Engineering elephant models of cold-adaptation and cancer resistance

Emil Karpinski* [a], Nikil Badey [b], Esther Mintzer [a], Asaf Ashkenazy-Titelman [a], Li Li [a,b], 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 – emil_karpinski@hms.harvard.edu
George Church – gchurch@genetics.hms.harvard.edu

Abstract

Throughout their evolution, proboscideans have gotten larger and spread to occupy novel environments throughout the world. These phenotypic changes have been accompanied by 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. Additionally, 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 role in the tumor microenvironment and mitigating metastatic growth. This highlights the usability of cell culture models to disentangle the effects of regulatory mutations or complex genetic systems, especially in non-model species where the generation of animal models is not possible.

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

1. Introduction

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

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

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

Peto's paradox is the observation that cancer incidence in a species does not always scale with body mass and/or lifespan [11,12]. 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 [12,13]. Genomic analyses have identified the duplication of many tumor suppressor genes across extant elephants, including multiple retrogene copies of TP53 [12,13]. 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

64 responses to DNA damage [12–14]. However, most previous studies have largely focused only on African
65 elephant TP53 retrogenes and often in non-native genetic (e.g. plasmid expression) or cellular
66 backgrounds (e.g. human or mouse cells).
67 In this study, we examine the regulatory changes responsible for adaptation and tumor
68 resistance in woolly mammoths and Asian elephants in vitro. We have engineered a handful of
69 mammoth-specific non-coding mutations into Asian elephant cells, and observed changes implicated in
70 cardiac development, hair and skin function, and lipid metabolism. We further validate our cell culture
71 models by examining the best-studied elephant genetic pathway – their expanded TP53 repertoire. We
72 used CRISPR-Cas9 to disrupt TP53, all TP53 retrogenes, or both in combination in Asian elephant
73 fibroblasts, and examined the transcriptional profiles of these knockouts with and without DNA damage
74 stress, to resolve the protective roles of TP53 retrogenes (Fig. 1). Our results suggest that these cell
75 culture models are tremendously useful for investigating the potential effects of both adaptive
76 regulatory changes and tumor suppressor duplications, and generate a strong foundation of testable
77 hypotheses for future research.
78
79 **2. Results**
80 *2.1 – Identification of mammoth deletions*
81 Our in silico TFBS annotation predicted ~1.98 million transcription factor binding sites (TFBS)
82 upstream of 21,808 genes (mean 90.6/gene). We originally identified ~235k mammoth-specific
83 deletions, with an average size of 236 bp (min: 25; max: 13.2kb), of which 1,798 intersected our
84 predicted transcription factor binding sites (TFBS; Supplementary Table 3). Using previously established
85 methods we also predicted 629 large mammoth-specific deletions (i.e. >=500 bp; Supplementary Table
86 4) of which 46 were predicted to have a genome-unique guide within 25bp of their start and end
87 coordinates [15]. Notably we find strong concordance between both methods, with deletions identified
88 using our own methodology intersecting 88.1% of the large mammoth deletions identified using the
89 established pipeline.
90
91 *2.2 – Mammoth-specific and Synthetic Regulatory Editing*
92 We attempted to introduce one mammoth specific deletion upstream of PCNT (pericentrin), and
93 one synthetic deletion upstream of EIF3I (eukaryotic translation initiation factor 3 subunit I) due to its
94 high expression in our mesenchymal stem cell (MSC) line. PCNT mutations have implicated in the
95 development of smaller cranial features and olfactory development [8,16], which may contribute to arctic

adaptations present in woolly mammoths. While we were successful in generating mutations in these regions (Fig 2A), we didn't recover differential expression of PCNT in either our pilot or verification experiments. We did however observe differential expression of EIF3I, but only with 50% of the guide combinations, one of which we successfully verified. Our lack of signal in PCNT and sometimes in EIF3I underscores the effect tissue or timing has on gene regulation, and may be why we fail to observe a signal in all cases.

2.3 – Mammoth-specific deletion differentially expressed genes

We also generated three large mammoth deletions (Deletions 3, 51, and 89) with comparable editing efficiency in both fibroblasts and mesenchymal stem cells (Fig. 2A). Following bulk RNA-seq we recovered a large number of differentially expressed genes (DEGs), although a significant portion of them had very small log2 fold changes. To focus our analysis on those that were likely to be most impacted we filtered our dataset to a minimum log2 fold change of 1 before further analysis. Interestingly, we see a high level of variability in the DEGs affected between cell types, even given the similarity of fibroblast and MSCs [17] (Fig. 2B), with the majority of DEGs being unique to each cell type. This is also true even if we remove the minimum log2 fold change (Fig. S2). This suggests that these deletions likely have cell-type specific effects and may result in very different phenotypes depending on the cell's specific context.

2.4 – Potential impact on mammoth phenotypes

We next wanted to see if we could infer some of the functions of these deletions, particularly as they may relate to arctic adaptations in woolly mammoths. To do this, we first screened our filtered DEGs to identify genes within pathways of interest (Fig. 2C), and then investigated what was known about perturbations in those genes in various cell and animal model analyses. We identified changes that broadly fall under three phenotypes: skin and hair adaptations; vascular adaptations; and metabolism and thermogenesis.

2.5 – Skin and Hair Adaptations

In comparison to their closest living relatives, Asian elephants, woolly mammoths have both denser hair coverage and multiple coats (layers of hair) to protect them from the arctic cold [8]. We find DEGs which may be related to skin and hair adaptions within each deletion, although not in every cell type – MSC Deletion 3 (SOSTDC1 and BCL2); MSC Deletion 51 (FABP5); and Fibroblast Deletion 89

128 (IGFBP5). SOSTDC1 is involved in the regulation of the number, size and/or shape of hair, and is
129 upregulated during the anagen phase in response to depilation-induced hair regeneration [18,19]. BCL2
130 prevents apoptosis, and its upregulation has been shown to protect against hair cell death in the utricle
131 [20]. Likewise, we observe downregulation of IGFBP5 in our fibroblast Deletion 89 cells, which is a strong
132 negative regulator of hair development and differentiation [21]. Previous studies have also found that
133 overexpression of IGFBP5 produces thinner and shorter hair with aberrant morphologies [21]. The
134 upregulation of SOSTDC1 of these in MSC Deletion 3, and downregulation of IGFBP5 in Fibroblast
135 Deletion 89, suggest these deletions may contribute to the abundance and robustness of the
136 mammoth's coats.
137 We also identify FABP5 upregulated in MSC Deletion 51, which is involved in regulating water
138 permeability. Knockouts of FABP5 result in much lower fatty acid content in the epidermis and impaired
139 water-barrier function and keratinocyte migration [22]. Likewise, this gene is expressed in sebaceous
140 epithelial cells where it plays a role in sebaceous wax production [23]. Sebaceous wax is believed to play a
141 role in waterproofing and water loss, as well as thermoregulation [8,24]. Mice which have been engineered
142 to lack or have nonfunctional sebaceous glands suffer from abnormal hair growth/loss, skin lesions, and
143 compromised thermoregulation [24]. Notably, histological analyses of the few mammoth soft tissue
144 remains that have preserved have revealed the presence of sebaceous cells, which are almost entirely
145 absent in extant elephants [8,25].
146
147 *2.6 – Vascular Adaptations*
148 Vascular adaptations are also frequently observed in many arctic species. One of the best
149 studied adaptations within this category is an increase in heart size, likely due to the need for more
150 blood flow to maintain temperature and oxygen homeostasis [26,27]. Although mammoth soft tissue has
151 poor taphonomy, there are claims that mammoth hearts were larger than similar warm-adapted
152 elephants of their age and size [28]. Correspondingly, we observe genes involved in heart remodeling and
153 angiogenesis in four experimental conditions – MSC Deletion 3 (IGF2BP2); MSC Deletion 51 (NRG1 and
154 LAMA4); and Fibroblast Deletion 51 (HBEGF and PPP1R3A). Previous research has implicated IGF2BP2,
155 NRG1, LAMA4, HBEGF, and PPP1R3A in heart morphology and metabolism, with the overexpression of
156 many of these genes resulting in heart remodeling and hypertrophy [29–33]. Additionally, IGFBP2 and
157 PPP1R3A have also been shown to play a role in metabolic changes that accompany abnormal heart
158 morphologies in mice [31,32].

159 Accompanying these changes, we also observe differential expression in genes related to
160 angiogenesis – MSC Deletion 51 (LAMA4); and Fibroblast Deletion 89 (SEC14L3). SEC14L3, in conjunction
161 with VEGF, has a role in promoting the migration of endothelial and vein progenitor cells and
162 consequently angiogenesis, and its overexpression has been implicated in the proliferation and
163 metastasis of tumors [34]. Likewise, LAMA4 appears to be highly involved in blood vessel development.
164 Mouse LAMA4 knockouts appear generally phenotypically normal, but exhibit frequent hemorrhages,
165 and in cornea angiogenesis assays developed grossly distorted blood vessel architectures [35].
166 Interestingly, while both genes are strong drivers of angiogenesis, we observe downregulation of
167 SEC14L3 and upregulation of LAMA4 in their respective experimental conditions. Although the general
168 trend here is still unclear, it's possible these deletions are implicated in fine tuning angiogenesis in
169 mammoths.
170
171 *2.7 – Metabolism and Thermogenesis*
172 The ability to sense and respond to cold temperatures via metabolic changes is probably
173 amongst the most important cold adaptations. There is a suite of metabolic cascades that are triggered
174 in response to cold nociception, and the ability to respond to cold stimuli via thermogenesis is highly
175 dependent on the regulation of both glucose and lipid metabolic reserves [36]. Not surprisingly, we
176 recover many differentially expressed genes involved in adipocyte proliferation and lipid and glucose
177 metabolism in four out of our six experimental conditions.
178 The bulk of our differentially expressed genes are involved in adipocyte differentiation and/or
179 lipid metabolism – MSC Deletion 3 (IGF2BP2, PID1, and NPC1L1); MSC Deletion 51 (IGF2BP2, MSR1,
180 DGAT2, DUSP26, and DDIT3); Fibroblast Deletion 51 (NR4A1, EGR2, and MAFB); and Fibroblast Deletion
181 89 (IRF4 and IGFBP5). We observe upregulation of IGF2BP2, PID1, DDIT3, and IRF4 in their respective
182 experimental conditions, all of which except DDIT3 are positive regulators of adipocyte
183 differentiation/proliferation [37–40]. Likewise, we observe downregulation of NR4A1, EGR2, and MAFB all
184 of which have previously been shown to act as negative regulators of adipocyte differentiation or size [41–
185 43]. Unsurprisingly, almost all these genes and additionally MSR1, DGAT2, NPC1L1, DUSP26, and IGFBP5
186 are also involved in lipid metabolism and controlling the rates of stored versus circulating lipids [44–48]. For
187 example, mice induced to overexpress DGAT2 do not appear to significantly different fat pads or body
188 weights relative to controls but exhibit much higher levels of circulating plasma triglycerides [47]. Likewise,
189 MAFB deficient mice have higher body weights and amounts of body, due to large adipocytes and higher
190 serum cholesterol levels [43].

191 Accordingly, many of the above genes have been shown to be implicated in the ability to sense
192 and respond to cold in mouse models. GRIK2 (MSC Deletion 3) knockouts are unable to sense cold, but
193 not cool, temperatures and have a sever reduction in cold-specific neurons [49]. IGF2BP2 knockout mice
194 have better thermoregulation than controls in the short term, but have fewer and smaller adipocytes
195 and adipocyte precursors, and have diminished UCP1 levels, a key thermogenic protein [38]. Interestingly,
196 IRF4 knockouts have been shown to be more obese than wild type mice but exhibit increased cold
197 intolerance likely due to the cold responsive nature of IRF4 [37].
198
199 *2.8 – The Asian elephant genome contains 29 TP53 retrogenes*
200 Although our RNA-seq results suggest that the deletions we engineered are likely involved in
201 woolly mammoth cold adaptations, many of these phenotypes would only be measurable in organismal
202 models. Towards this end we have primarily referenced literature on perturbations in these genes
203 within mouse models, though we note that elephants are quite diverged from all common mammalian
204 model species [50]. To both validate our cellular models, and expand our understanding of elephant cancer
205 resistance, we investigated the best-studied elephant genetic system – the expanded TP53 repertoire.
206 We searched the annotations of the Asian elephant reference genome and identified 29
207 retrogene copies of TP53 (RTGs). 27 of these RTGs are found on chromosome 27, while the remaining
208 two are found on chromosomes 24 and 26. By comparison, the canonical copy of TP53 is located on
209 chromosome 19. Most of the RTGs are approximately 1.1 kb in length, roughly the same length as the
210 CDS of TP53 as would be expected of retrogenes, except for three (LOC126068247 – 64.1 kb;
211 LOC126068267 – 32.5 kb; LOC126068248 – 4.9 kb; Fig. S3). When aligned and translated in the same
212 frame as TP53, each RTG carries a premature stop codon, which primarily cluster in five main areas
213 across the homology region (Fig. 3A).
214
215 *2.9 – TP53 and RTG editing and validation*
216 We attempted to design paired Cas9 guide RNAs (sgRNAs) to produce large, disruptive deletions
217 in TP53 and all the retrogenes. We managed to identify two sgRNAs that specifically target TP53, which
218 result in the partial loss of exon 1, the entirety of exon 2, and a premature stop codon after read
219 through into intron 2 (Fig. 3A). We could not identify just two guides that were both predicted to have
220 activity across all 29 retrogenes, so we designed a third guide to ensure each RTG was targetable by at
221 least two sgRNAs (Fig. 3B; Fig. S4).

222 We electroporated fibroblast cells with Cas9 RNPs consisting of either the two TP53 guides, the
223 three RTG guides, all five guides together (All), or a guide predicted to have no activity in the Asian
224 elephant genome (Non Targ.). After approximately four days, we then collected cells, extracted DNA,
225 and amplified a portion of TP53 and RTG genes flanking the expected editing sites. We observed high
226 levels of editing across all three conditions when we examined the resulting amplicons on an agarose gel
227 (Fig. S5). However, due to our PCR reaction amplifying all 29 loci simultaneously, we couldn't assess
228 which and how many of the RTGs were being edited.
229 To address this issue, we sent both PCR reactions for end-to-end nanopore sequencing. We
230 quantified the number of observed deletions, using a custom script which marks the presence of a
231 deletion if 90% of the bases within a defined region were absent. We note that this is likely an
232 underestimate of the true editing efficiency, as this will only identify the perfect deletion product and
233 miss any indels and/or incomplete loss of the flanked segment of DNA. We observed high editing
234 efficiency in all conditions using either the TP53 guides (TP53 KO – 67%; All KO – 66%) or the RTG guides
235 (RTG KO – 77%; All KO – 82%), and almost no editing in the Non Targ. condition (<0.1% at both loci) (Fig.
236 3C). We further examined the proportion of each deletion product observed in the RTG PCRs, and noted
237 a clear pattern, wherein deletions between guides 2 and 3 were the most frequent, in line with
238 expectations about the targeting scope of those two guides (Fig. 3D).
239 Next, we wanted to disentangle the editing on a per-locus level for the RTG data. We mapped
240 the data to each unique RTG (collapsing those with identical sequences outside the editing region) and
241 identified the proportion of each deletion at each locus. Although we generally observe a mix of all
242 three deletions at each locus (some of which is likely due to cross mapping between the RTGs), we
243 generally observe deletions in line with our expectations based on their targeting scope (Fig. 3E). Only
244 two loci buck this trend, LOC126068338 and LOC126068461, both of which have a single point mutation
245 at position 10 within the seed region of the guide (Fig. S4). Due to the importance of this region for
246 guide specificity, we filtered these targets out; however, we note that at least at these two loci the
247 guides appear to be permissive for editing. When combined with the per-locus reads that were not
248 identified as containing a deletion (Fig. 3F), we observe an average of 76.8% editing efficiency across all
249 unique loci.
250
251 *2.10 – RNA-seq analysis reveals overlapping and unique TP53 and RTG responses*
252 Given most literature on elephant RTGs has largely focused on how these retrogenes may
253 interact with canonical TP53 and MDM2, we first decided to look at changes in expression of all direct

254 downstream targets of TP53 (Fig. 4). In the absence of stress (i.e. 10µM Mitomycin C), we see relatively
255 small changes in expression within most downstream targets with the exception of MDM2 and CDKN1A
256 in both the TP53 and All KOs. Interestingly, we note that TP53 is not identified as significantly
257 downregulated in the TP53 KO, suggesting maybe that nonsense-mediated decay is less pronounced in
258 Asian elephants. In comparison, we observe very few differentially expressed genes in the RTG KO, and
259 only TP53 that has a log2 fold change greater than 1. This suggests that the RTGs may play some role in
260 either stabilizing or increasing the abundance of p53, although likely not directly by acting as a
261 transcription factor as none of the RTGs have an intact DNA-binding domain (Fig. 3).
262 In the presence of Mitomycin C, all three conditions have similar patterns in DEGs (Fig. 4),
263 although typically smaller expression changes in the RTG KO. As before, we also observe that TP53 is
264 significantly less expressed in the RTG KO and not in the TP53 KO. Additionally, IGFBP-3, a gene which
265 can contribute to cell growth and apoptosis [51,52], is also downregulated in the RTG KO. We also observe
266 significant downregulation of SFN across all three conditions in the MMC-treated cells, in line with this
267 gene's known tumor suppressor function of arresting cellular replication in response to DNA damage
268 [53,54].
269 Next, we wanted to examine the different pathways enriched within each condition. Due to the
270 large number of genes with small changes we filtered the dataset to a minimum absolute log2 fold
271 change of at least 1. We then extracted the DEG lists for each condition and analyzed the overlap
272 between them in either the non-stressed or MMC-treated cells (Fig. 5A). We performed BioPlanet
273 pathway enrichment [55] on the 763 DEGs common to all three conditions treated with Mitomycin C. We
274 observed significant enrichment in p53 signaling pathways, as would be expected of cells exposed to a
275 DNA damaging agent, as well as enrichment in many DNA replication and cell cycle pathways (Fig. 5B).
276 We also examined the unique DEGs in the TP53 and RTG KO cells with Mitomycin C treatment
277 (Fig. 5C). For the TP53 KO, we see enrichment in many pathways that would be expected given the
278 known checkpoint and cell stress responses coordinated by TP53 (e.g. cell cycle and proliferation, DNA
279 repair, etc; Fig. 4D) [56–58]. By comparison, unique RTG KO DEGs are enriched in extracellular organization
280 and signaling/interaction pathways (Fig. 5E), with many genes that contribute to multiple extracellular
281 matrix related processes (Fig. 5F). This suggests that RTGs may play a unique protective role in
282 regulating the tumor microenvironment and limiting metastasis [59]. Additionally, the largest number of
283 DEGs appear to belong to the apoptotic signaling in response to DNA damage pathway, a function that
284 has previously been suggested for African elephant RTGs [12,14].
285

2.11 – Very few RTGs are likely to be expressed or functional

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

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

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

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

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

350 likely to be functional are closely related to African elephant RTG9, which has been shown to induce
351 strong cellular responses to DNA damage [12–14] and apoptosis in mouse fibroblasts and human cancer cell
352 lines [12,14]. We also observe clear transcriptional differences between TP53 and RTG KO cells, including a
353 significant downregulation of TP53 expression in our RTG KO cells, suggesting a novel function for
354 elephant retrogenes. Additionally, we observe a differential enrichment of pathways when disrupting
355 RTGs versus TP53, indicating that while there is a high amount of overlap between these genes, they
356 also regulate many unique pathways. In particular, genes enriched within many extracellular
357 organization and signaling pathways were observed in the RTG KO population. Some of these (e.g.
358 extracellular matrix organization, angiogenesis) have been implicated as contributing to the
359 establishment of tumor microenvironments, and are gaining increased importance as vital to cancer
360 growth and spread [59]. This may also explain the low prevalence of malignant cancers in Asian elephants
361 [62].
362 Our work highlights the utility of cellular models to begin disentangling the complex effects of
363 regulatory noncoding deletions and cancer resistance genetic networks in non-model species. However,
364 in addition to the tissue-specific response limitations mentioned above, it is further hindered by our
365 reliance on primary, non-immortalized cells, although this will likely be addressed as genetic resources
366 for non-model species become more abundant [63,64]. Of particular importance would be stable and
367 readily available iPSCs, which would allow for differentiation of various tissues and more complex
368 organoid models to better investigate the effects of both adaptive and cancer resistance mutations,
369 especially where the generation of live animal models is not possible.
370
371 **4. Methods**
372 *4.1 – Elephant Genomic Dataset*
373 All bioinformatics analyses for deletion identification used previously published datasets. We
374 used datasets for eleven mammoths (*Mammuthus*), thirteen Asian elephants (*Elephas*), and twenty-two
375 African elephants (*Loxodonta*) (Supplementary Table 1). Published datasets were aligned to the *Elephas*
376 *maximus* reference genome (GCF_024166365.1) using BWA aln [65].
377
378 *4.2 – Mammoth Deletion Identification*
379 A summary of the mammoth deletion pipeline is available in Fig S1. Briefly, the *Elephas maximus*
380 reference genome was split into individual contigs. These were then fed into GenerateWindows.awk
381 (WinSize = 25; WinSlide = 10; AmbProp = 0.2), to generate windows across the genome and filter out

382 any with more than 20% missing data (i.e. N's). We then generated read depth files for each
383 mammoth/elephant alignment using samtools depth (-a to output all positions) [66], and split these by
384 contig. Windows were then cross referenced with depth files using another CoverageCalc.awk to extract
385 windows containing no mapped reads (StdevMultiplier=99), and adjacent windows concatenated using
386 ConcatonateWindows.awk. Windows with zero coverage identified in *Loxodonta* and *Elephas* were then
387 combined using bedops [67], then were used to filter the identified mammoth deletions using bedtools
388 intersect (-v flag to retain windows which are uniquely in the mammoth dataset) [68].
389 We also attempted to predict transcription factor binding sites (TFBS) upstream of elephant
390 genes. We first extracted the transcription start side and direction of each protein coding gene in the
391 *Elephas maximus* reference genome, then generated intervals for 1kb upstream of each gene. These
392 intervals were extracted using the samtools faidx program [66], and TFBS predicted using TFBStools [69] in
393 custom Rscript (TFBSTools_JASPAR2014.R). Identified TFBS were then cross referenced against filtered
394 mammoth deletions using bedtools intersect [68] to identify putative cis-regulatory deletions present in
395 woolly mammoths.
396 During the course of this project, a separate ancient DNA deletion pipeline was published [15]. We
397 also performed deletion prediction using the published pipeline using the same datasets above. We
398 examined the overlap between our two datasets using bedtools intersect [68].
399
400 *4.3 – TP53 identification and primer design*
401 We extracted and aligned the canonical TP53 locus as well as all 29 retrogenes (RTGs) from the
402 *Elephas maximus* reference genome (GCF_024166365.1). For the RTGs, we trimmed the alignment to
403 include only the ~1.1kb segment overlapping the exons in TP53. We then designed two primers to
404 specifically amplify the TP53 gene (Supplementary Table 2). For the RTGs, we designed seven forward
405 and four reverse primer variations, all amplifying the same locus, but containing all possible binding site
406 mutations to minimize amplification bias (Supplementary Table 2). These primers were made up to
407 10µM stock solutions and equal volumes of all forward and all reverse primers combined to produce
408 master forward and reverse stock solutions.
409 For all other deletions, we generated at least four primers flanking the region containing the
410 deletion ensuring similar melting temperatures across target sites, and that all primers were outside of
411 the region targeted by Cas9. For each locus, we tested all possible combinations of forward and reverse
412 primers and chose those that produced the cleanest and most reproducible amplifications
413 (Supplementary Table 2).

4.4 – 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 [70]. 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 [66] 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.

4.5 – Cell culturing and media

445 Fibroblast cells were grown in a custom media formulation with the final concentrations: 50%
446 FGM-2 (Lonza FBM Basal Media (CC-3131) supplemented with FGM-2 SingleQuots supplements (CC-
447 4126)); 20% heat-inactivated fetal bovine serum (ThermoFisher, A5670801); and 30% Mem Alpha
448 (Gibco, 12571063). Mesenchymal stem cells were grown in a custom media formulation with final
449 concentrations: 79% DMEM High Glucose + GlutaMax (ThermoFisher 10566016); 20% heat-inactivated
450 fetal bovine serum (ThermoFisher, A5670801); and 1% MEM Non-Essential Amino Acids Solution
451 (ThermoFisher 11140050). Epithelial cells were grown in a custom media formulation with final
452 concentrations: EGM-2 (Lonza Basal Medium (CC-3156) supplemented with EGMTM-2 SingleQuotsTM
453 Supplements (CC-4176)); 40% DMEM High Glucose + GlutaMax (ThermoFisher 10566016); and 10%
454 heat-inactivated fetal bovine serum (ThermoFisher, A5670801). Cells were grown in an incubator at 37°C
455 with 5% CO_2.
456
457 *4.6 – Cas9 RNP Electroporation*
458 All gRNAs were ordered from Synthego and resuspended to 100µM in 1X TE buffer. Cas9
459 nuclease was ordered from IDT (catalog # 10007807). All nucleofections were conducted using the Lonza
460 P4 Primary Cell 96-well Kit (catalog # V4SP-4096) using codes CA-137 (fibroblasts) or EN-150 (epithelial
461 cells), or the P3 Primary Cell 96-well Kit (catalog # V4SP-3096) using code CA-137 (MSCs).
462 For the TFBS, synthetic, and large-mammoth deletions, approximately 100-200k cells were used
463 for each 20µl transfection reaction, with a 3µM final concentration of each guide (i.e. both upstream
464 and downstream of the deletion) and 1µM of Cas9 nuclease per guide (i.e. a 3:1 ratio of gRNA:Cas9).
465 RNPs were pre-complexed in nucleofection solution for 20 minutes at room temperature then
466 electroporated into cells using the above codes.
467 We conducted pilot studies for the seven identified TFBS deletions upstream of genes that
468 showed consistent expression in our MSC line, although only one guide was used for deletions upstream
469 of LOC126074853 and LOC126068891 due to the small size of these deletions. We also conducted pilot
470 studies using all three synthetic deletion genes (FN1, CD109, EIF3I), as well as for the six smallest
471 intragenic and seven smallest intergenic large mammoth-specific deletions. Deletions were carried out
472 alongside negative control containing just Cas9 and/or a non-targeting gRNA RNP which has no
473 significant homology to the *E. maximus* genome. All experimental conditions for pilot work were
474 conducted in duplicate and in MSCs. We subsequently chose a subset of these deletions which
475 generated a high number of DEGs for our final analysis. These deletions were done as above, except
476 with four replicates each and accompanied by 3-4 negative controls using a non-targeting gRNA. For

477 Deletions 3, 51, and 89 we also attempted to introduce these deletions into fibroblast and epithelial
478 cells, however the epithelial cells failed to recover from the nucleofections and were ultimately
479 discarded.

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

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

493

494 *4.7 – Editing Efficiency*

495 During the first passage (approximately 4 days post nucleofection), about 10% of the cells were
496 collected and DNA extracted using the DNeasy Blood and Tissue Kit (Qiagen) using the manufacturer's
497 recommended protocol. For the TP53-editing extractions we amplified both the TP53 locus spanning the
498 expected knockout deletion, and the retrogene loci. For all other extractions we amplified the locus
499 spanning the targeted deletion and examined these loci across the negative controls as well. We
500 preformed 25µL PCR reactions with final concentrations of 1X Phusion Hot Start Flex 2X Master Mix and
501 500nM of each primer, using between 1-3µL of template. Cycling conditions were: 98°C for 3 minutes;
502 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
503 seconds, then 72°C for 45 seconds; followed by a final extension at 72°C for 5 minutes. Resulting PCR
504 amplicons were purified over MinElute PCR Purification columns (Qiagen) using the recommended
505 protocol and visualized on E-gel EX 2% Agarose gels (Invitrogen) to confirm the presence or absence of
506 any large deletions. We examined the relative fluorescence of our products using the Image Lab 6.1 (Bio-
507 Rad). We then calculated a minimum editing efficiency (the relative fluorescence of our expected

508 deleted product, relative to everything else in the lane) and a maximum editing efficiency (the relative
509 fluorescence of all non-wild-type bands within each lane).
510 To confirm efficiency estimates and examine RTG deletion distribution in our TP53-editing
511 experiments, we sent each of our TP53 and RTG amplicon pools for nanopore sequencing using the
512 Premium PCR service at Primordium to get full length sequences of our DNA fragments. For each
513 sample, raw reads were first filtered using a custom python script (FastqFilter.py) to between the
514 maximum and minimum expected sizes +/- 10% respectively. Mapping was carried out using BBMap [71]
515 using default parameters against either the TP53 amplicon, an RTG consensus amplicon, or a multi-fasta
516 with 26 RTGs (RTG amplicons with identical sequences outside of the editing region collapsed).
517 To calculate editing efficiency, we used a custom python script (CigarDel.py) which uses a region
518 file containing the deletion regions of interest, a minimum deletion proportion (i.e. the proportion of
519 bases mapping to the deleted region that need to be missing), and the CIGAR string of each read to
520 evaluate the presence or absence of various deletions in the data. For TP53 we analyzed the presence of
521 the expected deletion if Cas9 induced a double-strand break at both sgRNA targeting sites. For the RTG
522 we examined the presence of any of the three possible deletions between two guides. All analyses were
523 carried out with a minimum deletion proportion of 0.9.
524 To examine the distribution of deletions at each locus we identified the names of reads
525 containing deletions and extracted those that mapped to each locus. We then cross-referenced this with
526 the deletions identified against the RTG consensus to obtain read distributions for reads mapped to all
527 RTGs or the collapsed RTG loci.
528
529 *4.8 – Bulk RNA-seq*
530 Approximately 300-500k cells were collected, flash frozen on dry ice, and shipped to Genewiz
531 for bulk RNA sequencing within ten days of transfection (TP53) or 2-3 weeks (all other deletions).
532 Expression of transcripts was done using Salmon (v1.8.0 [72]) using a custom index consisting of all mRNAs
533 and non-coding RNAs annotated within the *E. maximus* reference genome (GCF_024166365.1), with all
534 29 TP53 retrogene sequences added. The salmon index was generated using the keepDuplicates flag to
535 avoid collapsing any RTGs. Salmon quant files were analyzed using DESeq2 [73] as part of a custom Rscript
536 to analyze samples in a pairwise fashion (i.e. one set of biological replicates against the controls at a
537 time). Differentially expressed genes (DEGs) were identified with an adjusted p value of less than 0.05
538 (padj < 0.05). No minimum log2 fold change was applied for initial analysis of TP53 pathway genes. A
539 minimum absolute log2 fold change of 1 was used for all other analyses. Venn diagrams and DEG

540 overlap was generated using a web based custom venn diagram tool
541 (https://bioinformatics.psb.ugent.be/webtools/Venn/), and pathway enrichment preformed using
542 GeneCodis4 [74].
543
544 *4.9 – Retrogene expression in public datasets*
545 To ensure the expression patterns we observed were not tissue/cell type specific, we examined
546 the expression of our retrogenes in some preliminary shared data from mesenchymal stem cells and
547 epithelial cells. These cells contained deletions generated using the same guides as in this study, but the
548 epithelial cells were also subject to various small molecule treatments. We also downloaded public RNA-
549 seq datasets for Asian elephant ovary (SRX15554240), salivary glands (SRX15554239), lymph nodes
550 (SRX15554238), thyroid (SRX15554237), lung (SRX15554236), and lymphocytes (SRX1423033), and
551 analyzed it as above.
552
553 *4.10 – Depth plots*
554 To examine where reads from each of the three RTGs showing repeated expression were
555 originating, we mapped the data for our datasets and the public datasets back to the Asian elephant
556 reference genome using STAR [75]. We then extracted read depths along the full annotated length of RTGs
557 LOC126068247, LOC126068267, and LOC126068248 using samtools depth [76]. For our dataset we
558 averaged expression values within biological replicates and plotted the data in Microsoft Excel.
559
560 *4.11 – Maximum likelihood phylogeny*
561 We download African elephant retrogene 9 (KF715863.1) to investigate a possible relationship
562 between this retrogenes and the Asian elephant retrogenes. We then aligned the retrogene to the Asian
563 elephant TP53 CDS and truncated the RTG 9 sequence to contain only the homologous sequence to
564 TP53. We then aligned the truncated RTG 9 sequence, Asian elephant TP53, and all 29 Asian elephant
565 retrogenes using muscle (v5.1) [77]. We generated a maximum likelihood phylogeny using IQ-Tree [78], using
566 ModelFinder to identify the best supported substitution model [79], and 100 bootstraps (-b 100).
567
568 Supplementary References: [12,65,71,72,80–85]
569
570 **5. Author contributions**

571 EK and GMC conceived the study with feedback from EM and AAT. EK and NB designed and
572 conducted all wet-lab experiments with feedback and help from EM, AAT, and LL. EK and LL completed
573 all bioinformatic analyses. AAT provided additional RNA-seq data for epithelial cell lines. EK wrote the
574 first draft of the manuscript. All authors helped in the revision of subsequent drafts.
575
576 **6. Data availability**
577 Bulk RNA-seq data for all mammoth deletions as well as TP53 conditions with and without
578 mitomycin was deposited on NCBI (BioProject Accessions: PRJNA1147676 and PRJNA1217465).
579 Bioinformatic scripts and tools created for this work are available at:
580 https://github.com/EmilKarpinski/RTG_TP53_MammothDel_Analysis.
581
582 **7. Declaration of competing interests**
583 For a complete list of G.M.C.'s financial interests, please visit
584 http://arep.med.harvard.edu/gmc/tech.html. All other authors declare no other conflicts of interest.
585
586 **8. Acknowledgments**
587 We would like to thank the San Diego Frozen Zoo for sharing the fibroblast cell line. We would
588 like to thank Zhengkuan Tang for discussions on RNA-seq experimental setup and analysis, and Ana Luiza
589 Lemos Queiroz and Ramiro Perrota for assistance during early pilot experiments. We would also like to
590 thank George Chao for guidance on nanopore analysis, and Sergiy Velychko for help during
591 troubleshooting of the fibroblast cell culture techniques. Lastly, we would like to thank Roman Teo
592 Oliynyk and Chun-Ting Wu, as well as the rest of the Church lab for helpful discussions throughout this
593 project.
594 EK and AAT were supported by a structured research agreement between Harvard Medical
595 School and Colossal Biosciences.
596
597
598
599

References

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

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

3. Hu, L. *et al.* Arctic introgression and chromatin regulation facilitated rapid Qinghai-Tibet Plateau colonization by an avian predator. *Nat Commun* **13**, 6413 (2022).

4. Lin, Z. *et al.* Biological adaptations in the Arctic cervid, the reindeer (*Rangifer tarandus*). *Science (1979)* **364**, (2019).

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

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

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

8. Boeskorov, G. G. *et al.* Adaptation of the woolly mammoth Mammuthus primigenius (Blumenbach, 1799) to habitat conditions in the glacial period. *Contemp Probl Ecol* **9**, 544–553 (2016).

9. van der Valk, T. *et al.* Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* **591**, 265–269 (2021).

10. Lynch, V. J. *et al.* Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. *Cell Rep* **12**, 217–228 (2015).

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

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

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

15. van der Valk, T. *et al.* Evolutionary consequences of genomic deletions and insertions in the woolly mammoth genome. *iScience* **25**, 104826 (2022).

16. Endoh-Yamagami, S. *et al.* A mutation in the pericentrin gene causes abnormal interneuron migration to the olfactory bulb in mice. *Dev Biol* **340**, 41–53 (2010).

634 17. Denu, R. A. *et al.* Fibroblasts and Mesenchymal Stromal/Stem Cells Are Phenotypically
635 Indistinguishable. *Acta Haematol* **136**, 85–97 (2016).

636 18. Närhi, K. *et al.* Sostdc1 defines the size and number of skin appendage placodes. *Dev Biol* **364**,
637 149–161 (2012).

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

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

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

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

647 23. Owada, Y. *et al.* Altered Water Barrier Function in Epidermal-Type Fatty Acid Binding Protein-
648 Deficient Mice. *Journal of Investigative Dermatology* **118**, 430–435 (2002).

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

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

654 26. Martin, T. G. & Leinwand, L. A. Molecular regulation of reversible cardiac remodeling: lessons
655 from species with extreme physiological adaptations. *Journal of Experimental Biology* **227**,
656 (2024).

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

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

661 29. Grego-Bessa, J. *et al.* Nrg1 Regulates Cardiomyocyte Migration and Cell Cycle in Ventricular
662 Development. *Circ Res* **133**, 927–943 (2023).

663 30. Ushikoshi, H. *et al.* Local overexpression of HB-EGF exacerbates remodeling following myocardial
664 infarction by activating noncardiomyocytes. *Laboratory Investigation* **85**, 862–873 (2005).

665 31. Cordero, P. *et al.* Pathologic gene network rewiring implicates PPP1R3A as a central regulator in
666 pressure overload heart failure. *Nat Commun* **10**, 2760 (2019).

667 32. Krumbein, M. *et al.* RNA binding protein IGF2BP2 expression is induced by stress in the heart and
668 mediates dilated cardiomyopathy. *Commun Biol* **6**, 1229 (2023).

669 33. Wang, J. *et al.* Cardiomyopathy Associated with Microcirculation Dysfunction in Laminin α4
670 Chain-deficient Mice. *Journal of Biological Chemistry* **281**, 213–220 (2006).

671 34. Jiang, Z. *et al*. SEC14L3 knockdown inhibited clear cell renal cell carcinoma proliferation,
672 metastasis and sunitinib resistance through an SEC14L3/RPS3/NFκB positive feedback loop.
673 *Journal of Experimental & Clinical Cancer Research* **43**, 288 (2024).

674 35. Thyboll, J. *et al.* Deletion of the Laminin α4 Chain Leads to Impaired Microvessel Maturation. *Mol*
675 *Cell Biol* **22**, 1194–1202 (2002).

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

678 37. Eguchi, J. *et al.* Transcriptional Control of Adipose Lipid Handling by IRF4. *Cell Metab* **13**, 249–259
679 (2011).

680 38. Dai, N. *et al.* IGF2BP2/IMP2-Deficient Mice Resist Obesity through Enhanced Translation of Ucp1
681 mRNA and Other mRNAs Encoding Mitochondrial Proteins. *Cell Metab* **21**, 609–621 (2015).

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

685 40. Wang, B. *et al.* Identification and characterization of NYGGF4, a novel gene containing a
686 phosphotyrosine-binding (PTB) domain that stimulates 3T3-L1 preadipocytes proliferation. *Gene*
687 **379**, 132–140 (2006).

688 41. Zhang, Y. *et al.* Targeting nuclear receptor NR4A1–dependent adipocyte progenitor quiescence
689 promotes metabolic adaptation to obesity. *Journal of Clinical Investigation* **128**, 4898–4911
690 (2018).

691 42. Wang, S.-S. *et al.* Early growth response 2 (Egr2) plays opposing roles in committing C3H10T1/2
692 stem cells to adipocytes and smooth muscle-like cells. *Int J Biochem Cell Biol* **45**, 1825–1832
693 (2013).

694 43. Tran, M. T. N. *et al.* MafB deficiency accelerates the development of obesity in mice. *FEBS Open*
695 *Bio* **6**, 540–547 (2016).

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

698 45. Herijgers, N. *et al.* Effect of human scavenger receptor class A overexpression in bone marrow-
699 derived cells on lipoprotein metabolism and atherosclerosis in low density lipoprotein receptor
700 knockout mice. *J Lipid Res* **41**, 1402–1409 (2000).

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

703 47. Hung, Y.-H., Carreiro, A. L. & Buhman, K. K. Dgat1 and Dgat2 regulate enterocyte triacylglycerol
704 distribution and alter proteins associated with cytoplasmic lipid droplets in response to dietary

fat. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1862**, 600–614 (2017).

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

49. Cai, W. *et al.* The kainate receptor GluK2 mediates cold sensing in mice. *Nat Neurosci* **27**, 679–688 (2024).

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

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

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

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

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

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

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

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

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

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

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

61. Sonawane, A. R. *et al.* Understanding Tissue-Specific Gene Regulation. *Cell Rep* **21**, 1077–1088 (2017).

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

63. Li, Z. *et al.* Establishment of Bactrian Camel Induced Pluripotent Stem Cells and Prediction of Their Unique Pluripotency Genes. *Int J Mol Sci* **24**, 1917 (2023).

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

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

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

67. Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

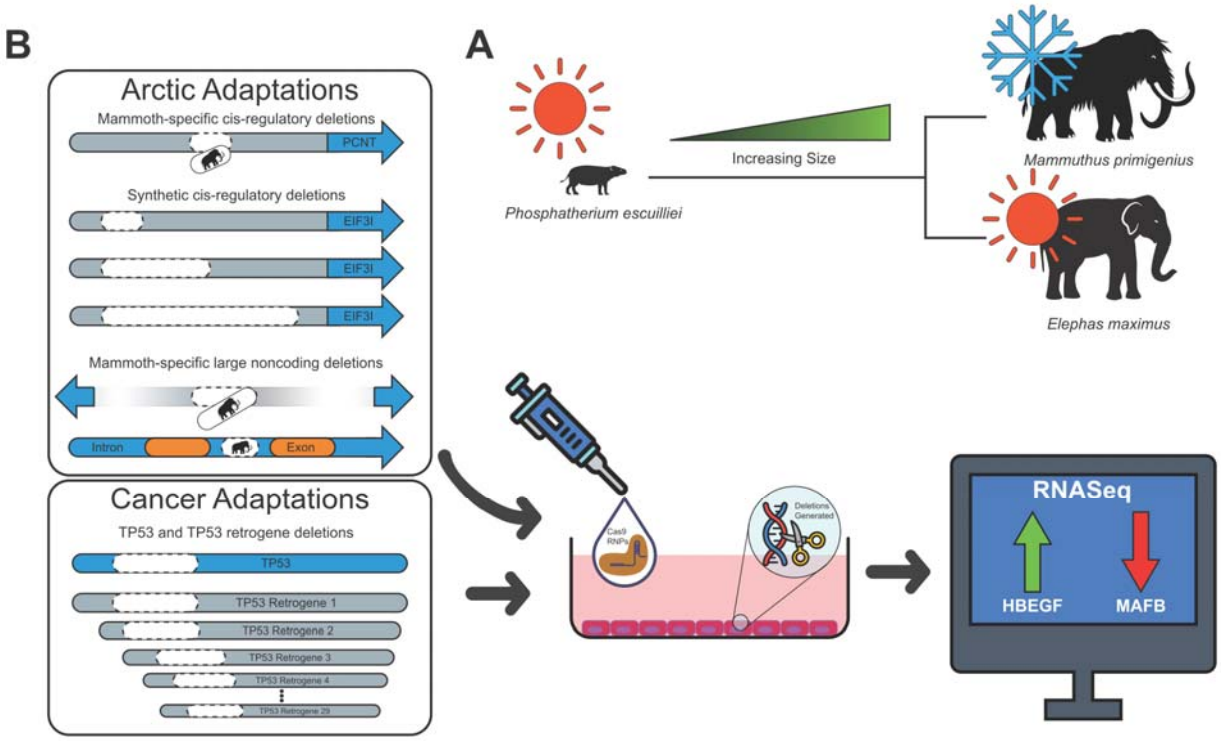

Figure 1. Experimental Design. A. Trends in proboscidean evolution include a general increase in body size, and adaptations to various environments including warm tropical grasslands and cold arctic tundra. **B.** We engineered mammoth specific or synthetic cis-regulatory deletions, large mammoth-specific 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.

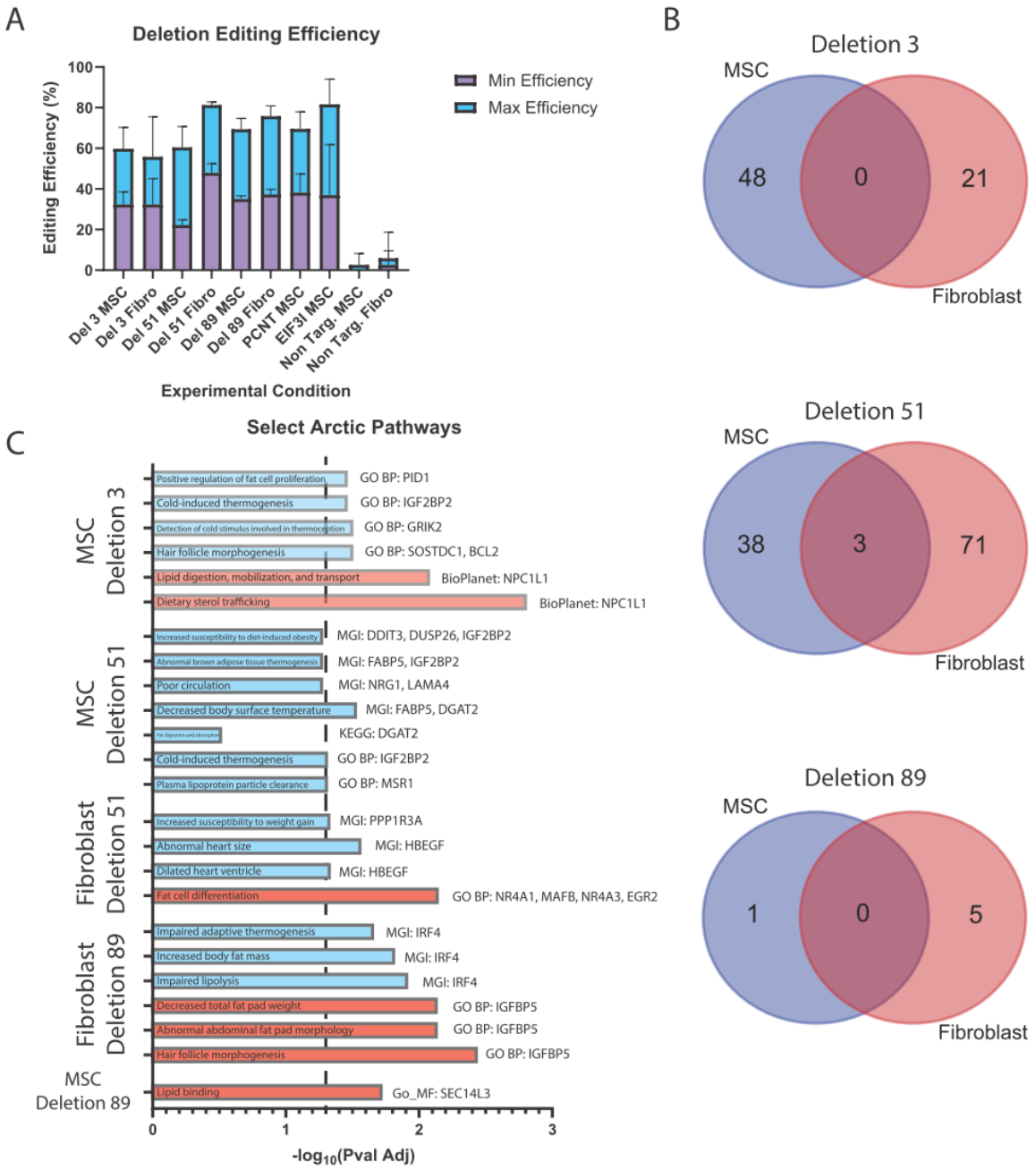

Figure 2. Mammoth deletion efficiency and effects. A. Editing efficiency observed for each deletion in both MSCs and fibroblasts estimated using gel electrophoresis. **B.** DEG overlap between MSCs and fibroblasts for each of the three large mammoth-specific noncoding deletions. **C.** A selection of 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.

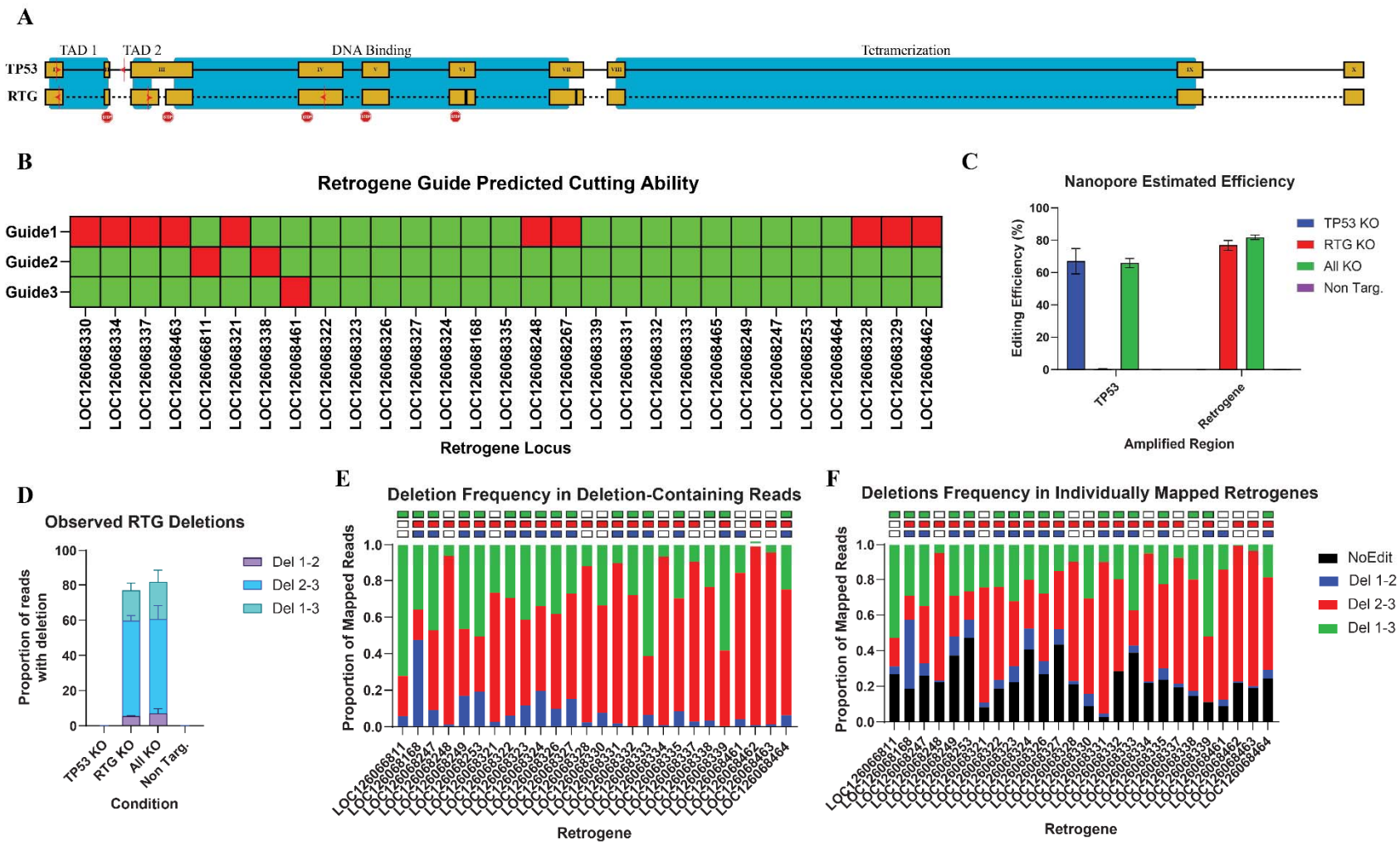

Figure 3. TP53 Editing. A. Schematic showing an alignment of Asian elephant TP53 and a retrogene (RTG) consensus, with domains labelled based on homology to human TP53. Exons are shown in yellow, with introns indicated by black lines. Missing sequence is indicated with a dashed line in the RTGs. Additionally, 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

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

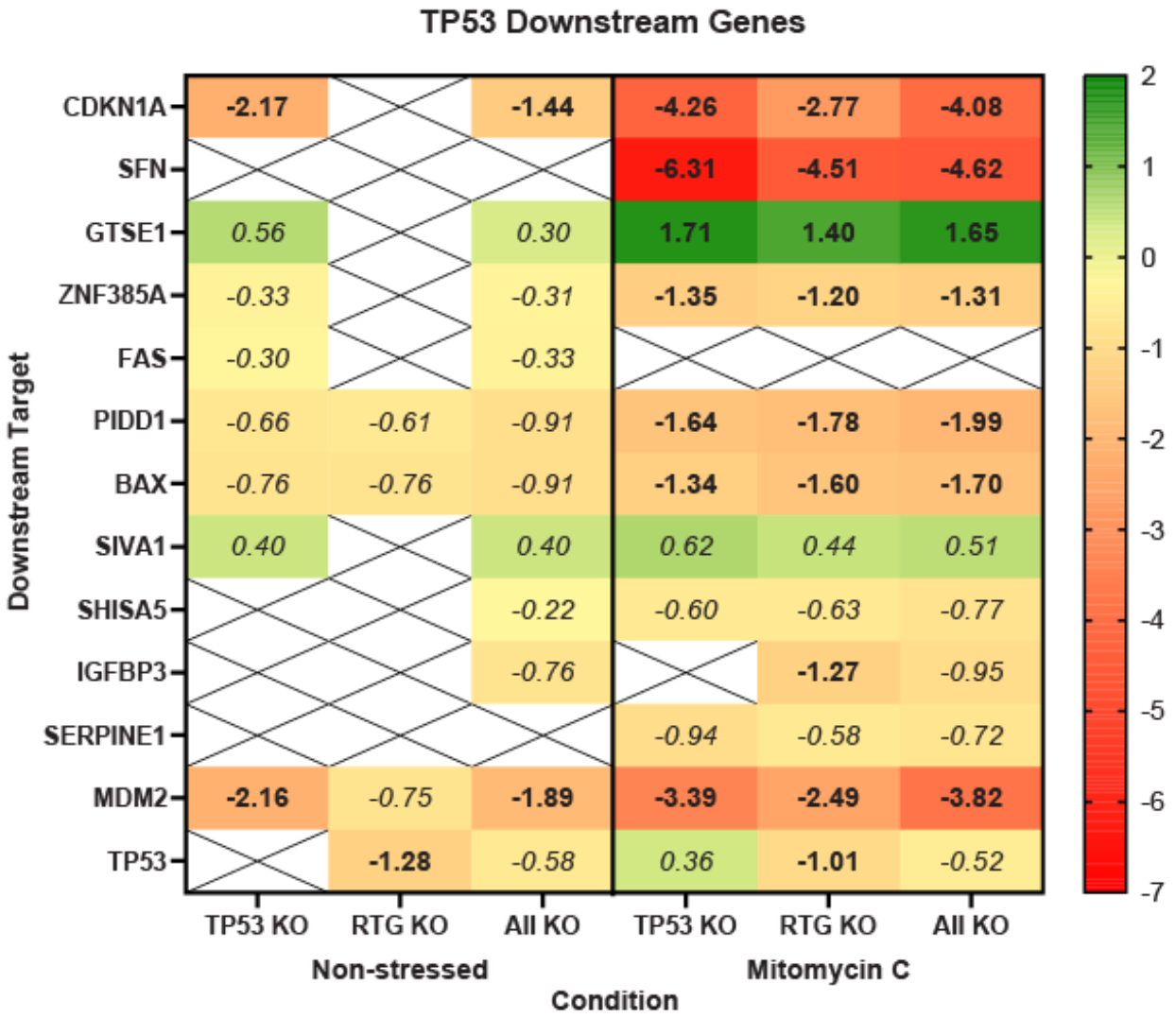

Figure 4. Expression of TP53 pathway genes. Log2 fold changes observed for genes immediately downstream of TP53 in the p53 signaling pathway (KEGG pathway hsa04115). Crosses indicate that no significant difference in expression was observed for that gene in a given condition. Genes in italics have an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

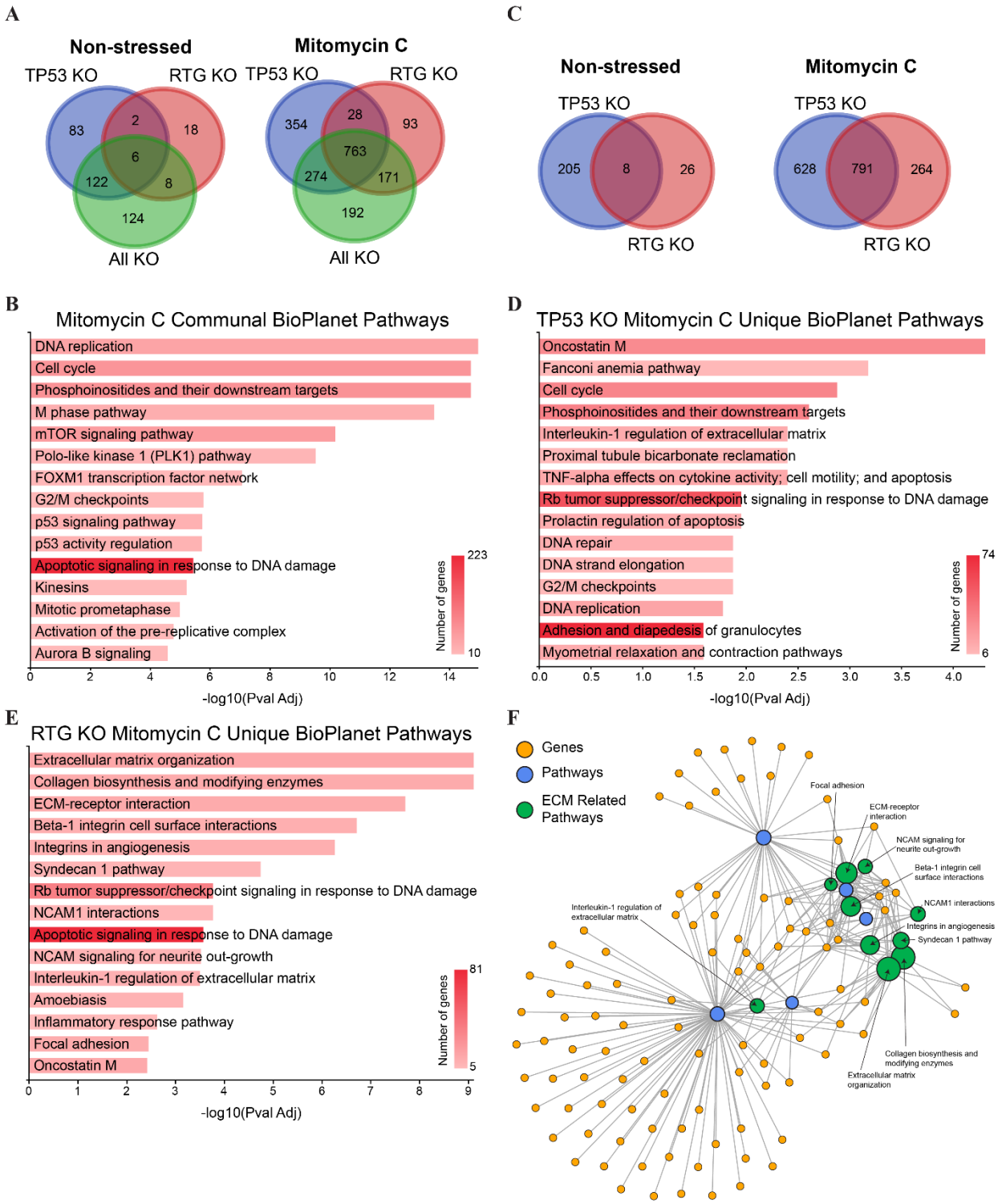

Figure 5. Pathway enrichment. A. Differentially expressed gene (DEG) overlap observed between treatment conditions for cells in normal media or exposed to 10µM mitomycin C. Genes were filtered for a minimum absolute log2 fold change greater than 1. **B.** BioPlanet pathway enrichment for the 763 DEGs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

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. **F.** Pathway/gene cluster analysis for DEGs in the RTG KO Mitomycin C treated cells. ECM related pathways are colored green and labelled. Pathway clusters are scaled by their p-value.