Engineering elephant models of cold adaptation and cancer resistance

Emil Karpinski* [1], Nikil Badey [2], Esther Mintzer [1], Asaf Ashkenazy-Titelman [1], Li Li [1,2], George M. Church* [1,2]

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

* Corresponding authors.

E-mail addresses:

Emil Karpinski – emil_karpinski@hms.harvard.edu
George Church – gchurch@genetics.hms.harvard.edu

Abstract

Proboscideans have developed a suite of genetic changes including those responsible for tumor suppression and environmental adaptation. Here we examine a handful of woolly mammoth-specific deletions, and their potential contributions to arctic adaptation, as well as the expanded TP53 genetic repertoire in elephants. We use CRISPR-Cas9 to introduce mammoth-specific noncoding deletions as well as knockouts of TP53, all 29 TP53 retrogenes, or both in combination in Asian elephant cell lines, and examine the transcriptomic response. We find that many of the mammoth-specific deletions likely contribute to various arctic phenotypes including vascular development, metabolism and thermogenesis, and hair and skin adaptations. We also find that while there is considerable overlap in DNA damage responses of the TP53 and retrogene knockouts, retrogenes knockouts also exhibit strong enrichment of many extracellular pathways suggesting they may play a role in the tumor microenvironment and mitigating metastatic growth.

Keywords: woolly mammoths; Asian elephants; arctic adaptation; TP53; retrogenes; CRISPR-Cas9; RNA-seq; gene regulation

28 **Introduction**

29 Gene regulation plays a significant role in the ability of species to adapt to novel environments
30 and is believed to contribute to speciation via hybrid dysfunction (*1*, *2*). Adaptive regulatory changes
31 have been identified across Animalia, including in primates, mice, birds, fish, and cervids (*2*–*5*). One of
32 reasons for the large impact of regulatory mutations is their ability to fine-tune expression of genes
33 during different developmental stages or tissues, in comparison to the near ubiquitous change that
34 mutating protein sequences would have (*5*). However, these tissue-specific effects, as well as the fact
35 that regulatory sequences may be very distant from their regulatory targets, make them difficult to
36 examine bioinformatically, and necessitate the use of cell or animal models (*2*).

37 Throughout their ~60-million-year evolutionary history, proboscidean evolution has largely
38 followed two main trends: a general increase in size from the earliest proboscideans, *Eritherium*
39 *azzouzorum* and *Phosphatherium escuillei* (~5 and 17 kg respectively); and adaptive radiation to expand
40 into and occupy many different environments from tropical wetlands to the cold tundra (*6*, *7*) (Fig. 1A).
41 These trends are especially apparent in the elephants (*Elephantidae*), which during the late Pleistocene
42 included one of the largest land mammals of all time, *Palaeoloxodon antiquus* (>13 tons), and species
43 found in various environments across Africa, Eurasia, and North America (*6*, *7*).

44 Woolly mammoths (*Mammuthus primigenius*) occupied dry, cold steppe-tundra environments
45 during the coldest periods of the Pleistocene (*6*, *8*). They evolved many physiological adaptations to
46 survive in this environment, including a multi-layered and thicker hair coat, changes in skin and fat
47 distribution, smaller ears and tails to minimize heat exchange, and a general increase in size and growth
48 rate. These changes are especially striking given their closest living relative, the Asian elephant (*Elephas*
49 *maximus*), is primarily found in warm topical forests and grasslands environments throughout Southeast
50 Asia (*6*, *8*). Previous studies have reported protein-coding changes between woolly mammoths and
51 extant elephants which may have contributed to some of these phenotypes (*9*, *10*). However, the
52 evolved regulatory landscape of woolly mammoths has been largely ignored.

53 Peto's paradox is the observation that cancer incidence in a species does not always scale with
54 body mass and/or lifespan (*11*, *12*). Elephants have emerged as a key species in attempts to resolve this
55 paradox due to their large size, long lifespan, and very low cancer risk (*12*, *13*). Genomic analyses have
56 identified the duplication of many tumor suppressor genes across extant elephants, including multiple
57 retrogene copies of TP53 (*12*, *13*). Although there is still considerable uncertainty about how or which of
58 these TP53 retrogenes (RTGs) may confer tumor resistance, previous works have shown that the
59 encoded protein of at least one of these retrogenes, African elephant RTG 9, induces strong cellular

60 responses to DNA damage (*12–14*). However, most previous studies have largely focused only on African
61 elephant TP53 retrogenes and often in non-native genetic (e.g. plasmid expression) or cellular
62 backgrounds (e.g. human or mouse cells).
63 In this study, we examine the regulatory changes responsible for adaptation and tumor
64 resistance in woolly mammoths and Asian elephants in vitro. We have engineered a handful of
65 mammoth-specific non-coding mutations into Asian elephant cells, and observed changes implicated in
66 cardiac development, hair and skin function, and lipid metabolism. We further validate our cell culture
67 models by examining the best-studied elephant genetic pathway – their expanded TP53 repertoire. We
68 used CRISPR-Cas9 to disrupt TP53, all TP53 retrogenes, or both in combination in Asian elephant
69 fibroblasts, and examined the transcriptional profiles of these knockouts with and without DNA damage
70 stress, to resolve the protective roles of TP53 retrogenes (Fig. 1). Our results suggest that these cell
71 culture models are tremendously useful for investigating the potential effects of both adaptive
72 regulatory changes and tumor suppressor duplications, and generate a strong foundation of testable
73 hypotheses for future research.
74
75 **Results**
76 ***Identification of mammoth deletions***
77 Our in silico TFBS annotation predicted ~1.98 million transcription factor binding sites (TFBS)
78 upstream of 21,808 genes (mean 90.6/gene). We originally identified ~235k mammoth-specific
79 deletions, with an average size of 236 bp (min: 25; max: 13.2kb), of which 1,798 intersected our
80 predicted transcription factor binding sites (TFBS; Supplementary Table 3). Using previously established
81 methods we also predicted 629 large mammoth-specific deletions (i.e. >=500 bp; Supplementary Table
82 4) of which 46 were predicted to have a genome-unique guide within 25bp of their start and end
83 coordinates (*15*). Notably we find strong concordance between both methods, with deletions identified
84 using our own methodology intersecting 88.1% of the large mammoth deletions identified using the
85 established pipeline.
86
87 ***Mammoth-specific and Synthetic Regulatory Editing***
88 We attempted to introduce one mammoth specific deletion upstream of PCNT (pericentrin), and
89 one synthetic deletion upstream of EIF3I (eukaryotic translation initiation factor 3 subunit I) due to its
90 high expression in our mesenchymal stem cell (MSC) line. PCNT mutations have implicated in the
91 development of smaller cranial features and olfactory development (*8*, *16*), which may contribute to

92 arctic adaptations present in woolly mammoths. While we were successful in generating mutations in
93 these regions (Fig. 2A), we didn't recover differential expression of PCNT in either our pilot or
94 verification experiments. We did however observe differential expression of EIF3I, but only with 50% of
95 the guide combinations, one of which we successfully verified. Our lack of signal in PCNT and sometimes
96 in EIF3I underscores the effect tissue or timing has on gene regulation, and may be why we fail to
97 observe a signal in all cases.
98
99 ***Mammoth-specific deletion differentially expressed genes***
100 We also generated three large mammoth deletions (Deletions 3, 51, and 89) with comparable
101 editing efficiency in both fibroblasts and mesenchymal stem cells (Fig. 2A). Following bulk RNA-seq we
102 recovered a large number of differentially expressed genes (DEGs), although a significant portion of
103 them had very small log2 fold changes. To focus our analysis on those that were likely to be most
104 impacted we filtered our dataset to a minimum log2 fold change of 1 before further analysis.
105 Interestingly, we see a high level of variability in the DEGs affected between cell types, even given the
106 similarity of fibroblast and MSCs (*17*) (Fig. 2B), with the majority of DEGs being unique to each cell type.
107 This is also true even if we remove the minimum log2 fold change (Fig. S2). This suggests that these
108 deletions likely have cell-type specific effects and may result in very different phenotypes depending on
109 the cell's specific context.
110
111 ***Potential impact on mammoth phenotypes***
112 We next wanted to see if we could infer some of the functions of these deletions, particularly as
113 they may relate to arctic adaptations in woolly mammoths. To do this, we first screened our filtered
114 DEGs to identify genes within pathways of interest (Fig. 2C), and then investigated what was known
115 about perturbations in those genes in various cell and animal model analyses. We identified changes
116 that broadly fall under three phenotypes: skin and hair adaptations; vascular adaptations; and
117 metabolism and thermogenesis.
118
119 ***Skin and Hair Adaptations***
120 In comparison to their closest living relatives, Asian elephants, woolly mammoths have both
121 denser hair coverage and multiple coats (layers of hair) to protect them from the arctic cold (*8*). We find
122 DEGs which may be related to skin and hair adaptions within each deletion, although not in every cell
123 type – MSC Deletion 3 (SOSTDC1 and BCL2); MSC Deletion 51 (FABP5); and Fibroblast Deletion 89

124 (IGFBP5). SOSTDC1 is involved in the regulation of the number, size and/or shape of hair, and is
125 upregulated during the anagen phase in response to depilation-induced hair regeneration (*18*, *19*). BCL2
126 prevents apoptosis, and its upregulation has been shown to protect against hair cell death in the utricle
127 (*20*). Likewise, we observe downregulation of IGFBP5 in our fibroblast Deletion 89 cells, which is a strong
128 negative regulator of hair development and differentiation (*21*). Previous studies have also found that
129 overexpression of IGFBP5 produces thinner and shorter hair with aberrant morphologies (*21*). The
130 upregulation of SOSTDC1 of these in MSC Deletion 3, and downregulation of IGFBP5 in Fibroblast
131 Deletion 89, suggest these deletions may contribute to the abundance and robustness of the
132 mammoth's coats.
133 We also identify FABP5 upregulated in MSC Deletion 51, which is involved in regulating water
134 permeability. Knockouts of FABP5 result in much lower fatty acid content in the epidermis and impaired
135 water-barrier function and keratinocyte migration (*22*). Likewise, this gene is expressed in sebaceous
136 epithelial cells where it plays a role in sebaceous wax production (*23*). Sebaceous wax is believed to play
137 a role in waterproofing and water loss, as well as thermoregulation (*8*, *24*). Mice which have been
138 engineered to lack or have nonfunctional sebaceous glands suffer from abnormal hair growth/loss, skin
139 lesions, and compromised thermoregulation (*24*). Notably, histological analyses of the few mammoth
140 soft tissue remains that have preserved have revealed the presence of sebaceous cells, which are almost
141 entirely absent in extant elephants (*8*, *25*).
142
143 ***Vascular Adaptations***
144 Vascular adaptations are also frequently observed in many arctic species. One of the best
145 studied adaptations within this category is an increase in heart size, likely due to the need for more
146 blood flow to maintain temperature and oxygen homeostasis (*26*, *27*). Although mammoth soft tissue
147 has poor taphonomy, there are claims that mammoth hearts were larger than similar warm-adapted
148 elephants of their age and size (*28*). Correspondingly, we observe genes involved in heart remodeling
149 and angiogenesis in four experimental conditions – MSC Deletion 3 (IGF2BP2); MSC Deletion 51 (NRG1
150 and LAMA4); and Fibroblast Deletion 51 (HBEGF and PPP1R3A). Previous research has implicated
151 IGF2BP2, NRG1, LAMA4, HBEGF, and PPP1R3A in heart morphology and metabolism, with the
152 overexpression of many of these genes resulting in heart remodeling and hypertrophy (*29–33*).
153 Additionally, IGFBP2 and PPP1R3A have also been shown to play a role in metabolic changes that
154 accompany abnormal heart morphologies in mice (*31*, *32*).

155 Accompanying these changes, we also observe differential expression in genes related to
156 angiogenesis – MSC Deletion 51 (LAMA4); and Fibroblast Deletion 89 (SEC14L3). SEC14L3, in conjunction
157 with VEGF, has a role in promoting the migration of endothelial and vein progenitor cells and
158 consequently angiogenesis, and its overexpression has been implicated in the proliferation and
159 metastasis of tumors (*34*). Likewise, LAMA4 appears to be highly involved in blood vessel development.
160 Mouse LAMA4 knockouts appear generally phenotypically normal, but exhibit frequent hemorrhages,
161 and in cornea angiogenesis assays developed grossly distorted blood vessel architectures (*35*).
162 Interestingly, while both genes are strong drivers of angiogenesis, we observe downregulation of
163 SEC14L3 and upregulation of LAMA4 in their respective experimental conditions. Although the general
164 trend here is still unclear, it's possible these deletions are implicated in fine tuning angiogenesis in
165 mammoths.
166
167 ***Metabolism and Thermogenesis***
168 The ability to sense and respond to cold temperatures via metabolic changes is probably
169 amongst the most important cold adaptations. There is a suite of metabolic cascades that are triggered
170 in response to cold nociception, and the ability to respond to cold stimuli via thermogenesis is highly
171 dependent on the regulation of both glucose and lipid metabolic reserves (*36*). Not surprisingly, we
172 recover many differentially expressed genes involved in adipocyte proliferation and lipid and glucose
173 metabolism in four out of our six experimental conditions.
174 The bulk of our differentially expressed genes are involved in adipocyte differentiation and/or
175 lipid metabolism – MSC Deletion 3 (IGF2BP2, PID1, and NPC1L1); MSC Deletion 51 (IGF2BP2, MSR1,
176 DGAT2, DUSP26, and DDIT3); Fibroblast Deletion 51 (NR4A1, EGR2, and MAFB); and Fibroblast Deletion
177 89 (IRF4 and IGFBP5). We observe upregulation of IGF2BP2, PID1, DDIT3, and IRF4 in their respective
178 experimental conditions, all of which except DDIT3 are positive regulators of adipocyte
179 differentiation/proliferation (*37–40*). Likewise, we observe downregulation of NR4A1, EGR2, and MAFB
180 all of which have previously been shown to act as negative regulators of adipocyte differentiation or size
181 (*41–43*). Unsurprisingly, almost all these genes and additionally MSR1, DGAT2, NPC1L1, DUSP26, and
182 IGFBP5 are also involved in lipid metabolism and controlling the rates of stored versus circulating lipids
183 (*44–48*). For example, mice induced to overexpress DGAT2 do not appear to significantly different fat
184 pads or body weights relative to controls but exhibit much higher levels of circulating plasma
185 triglycerides (*47*). Likewise, MAFB deficient mice have higher body weights and amounts of body, due to
186 large adipocytes and higher serum cholesterol levels (*43*).

187 Accordingly, many of the above genes have been shown to be implicated in the ability to sense
188 and respond to cold in mouse models. GRIK2 (MSC Deletion 3) knockouts are unable to sense cold, but
189 not cool, temperatures and have a sever reduction in cold-specific neurons (*49*). IGF2BP2 knockout mice
190 have better thermoregulation than controls in the short term, but have fewer and smaller adipocytes
191 and adipocyte precursors, and have diminished UCP1 levels, a key thermogenic protein (*38*).
192 Interestingly, IRF4 knockouts have been shown to be more obese than wild type mice but exhibit
193 increased cold intolerance likely due to the cold responsive nature of IRF4 (*37*).
194
195 ***The Asian elephant genome contains 29 TP53 retrogenes***
196 Although our RNA-seq results suggest that the deletions we engineered are likely involved in
197 woolly mammoth cold adaptations, many of these phenotypes would only be measurable in organismal
198 models. Towards this end we have primarily referenced literature on perturbations in these genes
199 within mouse models, though we note that elephants are quite diverged from all common mammalian
200 model species (*50*). To both validate our cellular models, and expand our understanding of elephant
201 cancer resistance, we investigated the best-studied elephant genetic system – the expanded TP53
202 repertoire.
203 We searched the annotations of the Asian elephant reference genome and identified 29
204 retrogene copies of TP53 (RTGs). 27 of these RTGs are found on chromosome 27, while the remaining
205 two are found on chromosomes 24 and 26. By comparison, the canonical copy of TP53 is located on
206 chromosome 19. Most of the RTGs are approximately 1.1 kb in length, roughly the same length as the
207 CDS of TP53 as would be expected of retrogenes, except for three (LOC126068247 – 64.1 kb;
208 LOC126068267 – 32.5 kb; LOC126068248 – 4.9 kb; Fig. S3). When aligned and translated in the same
209 frame as TP53, each RTG carries a premature stop codon, which primarily cluster in five main areas
210 across the homology region (Fig. 3A).
211
212 ***TP53 and RTG editing and validation***
213 We attempted to design paired Cas9 guide RNAs (sgRNAs) to produce large, disruptive deletions
214 in TP53 and all the retrogenes. We managed to identify two sgRNAs that specifically target TP53, which
215 result in the partial loss of exon 1, the entirety of exon 2, and a premature stop codon after read
216 through into intron 2 (Fig. 3A). We could not identify just two guides that were both predicted to have
217 activity across all 29 retrogenes, so we designed a third guide to ensure each RTG was targetable by at
218 least two sgRNAs (Fig. 3B; Fig. S4).

219 We electroporated fibroblast cells with Cas9 RNPs consisting of either the two TP53 guides, the
220 three RTG guides, all five guides together (All), or a guide predicted to have no activity in the Asian
221 elephant genome (Non Targ.). After approximately four days, we then collected cells, extracted DNA,
222 and amplified a portion of TP53 and RTG genes flanking the expected editing sites. We observed high
223 levels of editing across all three conditions when we examined the resulting amplicons on an agarose gel
224 (Fig. S5). However, due to our PCR reaction amplifying all 29 loci simultaneously, we couldn't assess
225 which and how many of the RTGs were being edited.
226 To address this issue, we sent both PCR reactions for end-to-end nanopore sequencing. We
227 quantified the number of observed deletions, using a custom script which marks the presence of a
228 deletion if 90% of the bases within a defined region were absent. We note that this is likely an
229 underestimate of the true editing efficiency, as this will only identify the perfect deletion product and
230 miss any indels and/or incomplete loss of the flanked segment of DNA. We observed high editing
231 efficiency in all conditions using either the TP53 guides (TP53 KO – 67%; All KO – 66%) or the RTG guides
232 (RTG KO – 77%; All KO – 82%), and almost no editing in the Non Targ. condition (<0.1% at both loci) (Fig.
233 3C). We further examined the proportion of each deletion product observed in the RTG PCRs, and noted
234 a clear pattern, wherein deletions between guides 2 and 3 were the most frequent, in line with
235 expectations about the targeting scope of those two guides (Fig. 3D).
236 Next, we wanted to disentangle the editing on a per-locus level for the RTG data. We mapped
237 the data to each unique RTG (collapsing those with identical sequences outside the editing region) and
238 identified the proportion of each deletion at each locus. Although we generally observe a mix of all
239 three deletions at each locus (some of which is likely due to cross mapping between the RTGs), we
240 generally observe deletions in line with our expectations based on their targeting scope (Fig. 3E). Only
241 two loci buck this trend, LOC126068338 and LOC126068461, both of which have a single point mutation
242 at position 10 within the seed region of the guide (Fig. S4). Due to the importance of this region for
243 guide specificity, we filtered these targets out; however, we note that at least at these two loci the
244 guides appear to be permissive for editing. When combined with the per-locus reads that were not
245 identified as containing a deletion (Fig. 3F), we observe an average of 76.8% editing efficiency across all
246 unique loci.
247
248 ***RNA-seq analysis reveals overlapping and unique TP53 and RTG responses***
249 Given most literature on elephant RTGs has largely focused on how these retrogenes may
250 interact with canonical TP53 and MDM2, we first decided to look at changes in expression of all direct

251 downstream targets of TP53 (Fig. 4). In the absence of stress (i.e. 10µM Mitomycin C), we see relatively
252 small changes in expression within most downstream targets with the exception of MDM2 and CDKN1A
253 in both the TP53 and All KOs. Interestingly, we note that TP53 is not identified as significantly
254 downregulated in the TP53 KO, suggesting maybe that nonsense-mediated decay is less pronounced in
255 Asian elephants. In comparison, we observe very few differentially expressed genes in the RTG KO, and
256 only TP53 that has a log2 fold change greater than 1. This suggests that the RTGs may play some role in
257 either stabilizing or increasing the abundance of p53, although likely not directly by acting as a
258 transcription factor as none of the RTGs have an intact DNA-binding domain (Fig. 3).
259 In the presence of Mitomycin C, all three conditions have similar patterns in DEGs (Fig. 4),
260 although typically smaller expression changes in the RTG KO. As before, we also observe that TP53 is
261 significantly less expressed in the RTG KO and not in the TP53 KO. Additionally, IGFBP-3, a gene which
262 can contribute to cell growth and apoptosis (*51*, *52*), is also downregulated in the RTG KO. We also
263 observe significant downregulation of SFN across all three conditions in the MMC-treated cells, in line
264 with this gene's known tumor suppressor function of arresting cellular replication in response to DNA
265 damage (*53*, *54*).
266 Next, we wanted to examine the different pathways enriched within each condition. Due to the
267 large number of genes with small changes we filtered the dataset to a minimum absolute log2 fold
268 change of at least 1. We then extracted the DEG lists for each condition and analyzed the overlap
269 between them in either the non-stressed or MMC-treated cells (Fig. 5A). We performed BioPlanet
270 pathway enrichment (*55*) on the 763 DEGs common to all three conditions treated with Mitomycin C.
271 We observed significant enrichment in p53 signaling pathways, as would be expected of cells exposed to
272 a DNA damaging agent, as well as enrichment in many DNA replication and cell cycle pathways (Fig. 5B).
273 We also examined the unique DEGs in the TP53 and RTG KO cells with Mitomycin C treatment
274 (Fig. 5C). For the TP53 KO, we see enrichment in many pathways that would be expected given the
275 known checkpoint and cell stress responses coordinated by TP53 (e.g. cell cycle and proliferation, DNA
276 repair, etc; Fig. 4D) (*56*–*58*). By comparison, unique RTG KO DEGs are enriched in extracellular
277 organization and signaling/interaction pathways (Fig. 5E), with many genes that contribute to multiple
278 extracellular matrix related processes (Fig. 5F). This suggests that RTGs may play a unique protective role
279 in regulating the tumor microenvironment and limiting metastasis (*59*). Additionally, the largest number
280 of DEGs appear to belong to the apoptotic signaling in response to DNA damage pathway, a function
281 that has previously been suggested for African elephant RTGs (*12*, *14*).
282

283 ***Very few RTGs are likely to be expressed or functional***

284 Given the unique function suggested by the RNA-seq of the RTG KO cells, we next wanted to see
285 if we could investigate which RTGs might be responsible for the changes we observed. In none of the
286 conditions, irrespective of treatment, did we detect differential expression of any of the RTGs.
287 Furthermore, we generally do not observe differences in the amount of RTGs expressed in response to
288 Mitomycin C treatment (Fig. S10A), with the exception of the All KO. We do observe an increase in the
289 amount of RTGs expressed in the RTG and All KO conditions, irrespective of Mitomycin C treatment
290 (significance values not shown), although this may be a lingering transcriptional effect caused by double-
291 strand break repair (*60*). However, we do note that all but three RTGs have very low read counts
292 (Supplementary Table 6).

293 We next wanted to confirm that this pattern in RTG expression is not due to cell type or tissue
294 specific. We download public RNA-seq datasets for Asian elephant ovary, salivary glands, lymph nodes,
295 lung, thyroid, and lymphocytes, as well as examined some preliminary data generated with the same
296 guides in epithelial and mesenchymal stem cells, and examined RTG expression patterns in these
297 datasets. We observe the same general pattern, where the only three RTGs which show consistent
298 expression are LOC126068247, LOC126068267, and LOC126068248 (Supplementary Table 7). However,
299 these three RTGs are the three largest RTG genes (LOC126068247 – 64.1 kb; LOC126068267 – 32.5 kb;
300 LOC126068248 – 4.9 kb) based on their annotations.

301 To examine what was driving the expression of these RTGs, we extracted RNA-seq mapped
302 reads across each of these three loci and examined the frequency of reads across each locus. For two of
303 these RTGs, we observe that most of the reads assigned to these RTGs map thousands of base pairs
304 upstream of the region homologous to the TP53 CDS. In LOC126068247 and LOC126068248, we observe
305 almost no mapped reads within the TP53 homology region either in our data or the publicly available
306 datasets (Figs. S11-S14). By comparison, LOC126068267 also has peaks of high coverage upstream of the
307 TP53 homology region but shows similar levels of coverage across the homology region in both datasets
308 (Figs. S15-S16).

309 Given the lack of expression of most of the annotated retrogenes, we decided to compare those
310 that might be functional in Asian elephants with African elephant retrogene 9 (RTG 9; KF715863), on
311 which most previous studies have focused (*12–14*). We aligned the TP53 homologous regions of all Asian
312 elephant RTGs, African elephant RTG 9, and the TP53 CDS, and built a maximum-likelihood phylogeny.
313 We observe that RTG 9 is most closely related to a clade containing two of our putatively expressed
314 RTGs, LOC126068267 and LOC126068248 (Fig. S10B), further suggesting that these two RTGs might be

315 responsible for the changes we observe. However, we do note that due to the high similarity of many of
316 these RTGs, the bootstrap support values within our phylogeny are quite weak.
317
318 **Discussion**
319 Proboscidean evolution can be characterized by a general increase in size from the 5 kg
320 *Eritherium* to the 13-ton *Palaeoloxodon*, and a dispersal from the tropical wetlands of Africa into a wide
321 variety of environments across the globe (*6*, *7*). Elephants offers a unique opportunity to examine these
322 trends as they contain a highly evolved tumor suppressor system (*12*), and have closely related taxa
323 adapted to two distinct environments – the arctic steppe tundra (woolly mammoths), and the tropical
324 forests and grasslands of southeast Asia (Asian elephants) (*6*). Consequently, these taxa represent key
325 candidates to understand both the adaptations necessary to expand into arctic environments and
326 maintain cellular and genomic integrity with increased body size.
327 Studies in other taxa have suggested that a lot of adaptive evolution takes place on the level of
328 gene regulation as opposed to protein function, possibly due to the reduced constraints on regulatory
329 motif mutations (*2*, *5*). In this study we have investigated the regulatory role of some mammoth-specific
330 deletions on arctic adaptations, and the expanded TP53 repertoire found in all elephants. We do note
331 that all of these results should be viewed as hypotheses and testing their true impact would require the
332 generation of animal models to fully examine their developmental and tissue-specific effects. However,
333 as this is both currently not possible and ethical in elephants, we have generated many of these
334 mutations within elephant cells to shed some light on their possible function.
335 We observe many differentially expressed genes that may be involved in arctic adaptations
336 when we introduce our identified mammoth-specific deletions. Amongst these we see genes involved
337 broadly with vascular system development, skin and hair, lipid metabolism, temperature sensing, and
338 thermogenesis. However, we often fail to observe differential expression of genes containing or near
339 our perturbations, although this result is likely affected by cell type (*61*). In fact, even in our larger
340 mammoth deletions, we observe very little overlap in DEGs between MSCs and fibroblasts. Taken
341 together, the results presented here are promising and provide some insight into mammoth adaptations
342 but likely fail to fully capture the regulatory extent of these mutations, and which likely will not become
343 evident until transgenic non-model species become commonplace or our ability to bioinformatically
344 predict causative effects within a tissue-specific context drastically increases.
345 We also find that only three of the twenty-nine TP53 retrogenes are expressed across an
346 extensive list of tissues and likely functional. Interestingly, two of the Asian elephant retrogenes most

347 likely to be functional are closely related to African elephant RTG9, which has been shown to induce
348 strong cellular responses to DNA damage (*12–14*) and apoptosis in mouse fibroblasts and human cancer
349 cell lines (*12*, *14*). We also observe clear transcriptional differences between TP53 and RTG KO cells,
350 including a significant downregulation of TP53 expression in our RTG KO cells, suggesting a novel
351 function for elephant retrogenes. Additionally, we observe a differential enrichment of pathways when
352 disrupting RTGs versus TP53, indicating that while there is a high amount of overlap between these
353 genes, they also regulate many unique pathways. In particular, genes enriched within many extracellular
354 organization and signaling pathways were observed in the RTG KO population. Some of these (e.g.
355 extracellular matrix organization, angiogenesis) have been implicated as contributing to the
356 establishment of tumor microenvironments, and are gaining increased importance as vital to cancer
357 growth and spread (*59*). This may also explain the low prevalence of malignant cancers in Asian
358 elephants (*62*).
359 Our work highlights the utility of cellular models to begin disentangling the complex effects of
360 regulatory noncoding deletions and cancer resistance genetic networks in non-model species. However,
361 in addition to the tissue-specific response limitations mentioned above, it is further hindered by our
362 reliance on primary, non-immortalized cells, although this will likely be addressed as genetic resources
363 for non-model species become more abundant (*63*, *64*). Of particular importance would be stable and
364 readily available iPSCs, which would allow for differentiation of various tissues and more complex
365 organoid models to better investigate the effects of both adaptive and cancer resistance mutations,
366 especially where the generation of live animal models is not possible.
367
368
369
370
371
372
373
374
375

References

1. K. L. Mack, M. W. Nachman, Gene Regulation and Speciation. *Trends in Genetics* **33**, 68–80 (2017).

2. I. G. Romero, I. Ruvinsky, Y. Gilad, Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* **13**, 505–516 (2012).

3. L. Hu, J. Long, Y. Lin, Z. Gu, H. Su, X. Dong, Z. Lin, Q. Xiao, N. Batbayar, B. Bold, L. Deutschová, S. Ganusevich, V. Sokolov, A. Sokolov, H. R. Patel, P. D. Waters, J. A. M. Graves, A. Dixon, S. Pan, X. Zhan, Arctic introgression and chromatin regulation facilitated rapid Qinghai-Tibet Plateau colonization by an avian predator. *Nat Commun* **13**, 6413 (2022).

4. Z. Lin, L. Chen, X. Chen, Y. Zhong, Y. Yang, W. Xia, C. Liu, W. Zhu, H. Wang, B. Yan, Y. Yang, X. Liu, K. Sternang Kvie, K. H. Røed, K. Wang, W. Xiao, H. Wei, G. Li, R. Heller, M. T. P. Gilbert, Q. Qiu, W. Wang, Z. Li, Biological adaptations in the Arctic cervid, the reindeer (*Rangifer tarandus*). *Science (1979)* **364** (2019).

5. K. L. Mack, M. A. Ballinger, M. Phifer-Rixey, M. W. Nachman, Gene regulation underlies environmental adaptation in house mice. *Genome Res* **28**, 1636–1645 (2018).

6. J. Shoshani, Understanding proboscidean evolution: a formidable task. *Trends Ecol Evol* **13**, 480–487 (1998).

7. A. Larramendi, Proboscideans: Shoulder Height, Body Mass and Shape. *Acta Palaeontol Pol* **61**, 537–574 (2015).

8. G. G. Boeskorov, E. N. Mashchenko, V. V. Plotnikov, M. V. Shchelchkova, A. V. Protopopov, N. G. Solomonov, Adaptation of the woolly mammoth Mammuthus primigenius (Blumenbach, 1799) to habitat conditions in the glacial period. *Contemp Probl Ecol* **9**, 544–553 (2016).

9. T. van der Valk, P. Pečnerová, D. Díez-del-Molino, A. Bergström, J. Oppenheimer, S. Hartmann, G. Xenikoudakis, J. A. Thomas, M. Dehasque, E. Sağlıcan, F. R. Fidan, I. Barnes, S. Liu, M. Somel, P. D. Heintzman, P. Nikolskiy, B. Shapiro, P. Skoglund, M. Hofreiter, A. M. Lister, A. Götherström, L. Dalén, Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* **591**, 265–269 (2021).

10. V. J. Lynch, O. C. Bedoya-Reina, A. Ratan, M. Sulak, D. I. Drautz-Moses, G. H. Perry, W. Miller, S. C. Schuster, Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. *Cell Rep* **12**, 217–228 (2015).

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

12. M. Sulak, L. Fong, K. Mika, S. Chigurupati, L. Yon, N. P. Mongan, R. D. Emes, V. J. Lynch, TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *Elife* **5** (2016).

13. L. M. Abegglen, A. F. Caulin, A. Chan, K. Lee, R. Robinson, M. S. Campbell, W. K. Kiso, D. L. Schmitt, P. J. Waddell, S. Bhaskara, S. T. Jensen, C. C. Maley, J. D. Schiffman, Potential

413 Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA
414 Damage in Humans. *JAMA* **314**, 1850 (2015).

415 14. A. J. Preston, A. Rogers, M. Sharp, G. Mitchell, C. Toruno, B. B. Barney, L. N. Donovan, J. Bly, R.
416 Kennington, E. Payne, A. Iovino, G. Furukawa, R. Robinson, B. Shamloo, M. Buccilli, R. Anders, S.
417 Eckstein, E. A. Fedak, T. Wright, C. C. Maley, W. K. Kiso, D. Schmitt, D. Malkin, J. D. Schiffman, L.
418 M. Abegglen, Elephant TP53-RETROGENE 9 induces transcription-independent apoptosis at the
419 mitochondria. *Cell Death Discov* **9**, 66 (2023).

420 15. T. van der Valk, M. Dehasque, J. C. Chacón-Duque, N. Oskolkov, S. Vartanyan, P. D. Heintzman, P.
421 Pečnerová, D. Díez-del-Molino, L. Dalén, Evolutionary consequences of genomic deletions and
422 insertions in the woolly mammoth genome. *iScience* **25**, 104826 (2022).

423 16. S. Endoh-Yamagami, K. M. Karkar, S. R. May, I. Cobos, M. T. Thwin, J. E. Long, A. M. Ashique, K.
424 Zarbalis, J. L. R. Rubenstein, A. S. Peterson, A mutation in the pericentrin gene causes abnormal
425 interneuron migration to the olfactory bulb in mice. *Dev Biol* **340**, 41–53 (2010).

426 17. R. A. Denu, S. Nemcek, D. D. Bloom, A. D. Goodrich, J. Kim, D. F. Mosher, P. Hematti, Fibroblasts
427 and Mesenchymal Stromal/Stem Cells Are Phenotypically Indistinguishable. *Acta Haematol* **136**,
428 85–97 (2016).

429 18. K. Närhi, M. Tummers, L. Ahtiainen, N. Itoh, I. Thesleff, M. L. Mikkola, Sostdc1 defines the size
430 and number of skin appendage placodes. *Dev Biol* **364**, 149–161 (2012).

431 19. S.-Y. Yoon, M. Detmar, Sostdc1 Secreted from Cutaneous Lymphatic Vessels Acts as a Paracrine
432 Factor for Hair Follicle Growth. *Curr Issues Mol Biol* **44**, 2167–2174 (2022).

433 20. L. L. Cunningham, J. I. Matsui, M. E. Warchol, E. W. Rubel, Overexpression of *Bcl-2* prevents
434 neomycin-induced hair cell death and caspase-9 activation in the adult mouse utricle *in vitro*. *J*
435 *Neurobiol* **60**, 89–100 (2004).

436 21. T. Schlake, Segmental Igfbp5 expression is specifically associated with the bent structure of zigzag
437 hairs. *Mech Dev* **122**, 988–997 (2005).

438 22. D. Khnykin, J. H. Miner, F. Jahnsen, Role of fatty acid transporters in epidermis.
439 *Dermatoendocrinol* **3**, 53–61 (2011).

440 23. Y. Owada, I. Suzuki, R. Suzuki, H. Kondo, H. Takano, H. Yamanaka, H. Kobayashi, Y. Sugitani, Y.
441 Tomioka, T. Terui, M. Mizugaki, H. Tagami, T. Noda, Altered Water Barrier Function in Epidermal-
442 Type Fatty Acid Binding Protein-Deficient Mice. *Journal of Investigative Dermatology* **118**, 430–
443 435 (2002).

444 24. K. Vanderwolf, C. Kyle, C. Davy, A review of sebum in mammals in relation to skin diseases, skin
445 function, and the skin microbiome. *PeerJ* **11**, e16680 (2023).

446 25. V. E. Repin, O. S. Taranov, E. I. Ryabchikova, A. N. Tikhonov, V. G. Pugachev, Sebaceous Glands of
447 the Woolly Mammoth, Mammothus primigenius Blum.: Histological Evidence. *Doklady Biological*
448 *Sciences* **398**, 382–384 (2004).

449 26. T. G. Martin, L. A. Leinwand, Molecular regulation of reversible cardiac remodeling: lessons from
450 species with extreme physiological adaptations. *Journal of Experimental Biology* **227** (2024).

451 27. A. N. Keen, J. M. Klaiman, H. A. Shiels, T. E. Gillis, Temperature-induced cardiac remodeling in
452 fish. *Journal of Experimental Biology* **220**, 147–160 (2016).

453 28. C. A. Salsberg, Resurrecting the woolly mammoth: science, law, ethics, politics, and religion. *Stan.*
454 *Tech. L. Rev.* (2000).

455 29. J. Grego-Bessa, P. Gómez-Apiñaniz, B. Prados, M. J. Gómez, D. MacGrogan, J. L. de la Pompa,
456 Nrg1 Regulates Cardiomyocyte Migration and Cell Cycle in Ventricular Development. *Circ Res*
457 **133**, 927–943 (2023).

458 30. H. Ushikoshi, T. Takahashi, X. Chen, N. C. Khai, M. Esaki, K. Goto, G. Takemura, R. Maruyama, S.
459 Minatoguchi, T. Fujiwara, S. Nagano, K. Yuge, T. Kawai, Y. Murofushi, H. Fujiwara, K. Kosai, Local
460 overexpression of HB-EGF exacerbates remodeling following myocardial infarction by activating
461 noncardiomyocytes. *Laboratory Investigation* **85**, 862–873 (2005).

462 31. P. Cordero, V. N. Parikh, E. T. Chin, A. Erbilgin, M. J. Gloudemans, C. Shang, Y. Huang, A. C. Chang,
463 K. S. Smith, F. Dewey, K. Zaleta, M. Morley, J. Brandimarto, N. Glazer, D. Waggott, A. Pavlovic, M.
464 Zhao, C. S. Moravec, W. H. W. Tang, J. Skreen, C. Malloy, S. Hannenhalli, H. Li, S. Ritter, M. Li, D.
465 Bernstein, A. Connolly, H. Hakonarson, A. J. Lusis, K. B. Margulies, A. A. Depaoli-Roach, S. B.
466 Montgomery, M. T. Wheeler, T. Cappola, E. A. Ashley, Pathologic gene network rewiring
467 implicates PPP1R3A as a central regulator in pressure overload heart failure. *Nat Commun* **10**,
468 2760 (2019).

469 32. M. Krumbein, F. Oberman, Y. Cinnamon, M. Golomb, D. May, G. Vainer, V. Belzer, K. Meir, I.
470 Fridman, J. Haybaeck, G. Poelzl, I. Kehat, R. Beeri, S. M. Kessler, J. K. Yisraeli, RNA binding protein
471 IGF2BP2 expression is induced by stress in the heart and mediates dilated cardiomyopathy.
472 *Commun Biol* **6**, 1229 (2023).

473 33. J. Wang, M. Hoshijima, J. Lam, Z. Zhou, A. Jokiel, N. D. Dalton, K. Hultenby, P. Ruiz-Lozano, J. Ross,
474 K. Tryggvason, K. R. Chien, Cardiomyopathy Associated with Microcirculation Dysfunction in
475 Laminin α4 Chain-deficient Mice. *Journal of Biological Chemistry* **281**, 213–220 (2006).

476 34. Z. Jiang, G. Yang, G. Wang, J. Wan, Y. Zhang, W. Song, H. Zhang, J. Ni, H. Zhang, M. Luo, K. Wang,
477 B. Peng, SEC14L3 knockdown inhibited clear cell renal cell carcinoma proliferation, metastasis
478 and sunitinib resistance through an SEC14L3/RPS3/NFκB positive feedback loop. *Journal of*
479 *Experimental & Clinical Cancer Research* **43**, 288 (2024).

480 35. J. Thyboll, J. Kortesmaa, R. Cao, R. Soininen, L. Wang, A. Iivanainen, L. Sorokin, M. Risling, Y. Cao,
481 K. Tryggvason, Deletion of the Laminin α4 Chain Leads to Impaired Microvessel Maturation. *Mol*
482 *Cell Biol* **22**, 1194–1202 (2002).

483 36. G. Wade, A. McGahee, J. M. Ntambi, J. Simcox, Lipid Transport in Brown Adipocyte
484 Thermogenesis. *Front Physiol* **12** (2021).

485 37. J. Eguchi, X. Wang, S. Yu, E. E. Kershaw, P. C. Chiu, J. Dushay, J. L. Estall, U. Klein, E. Maratos-Flier,
486 E. D. Rosen, Transcriptional Control of Adipose Lipid Handling by IRF4. *Cell Metab* **13**, 249–259
487 (2011).

488 38. N. Dai, L. Zhao, D. Wrighting, D. Krämer, A. Majithia, Y. Wang, V. Cracan, D. Borges-Rivera, V. K.
489 Mootha, M. Nahrendorf, D. R. Thorburn, L. Minichiello, D. Altshuler, J. Avruch, IGF2BP2/IMP2-
490 Deficient Mice Resist Obesity through Enhanced Translation of Ucp1 mRNA and Other mRNAs
491 Encoding Mitochondrial Proteins. *Cell Metab* **21**, 609–621 (2015).

492 39. B. Song, D. Scheuner, D. Ron, S. Pennathur, R. J. Kaufman, Chop deletion reduces oxidative stress,
493 improves β cell function, and promotes cell survival in multiple mouse models of diabetes.
494 *Journal of Clinical Investigation* **118**, 3378–3389 (2008).

495 40. B. Wang, M. Zhang, Y. Ni, F. Liu, H. Fan, L. Fei, X. Pan, M. Guo, R. Chen, X. Guo, Identification and
496 characterization of NYGGF4, a novel gene containing a phosphotyrosine-binding (PTB) domain
497 that stimulates 3T3-L1 preadipocytes proliferation. *Gene* **379**, 132–140 (2006).

498 41. Y. Zhang, A. J. Federation, S. Kim, J. P. O'Keefe, M. Lun, D. Xiang, J. D. Brown, M. L. Steinhauser,
499 Targeting nuclear receptor NR4A1–dependent adipocyte progenitor quiescence promotes
500 metabolic adaptation to obesity. *Journal of Clinical Investigation* **128**, 4898–4911 (2018).

501 42. S.-S. Wang, H.-Y. Huang, S.-Z. Chen, X. Li, Y. Liu, W.-T. Zhang, Q.-Q. Tang, Early growth response 2
502 (Egr2) plays opposing roles in committing C3H10T1/2 stem cells to adipocytes and smooth
503 muscle-like cells. *Int J Biochem Cell Biol* **45**, 1825–1832 (2013).

504 43. M. T. N. Tran, M. Hamada, M. Nakamura, H. Jeon, R. Kamei, Y. Tsunakawa, K. Kulathunga, Y. Lin,
505 K. Fujisawa, T. Kudo, S. Takahashi, MafB deficiency accelerates the development of obesity in
506 mice. *FEBS Open Bio* **6**, 540–547 (2016).

507 44. Z. Xiao, Y. Chu, W. Qin, IGFBP5 modulates lipid metabolism and insulin sensitivity through
508 activating AMPK pathway in non-alcoholic fatty liver disease. *Life Sci* **256**, 117997 (2020).

509 45. N. Herijgers, M. P. J. de Winther, M. Van Eck, L. M. Havekes, M. H. Hofker, P. M. Hoogerbrugge, T.
510 J. C. Van Berkel, Effect of human scavenger receptor class A overexpression in bone marrow-
511 derived cells on lipoprotein metabolism and atherosclerosis in low density lipoprotein receptor
512 knockout mice. *J Lipid Res* **41**, 1402–1409 (2000).

513 46. E. M. Thompson, A. W. Stoker, A Review of DUSP26: Structure, Regulation and Relevance in
514 Human Disease. *Int J Mol Sci* **22**, 776 (2021).

515 47. Y.-H. Hung, A. L. Carreiro, K. K. Buhman, Dgat1 and Dgat2 regulate enterocyte triacylglycerol
516 distribution and alter proteins associated with cytoplasmic lipid droplets in response to dietary
517 fat. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1862**, 600–614
518 (2017).

519 48. J. L. Betters, L. Yu, NPC1L1 and cholesterol transport. *FEBS Lett* **584**, 2740–2747 (2010).

520 49. W. Cai, W. Zhang, Q. Zheng, C. C. Hor, T. Pan, M. Fatima, X. Dong, B. Duan, X. Z. S. Xu, The kainate
521 receptor GluK2 mediates cold sensing in mice. *Nat Neurosci* **27**, 679–688 (2024).

522 50. J. A. Esselstyn, C. H. Oliveros, M. T. Swanson, B. C. Faircloth, Investigating Difficult Nodes in the
523 Placental Mammal Tree with Expanded Taxon Sampling and Thousands of Ultraconserved
524 Elements. *Genome Biol Evol* **9**, 2308–2321 (2017).

525 51. K. A. Marzec, M. Z. Lin, J. L. Martin, R. C. Baxter, Involvement of p53 in insulin-like growth factor
526 binding protein-3 regulation in the breast cancer cell response to DNA damage. *Oncotarget* **6**,
527 26583–26598 (2015).

528 52. A. Grimberg, C. M. Coleman, T. F. Burns, B. P. Himelstein, C. J. Koch, P. Cohen, W. S. El-Deiry, p53-
529 Dependent and p53-Independent Induction of Insulin-Like Growth Factor Binding Protein-3 by
530 Deoxyribonucleic Acid Damage and Hypoxia. *J Clin Endocrinol Metab* **90**, 3568–3574 (2005).

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

532 54. H. Hermeking, C. Lengauer, K. Polyak, T.-C. He, L. Zhang, S. Thiagalingam, K. W. Kinzler, B.
533 Vogelstein, 14-3-3σ Is a p53-Regulated Inhibitor of G2/M Progression. *Mol Cell* **1**, 3–11 (1997).

534 55. R. Huang, I. Grishagin, Y. Wang, T. Zhao, J. Greene, J. C. Obenauer, D. Ngan, D.-T. Nguyen, R.
535 Guha, A. Jadhav, N. Southall, A. Simeonov, C. P. Austin, The NCATS BioPlanet – An Integrated
536 Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems
537 Biology, and Chemical Genomics. *Front Pharmacol* **10** (2019).

538 56. G. Dey, A. Radhakrishnan, N. Syed, J. K. Thomas, A. Nadig, K. Srikumar, P. P. Mathur, A. Pandey,
539 S.-K. Lin, R. Raju, T. S. K. Prasad, Signaling network of Oncostatin M pathway. *J Cell Commun*
540 *Signal* **7**, 103–108 (2013).

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

543 58. W. Du, J. Searle, The Rb Pathway and Cancer Therapeutics. *Curr Drug Targets* **10**, 581–589
544 (2009).

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

547 60. S. Guha, S. R. Bhaumik, Transcription-coupled DNA double-strand break repair. *DNA Repair*
548 *(Amst)* **109**, 103211 (2022).

549 61. A. R. Sonawane, J. Platig, M. Fagny, C.-Y. Chen, J. N. Paulson, C. M. Lopes-Ramos, D. L. DeMeo, J.
550 Quackenbush, K. Glass, M. L. Kuijjer, Understanding Tissue-Specific Gene Regulation. *Cell Rep* **21**,
551 1077–1088 (2017).

552 62. M. Tollis, E. Ferris, M. S. Campbell, V. K. Harris, S. M. Rupp, T. M. Harrison, W. K. Kiso, D. L.
553 Schmitt, M. M. Garner, C. A. Aktipis, C. C. Maley, A. M. Boddy, M. Yandell, C. Gregg, J. D.
554 Schiffman, L. M. Abegglen, Elephant Genomes Reveal Accelerated Evolution in Mechanisms
555 Underlying Disease Defenses. *Mol Biol Evol* **38**, 3606–3620 (2021).

556 63. Z. Li, Y. Li, Q. Zhang, W. Ge, Y. Zhang, X. Zhao, J. Hu, L. Yuan, W. Zhang, Establishment of Bactrian
557 Camel Induced Pluripotent Stem Cells and Prediction of Their Unique Pluripotency Genes. *Int J*
558 *Mol Sci* **24**, 1917 (2023).

559 64. E. Appleton, H. Kyunghee, C. Rodríguez-Caycedo, Y. Tanaka, A. Ashkenazy-Titelman, K. Bhide, C.
560 Rasmussen-Ivey, X. Ambriz-Peña, N. Korover, H. Bai, A. Quieroz, J. Nelson, G. Rathod, G. Knox, M.
561 Morgan, N. Malviya, K. Zhang, B. Mcnutt, J. Kehler, A. Kowalczyk, A. Bow, B. Mclendon, B.
562 Cantarel, M. James, C. E. Mason, C. Gray, K. R. Koehler, V. Pearson, B. Lamm, G. Church, E. Hysolli,
563 Derivation of elephant induced pluripotent stem cells. *bioRxiv*, doi: 10.1101/2024.03.05.583606
564 (2024).

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

567 66. B. Bushnell, BBMap (2022). http://sourceforge.net/projects/bbmap/.

568 67. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform.
569 *Bioinformatics* **25**, 1754–1760 (2009).

570 68. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R.
571 Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

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

574 70. T. Fukuda, Y. Iino, M. Onuma, B. Gen, M. Inoue-Murayama, T. Kiyono, Expression of human cell
575 cycle regulators in the primary cell line of the African savannah elephant (loxodonta africana)
576 increases proliferation until senescence, but does not induce immortalization. *In Vitro Cell Dev*
577 *Biol Anim* **52**, 20–26 (2016).

578 71. V. R. Pearson, J. B. Bosse, O. O. Koyuncu, J. Scherer, C. Toruno, R. Robinson, L. M. Abegglen, J. D.
579 Schiffman, L. W. Enquist, G. F. Rall, Identification of African Elephant Polyomavirus in wild
580 elephants and the creation of a vector expressing its viral tumor antigens to transform elephant
581 primary cells. *PLoS One* **16**, e0244334 (2021).

582 72. B. E. Housden, A. J. Valvezan, C. Kelley, R. Sopko, Y. Hu, C. Roesel, S. Lin, M. Buckner, R. Tao, B.
583 Yilmazel, S. E. Mohr, B. D. Manning, N. Perrimon, Identification of potential drug targets for
584 tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi.
585 *Sci Signal* **8** (2015).

586 73. R. Munroe, J. Schimenti, "Mutagenesis of Mouse Embryonic Stem Cells with
587 Ethylmethanesulfonate" (2009; http://link.springer.com/10.1007/978-1-59745-471-1_7), pp.
588 131–138.

589 74. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin,
590 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

591 75. S. Neph, M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman, A. K. Johnson, E. Rynes, M. T.
592 Maurano, J. Vierstra, S. Thomas, R. Sandstrom, R. Humbert, J. A. Stamatoyannopoulos, BEDOPS:
593 high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).

594 76. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features.
595 *Bioinformatics* **26**, 841–842 (2010).

596 77. G. Tan, B. Lenhard, TFBSTools: an R/bioconductor package for transcription factor binding site
597 analysis. *Bioinformatics* **32**, 1555–1556 (2016).

598 78. H. Schmidt, M. Zhang, H. Mourelatos, F. J. Sanchez-Rivera, S. W. Lowe, A. Ventura, C. S. Leslie, Y.
599 Pritykin, D. H. Koch, Genome-wide CRISPR guide RNA design and specificity analysis with
600 GuideScan2. doi: 10.1101/2022.05.02.490368 (2022).

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

603 80. A. Garcia-Moreno, R. López-Domínguez, J. A. Villatoro-García, A. Ramirez-Mena, E. Aparicio-
604 Puerta, M. Hackenberg, A. Pascual-Montano, P. Carmona-Saez, Functional Enrichment Analysis of
605 Regulatory Elements. *Biomedicines* **10**, 590 (2022).

606 81. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R.
607 Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

608 82. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin,
609 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

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

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

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

616

Acknowledgments

We would like to thank the San Diego Frozen Zoo for sharing the fibroblast cell line. We would like to thank Zhengkuan Tang for discussions on RNA-seq experimental setup and analysis, and Ana Luiza Lemos Queiroz and Ramiro Perrota for assistance during early pilot experiments. We would also like to thank George Chao for guidance on nanopore analysis, and Sergiy Velychko for help during troubleshooting of the fibroblast cell culture techniques. Lastly, we would like to thank Roman Teo Oliynyk and Chun-Ting Wu, as well as the rest of the Church lab for helpful discussions throughout this project. **Funding**: EK and AAT were supported by a structured research agreement between Harvard Medical School and Colossal Biosciences. **Author contributions**: EK and GMC conceived the study with feedback from EM and AAT. EK and NB designed and conducted all wet-lab experiments with feedback and help from EM, AAT, and LL. EK and LL completed all bioinformatic analyses. AAT provided additional RNA-seq data for epithelial cell lines. EK wrote the first draft of the manuscript. All authors helped in the revision of subsequent drafts. **Competing interests**: For a complete list of G.M.C.'s financial interests, please visit http://arep.med.harvard.edu/gmc/tech.html. All other authors declare no other conflicts of interest. **Data and materials availability**: Bulk RNA-seq data for all mammoth deletions as well as TP53 conditions with and without mitomycin was deposited on NCBI (BioProject Accessions: PRJNA1147676 and PRJNA1217465). Bioinformatic scripts and tools created for this work are available at: https://github.com/EmilKarpinski/RTG_TP53_MammothDel_Analysis.

Supplementary Materials

Materials and Methods

Supplementary Tables 1 to 7

Figs. S1 to S16

Table S1

References *(65-85)*

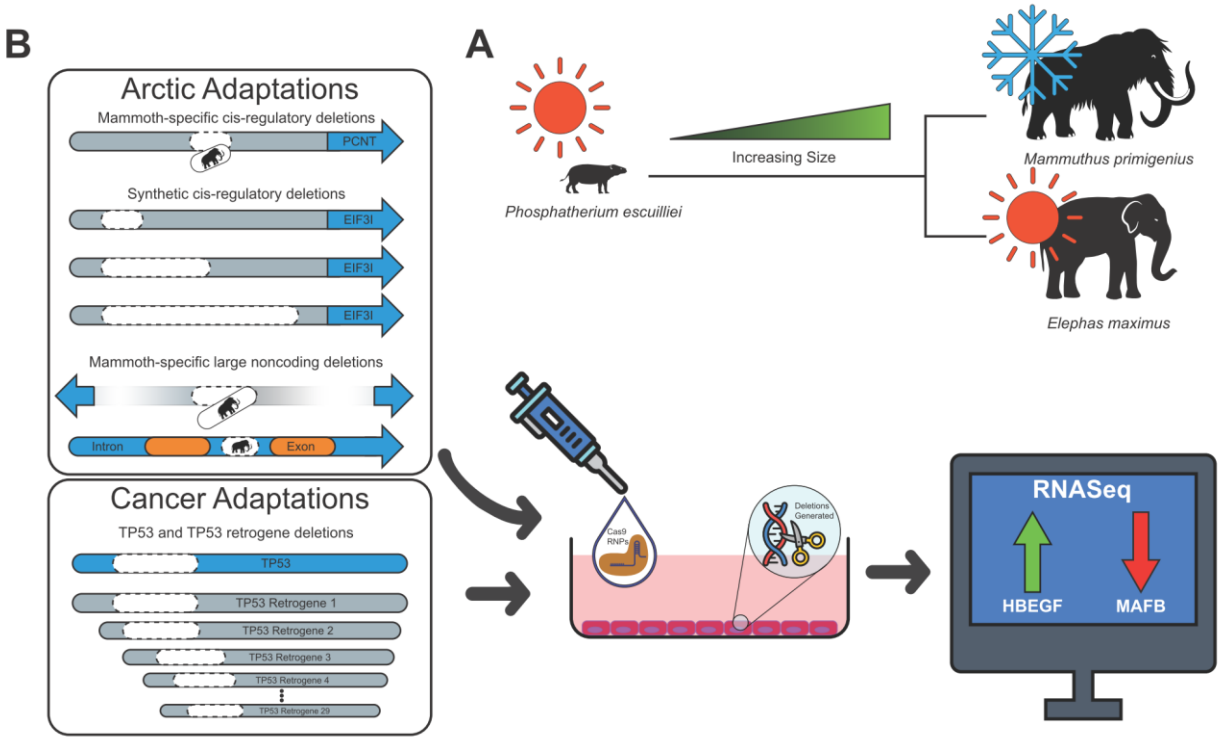

Figure 1. Experimental Design. (A) Trends in proboscidean evolution include a general increase in body size, and adaptations to various environments including warm tropical grasslands and cold arctic tundra. **(B)** We engineered mammoth specific or synthetic cis-regulatory deletions, large mammoth-specific deletions in both intra and intergenic regions, as well as deletions within TP53 and/or all TP53 RTGs using Cas9 RNPs within various Asian elephant cells in culture, before examining their effects via RNA-seq.

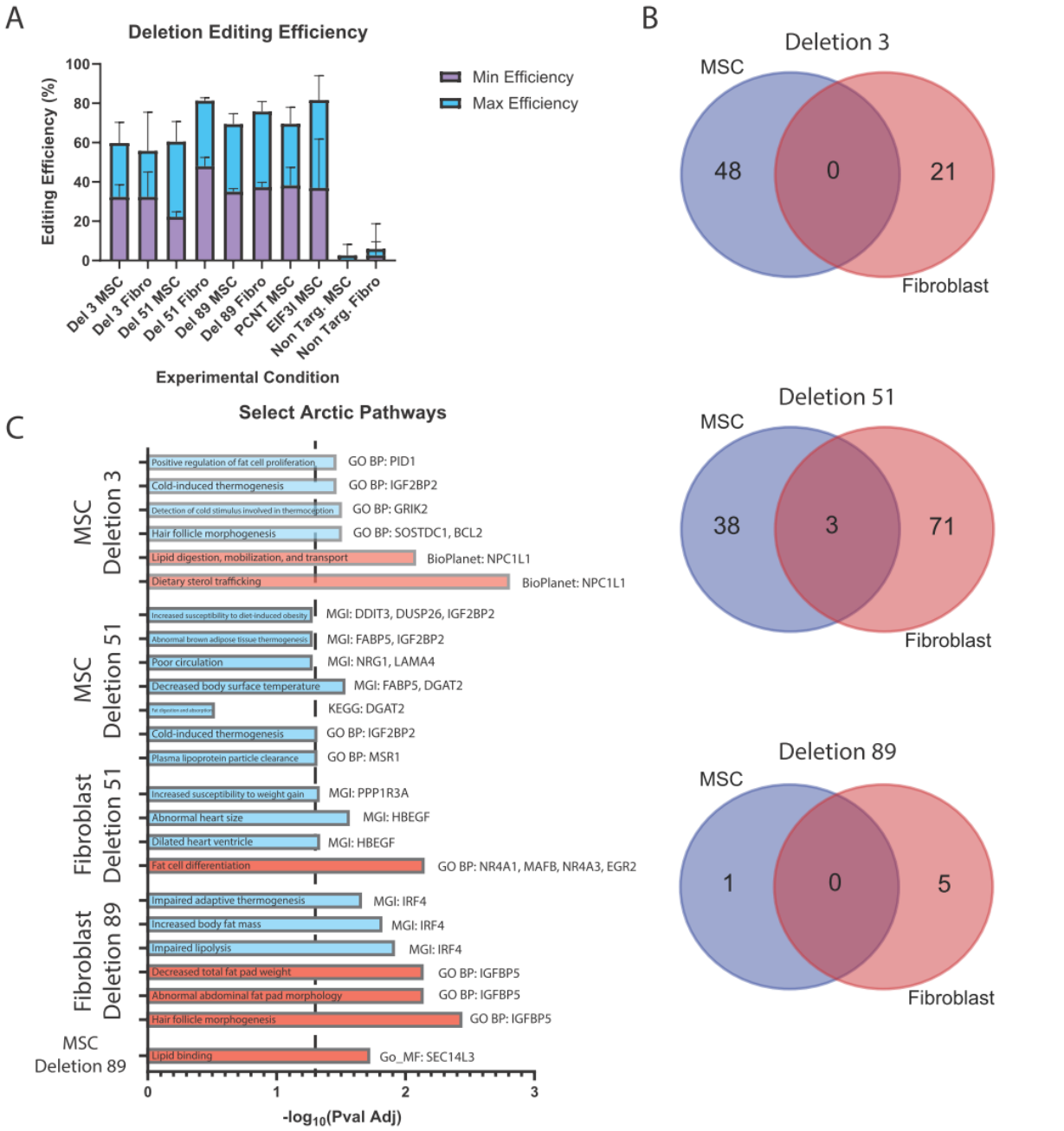

Figure 2. Mammoth deletion efficiency and effects. (A) Editing efficiency observed for each deletion in both MSCs and fibroblasts estimated using gel electrophoresis. **(B)** DEG overlap between MSCs and fibroblasts for each of the three large mammoth-specific noncoding deletions. **(C)** A selection of pathways potentially related to arctic adaptation. Bars are colored blue/red for upregulated and downregulated genes respectively. The database and key genes are listed beside each bar. A vertical dashed line indicates a <0.05 significance level.

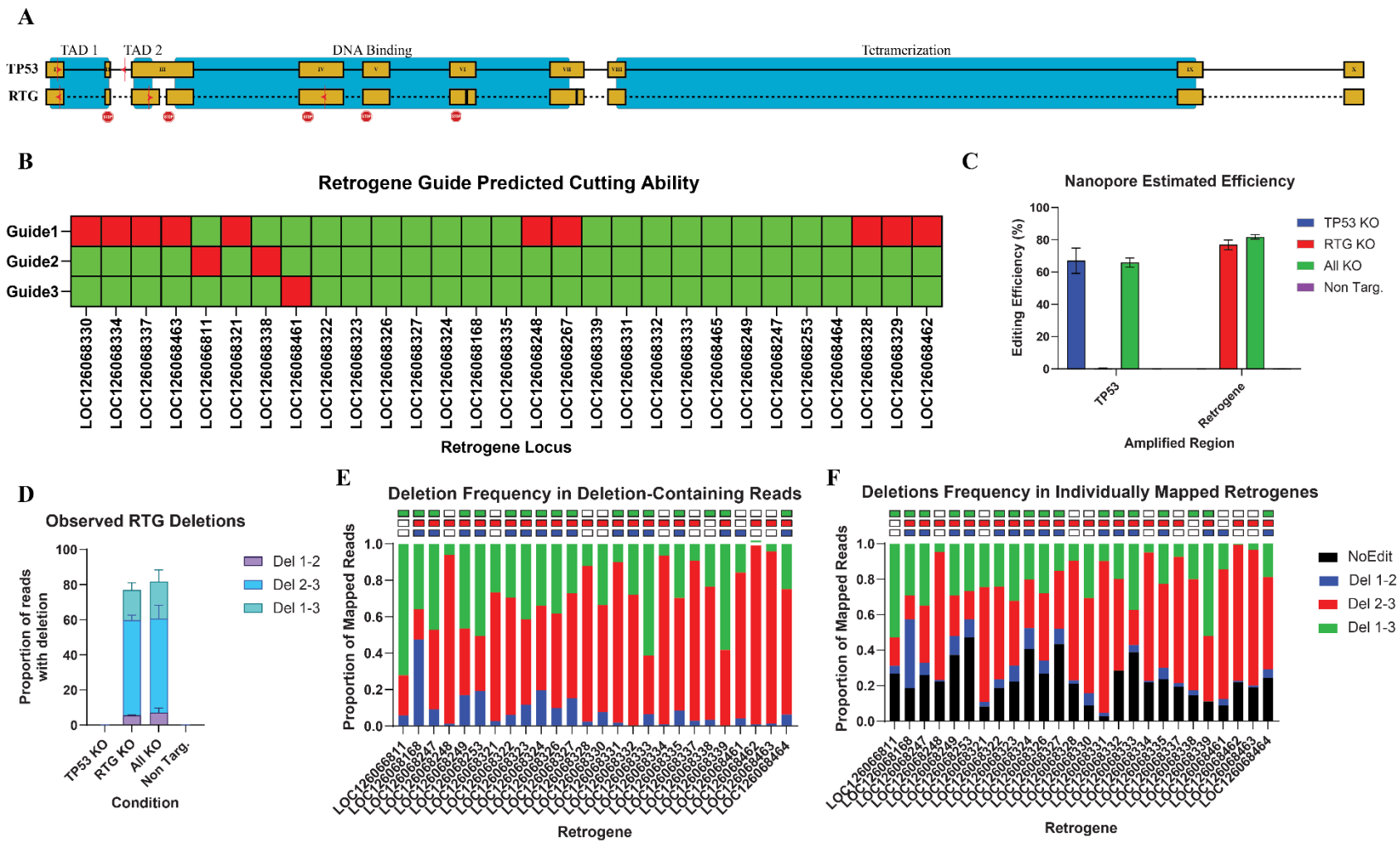

Figure 3. TP53 Editing. (A) Schematic showing an alignment of Asian elephant TP53 and a retrogene (RTG) consensus, with domains labelled based on homology to human TP53. Exons are shown in yellow, with introns indicated by black lines. Missing sequence is indicated with a dashed line in the RTGs. Additionally, 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 the retrogenes. **(B)** Predicted guide activity against each of the 29 Asian elephant retrogenes, with green indicating the guide is expected to be

662 functional against that target and red indicating non-functionality. **(C)** Observed editing efficiency of the expected deletions based on nanopore
663 data. (**D)** The frequency of deletions observed in all reads mapped against the RTG consensus across different treatment conditions. **E.** The
664 frequency of each deletion across the individual Asian elephant retrogenes as a proportion of all deletion-containing reads mapped to that locus,
665 or all mapped read **(F)**. Amplicons which were identical across the non-edited regions were collapsed as they are bioinformatically
666 undistinguishable. Error bars represent standard deviation in all plots.

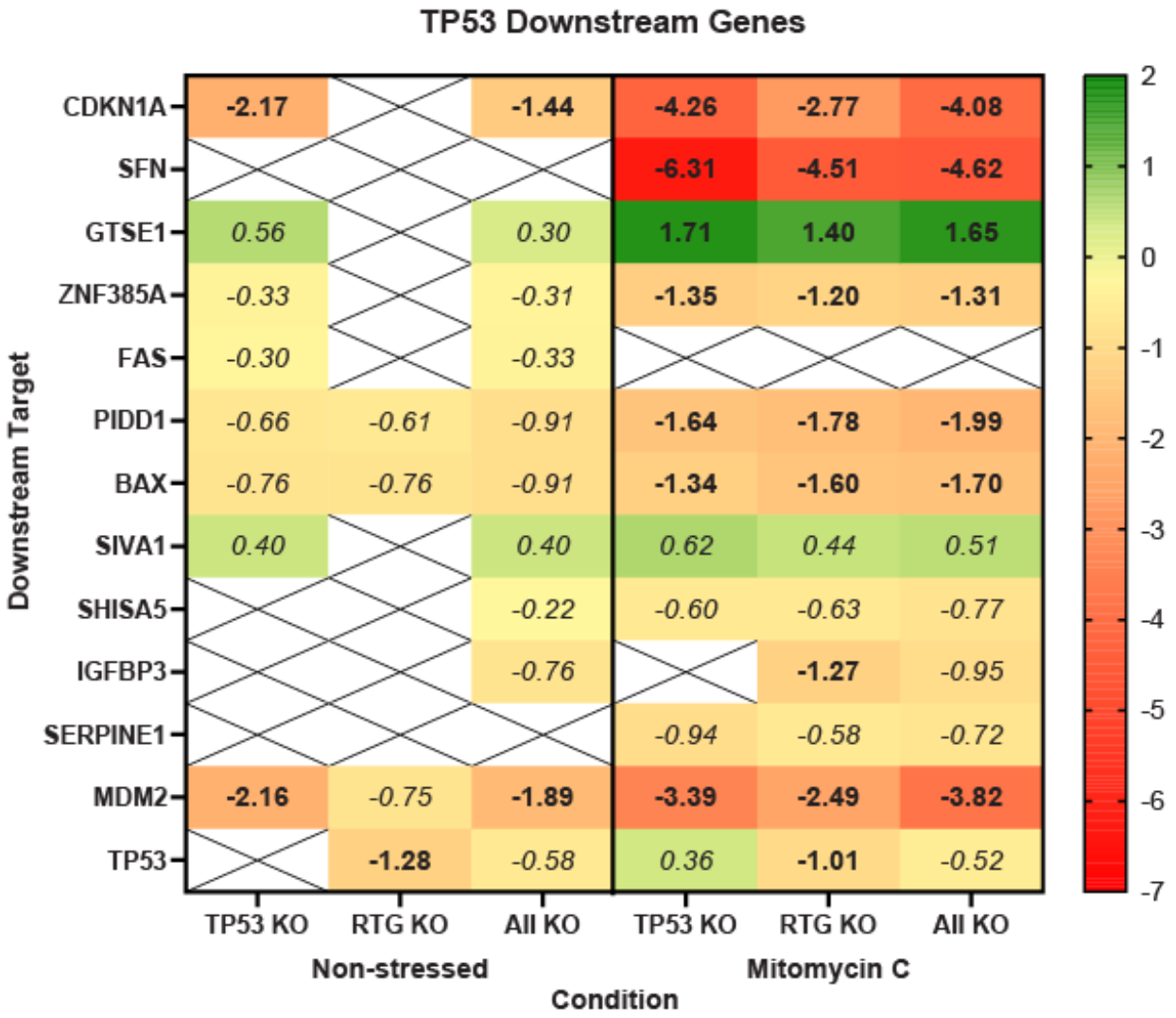

Figure 4. Expression of TP53 pathway genes. Log2 fold changes observed for genes immediately downstream of TP53 in the p53 signaling pathway (KEGG pathway hsa04115). Crosses indicate that no significant difference in expression was observed for that gene in a given condition. Genes in italics have an 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.

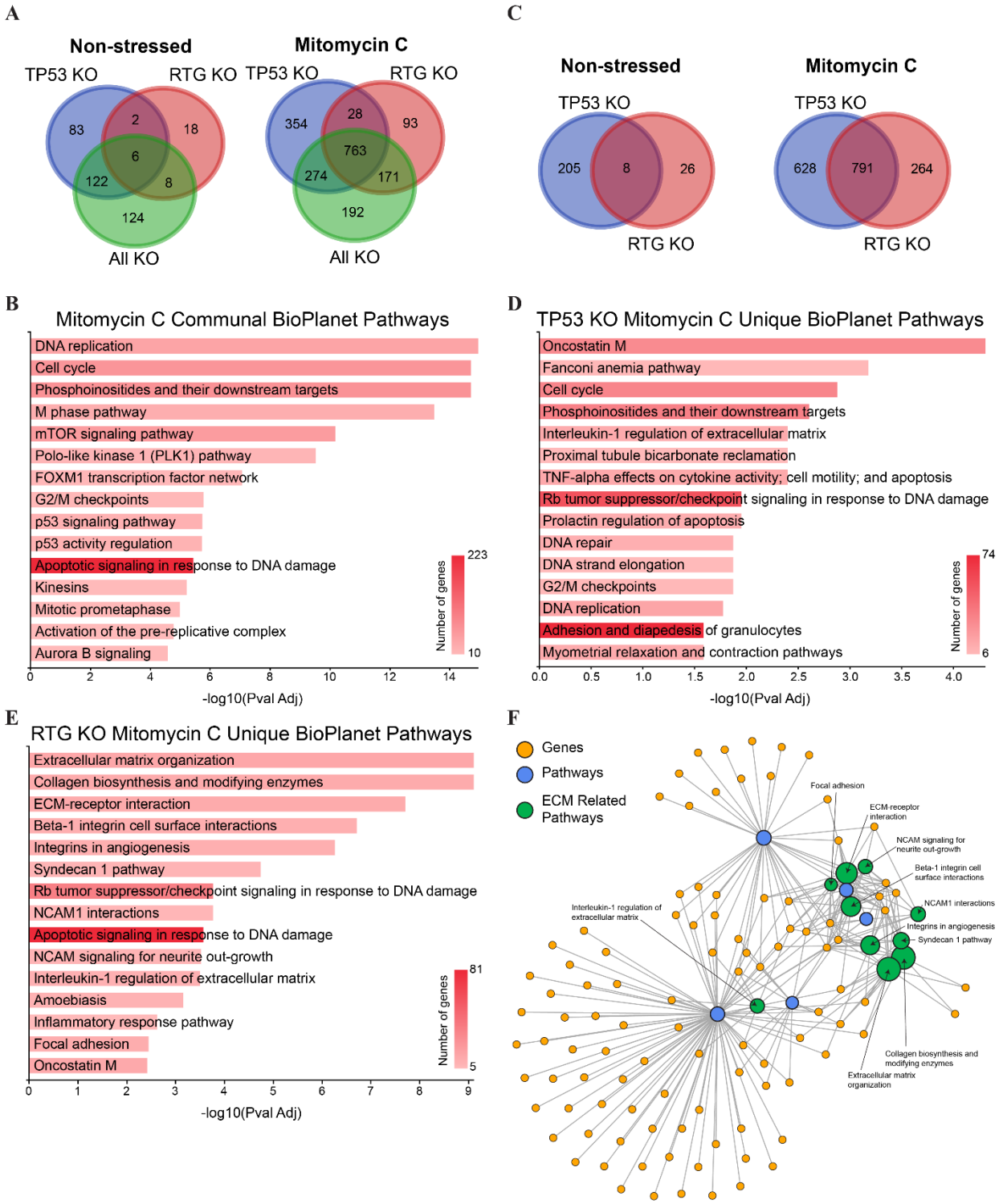

Figure 5. Pathway enrichment. (A) Differentially expressed gene (DEG) overlap observed between treatment conditions for cells in normal media or exposed to 10µM mitomycin C. Genes were filtered for a minimum absolute log2 fold change greater than 1. **(B)** BioPlanet pathway enrichment for the 763 DEGs 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

679 enrichment of unique DEGs in TP53 KO Mitomycin C treated cells. **(E)** BioPlanet pathway enrichment of
680 unique DEGs in RTG KO Mitomycin C treated cells. **(F)** Pathway/gene cluster analysis for DEGs in the RTG
681 KO Mitomycin C treated cells. ECM related pathways are colored green and labelled. Pathway clusters
682 are scaled by their p-value.

Methods

Elephant Genomic Dataset

All bioinformatics analyses for deletion identification used previously published datasets. We used datasets for eleven mammoths (*Mammuthus*), thirteen Asian elephants (*Elephas*), and twenty-two African elephants (*Loxodonta*) (Supplementary Table 1). Published datasets were aligned to the *Elephas maximus* reference genome (GCF_024166365.1) using BWA aln (*67*).

Mammoth Deletion Identification

A summary of the mammoth deletion pipeline is available in Fig. S1. Briefly, the *Elephas maximus* reference genome was split into individual contigs. These were then fed into GenerateWindows.awk (WinSize = 25; WinSlide = 10; AmbProp = 0.2), to generate windows across the genome and filter out any with more than 20% missing data (i.e. N's). We then generated read depth files for each mammoth/elephant alignment using samtools depth (-a to output all positions) (*74*), and split these by contig. Windows were then cross referenced with depth files using another CoverageCalc.awk to extract windows containing no mapped reads (StdevMultiplier=99), and adjacent windows concatenated using ConcatonateWindows.awk. Windows with zero coverage identified in *Loxodonta* and *Elephas* were then combined using bedops (*75*), then were used to filter the identified mammoth deletions using bedtools intersect (-v flag to retain windows which are uniquely in the mammoth dataset) (*76*).

We also attempted to predict transcription factor binding sites (TFBS) upstream of elephant genes. We first extracted the transcription start side and direction of each protein coding gene in the *Elephas maximus* reference genome, then generated intervals for 1kb upstream of each gene. These intervals were extracted using the samtools faidx program (*74*), and TFBS predicted using TFBStools (*77*) in custom Rscript (TFBSTools_JASPAR2014.R). Identified TFBS were then cross referenced against filtered mammoth deletions using bedtools intersect (*76*) to identify putative cis-regulatory deletions present in woolly mammoths.

During the course of this project, a separate ancient DNA deletion pipeline was published (*15*). We also performed deletion prediction using the published pipeline using the same datasets above. We examined the overlap between our two datasets using bedtools intersect (*76*).

TP53 identification and primer design

We extracted and aligned the canonical TP53 locus as well as all 29 retrogenes (RTGs) from the *Elephas maximus* reference genome (GCF_024166365.1). For the RTGs, we trimmed the alignment to include only the ~1.1kb segment overlapping the exons in TP53. We then designed two primers to specifically amplify the TP53 gene (Supplementary Table 2). For the RTGs, we designed seven forward and four reverse primer variations, all amplifying the same locus, but containing all possible binding site mutations to minimize amplification bias (Supplementary Table 2). These primers were made up to 10µM stock solutions and equal volumes of all forward and all reverse primers combined to produce master forward and reverse stock solutions.

For all other deletions, we generated at least four primers flanking the region containing the deletion ensuring similar melting temperatures across target sites, and that all primers were outside of the region targeted by Cas9. For each locus, we tested all possible combinations of forward and reverse primers and chose those that produced the cleanest and most reproducible amplifications (Supplementary Table 2).

Guide Design

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

To design guides for the RTG knockouts, we generated all possible guides against the trimmed RTG alignment for all 29 retrogenes using the generate_kmery.py script as before. We then screened all guides for those that were predicted to have activity to the most possible targets and identified two gRNAs that were each predicted to have activity against all but three retrogenes, which were only targetable by either one of the two guides (Supplementary Table 2; Fig. 3). We then subsequently chose one additional guide that targeted those three remaining retrogenes to ensure that each retrogene had at least two gRNAs.

We also identified three genes that had consistent expression across our MSC and epithelial cell line: FN1, EIF3I, and CD109. For each of these we designed all possible guides as above and filtered for off-targets against the *E. maximus* reference genome using GuideScan2 and filtered as above. We then selected four guides for each spaced roughly equally apart across a 1kb upstream region of each gene.

For all other deletions we extracted the *E. maximus* reference sequence ±25bp from the start or end point of each deletion using samtools faidx (*74*) and generated all possible guides within those regions using generate_kmery.py. We then filtered the automatically generated guides for off-targets using GuideScan2 and the python script as above. In the case multiple guides were possible for each deletion we chose one that would result in a double-stranded break closest to the original endpoint of each deletion.

Cell culturing and media

Fibroblast cells were grown in a custom media formulation with the final concentrations: 50% FGM-2 (Lonza FBM Basal Media (CC-3131) supplemented with FGM-2 SingleQuots supplements (CC-4126)); 20% heat-inactivated fetal bovine serum (ThermoFisher, A5670801); and 30% Mem Alpha (Gibco, 12571063). Mesenchymal stem cells were grown in a custom media formulation with final concentrations: 79% DMEM High Glucose + GlutaMax (ThermoFisher 10566016); 20% heat-inactivated fetal bovine serum (ThermoFisher, A5670801); and 1% MEM Non-Essential Amino Acids Solution (ThermoFisher 11140050). Epithelial cells were grown in a custom media formulation with final concentrations: EGM-2 (Lonza Basal Medium (CC-3156) supplemented with EGMTM-2 SingleQuotsTM Supplements (CC-4176)); 40% DMEM High Glucose + GlutaMax (ThermoFisher 10566016); and 10% heat-inactivated fetal bovine serum (ThermoFisher, A5670801). Cells were grown in an incubator at 37°C with 5% CO_2.

Cas9 RNP Electroporation

All gRNAs were ordered from Synthego and resuspended to 100µM in 1X TE buffer. Cas9 nuclease was ordered from IDT (catalog # 10007807). All nucleofections were conducted using the Lonza P4 Primary Cell 96-well Kit (catalog # V4SP-4096) using codes CA-137 (fibroblasts) or EN-150 (epithelial cells), or the P3 Primary Cell 96-well Kit (catalog # V4SP-3096) using code CA-137 (MSCs).

For the TFBS, synthetic, and large-mammoth deletions, approximately 100-200k cells were used for each 20µl transfection reaction, with a 3µM final concentration of each guide (i.e. both upstream and downstream of the deletion) and 1µM of Cas9 nuclease per guide (i.e. a 3:1 ratio of gRNA:Cas9).

RNPs were pre-complexed in nucleofection solution for 20 minutes at room temperature then electroporated into cells using the above codes.

We conducted pilot studies for the seven identified TFBS deletions upstream of genes that showed consistent expression in our MSC line, although only one guide was used for deletions upstream of LOC126074853 and LOC126068891 due to the small size of these deletions. We also conducted pilot studies using all three synthetic deletion genes (FN1, CD109, EIF3I), as well as for the six smallest intragenic and seven smallest intergenic large mammoth-specific deletions. Deletions were carried out alongside negative control containing just Cas9 and/or a non-targeting gRNA RNP which has no significant homology to the *E. maximus* genome. All experimental conditions for pilot work were conducted in duplicate and in MSCs. We subsequently chose a subset of these deletions which generated a high number of DEGs for our final analysis. These deletions were done as above, except with four replicates each and accompanied by 3-4 negative controls using a non-targeting gRNA. For Deletions 3, 51, and 89 we also attempted to introduce these deletions into fibroblast and epithelial cells, however the epithelial cells failed to recover from the nucleofections and were ultimately discarded.

TP53 deletions were done as above, except with 400-500k cells per transfection. Cells were transfected with either both TP53 gRNAs, all three RTG gRNAs, both TP53 and RTG gRNAs ("All" condition), or a non-targeting gRNA (Supplementary Table 1). For the All nucleofection, gRNAs were first combined and then concentrated using a Savant SpeedVac DNA 120 Concentrator (Thermo Scientific), before being resuspended in premade nucleofector and supplement solution. Additionally, due to the presence of possible TP53 mutations that accumulate with prolonging passaging, a second separate nucleofection was conducted as above prior to the Mitomycin C stressed RNA-seq. In both cases, all treatment conditions were conducted using three biological replicates. Due to their greater robustness, their ease of culture and use, and the availability of younger passage cells, we elected to conduct these experiments only in our fibroblast line.

Following nucleofection cells were immediately transferred into 6-well plates containing prewarmed media and allowed to recover for 4 days. Media was changed 24 hours post nucleofection and every 2-3 days thereafter.

Editing Efficiency

During the first passage (approximately 4 days post nucleofection), about 10% of the cells were collected and DNA extracted using the DNeasy Blood and Tissue Kit (Qiagen) using the manufacturer's

recommended protocol. For the TP53-editing extractions we amplified both the TP53 locus spanning the expected knockout deletion, and the retrogene loci. For all other extractions we amplified the locus spanning the targeted deletion and examined these loci across the negative controls as well. We preformed 25µL PCR reactions with final concentrations of 1X Phusion Hot Start Flex 2X Master Mix and 500nM of each primer, using between 1-3µL of template. Cycling conditions were: 98°C for 3 minutes; 12 cycles of 98°C for 15 seconds, then 69°C for 60 seconds; 25 cycles of 98°C for 15 seconds, 67°C for 30 seconds, then 72°C for 45 seconds; followed by a final extension at 72°C for 5 minutes. Resulting PCR amplicons were purified over MinElute PCR Purification columns (Qiagen) using the recommended protocol and visualized on E-gel EX 2% Agarose gels (Invitrogen) to confirm the presence or absence of any large deletions. We examined the relative fluorescence of our products using the Image Lab 6.1 (Bio-Rad). We then calculated a minimum editing efficiency (the relative fluorescence of our expected deleted product, relative to everything else in the lane) and a maximum editing efficiency (the relative fluorescence of all non-wild-type bands within each lane).

To confirm efficiency estimates and examine RTG deletion distribution in our TP53-editing experiments, we sent each of our TP53 and RTG amplicon pools for nanopore sequencing using the Premium PCR service at Primordium to get full length sequences of our DNA fragments. For each sample, raw reads were first filtered using a custom python script (FastqFilter.py) to between the maximum and minimum expected sizes +/- 10% respectively. Mapping was carried out using BBMap (*66*) using default parameters against either the TP53 amplicon, an RTG consensus amplicon, or a multi-fasta with 26 RTGs (RTG amplicons with identical sequences outside of the editing region collapsed).

To calculate editing efficiency, we used a custom python script (CigarDel.py) which uses a region file containing the deletion regions of interest, a minimum deletion proportion (i.e. the proportion of bases mapping to the deleted region that need to be missing), and the CIGAR string of each read to evaluate the presence or absence of various deletions in the data. For TP53 we analyzed the presence of the expected deletion if Cas9 induced a double-strand break at both sgRNA targeting sites. For the RTG we examined the presence of any of the three possible deletions between two guides. All analyses were carried out with a minimum deletion proportion of 0.9.

To examine the distribution of deletions at each locus we identified the names of reads containing deletions and extracted those that mapped to each locus. We then cross-referenced this with the deletions identified against the RTG consensus to obtain read distributions for reads mapped to all RTGs or the collapsed RTG loci.

Bulk RNA-seq

Approximately 300-500k cells were collected, flash frozen on dry ice, and shipped to Genewiz for bulk RNA sequencing within ten days of transfection (TP53) or 2-3 weeks (all other deletions). Expression of transcripts was done using Salmon (v1.8.0 (*69*)) using a custom index consisting of all mRNAs and non-coding RNAs annotated within the *E. maximus* reference genome (GCF_024166365.1), with all 29 TP53 retrogene sequences added. The salmon index was generated using the keepDuplicates flag to avoid collapsing any RTGs. Salmon quant files were analyzed using DESeq2 (*79*) as part of a custom Rscript to analyze samples in a pairwise fashion (i.e. one set of biological replicates against the controls at a time). Differentially expressed genes (DEGs) were identified with an adjusted p value of less than 0.05 (padj < 0.05). No minimum log2 fold change was applied for initial analysis of TP53 pathway genes. A minimum absolute log2 fold change of 1 was used for all other analyses. Venn diagrams and DEG overlap was generated using a web based custom venn diagram tool (https://bioinformatics.psb.ugent.be/webtools/Venn/), and pathway enrichment preformed using GeneCodis4 (*80*).

Retrogene expression in public datasets

To ensure the expression patterns we observed were not tissue/cell type specific, we examined the expression of our retrogenes in some preliminary shared data from mesenchymal stem cells and epithelial cells. These cells contained deletions generated using the same guides as in this study, but the epithelial cells were also subject to various small molecule treatments. We also downloaded public RNA-seq datasets for Asian elephant ovary (SRX15554240), salivary glands (SRX15554239), lymph nodes (SRX15554238), thyroid (SRX15554237), lung (SRX15554236), and lymphocytes (SRX1423033), and analyzed it as above.

Depth plots

To examine where reads from each of the three RTGs showing repeated expression were originating, we mapped the data for our datasets and the public datasets back to the Asian elephant reference genome using STAR (*81*). We then extracted read depths along the full annotated length of RTGs LOC126068247, LOC126068267, and LOC126068248 using samtools depth (*82*). For our dataset we averaged expression values within biological replicates and plotted the data in Microsoft Excel.

Maximum likelihood phylogeny

We download African elephant retrogene 9 (KF715863.1) to investigate a possible relationship between this retrogenes and the Asian elephant retrogenes. We then aligned the retrogene to the Asian elephant TP53 CDS and truncated the RTG 9 sequence to contain only the homologous sequence to TP53. We then aligned the truncated RTG 9 sequence, Asian elephant TP53, and all 29 Asian elephant retrogenes using muscle (v5.1) (*83*). We generated a maximum likelihood phylogeny using IQ-Tree (*84*), using ModelFinder to identify the best supported substitution model (*85*), and 100 bootstraps (-b 100).