR were obtained using DirectPCR lysis reagent
163 (Viagenbiotech, 301C). Positive clones were seeded at 10^3 cells per 96-well 24 h prior to treatment
164 with 2.5 µM TAT-Cre (Excellgen, EG1001). 24 h after treatment, the medium was aspirated and
165 replaced with fresh growth medium. Four days after TAT-Cre treatment, cells were subjected to a
166 second round of single cell seeding and PCR screening. Clones showing aberrant growth rates or
167 morphology were omitted for further characterization. PCR-positive clones were subjected to
168 Western blot using 1/1000 diluted anti-p53 antibody (Calbiochem, OP43T) and near-infrared
169 fluorescent secondary antibodies (LI-COR). Additionally, Southern blot analysis was performed to
170 verify single genomic integration of the floxed neomycin resistance cassette. [α-^{32}P] Southern blot
171 probe was generated as described before.[41] For every retained clone, engineered loci were verified
172 using Sanger sequencing as described earlier. Primers used for screening, sequencing and generation
173 of the Southern blot probe can be found in table S1.

174 **T2A mutagenesis**

175 7.2×10^5 HEK293T cells were seeded in 6-well plates the day prior to transfection. A total amount of
176 800 ng DNA was mixed with 4 µl polyethylenime (PEI) for every transfection condition and incubated
177 for 10 min at room temperature. Meanwhile cells were placed on 1.5 ml DMEM medium
178 supplemented with 2% FCS. 6 h after addition of the transfection mix, the medium was removed and

179 replaced by DMEM with additives after washing once with Phosphate Buffered Saline (PBS) (Gibco,
180 14190094). 72 h post transfection cells were washed with PBS before detachment using a cell scraper
181 (Greiner, 541070) and lysis in 2X SDS-PAGE XT sample buffer (Bio-Rad, 1610791) and XT reducing agent
182 (Bio-Rad, 1610792) by heating for 10 min at 95°C. Afterwards, 100 µl MilliQ water was added to dilute
183 samples prior loading on a QIAshredder homogenizer spin-column (Qiagen, 79656) and running them
184 on a Criterion XT 4-12% BisTris gel (Bio-Rad, 3450123). After blotting, the PVDF membrane was
185 incubated with anti-FLAG (Sigma, F1804) and anti-Myc (in-house produced) primary antibodies and
186 detected with LI-COR near-infrared fluorescent secondary antibodies. For endogenous mutagenesis,
187 $BE3_{puro}$ vector and gRNA plasmids were transfected using FugeneHD in HCT116 *TP53*$^{+/T2A\text{-}BirA*}$ cells. 24
188 h after transfection, 2 µg/ml puromycin was added to the growth medium for 48 h to enrich for $BE3_{puro}$
189 transfected cells prior to expansion for proteomics experiments.

190 **Illumina targeted sequencing**

191 For targeted sequencing, subconfluent 25 cm^2 bottles containing HCT116 *TP53*$^{+/T2A\text{-}BirA*}$ cells after
192 $BE3_{puro}$ treatment and subsequent enrichment were pelleted by centrifugation at 500 g for 5 min and
193 washed once with PBS. To each pellet, 500 µl lysis buffer [20 mM Tris-HCl pH 7.5, 5 mM EDTA, 150
194 mM NaCl, 400 U Proteinase K (Sigma, P4850), 0.2% SDS] was added and incubated for 16 h at 37°C.
195 Afterwards, genomic DNA was obtained by phenol-chloroform extraction and ethanol precipitation.
196 Next, the T2A locus was amplified and provided with Nextera overhangs by PCR using Herculase
197 polymerase (Agilent, 600675). Sequencing was performed using the MiSeq Nano v2 kit (Illumina,
198 MS1022002) by the VIB Nucleomics core facility (http://www.nucleomics.be/). FASTQ files were
199 uploaded to CRISPRESSO [43] to analyse results. The CRISPRESSO allele frequency output file (Table
200 S3) was used as input to generate the sequence logo using the online tool WebLogo [44].

201 **Nuclear and cytoplasmic extraction**

202 For every cell line, cells were cultured in McCoys medium supplemented with a final concentration of
203 10% dialyzed FBS as described above. 1 mM doxorubicin (Sigma, D1515) and 50 µM biotin (Sigma,

204 B4639) was added to the growth medium 24 h prior to the start of the experiment at which $7x10^6$ cells
205 were pelleted by centrifugation at 500 g for 5 min. After washing once with ice-cold PBS and
206 centrifugation at 500 g for 5 min, the supernatant was aspirated from the cell pellet. Extraction was
207 done using the NE-PER kit (Thermo Fisher Scientific, 78833) according to the instructions provided by
208 the manufacturer. In brief, 500 µl ice-cold CER 1 buffer was added to every cell pellet before
209 resuspension by vortexing. After 10 min incubation on ice, 27.5 µl ice-cold CER 2 buffer was added to
210 the mix and samples were centrifuged at 16000 g on 4°C for 5 min. Supernatant resulting from this
211 centrifugation step was stored as cytoplasmic extract. The pellet fraction was resolved in 250 µl ice-
212 cold NER and vortexed vigorously every 10 min for 15 s during an incubation period on ice for a total
213 of 40 min. Afterwards, nuclear extract was obtained by centrifugation at 16000 g for 10 min at 4°C and
214 the supernatant containing this nuclear extract was transferred to a new tube.
215 In parallel, 10^6 cells were washed once with PBS prior to resuspension in 150 µl ice-cold RIPA lysis
216 buffer [50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 1 mM EDTA, 1 mM EGTA, 0.1% SDS, 0.5%
217 sodium deoxycholate (DOC) and complete protease inhibitor cocktail (Roche, 11873580001)].
218 Following 10 min incubation on ice, lysates were centrifuged at 16000 g for 15 min at 4°C and the
219 supernatant was transferred to a new tube (= total cell lysate sample). For every fraction, the protein
220 concentrations as determined by a Bradford assay and 40 µg was loaded on ExpressPlus SDS-PAGE
221 (GenScript, M42015) and afterwards blotted on a PVDF membrane. α-tubulin was used as cytoplasmic
222 marker (Sigma, T5168).

223 **BioID affinity purification**

224 For every sample three 145 cm^2 culture dishes (Nunc) were seeded with $4.5x10^6$ HCT116 *TP53*$^{+/T2A\text{-}BirA*}$
225 cells cultured on McCoy's 5A modified medium supplemented with dialyzed FBS. 24 h after seeding 1
226 µM doxorubicin was added. To every culture dish biotin was added to a final concentration of 50 µM
227 at the start of the treatment. 24 h after treatment the growth medium was replenished to deprive
228 free biotin from the culture. 3 h later, the growth medium was aspirated and cells were washed once
229 with cold PBS before detachment in PBS using a cell scraper. After pelleting cells by centrifugation for

230 5 min at 500 g on 4°C, a second PBS washing step occurred prior to cell lysis in a 15 ml tube. 2 ml of
231 ice-cold RIPA lysis buffer containing 250 U of benzonase (Sigma, E1014) was added to each pellet. Lysis
232 was obtained by incubation with agitation for 1 h on 4°C prior to sonication (30% amplitude, 5 x 6 s
233 burst, 2 s interruption) on ice. The insoluble fraction of the lysate was removed by centrifugation for
234 15 min at 16000 g on 4°C and the supernatant was transferred to a new 15 ml tube. The total protein
235 concentration was determined by the Bradford protein assay to normalize input material to the
236 maximum shared total protein amount throughout the samples. For every sample, 90 µl Streptavidin
237 Sepharose High Performance bead suspension (GE Healthcare, 17-5113-01) was used for enrichment
238 of biotinylated proteins. Beads were pelleted by centrifugation at 500 g for 1 min and washed once in
239 wash buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 1 mM EDTA, 1 mM EGTA, 0,1% SDS)
240 prior to addition to the lysate. Samples were incubated for affinity purification on a rotator at 4°C for
241 3 h. Afterwards, beads were recovered by centrifugation of the samples at 500 g for 1 min at 4 °C and
242 aspirating the supernatant. Beads were washed three times in wash buffer, twice in 50 mM
243 ammonium bicarbonate pH 8.0 and once with 20 mM Tris-HCl pH 8.0, 2 mM $CaCl_2$ prior to
244 resuspension in 20 µl 20 mM Tris-HCl pH 8.0. Trypsin digest occurred overnight by addition of 1 µg
245 trypsin (Promega, V5111) to each sample. Digestion mixtures were depleted of beads upon
246 centrifugation at 500 g for 1 min, after which the supernatant was transferred to a mass spectrometry
247 vial. To ensure complete digestion, peptide mixtures were incubated with an additional 500 ng trypsin
248 for 3 h before addition of formic acid to a final concentration of 2%. All experiments were performed
249 in biological triplicate for downstream label-free quantitative proteome analysis.
250 For transient transfection BioID experiments, $8.5x10^6$ HCT116 cells were seeded in a 56 cm^2 culture
251 dish (Nunc) and transfected the day after with pMet7-FLAG-PQS1-TP53-T2A-BirA*-Myc, the pMet7-
252 FLAG-PQS1-TP53-MUTT2A-BirA*-Myc or pSV-SPORT mock vector using FugeneHD. 24 h post
253 transfection, cells were transferred to 3 T145 cm^2 dishes per sample. 24 h after transfer, 1 µM
254 doxorubicin and 50 µM biotin were added 24 h prior to harvest. Affinity purification of biotinylated
255 proteins was performed as described above.

256 **LC-MS/MS instrument analysis**

257 Peptide mixtures were analyzed by LC−MS/MS on an Ultimate 3000 RSLC nano LC (Thermo Fisher
258 Scientific) in-line connected to a Q-Exactive mass spectrometer (Thermo Fisher Scientific). The
259 peptides were first loaded on a trapping column (made in-house, 100 µm internal diameter (I.D.) × 20
260 mm, 5 µm beads C18 Reprosil-HD, Dr. Maisch, Ammerbuch-Entringen, Germany). After flushing the
261 trapping column, peptides were loaded in solvent A (0.1% formic acid) on a reverse-phase column
262 (made in-house, 75 µm I.D. x 250 mm, 3 µm Reprosil-Pur-basic-C18-HD beads packed in the needle,
263 Dr. Maisch, Ammerbuch-Entringen, Germany) and eluted by an increasing concentration solvent B
264 (0.1% formic acid in acetonitrile) using a linear gradient from 2% solvent B up to 55% solvent B in 120
265 min, followed by a washing step with 99% solvent B, all at a constant flow rate of 300 nl/min. The mass
266 spectrometer was operated in data-dependent acquisition (DDA), positive ionization mode,
267 automatically switching between MS and MS/MS acquisition for the 5 most abundant peaks in a given
268 MS spectrum. Source voltage was set at 3.4 kV, with a capillary temperature of 275°C. One MS1 scan
269 (m/z 400−2000, AGC target 3×10^6 ions, maximum ion injection time 80 ms), acquired at a resolution
270 of 70000 (at 200 m/z), was followed by up to 5 tandem MS scans (resolution 17500 at 200 m/z) of the
271 most intense ions fulfilling predefined selection criteria (AGC target 5×10^4 ions, maximum ion
272 injection time 80 ms, isolation window 2 Da, fixed first mass 140 m/z, spectrum data type: centroid,
273 underfill ratio 2%, intensity threshold $1.3xE^4$, exclusion of unassigned, 1, 5-8 and >8 positively charged
274 precursors, peptide match preferred, exclude isotopes on, dynamic exclusion time 12 s). HCD collision
275 energy was set to 25% normalized collision energy and the polydimethylcyclosiloxane background ion
276 at 445.120025 Da was used for internal calibration (lock mass).

277 **Mass spectrometry data processing and interpretation**

278 Obtained Xcalibur Raw Files (.raw) were analyzed using MaxQuant and maxLFQ algorithms (MaxQuant
279 version 1.5.8.3) [45, 46]. Spectra were searched against the human UniProt sequence database (as
280 available on 01/09/2017). Methionine oxidation, N-terminal acetylation and lysine biotinylation were
281 set as variable modifications, with a maximum of 5 modifications per peptide. The minimum peptide

282 length was set at 7 amino acids and a maximum peptide mass of 4600 Da was used. PSM, protein and
283 site false discovery rates were set at 0.01. The minimum Label-Free Quantitation (LFQ) ratio count was
284 2 and the Fast LFQ option was disabled. 20 ppm and 4.5 ppm mass accuracies were used for the first
285 and main search respectively. After completion of searches, LFQ intensities were loaded in Perseus
286 (version 1.5.5.3) [47] for further analysis.
287 Samples were annotated on condition for analysis. Proteins only identified by site, reverse hits and
288 Perseus contaminant proteins were removed from the data matrix. Retained LFQ intensities were
289 transformed to log2 scale and three valid values in at least one of the two sample groups were needed
290 in order for the protein to be retained for further analysis. Missing values were imputed from a normal
291 distribution of intensities (0.3 width, 2.2 downshift). A two-sided t-test (0.05 FDR, 1000
292 randomizations) was performed to reveal differentially enriched proteins in the volcano plot. Gene
293 ontology (GO) analyses were performed using the database for annotation, visualization and
294 integrated discovery (DAVID 6.8) [48]. The mass spectrometry proteomics data have been deposited
295 to the ProteomeXchange Consortium via the PRIDE [49] partner repository with the dataset identifier
296 PXD010771.

297

298 # RESULTS

299 **BioID by forced expression of a p53-BirA* fusion protein in colon carcinoma cells**

300 Proximity biotinylation is an exciting development of recent years to identify vicinal proteins of a
301 target protein which includes the associated protein complexes [19, 22]. While the downstream
302 purification of biotinylated proteins of the *in vivo* biotinylation reaction is a powerful and sensitive
303 procedure, great care should be taken with the design of the experiment and the levels of BirA* [50].
304 Most BioID studies rely on inducible expression systems based on Flp recombinase integration (Flp-
305 In™ T-REx™ system). This powerful system allows low expression which can be tuned to match
306 endogenous levels. Despite being around for well over 10 years, only few engineered Flp-In cell lines
307 are available and virtually all BioID studies using this system are therefore performed in 293 cells [50-
308 52].

309 We set out to explore the so-called 'proxisome' of p53 in well-known and well-characterized colon
310 carcinoma HCT116 cells. The model system has been used extensively for the study of p53 function,
311 in part because of the availability of many knock out cell lines [53]. We were faced with the lack of an
312 inducible HCT116 cell line and we therefore decided to first explore an experiment with forced
313 expression of a p53-BirA* construct. We transiently transfected HCT116 cells with a fusion construct
314 that contains a C-terminal BirA* module cloned in frame through an inactivated T2A peptide that
315 served as a long linker (p53-MUTT2A-BirA*). As control experiments we used either mock-transfected
316 cells or cells transfected with a bi-cistronic p53-T2A-BirA* construct (Figure 1A and 1B). This last
317 construct ensured equal levels of the free BirA* module in the control cells, providing a distinct
318 advantage over expression of a free unfused BirA* protein which often requires tuning to obtain equal
319 expression levels. We performed triplicate experiments to allow LFQ analysis using MaxQuant and
320 Perseus workflows. Using the mock control, an extensive list of significant proteins for p53-BirA* was
321 obtained by LFQ analysis (Figure 1C). Of these, 17.1% were known p53 interaction partners in BioGRID.
322 Some well-known interaction partners can be discerned in the list, but despite the comprehensive list,
323 some key interaction partners such as MDM2 are lacking. This extensive list can be expected because

324 the control conditions lack any BirA*. When assessing the T2A-BirA* control condition we found some
325 distinct differences, pointing to removal of contaminants by the presence of BirA* in this control.
326 However, the retrieved lists were still surprisingly long with 408 significant proteins (Figure 1D), of
327 which 20.3% were known BioGRID p53 interaction partners. This large number of hits is likely due to
328 the excessive levels of BirA*, resulting in much background biotinylation in both transient expression
329 approaches. Levels of the fusion protein are well beyond the endogenous levels in the cells (Figure
330 1B). In addition, GO analysis on 'molecular function' showed very few links to p53 function, with 'p53
331 binding' only retrieved at position 38 and 88 for the mock and T2A-BirA* control respectively. GO
332 analysis for 'cellular component' showed strong enrichment of cytoplasmic and membrane proteins
333 (Table S4 and S5) which contradicts the predominant nuclear function of p53. Thus the results of these
334 analyses hint to the presence of substantial amounts of biologically irrelevant, false positive prey
335 proteins.

336 **Generation of a control cell line for BioID on endogenous p53**

337 The long lists in the overexpression experiment are likely caused by the high expression levels of the
338 bait protein. While this can be solved by introducing an inducible expression cassette containing a
339 p53-BirA* fusion construct through lentiviral transduction [54], we decided to explore the exciting
340 developments in the genome editing field to engineer the BirA* module in one allele of the *TP53* gene
341 in HCT116 cells. We started by first creating a control T2A-BirA* cell line as we reasoned that the best
342 way to ensure similar levels in the control condition is to use a bi-cistronic approach (Figure 2A). Similar
343 BirA* control expression levels are also maintained when p53 protein levels are changed due to stress
344 conditions on the cells (e.g. DNA damage). We used CRISPR/Cas9 to introduce a double stranded DNA
345 (dsDNA) break in the last coding exon of *TP53* and provided a single strand DNA (ssDNA) repair
346 template containing a T2A-BirA* module and a floxed expression cassette flanked by two homology
347 regions. Generation of this control cell line was performed using a two-step engineering approach as
348 described earlier ([40], Figure S1). Clones bearing insertions were identified by PCR (Figure S2). In the
349 second step, the floxed selection cassette is removed by CRE recombinase, again followed by PCR

350 screening of single cell clones (Figure S3). Single integration for these clones was shown by Southern
351 blot (Figure S4A), and by Sanger sequencing of both alleles ensuring an intact second allele
352 (Supplementary Figure S4B and S4C). Because of current limits in knock-in efficiencies, single cell
353 clones were selected and evaluated for the introduction of the knock-in construct and for CRE-based
354 removal of the selection cassette.

355 **Deriving an experimental cell line using a base editor approach**

356 An experimental cell line containing an engineered *TP53* locus with a C-terminal BirA* module can be
357 obtained using a similar genome engineering approach as for the control cell line, but there are a
358 number of things to consider. Firstly, the seeding, culturing and screening of single cell clones related
359 to current knock-in efficiencies requires a significant effort effectively doubling genome engineering
360 hands-on time to perform a BioID experiment. Secondly, the generation of the experimental cell line
361 would also require two selection steps of single cell clones. This could introduce clonal variation and
362 may introduce artefacts in the data. To solve these issues, we sought to derive an isogenic population
363 from the control cell line wherein the T2A site would not be present. This can be obtained by another
364 CRISPR/Cas9 knock-in approach using a template with an alternative linker (e.g. an inactive T2A site).
365 However, the efficiencies currently obtained for knock-in are too low to allow such an approach and
366 would still require clonal selection (see higher). As an alternate way to achieve T2A inactivation, we
367 turned to a recently reported Cas9-cytidine deaminase base editor BE3 [27]. Targeted inactivation of
368 the 2A sequence would ensure restoration of bait-BirA* transgene expression (Figure 2B). 2A peptides
369 have conserved amino acid residues that are critical for peptidyl transferase inhibition that causes
370 ribosomal skipping (Figure 2C) [55, 56]. Therefore, the targeted mutagenesis of one of these key
371 residues should enable efficient abrogation of 2A ribosomal skipping activity. To explore this, we
372 generated an expression vector that expresses a fusion construct of the tumor suppressor protein p53
373 linked to BirA* by a T2A sequence. Appropriate codon usage of the T2A sequence was verified to
374 ensure the presence of several target nucleotides for BE3 (Figure 2C). Varying amounts of a FLAG-
375 TP53-T2A-BirA*-MYC expression construct were co-transfected with the BE3 construct and T2A-

376 specific gRNA. Two different gRNAs were assessed for base editing activity (Table S2). Conversion of
377 the auto-cleavage site was clearly detectable by the presence of a high molecular weight band
378 corresponding to a p53-BioID fusion construct on Western blot analysis (Figure 2D).

379 **T2A conversion after integration by genome engineering**

380 A puromycin selection cassette was first cloned into the original BE3 vector allowing for efficient
381 enrichment of transfected cells. This ensures that non-transfected cells, and thus non-edited cells, are
382 efficiently removed in a short puromycin selection step. These non-edited cells would generate
383 additional noise to downstream LFQ proteomics analysis [45]. The gRNA directly targets the conserved
384 amino acid residues on the T2A sequence while leaving p53 and BirA* intact (inactivated T2A is
385 hereafter indicated as MUTT2A). When assessing p53 translation products before or after BE3
386 transfection, a clear signal for p53-MUTT2A-BirA* was observed in puromycin-resistant cells (Figure
387 3A), a result which was obtained for multiple clones. Notably, an additional band with an apparent
388 molecular weight corresponding to WT p53 was also present. This extra band can be explained by the
389 heterozygous knock-in of T2A-BirA*. In HCT116 *TP53*$^{+/\text{T2A-BirA*}}$ the modified p53 allele has 22 extra
390 amino acids due to translation of the GSG linker and the T2A peptide. This band could still be observed
391 in the enriched cell population post-transfection, hinting at a portion of cells that did not undergo T2A
392 sequence conversion. This partial conversion can be attributed to the intrinsic activity of BE3 that is
393 below 100%. Furthermore, it is likely that not all editing will result in T2A inactivation as BE3
394 mutagenesis activity is centered around positions 4 – 8 in the protospacer sequence [27], therefore
395 mutations might be introduced that have no impact on T2A activity or only attenuate its potency. To
396 quantify the mutagenesis rate we performed next-generation sequencing on the resulting cell
397 population showing a conversion rate over 40% (Figure 3B). In agreement with earlier reports on base
398 editing, CRISPRESSO analysis showed that the rate of indel formation in the resulting population was
399 negligible and the mutagenesis window was restricted to the predicted nucleotides (Figure S5).
400 Next to efficient T2A inactivation, proper localization of p53-MUTT2A-BirA* and BirA* is clearly crucial
401 for the effectiveness and reliability of the system. Ideally, both the bait and the biotinylation activity

402 of the engineered cell lines should follow the same subcellular distribution before and after BE3
403 treatment to provide relevant background protein signals. To evaluate this, we derived nuclear and
404 cytoplasmic fractions of the engineered cells upon p53 activation by doxorubicin. We could observe a
405 clear nuclear enrichment for p53 in both parental and engineered cells, indicative that proper p53
406 localization is not impaired by knock-in of T2A-BirA* or base editor treatment (Figure 3C). As expected,
407 only in the BE3 treated engineered cells a strong signal for p53-T2A-BirA* could be observed in the
408 nuclear fraction. Furthermore, assessment of the biotinylation pattern showed an intense laddering
409 pattern in engineered cells when compared to the parental HCT116 cells (Figure 3C). Both before and
410 after BE3 treatment, biotinylation was mainly detectable in the nuclear fraction, which indicates that
411 the engineered cells with cleaved BirA* (p53-T2A-BirA*) provide a good control condition.
412 Additionally, biotinylated WT p53 can be clearly observed in BE3 treated cells, hinting that fusion to
413 BirA* is not affecting the oligomerization properties of p53 (Figure 3C).

414 **Conversion of inserted T2A-BirA* as a potent tool to study the p53 proxisome**

415 We then performed endogenous BioID experiments using the engineered and converted HCT116
416 *TP53*$^{+/MUTT2A\text{-}BirA*}$ cells. Experiments were done in triplicate to allow MaxQuant LFQ analysis [45]. Cells
417 were harvested and lysed after provoking a p53 response by doxorubicin. Puromycin-selected HCT116
418 *TP53*$^{+/T2A\text{-}BirA*}$ cells transfected with BE3 in absence of T2A-directed gRNA served as control condition
419 (Figure 4A). Following a two-sided t-test (FDR: 0.05, s0: 0.1) we observed 186 potential p53 interactors
420 for the mutagenized cells (Figure 4B and table S6), resulting in a high-quality proxisome dataset for
421 p53. String pathway analysis showed a highly interconnected network around p53 with complexes
422 involved in transcription and chromatin remodelling (Figure S6). Among the most differential proteins
423 is p53 (p-value <0.00001), supporting that T2A is mutagenized in a way that abrogates ribosomal
424 skipping by the peptide. In the control condition treated with doxorubicin, p53 was detected at lower
425 levels, hinting at rare events whereby the T2A peptide did not cause interruption of translation in
426 accordance with the Western blot analysis (Figure 3A). Next to p53, 63 (34.2%) known direct p53
427 interaction partners were identified including several hallmark interactors such as MDM2 [57],

428 TP53BP1 [58], EP400 [59], and Sin3B [60]. Noteworthy, we also detected proteins that affect the post-
429 translational (PTM) signature of p53 such as the phosphatase PPP1R10 [61]. Several proteins that
430 could not directly be linked to p53 have been already implied in DNA damage pathways such as
431 DMAP1 [62] and EPC1 [63]. For virtually all significantly enriched proteins, multiple unique peptides
432 were detected, enabling a robust protein inference (Table S6). GO analysis showed highly specific
433 enrichment for p53-related 'cellular component' and 'molecular function' terms (Table S7). 'p53
434 binding' can be found at position 13 with a p-value of $1.21x10^{-10}$. These results support the concept
435 that BioID on the endogenous protein level can be exploited to further characterize dynamic p53
436 interactions in a physiologically relevant context.
437 To show the benefit of using a converted T2A design strategy, we also compared non-engineered WT
438 control cells (transfected with BE3 and selected on puromycin) to the p53-MUTT2A-BirA* cells as this
439 would mimic a simpler design strategy (Figure S7A). This analysis revealed 397 differential proteins
440 with a clear enrichment of RNA-binding proteins and chaperones (Figure S7B; Table S8) with 'p53
441 binding' found on position 17 with a p-value of $3.62x10^{-8}$. When compared to the T2A-BirA* control
442 condition, 232 proteins were exclusively significant for the analysis using transfected HCT116 WT cells
443 (Figure S7C). A part of this extended protein list is also found in the BioGRID database as interactors
444 of p53. Strikingly, when these additional p53 BioGRID hits were assessed for 'cellular component'
445 terms they were mainly assigned as cytoplasmic proteins (180 proteins; p-value $9.19x10^{-14}$), and their
446 'molecular function' was linked to functions such as RNA-binding (51 proteins; p-value $4.83x10^{-17}$),
447 ribosomal constituents (34 proteins; p-value $1.03x10^{-17}$) and poly(A) RNA or mRNA binding (142 and
448 15 proteins; p-values $2.02x10^{-67}$ and $1.14x10^{-6}$ respectively; Table S8). While these RNA-binding
449 proteins and chaperones can be *bona fide* interaction partners for p53, they likely turn up with many
450 different bait proteins (or directly with BirA*). Consequently, proteins with functions related to RNA
451 are typically enriched in contaminant repositories for affinity purification – mass spectrometry
452 experiments [64].

453 **DISCUSSION**

454 Since its initial emergence, BioID has been widely adopted by the proteomics community to detect
455 protein-protein interactions and vicinal proteins because of its distinct advantages over classic AP-MS
456 [22]. However, virtually all efforts rely on ectopic (over)expression of bait proteins fused to BioID [22].
457 To our knowledge, there is currently only one report where BioID is performed at the endogenous
458 level [65]. The endogenous application of BioID solves two important issues in typical MS-based
459 interactomics approaches. Firstly, the need for lysis is obviated since prey labelling occurs in intact
460 cells before the homogenization step, and secondly, overexpression artefacts are removed by using
461 the endogenous bait protein. This last issue is supported by our experiments with ectopic expression
462 of p53-BirA* wherein we clearly show extensive lists of candidate proximal proteins pointing to the
463 presence of many contaminants. Coupling expression levels of both bait-BirA* and control BirA* might
464 be of particular relevance for proteins, such as p53, that have varying expression levels depending on
465 the cellular state. One longstanding question in the p53 field is how p53 is able to elicit different
466 responses to separate stimuli or stress conditions in a cell. We believe that future studies with BioID
467 at the endogenous level may reveal some of the subtle interaction changes or altered feedback loops
468 that are potentially overlooked in overexpression studies, thereby providing missing pieces to the p53
469 puzzle. Furthermore, the majority of p53 research has been based on overexpression experiments and
470 therefore PPI databases may be riddled with interactions that are spatially and physiologically not
471 possible in an intact cell. Although highly specific anti-p53 antibodies are commercially available for
472 conducting endogenous immunoprecipitation MS experiments, these will always have the drawback
473 that protein complexes need to be kept intact during purification. While crosslinking of protein
474 complexes may improve conservation during purification, crosslinking is prone to inter-experimental
475 variation in efficiency, suffers from ambiguous experimental data and can lead to problematic
476 identification of crosslinked peptides by search algorithms [66].
477 We engineered a single cell line that can be used both as a control and an experimental cell line

478 without the need to go through an additional intermediary single cell stage. Our approach relies on
479 mutagenesis of a T2A sequence incorporated as a linker between a bait protein of interest and a
480 functional tag sequence, here BirA*. By mutagenizing the T2A site, ribosomal skipping is prevented
481 and a functional fusion construct is generated. We showed the concept of T2A mutagenesis on vectors
482 for overexpression and on a knock-in HCT116 *TP53*$^{+/T2A\text{-}BirA*}$ cell line. In addition, we show correct
483 localization of the BirA* enzyme in the modified cells, both before and after T2A inactivation. Based
484 on these results, we believe that our approach is suitable for cytoplasmic and nuclear bait proteins.
485 The current system does not allow to target the released BirA* to restricted compartments (e.g.
486 mitochondria), which may be solved by adding localization signals. Proper localization should be
487 carefully assessed for every bait protein, which is required for any interactome study.
488 As site-directed mutagenesis is accomplished by introduction of a CRISPR/nCas9-cytidine deaminase
489 fusion (BE3), a simple short selection step suffices to obtain a strong enrichment of cells transiently
490 expressing the BE3 base editor and therefore having a high chance of expressing p53-MUTT2A-BirA*.
491 In this way, there is no need to generate separate control cell lines, or to have a design wherein
492 engineered single cell knock in clones are compared to parental populations. Such a design would
493 inevitably lead to higher background levels as presented in this work (Figure S7, table S8). Until
494 genome engineering progresses to efficiencies wherein clonal screening is no longer needed, T2A
495 mutagenesis may be a valuable approach to obtain appropriate isogenic control cell lines. While the
496 devised T2A system allows rigorous control of expression levels, also when gene expression levels are
497 modulated by various stimuli (i.e. doxorubicin effects on p53 levels), the free BirA* cannot replicate
498 the exact localization of the bait protein. A similar localization would allow discrimination between
499 proteins in close proximity and interaction partners (both direct or indirect) [67]. As with all BioID
500 designs, downstream validation is still required to assess this aspect. While we opted for a classic well-
501 controlled knock-in approach, novel CRISPR/Cas9 strategies would also allow the selection of isogenic
502 populations (e.g. CRISPR co-selection [68], CRISPaint [69]).

503

504 **CONCLUSIONS**

505 Taken together, our elegant base editing approach to derive isogenic experimental cell lines from a
506 control cell line enables BioID experiments at the endogenous level. Bringing together the advantages
507 of proximal biotinylation in intact cells with endogenous bait expression levels and with quantitative
508 proteomics pushes protein complex analysis to the next level and constitutes a useful addition to study
509 hub proteins such as p53.

510 **SUPPORTING INFORMATION**

511 Table S1: Primers used for cloning or screening. (PDF)

512 Table S2: gRNA gblock sequences cloned in pCR-Blunt vector for transfection. (PDF)

513 Table S3: CRISPRESSO allele frequency output. (Excel)

514 Table S4: GO analysis of the results from an analysis of a transient transfection p53 BioID experiment

515 using a mock vector as control. (PDF)

516 Table S5: GO analysis of the results from an analysis of a transient transfection p53 BioID experiment

517 using a bi-cistronic p53-T2A-BirA* expression construct as control. (PDF)

518 Table S6: Perseus output endogenous T2A inactivation. (Excel)

519 Table S7: GO analysis of significantly enriched proteins from converted p53-MUTT2A-BirA* HCT116

520 cells when compared to isogenic non-converted p53-T2A-BirA* cells. (PDF)

521 Table S8: GO analysis of significantly enriched proteins from converted p53-MUTT2A-BirA* HCT116

522 cells when compared to parental HCT116 cells. (PDF)

523 Table S9: Perseus data matrices generated for all analyses. (Excel)

524 Figure S1: Targeting strategy and primer scheme used for PCR screening. (PDF)

525 Figure S2: PCR screening of neomycin resistant clonal populations. (PDF)

526 Figure S3: PCR for assessing the removal of the selection cassette upon TAT-Cre treatment. (PDF)

527

528 Figure S4: Validation of targeted T2A-BirA* insertion. (PDF)

529 Figure S5: CRISPRESSO analysis of the T2A region after BE3 base editing. (PDF)

530 Figure S6: STRING analysis of proximal proteins for endogenous p53. (PDF)

531 Figure S7: Evaluation of BE3 transfected non-engineered HCT116 WT cells as control for p53 BioID.

532 (PDF)

533 ## ACKNOWLEDGEMENTS

534 We thank the VIB Proteomics Core for running the MS samples. We thank Elise Wyseure and Noortje
535 Samyn for technical assistance. pCMV-BE3 was provided by David Liu (Addgene plasmid #73021).

536 ## AUTHOR CONTRIBUTIONS

537 G.V. designed and performed cell line experiments with help of D.D.S, performed BioID experiments,
538 analyzed data and wrote the manuscript. G.V. and D.D.S performed cell line engineering and related
539 characterization experiments with help of A.M. J.P. operated the MS and analyzed the samples. S.E.
540 conceived and managed the project and wrote the manuscript. K.G. is co-promotor of G.V. and read
541 and corrected the manuscript.

542 ## FUNDING

543 G.V. is a PhD student funded by the Fund for Research - Flanders (FWO grant 3G050913N to K.G. and
544 S.E.) and BOF-GOA (2016000602 to S.E). K.G. acknowledges support from the Flanders Agency for
545 innovation by Science and Technology (IWT, SBO grant 60839).

546 ## CONFLICT OF INTEREST

547 The authors declare no competing interests.

548

549 **REFERENCES**

550 [1] S. Petrakis, M.A. Andrade-Navarro, Editorial: Protein Interaction Networks in Health and Disease,
551 Front Genet 7 (2016) 111.
552 [2] K. Titeca, I. Lemmens, J. Tavernier, S. Eyckerman, Discovering cellular protein-protein
553 interactions: Technological strategies and opportunities, Mass Spectrom Rev (2018) 1-33.
554 [3] A. Chatr-Aryamontri, R. Oughtred, L. Boucher, J. Rust, C. Chang, N.K. Kolas, L. O'Donnell, S. Oster,
555 C. Theesfeld, A. Sellam, C. Stark, B.J. Breitkreutz, K. Dolinski, M. Tyers, The BioGRID interaction
556 database: 2017 update, Nucleic Acids Res 45(D1) (2017) D369-D379.
557 [4] K. Luck, G.M. Sheynkman, I. Zhang, M. Vidal, Proteome-Scale Human Interactomics, Trends
558 Biochem Sci 42(5) (2017) 342-354.
559 [5] T.J. Gibson, M. Seiler, R.A. Veitia, The transience of transient overexpression, Nat Methods 10(8)
560 (2013) 715-21.
561 [6] G. Prelich, Gene overexpression: uses, mechanisms, and interpretation, Genetics 190(3) (2012)
562 841-54.
563 [7] J.E. Dumont, S. Dremier, I. Pirson, C. Maenhaut, Cross signaling, cell specificity, and physiology,
564 Am J Physiol Cell Physiol 283(1) (2002) C2-28.
565 [8] Y.B. Zhang, J. Howitt, S. McCorkle, P. Lawrence, K. Springer, P. Freimuth, Protein aggregation
566 during overexpression limited by peptide extensions with large net negative charge, Protein Expr
567 Purif 36(2) (2004) 207-16.
568 [9] T.K. Kim, J.H. Eberwine, Mammalian cell transfection: the present and the future, Anal Bioanal
569 Chem 397(8) (2010) 3173-8.
570 [10] A. Malovannaya, R.B. Lanz, S.Y. Jung, Y. Bulynko, N.T. Le, D.W. Chan, C. Ding, Y. Shi, N. Yucer, G.
571 Krenciute, B.J. Kim, C. Li, R. Chen, W. Li, Y. Wang, B.W. O'Malley, J. Qin, Analysis of the human
572 endogenous coregulator complexome, Cell 145(5) (2011) 787-99.
573 [11] L. Fagerberg, B.M. Hallstrom, P. Oksvold, C. Kampf, D. Djureinovic, J. Odeberg, M. Habuka, S.
574 Tahmasebpoor, A. Danielsson, K. Edlund, A. Asplund, E. Sjostedt, E. Lundberg, C.A. Szigyarto, M.
575 Skogs, J.O. Takanen, H. Berling, H. Tegel, J. Mulder, P. Nilsson, J.M. Schwenk, C. Lindskog, F.
576 Danielsson, A. Mardinoglu, A. Sivertsson, K. von Feilitzen, M. Forsberg, M. Zwahlen, I. Olsson, S.
577 Navani, M. Huss, J. Nielsen, F. Ponten, M. Uhlen, Analysis of the human tissue-specific expression by
578 genome-wide integration of transcriptomics and antibody-based proteomics, Mol Cell Proteomics
579 13(2) (2014) 397-406.
580 [12] M.G. Weller, Quality Issues of Research Antibodies, Anal Chem Insights 11 (2016) 21-7.
581 [13] A. Bradbury, A. Pluckthun, Reproducibility: Standardize antibodies used in research, Nature
582 518(7537) (2015) 27-9.
583 [14] E. Marcon, H. Jain, A. Bhattacharya, H. Guo, S. Phanse, S. Pu, G. Byram, B.C. Collins, E. Dowdell,
584 M. Fenner, X. Guo, A. Hutchinson, J.J. Kennedy, B. Krastins, B. Larsen, Z.Y. Lin, M.F. Lopez, P.
585 Loppnau, S. Miersch, T. Nguyen, J.B. Olsen, M. Paduch, M. Ravichandran, A. Seitova, G. Vadali, M.S.
586 Vogelsang, J.R. Whiteaker, G. Zhong, N. Zhong, L. Zhao, R. Aebersold, C.H. Arrowsmith, A. Emili, L.
587 Frappier, A.C. Gingras, M. Gstaiger, A.G. Paulovich, S. Koide, A.A. Kossiakoff, S.S. Sidhu, S.J. Wodak, S.
588 Graslund, J.F. Greenblatt, A.M. Edwards, Assessment of a method to characterize antibody
589 selectivity and specificity for use in immunoprecipitation, Nat Methods 12(8) (2015) 725-31.
590 [15] D. Gawron, E. Ndah, K. Gevaert, P. Van Damme, Positional proteomics reveals differences in N-
591 terminal proteoform stability, Molecular systems biology 12(2) (2016) 858.
592 [16] J.Z. Torres, J.J. Miller, P.K. Jackson, High-throughput generation of tagged stable cell lines for
593 proteomic analysis, Proteomics 9(10) (2009) 2888-91.
594 [17] S. Senkel, C. Waldner, G.U. Ryffel, H. Thomas, Improved conditional expression systems
595 resulting in physiological level of HNF4alpha expression confirm HNF4alpha induced apoptosis in the
596 pancreatic beta-cell line INS-1, BMC Res Notes 2 (2009) 210.

597 [18] M.Y. Hein, N.C. Hubner, I. Poser, J. Cox, N. Nagaraj, Y. Toyoda, I.A. Gak, I. Weisswange, J.
598 Mansfeld, F. Buchholz, A.A. Hyman, M. Mann, A human interactome in three quantitative
599 dimensions organized by stoichiometries and abundances, Cell 163(3) (2015) 712-23.
600 [19] K.J. Roux, D.I. Kim, M. Raida, B. Burke, A promiscuous biotin ligase fusion protein identifies
601 proximal and interacting proteins in mammalian cells, J Cell Biol 196(6) (2012) 801-10.
602 [20] A. Chapman-Smith, J.E. Cronan, Jr., The enzymatic biotinylation of proteins: a post-translational
603 modification of exceptional specificity, Trends Biochem Sci 24(9) (1999) 359-63.
604 [21] A. Chapman-Smith, J.E. Cronan, Jr., Molecular biology of biotin attachment to proteins, J Nutr
605 129(2S Suppl) (1999) 477S-484S.
606 [22] R. Varnaite, S.A. MacNeill, Meet the neighbors: Mapping local protein interactomes by
607 proximity-dependent labeling with BioID, Proteomics 16(19) (2016) 2503-2518.
608 [23] Z. Liu, O. Chen, J.B.J. Wall, M. Zheng, Y. Zhou, L. Wang, H. Ruth Vaseghi, L. Qian, J. Liu,
609 Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector, Sci Rep 7(1)
610 (2017) 2193.
611 [24] A.L. Szymczak, C.J. Workman, Y. Wang, K.M. Vignali, S. Dilioglou, E.F. Vanin, D.A. Vignali,
612 Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral
613 vector, Nat Biotechnol 22(5) (2004) 589-94.
614 [25] R. Barrangou, J.A. Doudna, Applications of CRISPR technologies in research and beyond, Nat
615 Biotechnol 34(9) (2016) 933-941.
616 [26] C. Kuscu, M. Adli, CRISPR-Cas9-AID base editor is a powerful gain-of-function screening tool, Nat
617 Methods 13(12) (2016) 983-984.
618 [27] A.C. Komor, Y.B. Kim, M.S. Packer, J.A. Zuris, D.R. Liu, Programmable editing of a target base in
619 genomic DNA without double-stranded DNA cleavage, Nature 533(7603) (2016) 420-4.
620 [28] Y. Ma, J. Zhang, W. Yin, Z. Zhang, Y. Song, X. Chang, Targeted AID-mediated mutagenesis (TAM)
621 enables efficient genomic diversification in mammalian cells, Nat Methods 13(12) (2016) 1029-1035.
622 [29] N.M. Gaudelli, A.C. Komor, H.A. Rees, M.S. Packer, A.H. Badran, D.I. Bryson, D.R. Liu,
623 Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage, Nature 551(7681)
624 (2017) 464-471.
625 [30] C. Kuscu, M. Parlak, T. Tufan, J. Yang, K. Szlachta, X. Wei, R. Mammadov, M. Adli, CRISPR-STOP:
626 gene silencing through base-editing-induced nonsense mutations, Nat Methods 14(7) (2017) 710-
627 712.
628 [31] B. Vogelstein, D. Lane, A.J. Levine, Surfing the p53 network, Nature 408(6810) (2000) 307-10.
629 [32] J.A. Royds, B. Iacopetta, p53 and disease: when the guardian angel fails, Cell Death Differ 13(6)
630 (2006) 1017-26.
631 [33] D.P. Lane, Cancer. p53, guardian of the genome, Nature 358(6381) (1992) 15-6.
632 [34] A.M. Goh, C.R. Coffill, D.P. Lane, The role of mutant p53 in human cancer, J Pathol 223(2) (2011)
633 116-26.
634 [35] V.P. Almazov, D.V. Kochetkov, P.M. Chumakov, The use of p53 as a tool for human cancer
635 therapy, Mol Biol (Mosk) 41(6) (2007) 947-63.
636 [36] T. Riley, E. Sontag, P. Chen, A. Levine, Transcriptional control of human p53-regulated genes,
637 Nat Rev Mol Cell Biol 9(5) (2008) 402-12.
638 [37] M. Fischer, Census and evaluation of p53 target genes, Oncogene 36(28) (2017) 3943-3956.
639 [38] E.R. Kastenhuber, S.W. Lowe, Putting p53 in Context, Cell 170(6) (2017) 1062-1078.
640 [39] Y.C. Lin, M. Boone, L. Meuris, I. Lemmens, N. Van Roy, A. Soete, J. Reumers, M. Moisse, S.
641 Plaisance, R. Drmanac, J. Chen, F. Speleman, D. Lambrechts, Y. Van de Peer, J. Tavernier, N.
642 Callewaert, Genome dynamics of the human embryonic kidney 293 lineage in response to cell
643 biology manipulations, Nat Commun 5 (2014) 4767.
644 [40] G. Vandemoortele, A. Staes, G. Gonnelli, N. Samyn, D. De Sutter, E. Vandermarliere, E.
645 Timmerman, K. Gevaert, L. Martens, S. Eyckerman, An extra dimension in protein tagging by
646 quantifying universal proteotypic peptides using targeted proteomics, Sci Rep 6 (2016) 27220.

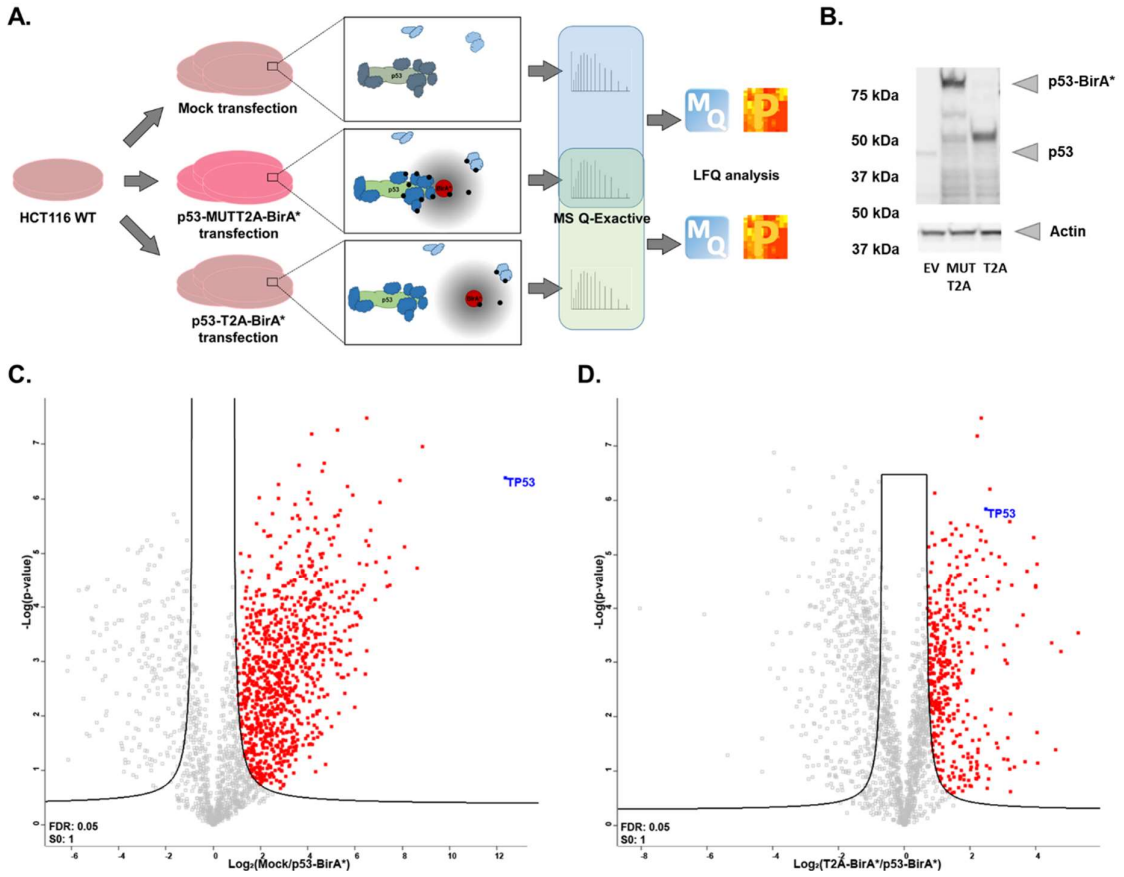
647 [41] G. Vandemoortele, D. De Sutter, S. Eyckerman, Robust Generation of Knock-in Cell Lines Using
648 CRISPR-Cas9 and rAAV-assisted Repair Template Delivery, Bio-protoc 7(7) (2017) e2211.
649 [42] F.A. Ran, P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, F. Zhang, Genome engineering using the
650 CRISPR-Cas9 system, Nat Protoc 8(11) (2013) 2281-2308.
651 [43] L. Pinello, M.C. Canver, M.D. Hoban, S.H. Orkin, D.B. Kohn, D.E. Bauer, G.C. Yuan, Analyzing
652 CRISPR genome-editing experiments with CRISPResso, Nature biotechnology 34(7) (2016) 695-7.
653 [44] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator,
654 Genome research 14(6) (2004) 1188-90.
655 [45] J. Cox, M.Y. Hein, C.A. Luber, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free
656 quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, Mol
657 Cell Proteomics 13(9) (2014) 2513-26.
658 [46] S. Tyanova, T. Temu, J. Cox, The MaxQuant computational platform for mass spectrometry-
659 based shotgun proteomics, Nat Protoc 11(12) (2016) 2301-2319.
660 [47] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus
661 computational platform for comprehensive analysis of (prote)omics data, Nat Methods 13(9) (2016)
662 731-40.
663 [48] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists
664 using DAVID bioinformatics resources, Nature protocols 4(1) (2009) 44-57.
665 [49] J.A. Vizcaino, A. Csordas, N. del-Toro, J.A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol,
666 F. Reisinger, T. Ternent, Q.W. Xu, R. Wang, H. Hermjakob, 2016 update of the PRIDE database and its
667 related tools, Nucleic Acids Res 44(D1) (2016) D447-56.
668 [50] J.P. Lambert, M. Tucholska, C. Go, J.D. Knight, A.C. Gingras, Proximity biotinylation and affinity
669 purification are complementary approaches for the interactome mapping of chromatin-associated
670 protein complexes, Journal of proteomics 118 (2015) 81-94.
671 [51] X. Liu, K. Salokas, F. Tamene, Y. Jiu, R.G. Weldatsadik, T. Ohman, M. Varjosalo, An AP-MS- and
672 BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular
673 localizations, Nature communications 9(1) (2018) 1188.
674 [52] E. Coyaud, M. Mis, E.M. Laurent, W.H. Dunham, A.L. Couzens, M. Robitaille, A.C. Gingras, S.
675 Angers, B. Raught, BioID-based Identification of Skp Cullin F-box (SCF)beta-TrCP1/2 E3 Ligase
676 Substrates, Molecular & cellular proteomics : MCP 14(7) (2015) 1781-95.
677 [53] P. Qiu, H. Guan, P. Dong, S. Li, C.T. Ho, M.H. Pan, D.J. McClements, H. Xiao, The p53-, Bax- and
678 p21-dependent inhibition of colon cancer cell growth by 5-hydroxy polymethoxyflavones, Molecular
679 nutrition & food research 55(4) (2011) 613-22.
680 [54] P. Samavarchi-Tehrani, H. Abdouni, R. Samson, A.C. Gingras, A versatile lentiviral delivery toolkit
681 for proximity-dependent biotinylation in diverse cell types, Molecular & cellular proteomics : MCP
682 (2018).
683 [55] M.L. Donnelly, G. Luke, A. Mehrotra, X. Li, L.E. Hughes, D. Gani, M.D. Ryan, Analysis of the
684 aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a
685 novel translational effect: a putative ribosomal 'skip', J Gen Virol 82(Pt 5) (2001) 1013-25.
686 [56] P.A. Radcliffe, K.A. Mitrophanous, Multiple gene products from a single vector: ‘self-cleaving’ 2A
687 peptides, Gene Therapy 11 (2004) 1673.
688 [57] U.M. Moll, O. Petrenko, The MDM2-p53 interaction, Molecular cancer research : MCR 1(14)
689 (2003) 1001-8.
690 [58] R. Cuella-Martin, C. Oliveira, H.E. Lockstone, S. Snellenberg, N. Grolmusova, J.R. Chapman,
691 53BP1 Integrates DNA Repair and p53-Dependent Cell Fate Decisions via Distinct Mechanisms, Mol
692 Cell 64(1) (2016) 51-64.
693 [59] N.L. Lill, S.R. Grossman, D. Ginsberg, J. DeCaprio, D.M. Livingston, Binding and modulation of
694 p53 by p300/CBP coactivators, Nature 387(6635) (1997) 823-7.
695 [60] N. Bansal, R. Kadamb, S. Mittal, L. Vig, R. Sharma, B.S. Dwarakanath, D. Saluja, Tumor
696 suppressor protein p53 recruits human Sin3B/HDAC1 complex for down-regulation of its target
697 promoters in response to genotoxic stress, PLoS One 6(10) (2011) e26156.

698 [61] S.J. Lee, C.J. Lim, J.K. Min, J.K. Lee, Y.M. Kim, J.Y. Lee, M.H. Won, Y.G. Kwon, Protein
699 phosphatase 1 nuclear targeting subunit is a hypoxia inducible gene: its role in post-translational
700 modification of p53 and MDM2, Cell Death Differ 14(6) (2007) 1106-16.
701 [62] T. Koizumi, M. Negishi, S. Nakamura, H. Oguro, K. Satoh, M. Ichinose, A. Iwama, Depletion of
702 Dnmt1-associated protein 1 triggers DNA damage and compromises the proliferative capacity of
703 hematopoietic stem cells, Int J Hematol 91(4) (2010) 611-9.
704 [63] Y. Wang, V. Alla, D. Goody, S.K. Gupta, A. Spitschak, O. Wolkenhauer, B.M. Putzer, D.
705 Engelmann, Epigenetic factor EPC1 is a master regulator of DNA damage response by interacting
706 with E2F1 to silence death and activate metastasis-related gene signatures, Nucleic Acids Res 44(1)
707 (2016) 117-33.
708 [64] D. Mellacheruvu, Z. Wright, A.L. Couzens, J.P. Lambert, N.A. St-Denis, T. Li, Y.V. Miteva, S. Hauri,
709 M.E. Sardiu, T.Y. Low, V.A. Halim, R.D. Bagshaw, N.C. Hubner, A. Al-Hakim, A. Bouchard, D. Faubert,
710 D. Fermin, W.H. Dunham, M. Goudreault, Z.Y. Lin, B.G. Badillo, T. Pawson, D. Durocher, B. Coulombe,
711 R. Aebersold, G. Superti-Furga, J. Colinge, A.J. Heck, H. Choi, M. Gstaiger, S. Mohammed, I.M. Cristea,
712 K.L. Bennett, M.P. Washburn, B. Raught, R.M. Ewing, A.C. Gingras, A.I. Nesvizhskii, The CRAPome: a
713 contaminant repository for affinity purification-mass spectrometry data, Nature methods 10(8)
714 (2013) 730-6.
715 [65] C.B. Mulholland, M. Smets, E. Schmidtmann, S. Leidescher, Y. Markaki, M. Hofweber, W. Qin, M.
716 Manzo, E. Kremmer, K. Thanisch, C. Bauer, P. Rombaut, F. Herzog, H. Leonhardt, S. Bultmann, A
717 modular open platform for systematic functional studies under physiological conditions, Nucleic
718 Acids Res 43(17) (2015) e112.
719 [66] J. Rappsilber, The beginning of a beautiful friendship: cross-linking/mass spectrometry and
720 modelling of proteins and multi-protein complexes, J Struct Biol 173(3) (2011) 530-40.
721 [67] B.T. Lobingier, R. Huttenhain, K. Eichel, K.B. Miller, A.Y. Ting, M. von Zastrow, N.J. Krogan, An
722 Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells, Cell 169(2) (2017)
723 350-360 e12.
724 [68] D. Agudelo, A. Duringer, L. Bozoyan, C.C. Huard, S. Carter, J. Loehr, D. Synodinou, M. Drouin, J.
725 Salsman, G. Dellaire, J. Laganiere, Y. Doyon, Marker-free coselection for CRISPR-driven genome
726 editing in human cells, Nature methods 14(6) (2017) 615-620.
727 [69] J.L. Schmid-Burgk, K. Honing, T.S. Ebert, V. Hornung, CRISPaint allows modular base-specific
728 gene tagging using a ligase-4-dependent mechanism, Nature communications 7 (2016) 12338.

729

730

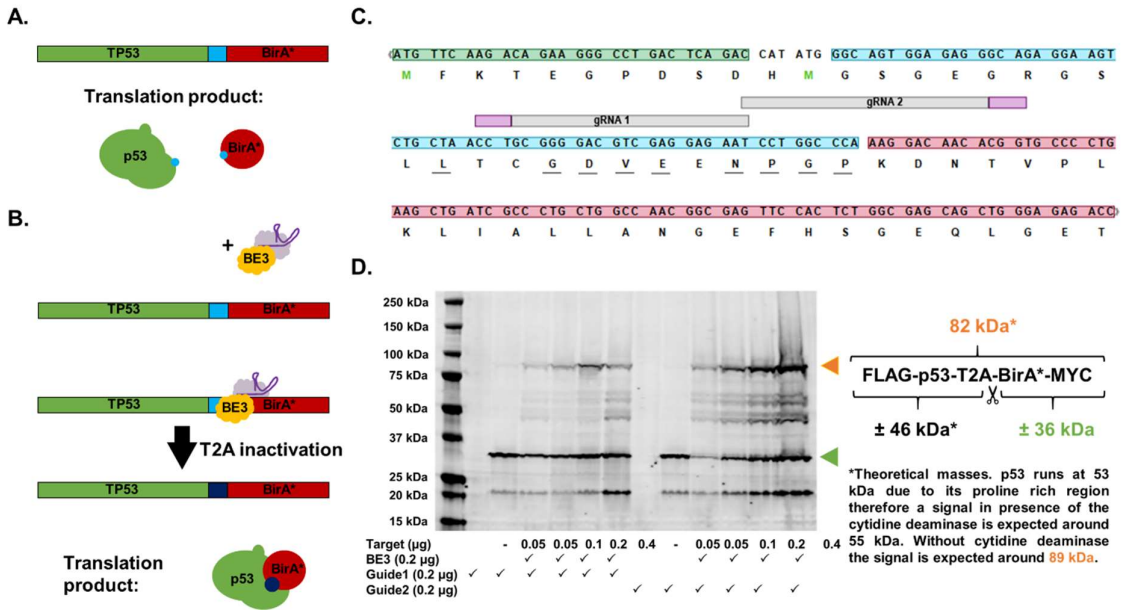
731 **FIGURES**



732

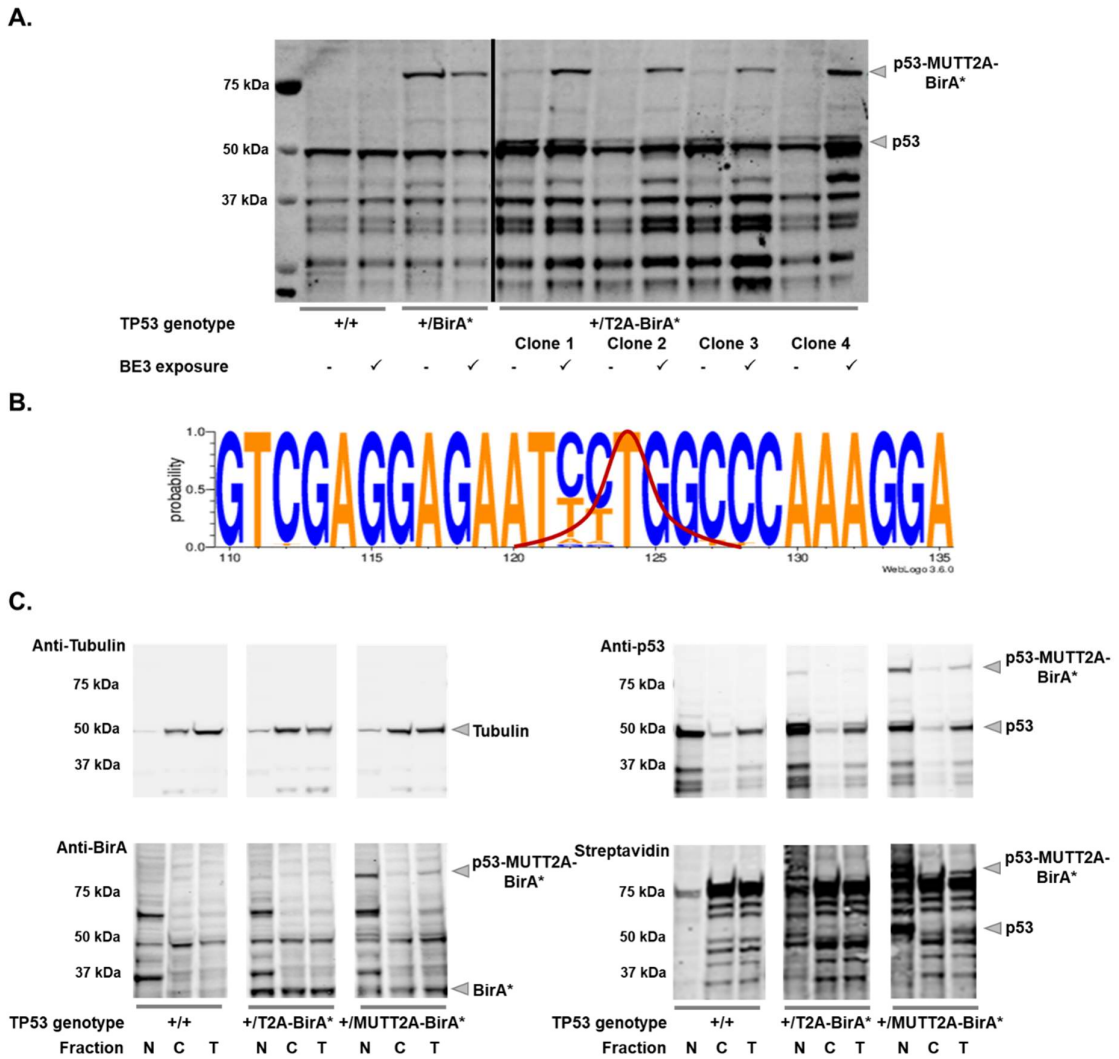
733 **Figure 1. Transfection p53 BioID experiments.** (A) Workflow used for transient expression LFQ
734 proteomics experiments. All experiments were conducted in presence of 1 µM doxorubicin. Top panel:
735 Control transfection with an empty vector. Middle panel: p53-MUTT2ABirA* transfection leading to
736 biotinylation of vicinal p53 proteins. Bottom panel: p53-T2A-BirA* control transfection leading to
737 biotinylation of background proteins. (B) Western blot for p53 in the transient expression experiment.
738 Different lanes depict p53 protein levels in the empty vector transfection (EV), in the fusion p53-
739 MUTT2ABirA* transfection (MUTT2A), and in the transfection with the bi-cistronic p53-T2A-BirA*
740 (T2A) control. (C) and (D) Volcano plots showing differential LFQ intensity levels (X-axis) and p-values
741 (Y-axis) in transient transfection BioID experiments using a mock and BirA* control transfection
742 respectively.

743



744 **Figure 2. Targeted T2A inactivation.** (A) Incorporation of a sequence coding for a T2A peptide (cyan)
745 between the bait (p53) and BirA* will lead to translation of two separate proteins. (B) After
746 transfection, the BE3 base editor will be guided to the T2A sequence by (a) specific gRNA(s) and will
747 convert cytidines mainly to thymidines. Dark blue: inactivated T2A (MUTT2A) (C) Layout of the T2A
748 region in the TP53-T2A-BirA* test construct. Conserved T2A residues are underlined. Green: p53 C-
749 terminus; cyan: linker and T2A sequence; red: BirA* N-terminus; grey: gRNA used; magenta: PAM-
750 motif. (D) Overexpression pilot experiment to assess T2A mutagenesis of a target plasmid. Cells were
751 lysed 72 h after transfection with a MYC-tagged p53-T2A-BirA* construct, and loaded on SDS-PAGE
752 for Western blotting using an anti-MYC antibody. Upon mutagenesis a band corresponding to the 82
753 kDa p53-MUTT2A-BirA* protein is visible. The two gRNAs targeting different residues of the T2A
754 sequence were tested in parallel for potency (left: gRNA 1; right: gRNA 2). In both cases, expression of
755 p53-BirA* could be observed at varying amounts of target concentrations. gRNA 2 was selected for
756 further endogenous use.

757

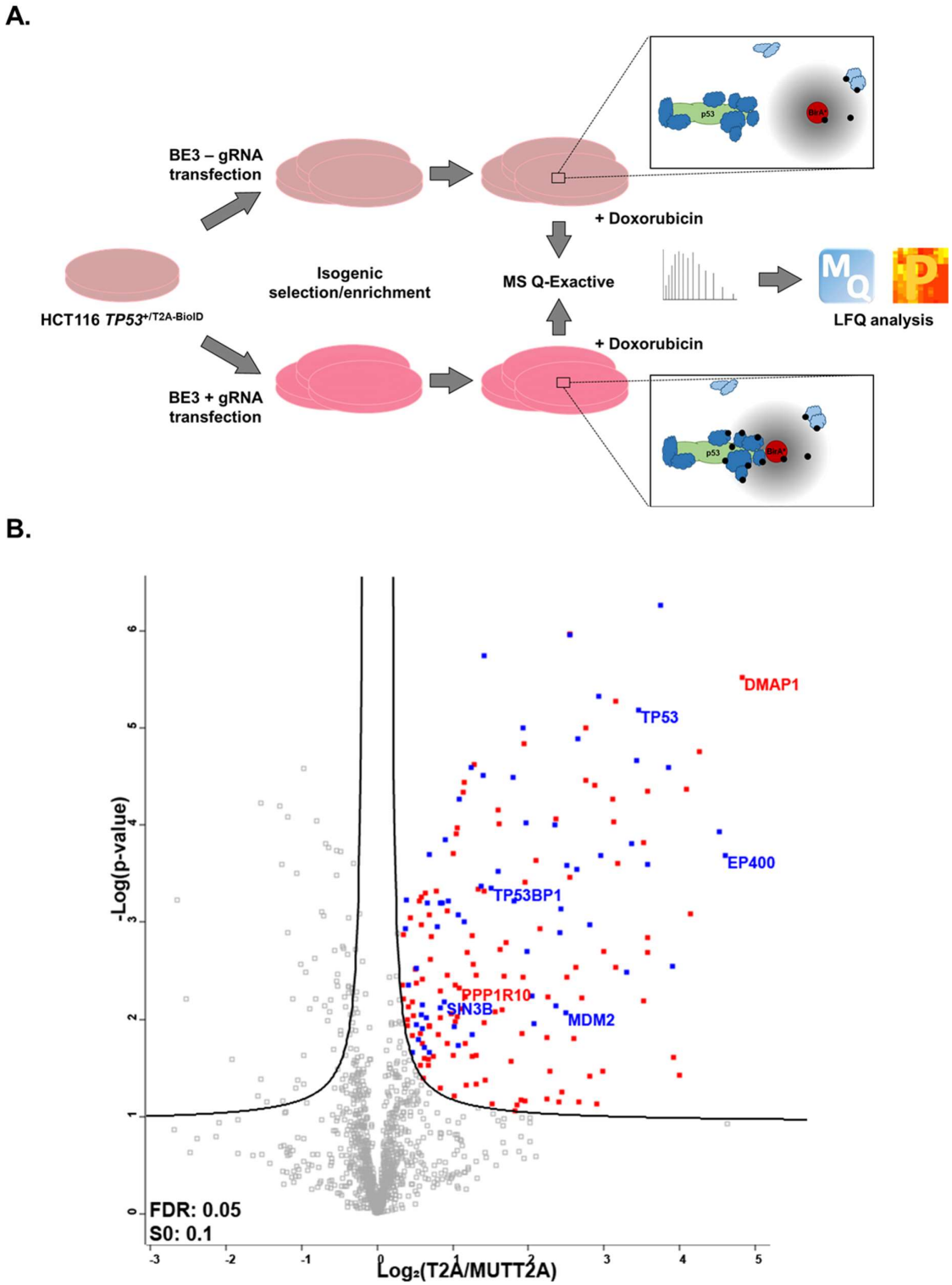


758

759 **Figure 3. Endogenous T2A mutagenesis and cell line characterization** (A) Endogenous T2A
760 mutagenesis. Cells were lysed 72 hours after BE3 transfection and loaded on SDS-PAGE for Western
761 blot analysis using an anti-p53 antibody. p53 or p53-BioID expression is not influenced by BE3
762 expression in parental HCT116 cells or in HCT116 *TP53*$^{+/BirA*}$ control cell lines. T2A inactivation could
763 be observed for all four HCT116 *TP53*$^{+/T2A\text{-}BirA*}$ clones tested resulting in expression of the p53-
764 MUTT2A-BirA* fusion protein. Clone 4 was used for further experiments. (B) Sequence logo depicting
765 mutation frequencies of BE3-treated cells. Paired-end sequencing was performed on the cell
766 population after BE3 treatment and enrichment. The relevant T2A region is depicted in the sequence
767 logo. C > T conversions could only be observed in the targeted residues. Red: expected BE3-induced

768 mutation prevalence according to the position of gRNA 2. (C) Nuclear and cytoplasmic extraction
769 experiments of doxorubicin-treated cell lines. p53 is mostly present in the nuclear extract of all cell
770 lines. Both before and after BE3 treatment, BirA* shows an identical localization in the nucleus, in
771 accordance with p53. In both populations, biotinylation is elevated when compared to parental
772 HCT116 cells, with the majority of biotinylated substrates detected in the nuclear extract. N: nuclear
773 extract; C: cytoplasmic extract; T: Total cell lysate.

774



775

776 **Figure 4. Mapping the endogenous p53 proxisome using BioID.** (A) Proteomics workflow used for

777 HCT116 *TP53*$^{+/T2A\text{-}BirA^*}$. Experiments were done in triplicate to support LFQ analysis. The puromycin

778 resistance cassette in the $BE3_{Puro}$ vector was used for enrichment of the transfected cells. Cells were
779 allowed to recover prior to doxorubicin treatment to activate p53. Zoom: graphic representation of
780 the promiscuous biotinylation in HCT116 *TP53*$^{+/T2A\text{-}BirA*}$ (top) and HCT116 *TP53*$^{+/MUTT2A\text{-}BirA*}$ (bottom).
781 Blue: p53-proximal proteins. Black: biotin. (B) Volcano plot showing the differential LFQ intensity levels
782 (X-axis) and the p-values (Y-axis) for HCT116 *TP53*$^{+/T2A\text{-}BirA*}$ cells before and after T2A conversion in
783 presence of doxorubicin. Known p53 interactors indexed in BioGRID are highlighted in blue. Significant
784 hits discussed in the main text are labeled.

785