Engineering Asian elephant TP53: TP53 retrogene knockouts activate common and unique cancer-relevant pathways

Emil Karpinski* [a], Nikil Badey [b], Esther Mintzer [a], Asaf Ashkenazy-Titelman [a], George M. Church* [a,b].

[a] *Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA*
[b] *Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02215, USA*

* Corresponding authors.

E-mail addresses: emil_karpinski@hms.harvard.edu; gchurch@genetics.med.harvard.edu

Abstract

TP53 functions as a central regulator in response to DNA damage and other cell stressors by inducing the expression of many protective pathways such as cell cycle arrest and apoptosis. Consequently, this gene is often found disrupted in human cancers. Elephants are a particularly interesting species for the study of cancer, by virtue of their large number of cell divisions and long lives yet low incidence of cancer. Elephants also possess multiple retrogene copies of TP53, which have previously been shown to induce strong cellular responses to DNA damage. However, most previous studies have largely focused only on African elephant TP53 retrogenes and often in non-native backgrounds. Here we generated CRISPR-Cas9 knockouts of TP53, all 29 TP53 retrogenes, or both in combination in Asian elephant fibroblasts. We find that while there is considerable overlap in the DNA damage responses of the TP53 and retrogene knockouts, there are also many unique pathways enriched in both. In particular, the retrogene knockouts exhibit strong enrichment of many extracellular pathways suggesting they may play a large role in the tumor microenvironment and mitigating metastatic growth. We also find that only a small fraction of these 29 retrogenes appear to be expressed across a variety of tissues and identify three loci that are likely driving this response. This work shows for the first time the transcriptomic effect of these retrogenes within their native background and establishes a foundation for future research into the relative contributions of these genes.

Keywords: Asian elephants; TP53; retrogenes; cancer; CRISPR-Cas9; RNA-seq

1. Introduction

The gene TP53 encodes the transcription factor p53 and acts as one of the central master regulators in response to various cell stresses, including DNA damage, hypoxia, and nutrient deprivation [1,2]. In response to stress, p53 degradation is inhibited, its tetramers are stabilized, and they migrate to the nucleus resulting in the upregulations of many downstream signaling cascades, including those related to cell cycle arrest and apoptosis [1,3]. Consequently, this gene is often implicated in cancer development, and is found mutated in most human cancers, undermining genomic integrity [1,2,4]. Likewise, individuals diagnosed with Li-Fraumeni Syndrome have germline mutations in TP53 and much greater risk and earlier onset of various cancers [2,3]. In addition, TP53 also plays a role in cellular senescence and reprogramming efficiency, making it important in fundamental cellular biology technologies, such as the generation of iPSCs and immortalization [1,2,5].

Peto's paradox is the observation that cancer incidence in a species does not always scale with body mass and/or lifespan [6,7]. Elephants have emerged as a key species in attempts to resolve this paradox due to their large size, long lifespan, and very low cancer risk [7,8]. Genomic analyses have identified the duplication of many tumor suppressor genes across extant elephants, including multiple retrogene copies of TP53 [7,8]. Although there is still considerable uncertainty about how or which of these TP53 retrogenes (RTGs) may confer tumor resistance, previous works have shown that the encoded protein of at least one of these retrogenes, African elephant RTG 9, induces strong cellular responses to DNA damage [7–9]. However, most prior studies have largely focused on their ability to induce apoptosis and have often examined the effect of retrogenes in different genetic backgrounds (i.e. human or mouse cells) [7–9]. Additionally, almost all functional analyses have focused on African elephant retrogenes despite the identification of TP53 retrogene expansion in other extinct and extant proboscideans [7].

Here, we establish a baseline for similar work in Asian elephants. We designed CRISPR-Cas9 guides to disrupt TP53, all TP53 retrogenes, or both in combination in Asian elephant fibroblasts. We then examined the transcriptional profiles of these knockouts with and without DNA damage stress, to resolve the protective roles TP53 retrogenes might play at a cellular level. Additionally, we compare our results with the transcriptomic patterns of these retrogenes across various Asian elephant tissues to identify which may be acting as tumor suppressors and which are likely nonfunctional. The work we present here suggests that these retrogenes may have a much larger and more varied response than previously anticipated and lays the groundwork for future studies focusing on specific retrogene targets.

3. Methods

61 *3.1 – TP53 identification and primer design*

62 We extracted and aligned the canonical TP53 locus as well as all 29 retrogenes (RTGs) from the
63 *Elephas maximus* reference genome (GCF_024166365.1). For the RTGs, we trimmed the alignment to
64 include only the ~1.1kb segment overlapping the exons in TP53. We then designed two primers to
65 specifically amplify the TP53 gene (Supplementary Table 1). For the RTGs, we designed seven forward
66 and four reverse primer variations, all amplifying the same locus, but containing all possible binding site
67 mutations to minimize amplification bias (Supplementary Table 1). These primers were made up to
68 10µM stock solutions and equal volumes of all forward and all reverse primers combined to produce
69 master forward and reverse stock solutions.

70

71 *3.2 – Guide Design*

72 We generated all possible guide RNAs (gRNAs) within TP53 from the start of first exon to the
73 end of the last exon using the generate_kmery.py script distributed as part of GuideScan2 [10]. All guides
74 were then screened for predicted off-targets against the entire *E. maximus* reference genome using
75 GuideScan2 with a maximum mismatch distance of 3 (-m 3), and with the option to output all guides (-t -
76 1). The output was then further filtered using a custom python script (Cas9_MMFiltering_V1_1.py) and
77 scoring matrix to filter potential gRNAs based on how many mismatches they possessed relative to
78 potential off-target sites and where along the spacer those mismatches occur. From the remaining
79 guides we selected two guides (Supplementary Table 1) that cut within exon 1 and between exon 2 and
80 3 respectively to generate TP53 knockouts as these disrupt the TAD1 domain and should interfere with
81 proper splicing resulting in a premature stop codon.

82 To design guides for the RTG knockouts, we generated all possible guides against the trimmed
83 RTG alignment for all 29 retrogenes using generate_kmery.py script as before. We then screened all
84 guides for those that were predicted to have activity to the most possible targets and identified two
85 gRNAs that were each predicted to have activity against all but three retrogenes, which were only
86 targetable by either one of the two guides (Supplementary Table 1; Fig. 1). We then subsequently chose
87 one additional guide that targeted those three remaining retrogenes to ensure that each retrogene had
88 at least two gRNAs.

89

90 *3.3 – Cell culturing and media*

91 Fibroblast cells were grown in a custom media formulation with the final concentrations: 50%
92 FGM-2 (Lonza FBM Basal Media (CC-3131) supplemented with FGM-2 SingleQuots supplements (CC-

93 4126)); 20% heat-inactivated fetal bovine serum (ThermoFisher, A5670801); and 30% Mem Alpha
94 (Gibco, 12571063). Cells were grown in an incubator at 37°C with 5% CO_2.
95
96 *3.4 – Cas9 RNP Electroporation*
97 All gRNAs used for both the TP53 and RTG knockouts were ordered from Synthego and
98 resuspended to 100µM in 1X TE buffer. Cas9 nuclease was ordered from IDT (catalog # 10007807). All
99 nucleofections were conducted using the Lonza P4 Primary Cell 96-well Kit (catalog # V4SP-4096) using
100 code CA-137. Approximately 400-500k cells were used for each transfection in 20µl reactions with 3µM
101 final concentrations of each guide and 1µM of Cas9 nuclease per guide (i.e. a 3:1 ratio of gRNA:Cas9).
102 RNPs were pre-complexed in nucleofection solution for 20 minutes at room temperature. Cells were
103 transfected with either both TP53 gRNAs, all three RTG gRNAs, both TP53 and RTG gRNAs ("All"
104 condition), or a non-targeting gRNA which has no significant homology to the *E. maximus* genome
105 (Supplementary Table 1). For the All nucleofection, gRNAs were first combined and then concentrated
106 using a Savant SpeedVac DNA 120 Concentrator (Thermo Scientific), before being resuspended in
107 premade nucleofector and supplement solution. Following nucleofection cells were immediately
108 transferred into 6-well plates containing prewarmed media and allowed to recover for 4 days. Media
109 was changed 24 hours post nucleofection and ever 2-3 days thereafter. Due to the presence of possible
110 TP53 mutations that accumulate with prolonging passaging, a second separate nucleofection was
111 conducted as above prior to the Mitomycin C stressed RNA-seq. In both cases, all treatment conditions
112 were conducted using three biological replicates.
113
114 *3.5 – Editing Efficiency*
115 During the first passage (approximately 4 days post nucleofection), about 10% of the cells were
116 collected and DNA extracted using the DNeasy Blood and Tissue Kit (Qiagen) using the manufacturer's
117 recommended protocol. From every extraction we amplified both the TP53 locus spanning the expected
118 knockout deletion, and the retrogene loci. 25µL PCR reactions were run with final concentrations of 1X
119 Phusion Hot Start Flex 2X Master Mix and 500nM of each primer, using between 1-3µL of template.
120 Cycling conditions were: 98°C for 3 minutes; 12 cycles of 98°C for 15 seconds, then 69°C for 60 seconds;
121 25 cycles of 98°C for 15 seconds, 67°C for 30 seconds, then 72°C for 45 seconds; followed by a final
122 extension at 72°C for 5 minutes. Resulting PCR amplicons were purified over MinElute PCR Purification
123 columns (Qiagen) using the recommended protocol and visualized on E-gel EX 2% Agarose gels
124 (Invitrogen) to confirm the presence or absence of any large deletions. We examined the relative

125 fluorescence of our products using the Image Lab 6.1 (Bio-Rad). We then calculated a minimum editing
126 efficiency (the relative fluorescence of our expected deleted product, relative to everything else in the
127 lane) and a maximum editing efficiency (the relative fluorescence of all non-wild-type bands within each
128 lane).
129 To confirm efficiency estimates and examine RTG deletion distribution, we sent each of our
130 amplicon pools for nanopore sequencing using the Premium PCR service at Primordium to get full length
131 sequences of our DNA fragments. For each sample, raw reads were first filtered using a custom python
132 script (FastqFilter.py) to between the maximum and minimum expected sizes +/- 10% respectively.
133 Mapping was carried out using BBMap [11] using default parameters against either the TP53 amplicon, an
134 RTG consensus amplicon, or a multi-fasta with 26 RTGs (RTG amplicons with identical sequences outside
135 of the editing region collapsed).
136 To calculate editing efficiency, we used a custom python script (CigarDel.py) which uses a region
137 file containing the deletion regions of interest, a minimum deletion proportion (i.e. the proportion of
138 bases mapping to the deleted region that need to be missing), and the CIGAR string of each read to
139 evaluate the presence or absence of various deletions in the data. For TP53 we analyzed the presence of
140 the expected deletion if Cas9 induced a double-strand break at both sgRNA targeting sites. For the RTG
141 we examined the presence of any of the three possible deletions between two guides. All analyses were
142 carried out with a minimum deletion proportion of 0.9.
143 To examine the distribution of deletions at each locus we identified the names of reads
144 containing deletions and extracted those that mapped to each locus. We then cross-referenced this with
145 the deletions identified against the RTG consensus to obtain read distributions for reads mapped to all
146 RTGs or the collapsed RTG loci.
147
148 *3.6 – Growth Curve*
149 Approximately four days post nucleofection, we seeded 50k cells from each sample into the
150 wells of four 24-well plate. 36 hours after seeding and every 24 hours thereafter, cells were detached
151 using TrypLE Express (Gibco) and cell counts,and viability estimates were obtained using a Countess II
152 cell counter (Thermo Fisher). For each data point we averaged two individual counts. These values were
153 then imported in GraphPad Prism 10 and analyzed using a two-way ANOVA for significance.
154
155 *3.7 – RT-qPCR*

156 50k cells were seeded into each well of four 48 well plates. Approximately 36 hours after
157 seeding, growth media was exchanged for one supplemented with either 0µM, 1µM, 10µM, or 50µM
158 Mitomycin C (MOLTOX). Cells were incubated for 18 hours in the presence of Mitomycin C-containing
159 media, after which RNA was immediately extracted using the RNeasy Micro kit (Qiagen) according to the
160 manufacturer's recommendations. We designed and validated four pairs of RT-qPCR primers for *E.*
161 *maximus* GAPDH, ATR, p21, and MDM2 (Supplementary Table 1). 20µL reactions were run with final
162 concentrations of 1X TaqPath 1-Step Multiplex Master Mix (No ROX), and 500nM of each primer, using
163 2µL of purified RNA template. Cycling conditions were: 25°C for 2 minutes; 53°C for 10 minutes
164 seconds; 95°C for 2 minutes; 40 cycles of 95°C for 15 seconds, 60°C for 30 seconds; followed by a final
165 extension at 60°C for 1 minute, and a melt curve between 60°C and 95°C. All samples for each treatment
166 condition were analyzed in duplicate, and relative expression standardized to GAPDH, using the $2^{-\Delta\Delta CT}$
167 method [12]. Data was combined across biological replicates within each treatment. Average results were
168 imported into GraphPad Prism 10 and analyzed for significance using a two-way ANOVA.
169
170 *3.8 – Bulk RNA-seq*
171 Approximately 300-500k cells were collected, flash frozen on dry ice, and shipped to Genewiz
172 for bulk RNA sequencing within ten days of transfection. Expression of transcripts was done using
173 Salmon (v1.8.0 [13]) using a custom index consisting of all mRNAs and non-coding RNAs annotated within
174 the *E. maximus* reference genome (GCF_024166365.1), with all 29 retrogene sequences added. The
175 salmon index was generated using the keepDuplicates flag to avoid collapsing any RTGs. Salmon quant
176 files were analyzed using DESeq2 [14] as part of a custom Rscript to analyze samples in a pairwise fashion
177 (i.e. one set of biological replicates against the controls at a time). Differentially expressed genes (DEGs)
178 were identified with an adjusted p value of less than 0.05 (padj < 0.05). No minimum log2 fold change
179 was applied for initial analysis of TP53 pathway genes. A minimum absolute log2 fold change of 1 was
180 used when analyzing DEG overlap between different conditions. Venn diagrams and DEG overlap was
181 generated using a web based custom venn diagram tool
182 (https://bioinformatics.psb.ugent.be/webtools/Venn/), and pathway enrichment preformed using
183 GeneCodis4 [15].
184
185 *3.9 – Retrogene expression in public datasets*
186 To ensure the expression patterns we observed were not tissue/cell type specific, we examined
187 the expression of our retrogenes in some preliminary shared data from mesenchymal stem cells and

188 epithelial cells. These cells contained deletions generated using the same guides as in this study, but the
189 epithelial cells were also subject to various small molecule treatments. We also downloaded public RNA-
190 seq datasets for Asian elephant ovary (SRX15554240), salivary glands (SRX15554239), lymph nodes
191 (SRX15554238), thyroid (SRX15554237), lung (SRX15554236), and lymphocytes (SRX1423033), and
192 analyzed it as above.
193
194 *3.10 – Depth plots*
195 To examine where reads from each of the three RTGs showing repeated expression were
196 originating, we mapped the data for our datasets and the public datasets back to the Asian elephant
197 reference genome using STAR [16]. We then extracted read depths along the full annotated length of RTGs
198 LOC126068247, LOC126068267, and LOC126068248 using samtools depth [17]. For our dataset we
199 averaged expression values within biological replicates and plotted the data in Microsoft Excel.
200
201 *3.11 – Maximum likelihood phylogeny*
202 We download African elephant retrogene 9 (KF715863.1) to investigate a possible relationship
203 between this retrogenes and the Asian elephant retrogenes. We then aligned the retrogene to the Asian
204 elephant TP53 CDS and truncated the RTG 9 sequence to contain only the homologous sequence to
205 TP53. We then aligned the truncated RTG 9 sequence, Asian elephant TP53, and all 29 Asian elephant
206 retrogenes using muscle (v5.1; Edgar, 2004). We generated a maximum likelihood phylogeny using IQ-
207 Tree [19], using ModelFinder to identify the best supported substitution model [20], and 100 bootstraps (-b
208 100).
209
210 *3.12 – EMS mutagenesis and clonal expansion*
211 Biological replicates were combined and 50k cells per condition were seeded into 6-well plates.
212 Cell were grown in media containing 0 µg/mL, 0.5 µg/mL, or 1 µg/mL of EMS (Thermo Scientific) for 16
213 hours, at which point EMS-containing media was replaced with fresh media [21]. Cells were then grown for
214 an additional 8 passages, to potentially outgrow any fast-growing and immortalized mutants. Following
215 8 passages, 1000 cells were seeded onto a 10cm^2 dish and observed over a period of 1-3 weeks. Colonies
216 were marked and labelled to observe growth. A subset of well-defined circular colonies were removed
217 by treating with 5µl of TrypLE Express (Gibco) twice, and transferring the detached cells to a 6-well plate
218 and then to a 10cm^2 dish.
219

220 **4. Results**

221 *4.1 – The Asian elephant genome contains 29 TP53 retrogenes*

222 We searched the annotations of the Asian elephant reference genome and identified 29
223 retrogene copies of TP53 (RTGs). 27 of these RTGs are found on chromosome 27, while the remaining
224 two are found on chromosomes 24 and 26. By comparison, the canonical copy of TP53 is located on
225 chromosome 19. Most of the RTGs are approximately 1.1 kb in length, roughly the same length as the
226 CDS of TP53 as would be expected of retrogenes, with the exception of three (LOC126068247 – 64.1 kb;
227 LOC126068267 – 32.5 kb; LOC126068248 – 4.9 kb; Fig. S1A).

228 When aligned and translated in the same frame as TP53, each RTG carries a premature stop
229 codon, which primarily cluster in five main areas across the homology region (Fig. 1A). We then
230 compared the percent identity of the DNA and amino acid sequences of the retrogenes to TP53. We find
231 significantly higher levels of homology at the DNA level than at the amino acid level, irrespective of
232 whether we compare the entire amino acid sequence, or only the truncated peptides (Fig. S1B).
233 Comparatively, across the same homologous region, the RTGs contain significantly higher levels of
234 identity (Fig. S1C).

235

236 *4.2 – TP53 and RTG editing and validation*

237 We attempted to design paired Cas9 guide RNAs (sgRNAs) to produce large, disruptive deletions
238 in TP53 and all the retrogenes. We managed to identify two sgRNAs that specifically target TP53, which
239 result in the partial loss of exon 1, the entirety of exon 2, and a premature stop codon after read
240 through into intron 2 (Fig. 1A). We could not identify just two guides that were both predicted to have
241 activity across all 29 retrogenes, so we designed a third guide to ensure each RTG was targetable by at
242 least two sgRNAs (Fig. 1B; Fig. S2).

243 We electroporated fibroblast cells with Cas9 RNPs consisting of either the two TP53 guides, the
244 three RTG guides, all five guides together (All), or a guide predicted to have no activity in the Asian
245 elephant genome (Non Targ.). After approximately four days, we then collected cells, extracted DNA,
246 and amplified a portion of TP53 and RTG genes flanking the expected editing sites. We observed high
247 levels of editing across all three conditions when we examined the resulting amplicons on an agarose gel
248 (Fig. S3). However, due to our PCR reaction amplifying all 29 loci simultaneously, we couldn't assess
249 which and how many of the RTGs were being edited.

250 To address this issue, we sent both PCR reactions for end-to-end nanopore sequencing. We
251 quantified the number of observed deletions, using a custom script which marks the presence of a

252 deletion if 90% of the bases within a defined region were absent. We note that this is likely an
253 underestimate of the true editing efficiency, as this will only identify the perfect deletion product, and
254 miss any indels and/or incomplete loss of the flanked segment of DNA. We observed high editing
255 efficiency in all conditions using either the TP53 guides (TP53 KO – 67%; All KO – 66%) or the RTG guides
256 (RTG KO – 77%; All KO – 82%), and almost no editing in the Non Targ. condition (<0.1% at both loci) (Fig.
257 1C). We further examined the proportion of each deletion product observed in the RTG PCRs, and noted
258 a clear pattern, wherein deletions between guides 2 and 3 were the most frequent, in line with
259 expectations about the targeting scope of those two guides (Fig. 1D).
260 Next, we wanted to disentangle the editing on a per-locus level for the RTG data. We mapped
261 the data to each unique RTG (collapsing those with identical sequences outside the editing region) and
262 identified the proportion of each deletion at each locus. Although we generally observe a mix of all
263 three deletions at each locus (some of which is likely due to cross mapping between the RTGs), we
264 generally observe deletions in line with our expectations based on their targeting scope (Fig. 1E). Only
265 two loci buck this trend, LOC126068338 and LOC126068461, both of which have a single point mutation
266 at position 10 within the seed region of the guide (Fig. S2). Due to the importance of this region for
267 guide specificity, we filtered these targets out; however, we note that at least at these two loci the
268 guides appear to be permissive for editing. When combined with the per-locus reads that were not
269 identified as containing a deletion (Fig. 1F), we observe an average of 76.8% editing efficiency across all
270 unique loci.
271
272 *4.3 – All editing conditions show similar growth and transcriptomic patterns*
273 Approximately four days following transfection, we seeded 50k cells from each condition into
274 four 24-well plates and observed the growth and viability of the cells over approximately 4.5 days. We
275 observed no significant difference in the growth (Fig. 2A) or viability (Fig. 2B) of the cells across this time.
276 However, we do note that the TP53 KO wells appeared to reach confluency a day before the other
277 conditions suggesting faster growth, which is consistent with our experiences handling these cells in
278 subsequent passages.
279 Next, we wanted to see if we could observe a functional effect of the knockouts we generated.
280 We cultured 50k cells in the presence of increasing concentrations of Mitomycin C (MMC), a DNA
281 damaging agent [7,22], and examined the expression of p21 (CDKN1A) and MDM2, two direct downstream
282 targets of p53, and an upstream regulator, ATR (Fig. 2C). We observed significant decreases in
283 expression of MDM2 and p21 at all non-zero Mitomycin C concentrations across each condition, and

284 almost no significant differences in ATR. However, we were surprised to see that the response in the
285 RTG KO was so similar to that of the TP53 and All KOs. We extracted DNA from the same line of cells one
286 passage later and PCR amplified the TP53 and RTG loci again. We observed a drop in Sanger sequence
287 quality across the TP53 amplicons in some of the biological replicates (Fig. S4), generally consistent with
288 the location of the RTG guides. This suggests there may be some off-target editing in the RTG samples
289 that becomes enriched during routine passaging of the RTG KO cells, particularly if these TP53 mutants
290 also possess a potential growth advantage as would be suggested by our growth curve.
291
292 *4.4 – RNA-seq analysis reveals overlapping and unique TP53 and RTG responses*
293 To minimize the effect of any off-target mutations on our RNA-seq results, we sent cells soon
294 after nucleofection (within ~10 days). Despite this we observe the presence of off-target edits at the
295 homologous location of the RTG guides within the TP53 transcripts (Fig. S5). Interestingly, we see
296 significant enrichment of these TP53 mutations in our MMC-treated cells. This suggests that these TP53
297 mutations might confer a significant growth advantage to the cells as was also suggested by our growth
298 curve analysis.
299 Given most literature on elephant RTGs has largely focused on how these retrogenes may
300 interact with TP53 and MDM2, we first decided to look at changes in expression of all direct
301 downstream targets of TP53 (Fig. 3). We find that in the absence of stress (i.e. 10µM Mitomycin C), we
302 see relatively small changes in expression within most downstream targets with the exception of MDM2
303 and CDKN1A in both the TP53 and All KOs. Interestingly, we note that TP53 is not identified as
304 significantly downregulated in the TP53 KO, suggesting maybe that nonsense-mediated decay is less
305 pronounced in Asian elephants. In comparison, we observe very few differentially expressed genes
306 (DEGs) in the RTG KO, and only TP53 that has a log2 fold change greater than 1. This suggests that the
307 RTGs may play some role in either stabilizing or increasing the abundance of p53, although likely not
308 directly by acting as a transcription factor as none of the RTGs have an intact DNA-binding domain (Fig.
309 1).
310 To confirm the results above were genuine and not due to mismapping between the RTGs and
311 the canonical TP53 gene, we generated a synthetic dataset, consisting of 1,000 2x150 bp reads from
312 TP53 and each of the RTG loci. We then used this dataset to compare the amount of reads assigned to
313 each locus when mapped back to the genome using BBMap [11], BWA-MEM [23], or STAR [16], followed by
314 quantification with Salmon [13], or via direct quantification by Salmon (Fig. S6). We find that all four
315 methods gave similar results, with BBMap performing the worst, possibly due to how it assigns multi-

316 mapped reads to the first locus in the reference. In light of these results, we conclude that the decrease
317 in TP53 expression is likely not a bioinformatics artifact due to poor mapping.
318 In the presence of Mitomycin C, all three conditions have similar patterns in DEGs (Fig. 3),
319 although typically less pronounced in the RTG KO. As before, we also observe that TP53 is significantly
320 less expressed in the RTG KO and not in the TP53 KO. Additionally, IGFBP-3, a gene which can contribute
321 to cell growth and apoptosis [24,25], is also downregulated in the RTG KO. We also observe significant
322 downregulation of SFN across all three conditions in the MMC-treated cells, in line with this gene's
323 known tumor suppressor function of arresting cellular replication in response to DNA damage [26,27].
324 Next, we wanted to examine the different pathways enriched within each condition. Due to the
325 large number of genes with small changes we filtered the dataset to a minimum absolute log2 fold-
326 change of at least 1. We then extracted the DEG lists for each condition and analyzed the overlap
327 between them in either the non-stressed or MMC-treated cells (Fig. 4A). We performed BioPlanet
328 pathway enrichment[28] on the 763 DEGs common to all three conditions treated with Mitomycin C. We
329 observed significant enrichment in p53 signaling pathways, as would be expected of cells exposed to a
330 DNA damaging agent, as well as enrichment in many DNA replication and cell cycle pathways (Fig. 4B).
331 We also examined the unique DEGs in the TP53 and RTG KO cells with Mitomycin C treatment
332 (Fig. 4C). For the TP53 KO, we see enrichment in many pathways that would be expected given the
333 known checkpoint and cell stress responses coordinated by TP53 (e.g. cell cycle and proliferation, DNA
334 repair, etc; Fig. 4D) [29–31]. By comparison, unique RTG KO DEGs are enriched in extracellular organization
335 and signaling/interaction pathways (Fig. 4E), suggesting that RTGs may play a unique protective role in
336 regulating the tumor microenvironment and limiting metastasis [32]. Additionally, the largest number of
337 DEGs appear to belong to the apoptotic signaling in response to DNA damage pathway, a function that
338 has previously been suggested for African elephant RTGs [7,9].
339
340 *4.5 – Very few RTGs are likely to be expressed or functional*
341 Given the unique function suggested by the RNA-seq of the RTG KO cells, we next wanted to see
342 if we could investigate which RTGs might be responsible for the changes we observed. In none of the
343 conditions, irrespective of treatment, did we detect differential expression of any of the RTGs.
344 Furthermore, we generally do not observe differences in the amount of RTGs expressed in response to
345 Mitomycin C treatment (Fig. 5A), with the exception of the All KO. We do observe an increase in the
346 amount of RTGs expressed in the RTG and All KO conditions, irrespective of Mitomycin C treatment
347 (significance values not shown), although this may be a lingering transcriptional effect caused by double-

348 strand break repair [33]. However, we do note that all but three RTGs have very low read counts
349 (Supplementary Table 2).
350 We next wanted to confirm that this pattern in RTG expression is not due to cell type or tissue
351 specific. We download public RNA-seq datasets for Asian elephant ovary, salivary glands, lymph nodes,
352 lung, thyroid, and lymphocytes, as well as examined some preliminary data generated with the same
353 guides in epithelial and mesenchymal stem cells, and examined RTG expression patterns in these
354 datasets. We observe the same general pattern, where the only three RTGs which show consistent
355 expression are LOC126068247, LOC126068267, and LOC126068248 (Supplementary Table 3). However,
356 these three RTGs are the three largest RTG genes (LOC126068247 – 64.1 kb; LOC126068267 – 32.5 kb;
357 LOC126068248 – 4.9 kb) based on their annotations.
358 To examine what was driving the expression of these RTGs, we extracted RNA-seq mapped
359 reads across each of these three loci and examined the frequency of reads across each locus. For two of
360 these RTGs, we observe that most of the reads assigned to these RTGs map thousands of base pairs
361 upstream of the region homologous to the TP53 CDS. In LOC126068247 and LOC126068248, we observe
362 almost no mapped reads within the TP53 homology region either in our data or the publicly available
363 datasets (Figs S7-S10). By comparison, LOC126068267 also has peaks of high coverage upstream of the
364 TP53 homology region but shows similar levels of coverage across the homology region in both datasets
365 (Figs S11-S12).
366 Given the lack of expression of most of the annotated retrogenes, we decided to compare those
367 that might be functional in Asian elephants with African elephant retrogene 9 (RTG 9; KF715863), on
368 which most previous studies have focused [7–9]. We aligned the TP53 homologous regions of all Asian
369 elephant RTGs, African elephant RTG 9, and the TP53 CDS, and built a maximum-likelihood phylogeny.
370 We observe that RTG 9 is most closely related to a clade containing both two of our putative expressed
371 RTGs, LOC126068267 and LOC126068248 (Fig. 5B), further suggesting that these two RTGs might be
372 responsible for the changes we observe. However, we do note that due to the high similarity of many of
373 these RTGs, the bootstrap support values within our phylogeny are quite weak.
374
375 *4.6 – RTG KO allows for partial clonal expansion*
376 Although compounded by a relative lack of studies and optimization relative to mouse and
377 human cell lines, elephant cells have proven difficult to immortalize and grow [34,35]. Our own efforts to
378 produce clonal lines of these cells using established protocols [36] have failed (data not shown). However,

379 we reasoned that since TP53 mutations often precede cancer development, and many cancers naturally
380 immortalize, we may be able to exploit this feature in our knockouts to produce isogenic lines.
381 To examine this, we treated cells with either 0µg/mL, 0.5µg/mL, or 1µg/mL EMS for 16 hours [21],
382 and then performed eight passages, seeding 10% of the detached cells, to outgrow fast growing and
383 potentially immortalized cells. All 1µg/mL cells died within the first few passages, as did all the TP53 KO
384 conditions and the Non Targ. 0µg/mL EMS treatment. After eight passages, the outgrown cells were
385 collected and 1,000 from each condition were seeded into a 10cm^2 dish, and colonies picked over 1-3
386 weeks. We observed between 4 and 26 well-defined circular colonies in the RTG and All KO conditions
387 treated with 0µg/mL or 0.5µg/mL, in comparison to only one in the Non Targ. 0.5µg/mL condition
388 (Supplementary Table 4). Of these we picked three colonies each from the All 0 µg/mL, All 0.5µg/mL,
389 and RTG 0 µg/mL plates, two colonies from the RTG 0.5µg/mL, and the only colony from the Non Targ.
390 0.5µg/mL plate. We attempted to expand colonies, but unfortunately most of the cells in most of these
391 lines appeared quite sick and exhibited very slow growth. One of the RTG 0.5µg/mL colonies managed to
392 propagate an additional four passages within the 10cm^2 dishes, but slowed down immensely thereafter
393 and became senescent.
394
395 **5. Discussion**
396 TP53 is tremendously important for the cellular response to stress and DNA damage, and has
397 been implicated in many biological important processes including cellular reprogramming and
398 immortalization [1,5,37,38]. Additional retrogene copies found in many proboscideans have been implicated
399 in this taxa's low cancer incidence rate [7,8,39,40], and likely also contribute to the difficulty in working with
400 and modifying these cells. With the publication of a new high quality Asian elephant reference genome
401 (as part of the VGP), we wanted to examine the high-level function of these retrogenes, both
402 independently and in combination with TP53 mutations.
403 We find that, similar to African elephants, Asian elephant retrogenes all contain premature stop
404 codons. Additionally, we note the putative encoded proteins are quite dissimilar, suggesting possibly
405 relaxed selection at these loci and perhaps non-functionality [41]. Despite this, we note that at least three
406 of these RTGs appear to be stably expressed in various cell types and tissues. Likewise, we observe clear
407 transcriptional differences between TP53 and RTG KO cells, including a significant downregulation of
408 TP53 expression in our RTG KO cells, suggesting a novel function for elephant retrogenes. Additionally,
409 we observe a differential enrichment of pathways when disrupting RTGs versus TP53, indicating that
410 while there is a high amount of overlap between these genes, they also regulate many unique pathways.

411 In particular, genes enriched within many extracellular organization and signaling pathways were
412 observed in the RTG KO population. Some of these (e.g. extracellular matrix organization, angiogenesis)
413 have been implicated as contributing to the establishment of tumor microenvironments, and are gaining
414 increased importance as vital to cancer growth and spread [32]. This may also explain the low prevalence
415 of malignant cancers in Asian elephants [39]. We also show that while these retrogenes are likely not
416 sufficient for clonal expansion, their knockout appears to permit the generation of single cell colonies.
417 This is congruent with current work which suggests that the knockdown of these retrogenes is also
418 important during the production of elephant iPSCs [42].
419 However, we do note that most of the annotated retrogenes with the Asian elephant genome
420 are not expressed across an extensive list of tissues, and only a handful may actually be functional.
421 Although this may be in line with previous estimates as to the amount of tumor suppressor evolution
422 necessary for the observed incidence rate in elephants [41]. Interestingly, two of the retrogenes we
423 observed most stably expressed across in different tissues appear to be closely related to African
424 elephant RTG 9, which has been shown to induce strong cellular responses to DNA damage [7–9] and
425 apoptosis in mouse fibroblasts and human cancer cell lines [7,9]. Future research should focus on
426 individual retrogenes, especially LOC126068267, to determine if these three are responsible for the
427 signal we observe, and possibly what the function of these highly expressed regions upstream of the
428 TP53 homologous sequences may be. Additionally, by extending the analysis of proboscidean RTGs into
429 Asian elephants we lay the groundwork for more comparative studies, to examine how much each
430 response is evolutionary conserved versus uniquely derived along each species' lineage.
431
432 **6. Author contributions**
433 EK and GMC conceived the study with feedback from EM and AAT. EK and NB designed and
434 conducted all wet-lab experiments with feedback and help from EM and AAT. EK completed all
435 bioinformatic analyses. AAT provided additional RNA-seq data for epithelial cell lines. EK wrote the first
436 draft of the manuscript. All authors helped in the revision of subsequent drafts.
437
438 **7. Data availability**
439 Bulk RNA-seq data for all conditions with and without mitomycin was deposited on NCBI
440 (BioProject Accession: PRJNA1147676). Bioinformatic scripts and tools created for this work are
441 available at: https://github.com/EmilKarpinski/RTG_TP53_Analysis.
442

443 **8. Declaration of competing interests**

444 For a complete list of G.M.C.'s financial interests, please visit

445 http://arep.med.harvard.edu/gmc/tech.html. All other authors declare no other conflicts of interest.

446

447 **9. Acknowledgments**

448 We would like to thank Li Li and Zhengkuan Tang for discussions on RNA-seq experimental setup

449 and analysis. We would also like to thank George Chao for guidance on nanopore analysis, and Sergiy

450 Velychko for suggestions and troubleshooting of elephant fibroblast cell culture. Lastly, we would like to

451 thank Roman Teo Oliynyk, Chun-Ting Wu, and Ahmed Mahas, as well as the rest of the Church lab for

452 helpful discussions throughout this project.

453 EK and AAT were supported by a structured research agreement between Harvard Medical

454 School and Colossal Biosciences.

455

456 **References**

457 1. Wang, H., Guo, M., Wei, H. & Chen, Y. Targeting p53 pathways: mechanisms, structures, and
458 advances in therapy. *Signal Transduct Target Ther* **8**, 92 (2023).

459 2. Hainaut, P. & Hollstein, M. p53 and Human Cancer: The First Ten Thousand Mutations. in 81–137
460 (1999). doi:10.1016/S0065-230X(08)60785-X.

461 3. Correa, H. Li–Fraumeni Syndrome. *J Pediatr Genet* **05**, 084–088 (2016).

462 4. Hainaut, P. & Pfeifer, G. P. Somatic TP53 Mutations in the Era of Genome Sequencing. *Cold Spring*
463 *Harb Perspect Med* **6**, a026179 (2016).

464 5. Rasmussen, M. A. *et al.* Transient p53 Suppression Increases Reprogramming of Human
465 Fibroblasts without Affecting Apoptosis and DNA Damage. *Stem Cell Reports* **3**, 404–413 (2014).

466 6. Peto, R., Roe, F. J., Lee, P. N., Levy, L. & Clack, J. Cancer and ageing in mice and men. *Br J Cancer*
467 **32**, 411–426 (1975).

468 7. Sulak, M. *et al.* TP53 copy number expansion is associated with the evolution of increased body
469 size and an enhanced DNA damage response in elephants. *Elife* **5**, (2016).

470 8. Abegglen, L. M. *et al.* Potential Mechanisms for Cancer Resistance in Elephants and Comparative
471 Cellular Response to DNA Damage in Humans. *JAMA* **314**, 1850 (2015).

472 9. Preston, A. J. *et al.* Elephant TP53-RETROGENE 9 induces transcription-independent apoptosis at
473 the mitochondria. *Cell Death Discov* **9**, 66 (2023).

474 10. Schmidt, H. *et al.* Genome-wide CRISPR guide RNA design and specificity analysis with
475 GuideScan2. (2022) doi:10.1101/2022.05.02.490368.

476 11. Bushnell, B. BBMap. http://sourceforge.net/projects/bbmap/ (2022).

477 12. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time
478 Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**, 402–408 (2001).

479 13. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-
480 aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).

481 14. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-
482 seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

483 15. Garcia-Moreno, A. *et al.* Functional Enrichment Analysis of Regulatory Elements. *Biomedicines* **10**,
484 590 (2022).

485 16. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

486 17. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
487 (2009).

488 18. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
489 *Nucleic Acids Res* **32**, 1792–1797 (2004).

490 19. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective
491 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* **32**, 268–274
492 (2015).

493 20. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder:
494 fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–589 (2017).

495 21. Munroe, R. & Schimenti, J. Mutagenesis of Mouse Embryonic Stem Cells with
496 Ethylmethanesulfonate. in 131–138 (2009). doi:10.1007/978-1-59745-471-1_7.

497 22. Lee, Y.-J., Park, S.-J., Ciccone, S. L. M., Kim, C.-R. & Lee, S.-H. An in vivo analysis of MMC-induced
498 DNA damage and its repair. *Carcinogenesis* **27**, 446–453 (2006).

499 23. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform.
500 *Bioinformatics* **25**, 1754–1760 (2009).

501 24. Marzec, K. A., Lin, M. Z., Martin, J. L. & Baxter, R. C. Involvement of p53 in insulin-like growth
502 factor binding protein-3 regulation in the breast cancer cell response to DNA damage.
503 *Oncotarget* **6**, 26583–26598 (2015).

504 25. Grimberg, A. *et al.* p53-Dependent and p53-Independent Induction of Insulin-Like Growth Factor
505 Binding Protein-3 by Deoxyribonucleic Acid Damage and Hypoxia. *J Clin Endocrinol Metab* **90**,
506 3568–3574 (2005).

507 26. Hermeking, H. The 14-3-3 cancer connection. *Nat Rev Cancer* **3**, 931–943 (2003).

508 27. Hermeking, H. *et al.* 14-3-3σ Is a p53-Regulated Inhibitor of G2/M Progression. *Mol Cell* **1**, 3–11
509 (1997).

510 28. Huang, R. *et al.* The NCATS BioPlanet – An Integrated Platform for Exploring the Universe of
511 Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics. *Front*
512 *Pharmacol* **10**, (2019).

513 29. Dey, G. *et al.* Signaling network of Oncostatin M pathway. *J Cell Commun Signal* **7**, 103–108
514 (2013).

515 30. Moldovan, G.-L. & D'Andrea, A. D. How the Fanconi Anemia Pathway Guards the Genome. *Annu*
516 *Rev Genet* **43**, 223–249 (2009).

517 31. Du, W. & Searle, J. The Rb Pathway and Cancer Therapeutics. *Curr Drug Targets* **10**, 581–589
518 (2009).

519 32. de Visser, K. E. & Joyce, J. A. The evolving tumor microenvironment: From cancer initiation to
520 metastatic outgrowth. *Cancer Cell* **41**, 374–403 (2023).

521 33. Guha, S. & Bhaumik, S. R. Transcription-coupled DNA double-strand break repair. *DNA Repair*
522 *(Amst)* **109**, 103211 (2022).

523 34. Fukuda, T. *et al.* Expression of human cell cycle regulators in the primary cell line of the African
524 savannah elephant (loxodonta africana) increases proliferation until senescence, but does not
525 induce immortalization. *In Vitro Cell Dev Biol Anim* **52**, 20–26 (2016).

526 35. Pearson, V. R. *et al.* Identification of African Elephant Polyomavirus in wild elephants and the
527 creation of a vector expressing its viral tumor antigens to transform elephant primary cells. *PLoS*
528 *One* **16**, e0244334 (2021).

529 36. Housden, B. E. *et al.* Identification of potential drug targets for tuberous sclerosis complex by
530 synthetic screens combining CRISPR-based knockouts with RNAi. *Sci Signal* **8**, (2015).

531 37. Liu, T. M. *et al.* Molecular Basis of Immortalization of Human Mesenchymal Stem Cells by
532 Combination of p53 Knockdown and Human Telomerase Reverse Transcriptase Overexpression.
533 *Stem Cells Dev* **22**, 268–278 (2013).

534 38. Jha, K. K., Banga, S., Palejwala, V. & Ozer, H. L. SV40-Mediated Immortalization. *Exp Cell Res* **245**,
535 1–7 (1998).

536 39. Tollis, M. *et al.* Elephant Genomes Reveal Accelerated Evolution in Mechanisms Underlying
537 Disease Defenses. *Mol Biol Evol* **38**, 3606–3620 (2021).

538 40. Vazquez, J. M. & Lynch, V. J. Pervasive duplication of tumor suppressors in Afrotherians during
539 the evolution of large bodies and reduced cancer risk. *Elife* **10**, 1–45 (2021).

540 41. Nunney, L. Cancer suppression and the evolution of multiple retrogene copies of TP53 in
541 elephants: A re-evaluation. *Evol Appl* **15**, 891–901 (2022).

542 42. Appleton, E. *et al.* Derivation of elephant induced pluripotent stem cells. *bioRxiv* (2024)
543 doi:10.1101/2024.03.05.583606.

544

545

546

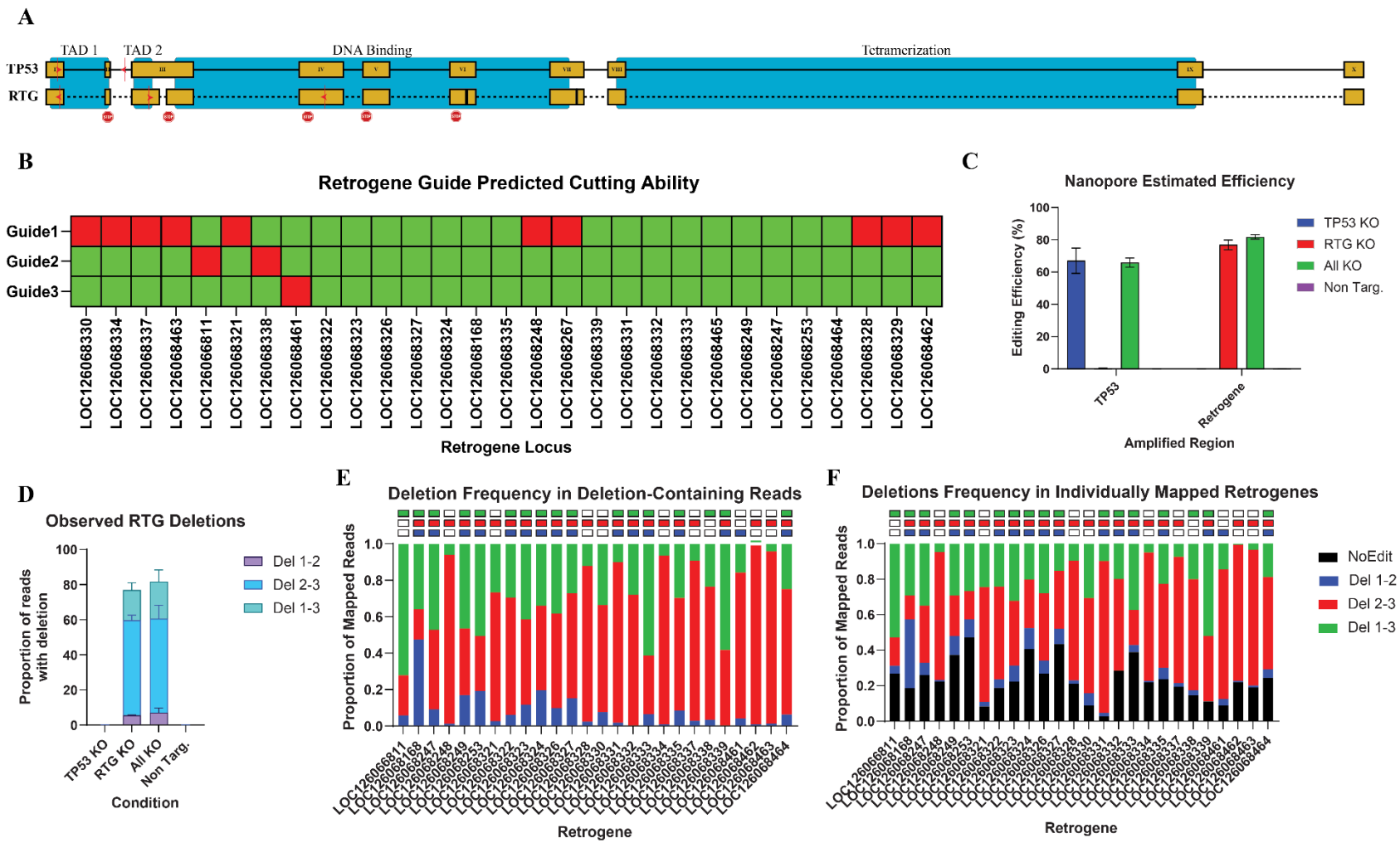
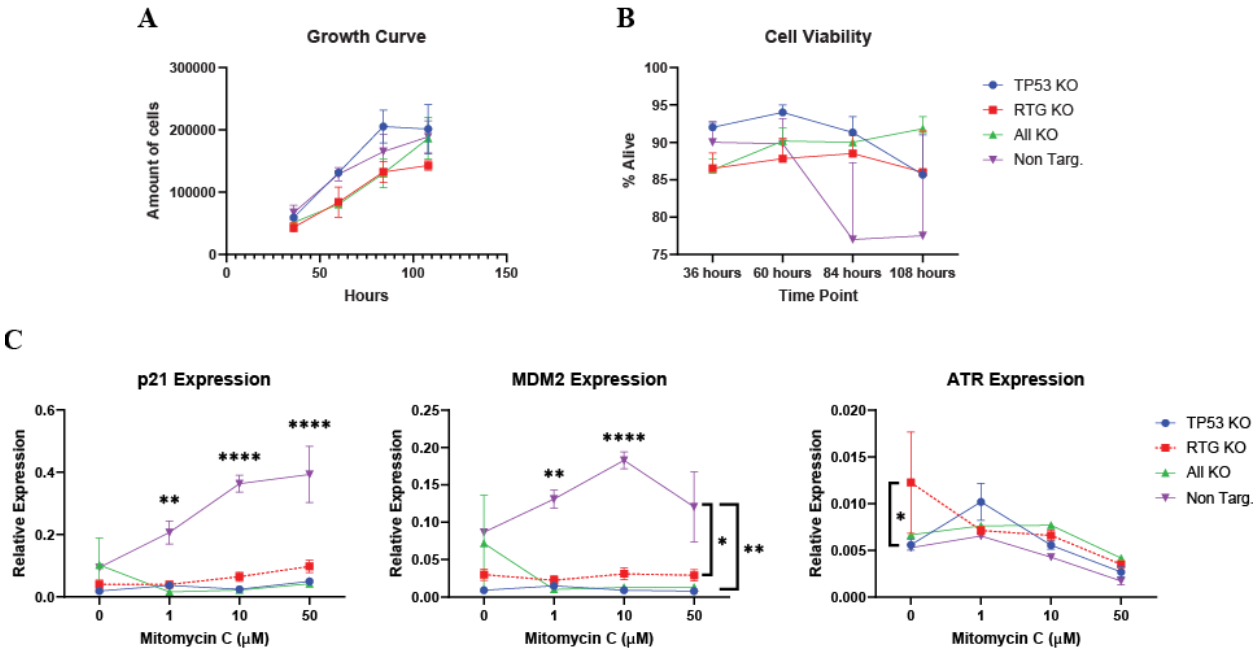


Figure 1. TP53 Editing. A. Schematic showing an alignment of Asian elephant TP53 and a retrogene (RTG) consensus, with domains labelled based on homology to human TP53. Exons are shown in yellow, with introns indicated by black lines. Missing sequence is indicated with a dashed line in the RTGs. Additionally, we've included the positions of deletions greater than 3bp in the RTG consensus, and the approximate locations of stop codons observed across the 29 Asian elephant retrogenes. The locations and directions of guides are shown for both TP53 and

552 the retrogenes. **B.** Predicted guide activity against each of the 29 Asian elephant retrogenes, with green indicating the guide is expected to be
553 functional against that target and red indicating non-functionality. **C.** Observed editing efficiency of the expected deletions based on nanopore
554 data. **D.** The frequency of deletions observed in all reads mapped against the RTG consensus across different treatment conditions. **E.** The
555 frequency of each deletion across the individual Asian elephant retrogenes as a proportion of all deletion-containing reads mapped to that locus,
556 or all mapped read (**F**). Amplicons which were identical across the non-edited regions were collapsed as they are bioinformatically
557 undistinguishable. Error bars represent standard deviation in all plots.

558



559

560 **Figure 2. Cell characterization and expression. A.** Growth rates of cells with TP53, RTG, or combined
561 knockouts, or treated with a non-targeting guide. **B.** Cell viability observed during the growth curve for
562 each condition. **C.** The expression of two downstream (p21 and MDM2) and one upstream (ATR) genes
563 within the TP53 pathway. Expression levels for the RTG KO are shown with a dashed line due to the
564 possible presence of TP53 off-target mutations. Significance levels: $p \leq 0.05$ (*); $p \leq 0.01$ (**); $p \leq 0.001$
565 (***); $p \leq 0.0001$ (****). Errors bars indicate standard error of the mean in all plots.

566

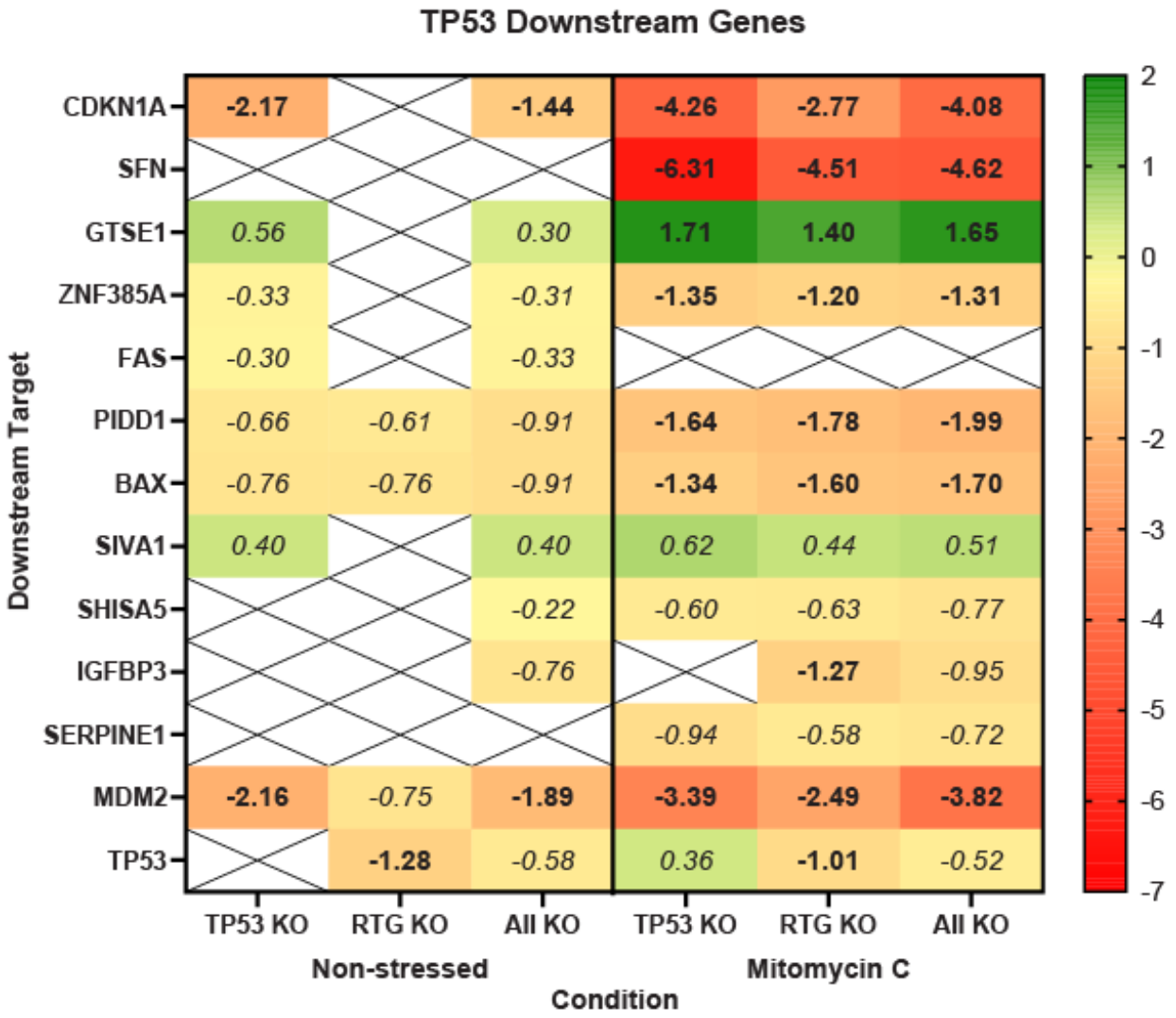


Figure 3. Expression of TP53 pathway genes. Log2 fold changes observed for genes immediately downstream of TP53 in the p53 signaling pathway (KEGG pathway hsa04115). Crosses indicate that no significant difference in expression was observed for that gene in a given condition. Genes in italics have an absolute log2 fold change of < 1 and were filtered out of enrichment analyses. Only downstream genes with a significant difference in at least one condition are shown.

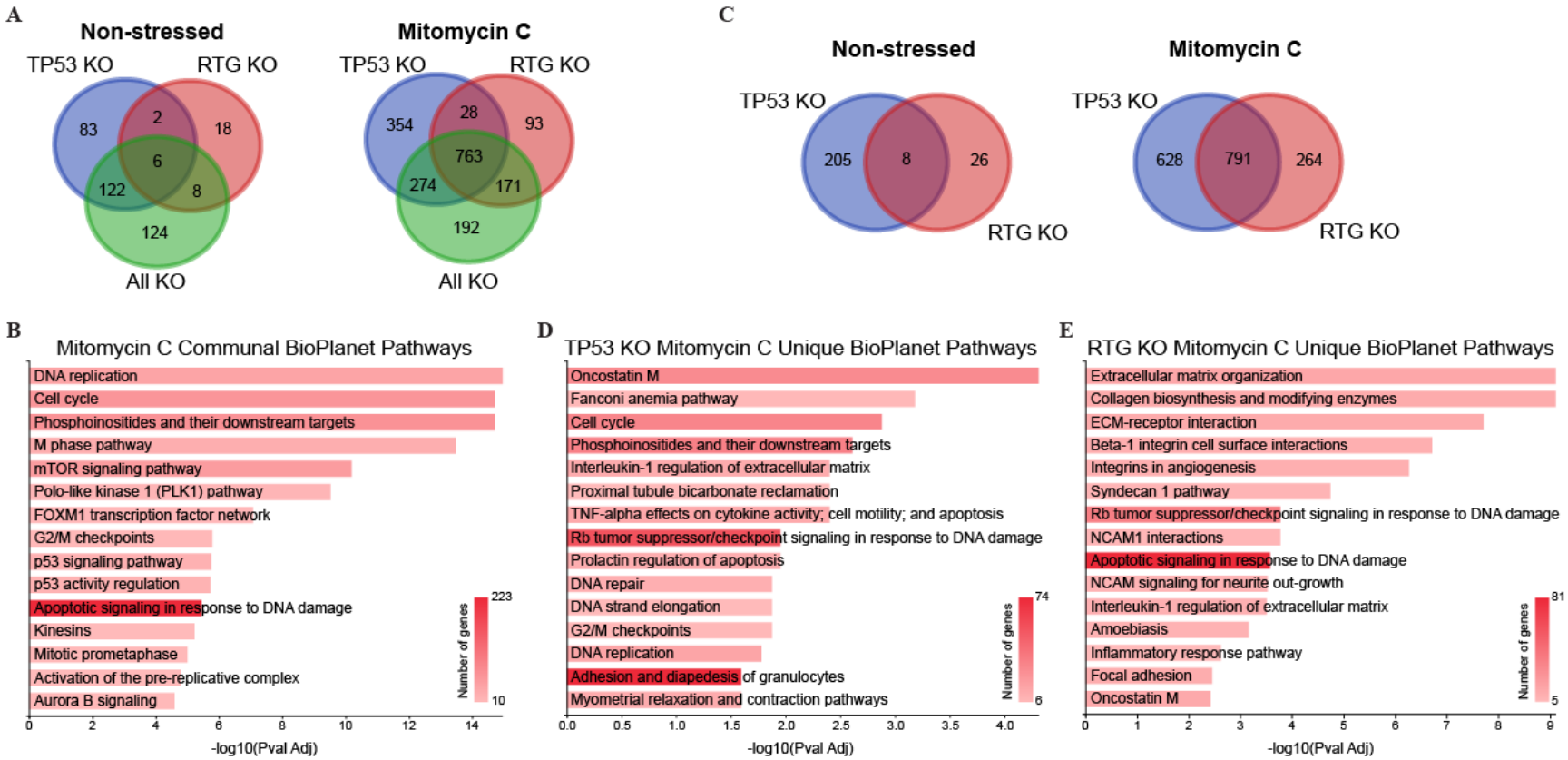


Figure 4. Pathway enrichment. A. Differentially expressed gene (DEG) overlap observed between treatment conditions for cells in normal media or exposed to 10µM Mitomycin C. Genes were filtered for a minimum absolute log2 fold change greater than 1. **B.** BioPlanet pathway enrichment for the 763 DEGs common to all three conditions in the Mitomycin C treatment cells. **C.** DEG overlap between TP53 KO and RTG KO cells in normal media or exposed to 10µM Mitomycin C. **D.** BioPlanet pathway enrichment of unique DEGs in TP53 KO Mitomycin C treated cells. **E.** BioPlanet pathway enrichment of unique DEGs in RTG KO Mitomycin C treated cells.

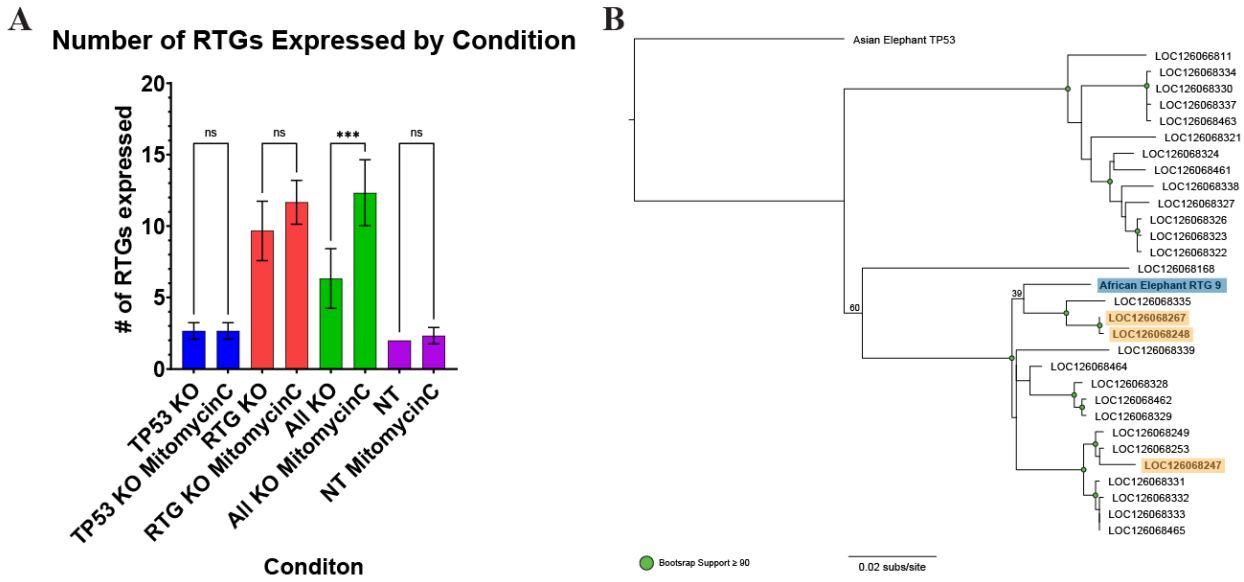


Figure 5. RTG expression. A. Expression of RTGs in each treatment condition in cells exposed to 10µM Mitomycin C or not. Errors bars indicate standard deviation. **B**. Maximum likelihood phylogeny showing the relationship between Asian elephant TP53 and the RTGs, as well as African elephant RTG9. Three tree was rooted using Asian elephant TP53. Green circles indicate nodes with bootstrap support ≥ 90%, while numbers above nodes represent bootstrap support for select nodes. The three Asian elephant RTGs which are most commonly expressed, and African elephant RTG9 are indicated in orange and blue respectively.