



RESEARCH ARTICLE

Chromosome-level Asian elephant genome assembly and comparative genomics of long-lived mammals reveal the common substitutions for cancer resistance

Xuanjing Li^{1,2} | Pengcheng Wang^{1,3} | Qi Pan^{1,2} | Gaoming Liu¹ | Weiqiang Liu^{1,2} | Olatunde Omotoso^{1,2} | Juan Du^{1,2} | Zihao Li^{1,2} | Yang Yu⁴ | Yun Huang⁴ | Pingfen Zhu¹ | Meng Li¹ | Xuming Zhou¹

¹CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Beijing, China

²University of Chinese Academy of Sciences, Beijing, China

³Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing, China

⁴Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China

Correspondence

Xuming Zhou, CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chaoyang District, Beijing 100101, China.

Email: zhouxuming@ioz.ac.cn

Abstract

The naked mole rat (*Heterocephalus glaber*), bats (e.g., genus *Myotis*), and elephants (family Elephantidae) are known as long-lived mammals and are assumed to be excellent cancer antagonists. However, whether there are common genetic changes underpinning cancer resistance in these long-lived species is yet to be fully established. Here, we newly generated a high-quality chromosome-level Asian elephant (*Elephas maximus*) genome and identified that the expanded gene families in elephants are involved in Ras-associated and base excision repair pathways. Moreover, we performed comparative genomic analyses of 12 mammals and examined genes with signatures of positive selection in elephants, naked mole rat, and greater horseshoe bat. Residues at positively selected sites of *CDR2L* and *ALDH6A1* in these long-lived mammals enhanced the inhibition of tumor cell migration compared to those in short-lived relatives. Overall, our study provides a new genome resource and a preliminary survey of common genetic changes in long-lived mammals.

KEYWORDS

cancer resistance, comparative genomics, elephants, long-lived mammals

Abbreviations: ALDH6A1, aldehyde dehydrogenase 6 family member A1; ALDHs, aldehyde dehydrogenase superfamily; BCC, basal cell carcinoma; bFGF, basic fibroblast growth factor; BUSCO, benchmarking universal single-copy orthologs; C19orf47, chromosome 19 open reading frame 47; CDR2L, cerebral degenerated-related protein 2-like; CDS, coding sequence; COP53, COP9 signalosome subunit 3; CSF3R, colony-stimulating factor-3 receptor; CYLD, cylindromatosis; DMEM, Dulbecco's modified Eagle's medium; DOK2, downstream of tyrosine kinase 2; DOK4, docking protein 4; ECI, early contact inhibition; EGF, epidermal growth factor; FBS, fetal bovine serum; GH, growth hormone; GH/IGF-1, growth hormone/insulin-like growth factor 1; GMPR, guanosine monophosphate reductase; GO, gene ontology; Hi-C, high-throughput chromosome conformation capture; KEGG, Kyoto Encyclopedia of Genes and Genomes; LIF, leukemia inhibitory factor; LINE, long interspersed nuclear elements; LRIT1, leucine rich repeat, Ig-like and transmembrane domains 1; LRT, likelihood ratio test; LTR, long terminal repeats; MAPK, mitogen activated protein kinases; MTF1, metal responsive transcription factor 1; MYO6, myosin-6; MYOT, myotilin; PBS, phosphate buffered saline; PCD, paraneoplastic cerebellar degeneration; PSGs, positively selected genes; Rap1, Ras-associated protein-1; RBM, RNA-binding motif; SD, standard deviation; SINE, short interspersed nuclear elements; stLFR, single-tube long fragment read; TES, transposable elements; TRAF6, TNF receptor associated factor 6; vHMW-HA, very high-molecular weight hyaluronan; WDR7, WD repeat domain 7; XPO1, exportin 1.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Aging Cell* published by Anatomical Society and John Wiley & Sons Ltd.



1 | INTRODUCTION

Cancer prevention is one of the observed traits in species with extended lifespan. For example, long-lived mammals, such as the naked mole rat (*Heterocephalus glaber*), bats (e.g., Brandt's bat *Myotis brandtii*), and elephants (family Elephantidae), were suggested to be resistant to cancer (Abegglen et al., 2015; Buffenstein, 2008; Podlutzky et al., 2005). Particularly, the naked mole rat, with an exceptionally maximum lifespan surpassing 37 years, has a low cancer mortality rate (Buffenstein, 2008; Buffenstein & Jarvis, 2002), compared to a similarly-sized mouse (*Mus musculus*) that only lives 4–5 years (de Magalhães et al., 2005). Previous research revealed several potential cancer resistant mechanisms in naked mole rat. The naked mole rat fibroblasts exhibit a remarkable hypersensitivity to contact inhibition, also referred to as early contact inhibition (ECI) (Seluanov et al., 2009). In addition, naked mole rat tissues can secrete very high-molecular weight hyaluronan (vHMW-HA), which is not only believed to be an adaptive response to subterranean habitat, but also serves as anti-cancer strategy (Tian et al., 2013) because vHMW-HA represses tumor angiogenesis, promotes apoptosis, and inhibits inflammatory responses (Liu et al., 2008; Zhao et al., 2019). The Brandt's bats were recaptured after 41 years and were considered the longest-lived mammals when adjusted for body size (Podlutzky et al., 2005). Low incidence of cancer in bats, especially in the *Myotis*, might involve bat-specific regulatory activities of microRNAs that function tumorigenesis pathways (Huang et al., 2016, 2019), down-regulation of the growth hormone/insulin-like growth factor 1 (GH/IGF-1) pathway, unique sequence change in the GH-receptor (Seim et al., 2013), repression of telomerase activities (Gomes et al., 2011), and positive selection in telomere-associated genes (Huang et al., 2019; Morgan et al., 2013). These findings received increased attention as they provide targets for clinical intervention employed in humans and mice (Guevara-Aguirre et al., 2011; Ikeno et al., 2009).

If an increase in cell division corresponds with an increased risk of cancer, then one would expect elephants to have higher cancer incidents compared to humans (Peto et al., 1975). Despite having ~1000 times more cells than humans, elephants have a cancer mortality rate of <5%, compared to humans with between 11 and 25% cancer mortality rate (Abegglen et al., 2015). The maximum lifespan of the African bush elephant (*Loxodonta africana*) is estimated at 65 years, while that of the Asian elephant (*Elephas maximus*) is ~80 years (de Magalhães et al., 2005). Elephant genome includes possession of extra copies of TP53 gene, a known tumor suppressor gene with roles in DNA damage response, apoptosis, and cell cycle, which may result in the lower cell death response to cellular stresses (Abegglen et al., 2015; Sulak et al., 2016; Tollis et al., 2021). In addition, *p53* is tightly regulated by the *Mdm2* (murine double minute-2) and their interaction generates a variety of BOX-I MDM2 binding motifs in 20 copies of TP53 isoforms, which would enhance sensitivity to DNA damage and suppress tumor (Padariya et al., 2022). In addition to extensive sequence conservation at TP53 retrogene loci found in elephants (Tollis et al., 2021), elephants also have expanded copies of the leukemia inhibitory factor (LIF) gene, a downstream

target of TP53 that induces apoptosis (Vazquez et al., 2018). These observations support the initiation of using elephant as model to investigate how evolution controls tumor in mammals.

Convergent phenotypes among unrelated taxa are commonly subjected to similar evolutionary pressure. Tumor suppression, though it evolutionary correlated with large body size and long lifespan across animals (Caulin & Maley, 2011; Gorbunova & Seluanov, 2009), the underlined genetic changes seems to have independently evolved across distant mammalian taxa (Omotoso et al., 2021). Here, using the high-quality genome resource of several long-lived mammals, we aimed to explore the “convergent” or common substitutions in long-lived mammals. To this end, we first generate a chromosome-level genome assembly of Asian elephants and then employed comparative genomic analysis of 12 mammalian genomes to identify positively selected genes in long-lived mammals. Our comparative analysis showed the signal of positive selection on *ALDH6A1* and on *CDR2L*. We further performed experimental assays to show that residues at the positively selected sites in *ALDH6A1* and *CDR2L* suppress the migration of tumor cells. To sum up, our results provide a new high-quality genome resource for elephant and partially indicate that there is common substitutions response to cancer resistance in long-lived mammals. These genomic resources and discoveries could be beneficial for the future researches in identifying effective cancer therapeutic approaches.

2 | MATERIALS AND METHODS

2.1 | Sampling and genome sequencing

The fresh blood sample of a male Asian elephant was gifted by wild elephant valley, Xishuangbanna, Yunnan Province, China. The sample was kept with anticoagulant at -80°C at the Institute of Zoology, Chinese Academy of Sciences. All collection and processing of blood samples were conducted in accordance with the guidelines of Institutional Animal Care and Use Committee of the Institute of Zoology, Chinese Academy of Sciences. DNA was extracted with phenol-chloroform method and its quality was evaluated with agar gelatin electrophoresis. After obtaining high-molecular-weight DNA, the single-tube long fragment read (stLFR) library (Wang et al., 2019) was constructed and then sequenced with a paired-end 100bp sequencing strategy on BGISEQ-500 high-throughput sequencing platform. To generate a chromosomal-level assembly of the Asian elephant genome, a Hi-C (high-throughput chromosome conformation capture) library (Burton et al., 2013) was further constructed and sequenced with BGISEQ-500 platform. To evaluate the quality of the assembly, we collected an additional placenta sample from a female Asian elephant in Xishuangbanna and used the same method described above to isolate DNA. The fragment library (insert length was 250bp) of this sample was constructed and sequenced with a paired-end 150bp sequencing strategy on BGISEQ-500.

To facilitate annotation of the Asian elephant genome, total RNA from the placenta sample was isolated and the fragment library



(insert length was 250bp) was built based on the Iso-Seq protocol. The fragment library of RNA was sequenced on the BGISEQ-500 platform with a paired-end 150bp sequencing strategy and aimed at 6 Gb data. The full-length transcripts were sequenced on Pacbio Sequel platform at BGI and aimed at about 30 Gb data.

2.2 | Genome assembly and evaluation

We employed the following criteria to filter the raw reads from the stLFR library: (1) the reads that have more than 50% bases with a quality of less than five; (2) the reads that have adapters; and (3) the reads that have more than 5% N bases. After filtering, we employed Supernova (Weisenfeld et al., 2017) to assemble the genome and then used Gapcloser (Luo et al., 2012) to close the gaps. The completeness was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão et al., 2015).

The Hi-C data were used for generating a chromosomal-level assembly. HiC-Pro v 2.8.0 (Servant et al., 2015) and bowtie2 v. 2.2.5 (Langmead et al., 2009) were used to filter the raw Hi-C data. Then, Juicer v. 1.5 (Durand et al., 2016) and 3D de novo assembly v. 170123 (Dudchenko et al., 2017) were employed to assign scaffolds to 28 pseudochromosomes.

To evaluate the accuracy and quality of the current version of Asian elephant genome assembly, reads from the fragment library in the Trim Galore v0.4.2 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) were filtered with default parameters, and then, the clean reads were aligned to the new assembly with BWA MEM (Li & Durbin, 2009). Samtools (Li et al., 2009) was employed to measure the mapping rate and the coverage.

2.3 | Genome annotation

We combined de novo and homology methods to identify the tandem repeats and interspersed repeats in the Asian elephant genome. We used RepeatModeler v. 2.0.1 (<https://github.com/Dfam-consortium/RepeatModeler>) with RepeatMasker v. 4.1.2 (Chen, 2004), RECON v. 1.08 (Bao & Eddy, 2002), RepeatScout v. 1.0.6 (Price et al., 2005), TRF v. 4.0.9 (Benson, 1999), and RMBlast v. 2.11.0 (Johnson et al., 2008) to predict the repetitive sequences. The LTR_Finder (Xu & Wang, 2007) was used to find full-length LTR retrotransposons in the genome. The RepeatMasker v. 4.1.2 (Chen, 2004) and repeat database, and Dfam v. 3.2 (Bao et al., 2015) were used to identify the repetitive sequences based on homology sequences. All predicted repetitive sequences were combined, and then, the short fragments (<20bp) and the overlapping repeats were removed.

To predict the gene structure in the Asian elephant genome, we combined de novo prediction, homology-based prediction, and transcriptome-based methods to analyze the genome. The AUGUSTUS (Stanke et al., 2006) was used to predict the gene structure. The protein sequences of cattle (*Bos taurus*), African bush elephant (*Loxodonta africana*), rock hyrax (*Procavia capensis*), aardvark

(*Orycteropus afer*), and Florida manatee (*Trichechus manatus*) were used as homology sequences in Genewise (Birney et al., 2004) to predict gene structure. The protein sequences of aardvark (accession numbers PRJNA237355) and Florida manatee (accession numbers: PRJNA189960) were downloaded from the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>), while sequences of other species were obtained from the Ensembl release 99 (<http://www.ensembl.org/>).

The SMRTanalysis v. 6.0.0 (<https://www.pacb.com/support/software-downloads/>) was used to produce the consensus transcripts from the iso-seq sequencing data. The reads from the fragment library were used to correct the consensus transcripts. The high-quality consensus transcripts were used as evidence to predict gene structure in Trinity v. 2.12.0 (Grabherr et al., 2011). The Glean v. 1.0.1 (Elsik et al., 2007) was used to combine the results from the above three methods and filter the results. Predicted genes were filtered out if (1) there is only one de novo evidence to support the gene structure; (2) the length of CDS is less than 150bp; and (3) the overlap length ratio with transposable elements (TEs) is less than 0.2. The mammalia_odb9 database from BUSCO was used to evaluate the quality of the gene set, and whole-genome synteny with cattle and human genome, k-mer analysis were used to estimate the genome quality. The database from SwissProt (Boeckmann et al., 2003), TrEMBL (Boeckmann et al., 2003), KEGG (Kanehisa & Goto, 2000), InterPro (Mitchell et al., 2019), and GO (Ashburner et al., 2000) were used to annotate the gene function. We also identified noncoding RNAs of the Asian elephant genome. The tRNA-scan-SE (Lowe & Eddy, 1997) was used to identify the tRNA, and the RNAmmer (Lagesen et al., 2007) was used to predict rRNA. The INFERNAL (Nawrocki & Eddy, 2013) and Rfam v. 14.5 (Griffiths-Jones et al., 2005) were used to predict miRNA and snRNA.

2.4 | Gene family clustering

Homologs among the following 12 mammals: Asian elephant (*Elephas maximus*), African bush elephant (*Loxodonta africana*), cattle (*Bos taurus*), dog (*Canis familiaris*), small Madagascar hedgehog (*Echinops telfairi*), naked mole rat (*Heterocephalus glaber*), human (*Homo sapiens*), gray short-tailed opossum (*Monodelphis domestica*), aardvark (*Orycteropus afer*), platypus (*Ornithorhynchus anatinus*), cape rock hyrax (*Procavia capensis*), and greater horseshoe bat (*Rhinolophus ferrumequinum*) were estimated. The greater horseshoe bat rather than Brandt's bat (*Myotis brandtii*) was selected because the greater horseshoe bat is also long-lived bats (with longevity estimated to be 30.5 years) and its genome assembly is chromosome level (Rhie et al., 2021). We extracted the longest transcripts of each gene and formatted the corresponding protein sequences from all 12 mammals into a Blastp database (Johnson et al., 2008). We further employed the protein sequence alignment of each species and confirm the homologous sequences in Blastp database with the e-value was 1×10^{-7} . Using the protein sequences alignment, OrthoMCL (Li et al., 2003) was used to cluster orthologous protein sequences.



The sequence alignment was performed for each ortholog in the MUSCLE program (Edgar, 2004), and the alignment was filtered using Gblocks v. 0.91b (Castresana, 2000).

Gene family expansions and contractions were analyzed using CAFÉ program (De Bie et al., 2006). The number of expanded and contracted genes for each branch and node of the phylogenetic tree was estimated and the significantly “expanded and contracted gene families” are identified when an exact p value (Viterbi method) ≤ 0.01 . Significantly overrepresented GO (Gene Ontology) terms were identified using the topGO (Alexa & Rahnenführer, 2009) package in R (<https://www.r-project.org/>), and the Benjamini and Hochberg FDR correction was applied. Significantly overrepresented GO terms were identified with corrected p values of ≤ 0.05 .

2.5 | Identification of positively selected genes

In order to measure selective pressures acting on protein-coding genes in long-lived mammals, we screened the signature of positive selection of all orthologs using the CodeML program implemented in PAML package v. 4.8 (Yang, 2007). The optimized branch-site model (Yang & Dos Reis, 2010) was used to detect signatures of positive selection along specific lineages. Several recent studies have called into question any evidence from codon-based models of sequence evolution that do not take into account simultaneous double mutations or variation in silent substitution rates (Venkat et al., 2018; Wisotsky et al., 2020). The Busted model in the Hyphy package was also used to confirm the selection signals of each ortholog on each lineage (Kosakovsky Pond et al., 2019). In this analysis, two groups of foreground branches were set: elephant lineage (Asian elephant + African bush elephant) and long-lived mammals (Asian elephant + African bush elephant + naked mole rat + greater horseshoe bat). The latter group was formulated due to these four mammals have been recognized to show the strong ability of anti-cancer and long-lived. Finally, the phylogenetic tree of 12 mammals was retrieved from TimeTree (<http://www.timetree.org/>) (Kumar et al., 2017).

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of these positively selected genes were implemented in clusterProfiler (Yu et al., 2012), with the following parameters, OrgDb = org.Hs.eg.db, fun = “enrichGO,” ont = “BP” (“MF” and “CC”), pvalueCutoff = 0.05, and qvalueCutoff = 0.2. The functional categories with p value less than 0.05 were considered to be statistically significant.

2.6 | Plasmid construction

To explore the roles of residues at positively selected sites in cancer resistance, we constructed several plasmids for functional cellular assays. Complete *ALDH6A1* coding sequence of elephants, greater horseshoe bat, naked mole rat (*ALDH6A1*^{AAC}, *ALDH6A1*^{AGT}, and *ALDH6A1*^{CCG}), and mouse (*ALDH6A1*^{ACA}) was used as wild

type. In addition, the positively selected site of *ALDH6A1* was mutated, with codons in long-lived mammals mutated to mouse genotype (*ALDH6A1*^{AACACA}, *ALDH6A1*^{AGTACA}, *ALDH6A1*^{CCGACA}) and codon in mouse was mutated to long-lived mammals genotype (*ALDH6A1*^{ACAAAC}, *ALDH6A1*^{ACAAGT}, *ALDH6A1*^{ACACCG}). Similarly, the complete *CDR2L* coding sequence of Asian elephant (*CDR2L*^{AGA}) and mouse (*CDR2L*^{AGT}) was used as wild type, and the positively selected site was mutated correspondingly (*CDR2L*^{AGAAGT} and *CDR2L*^{AGTAGA}). The wild and mutant sequences were synthesized by Beijing Genomics Institute (BGI) and cloned into pEGFP-N1 vectors (Clontech) separately.

2.7 | Cell culture and transfection

Dulbecco's modified Eagle's medium (DMEM, Gibco; Thermo Fisher Scientific, Inc.) with 10% fetal bovine serum (FBS, Gibco; Thermo Fisher Scientific, Inc.) and 1% penicillin-streptomycin were used to cultivate A549 cells (which is a widely used human lung adenocarcinoma cell line) and then cultured in an incubator containing 5% CO₂ at 37°C. The fresh culture medium was changed daily, and the logarithmic growth cells were collected for further experiments.

Cells were plated in 24-well plates, at 70% confluence; we performed cell transfection using Lipofectamine™ 3000 kit following the manufacturer's protocol. At 24 h post-transfection, cells were prepared for transwell migration assay.

2.8 | Transwell migration assay

In the migration experiment, A549 cells were diluted to 1×10^5 /mL with serum-free medium and seeded into the upper chamber of a transwell insert (8.0 μm pore size; CORNING Inc.), and a medium with 10% FBS was added to the lower chamber as a chemoattractant to induce A549 cell migration. After incubation at 37°C, 5% CO₂ for 48 h, the transwell chamber was taken out and the medium in the well was discarded and washed with PBS twice. The cells were then fixed in methanol for 30 min and stained with crystal violet (Beyotime) for 15 min. After staining, the upper immobile cells were slightly wiped off with a cotton swab, observed, and photographed by the microscope. The number of migrated cells was computed with ImageJ software.

2.9 | Tumor sphere formation assay

The validated *CDR2L/ALDH6A1* and their mutant plasmids were co-transfected with pMD2.G and psPAX2 into 293FT cells, and the supernatant was filtered after 2 days of culture. Then, A549 cells were infected with the supernatant and screened with puromycin. Stably expressed cells were selected for subsequent tumor sphere formation assay. For assessing the sphere-forming ability, 1×10^3 cells were seeded in six-well ultra-low attachment plates (Corning)



in serum-free medium containing DMEM (Gibco; Thermo Fisher Scientific, Inc.) supplemented with 20 ng/mL basic fibroblast growth factor (bFGF; HARVEYBIO), 20 ng/mL epidermal growth factor (EGF; Invitrogen), and B27 supplement (Invitrogen). Sphere size and number were measured after 7 days of seeding. Images and numbers of tumor spheres were taken and counted with the use of KEYENCE BZ-X800LE microscope (KEYENCE, Osaka, Japan). Tumor spheres greater than 75 μm were counted.

2.10 | Statistical analysis

After collecting the number of migration cells, statistical analyses were performed in GraphPad Prism software version 8.0 for Windows (GraphPad Prism Software, San Diego, CA, USA). The data were expressed as the means \pm standard deviation (SD). Statistical significance between the two groups was estimated by Student's *t*-test. Differences with *p* value <0.05 were considered to be statically significant.

3 | RESULTS

3.1 | Sequencing, assembly and annotation of the Asian elephant genome

A total of 2525.21 million (78.18 \times) stLFR clean reads were obtained for genome assembly. K-mer analysis ($k=17$) estimated size of 3.54 Gb genome (Figure S1, Table S1). In general, the average GC contents of the Asian elephant were similar to other mammals and the GC contents are $\sim 40\%$ (Figure S2). Then, Hi-C sequencing data were used to anchor the scaffolds and contigs into chromosomes, which yielded 733.58 million valid Hi-C pair reads with $\sim 97.12\%$ of the bases successfully aligned to 28 chromosomes ($2n=56$). Approximately 98.21% of reads from the fragment library could be aligned to the new assembly genome (Figure 1a, Table 1, Figure S3, Tables S2–S4). BUSCO analysis of genome assembly showed high BUSCO scores (94.1%) (Figure 1b, Table S5), reflecting that the majority of the assemblies were with high quality and continuity (Figure 1c). The new Asian elephant genome assembly shows a 20-fold increase of scaffold N50 length (Table 2) compared to with previous Asian elephant genome assembly generated using short reads (Tollis et al., 2021). Moreover, gaps in the new genome have been largely reduced and most evaluation indexes of genome quality show the new genome assembly is of better quality (Table 2).

Repetitive sequences account for 68.05% of the Asian elephant genome (Table S6), with DNA transposon, long interspersed nuclear elements (LINE), short interspersed nuclear elements (SINE), and long terminal repeats (LTR) in the genome was 1.29%, 47.59%, 4.54%, and 11.38%, respectively (Figure 1d, Table S7). Besides, the number of miRNA, rRNA, snRNA, and tRNA is 2429, 312, 1284, and 21,499 respectively (Table S8). A total of 21,955 genes, of which 96.48% on the 28 assembled chromosomes were predicted

(Figure 1e–g). In addition, approximately 93.07%, 84.70%, 94.38%, and 92.50% of protein-coding genes could be assigned functions based on Swissprot, KEGG, TrEMBL, and Interpro databases, respectively (Tables S9 and S10).

3.2 | Gene family evolution

We identified 318 significantly expanded and one contracted gene families in the ancestral branch leading to Asian elephant and Africa bush elephant (Figure 2a,b, Tables S11 and S12). Expanded families enriched in signaling pathways, such as Ras-associated protein-1 (Rap1) signaling pathway (corrected *p* value = 3.84×10^{-18}), Phospholipase D signaling pathway (corrected *p* value = 9.91×10^{-18}), calcium signaling pathway (corrected *p* value = 2.93×10^{-8}), pentose and glucuronate interconversions pathway (corrected *p* value = 1.01×10^{-6}), and B-cell receptor signaling pathway (corrected *p* value = 4.30×10^{-3}) (Figure 2c, Figure S5a, Tables S13–S15). *Rap1* is an important regulator of cellular migration and polarization as a small GTPase in the Ras-related protein family. Previous studies have suggested that *Rap1* has an important role in tumor development (Alemayehu et al., 2013; Lin et al., 2015; Maxson et al., 2013; Yang et al., 2015) and tumorigenesis (Shah et al., 2018; Yang et al., 2017). Active *Rap1* inhibits tumor invasion and metastasis in bladder, lung, and brain (Lyle et al., 2008; Vallés et al., 2004).

A total of 251 significantly expanded and 59 significantly contracted gene families were identified in Asian elephant (Figure 2b, Table S12). Expanded gene families in Asian elephant lineage were mostly enriched in similar categories that enriched by the expanded families in the ancestral lineages of elephants (Figure 2d, Figure S5b, Tables S16 and S17). However, there are some unique enriched pathways in Asian elephant (Table S17), for example, base excision repair pathway (corrected *p* value = 3.10×10^{-4}), apoptosis pathway (corrected *p* value = 2.30×10^{-3}), and longevity regulating pathway (corrected *p* value = 9.00×10^{-3}). In addition, we found functional KDM4 family duplicated in Asian elephant lineage (Table S18). The KDM4 family is one of KDMs subfamily and plays important roles in many different processes, including regulation of gene transcription, epigenetic silencing, and DNA repair (Kim et al., 2012). More so, irregularity of KDM4 protein can increase the risk of oncogenesis (Chi et al., 2010; Cloos et al., 2008).

3.3 | Positively selected genes in elephants

Of 4444 single-copy orthologous from 12 mammals, we identified 618 positively selected genes (PSGs) in the ancestral lineage leading to Asian elephant and African bush elephant (Table S19). Eight PSGs (*DOK4*, *C19orf47*, *TRAF6*, *XPO1*, *WDR7*, *DOK2*, *CYLD*, and *LRIT1*) involved in mitogen activated protein kinases (MAPK) signaling pathway (*p* value = 0.04) (Figure S6a,b, Table S20). Interestingly, MAPKs are activated in response to a variety of stimuli like UV damage and oxidative stress (Chang & Karin, 2001; Corre et al., 2017).

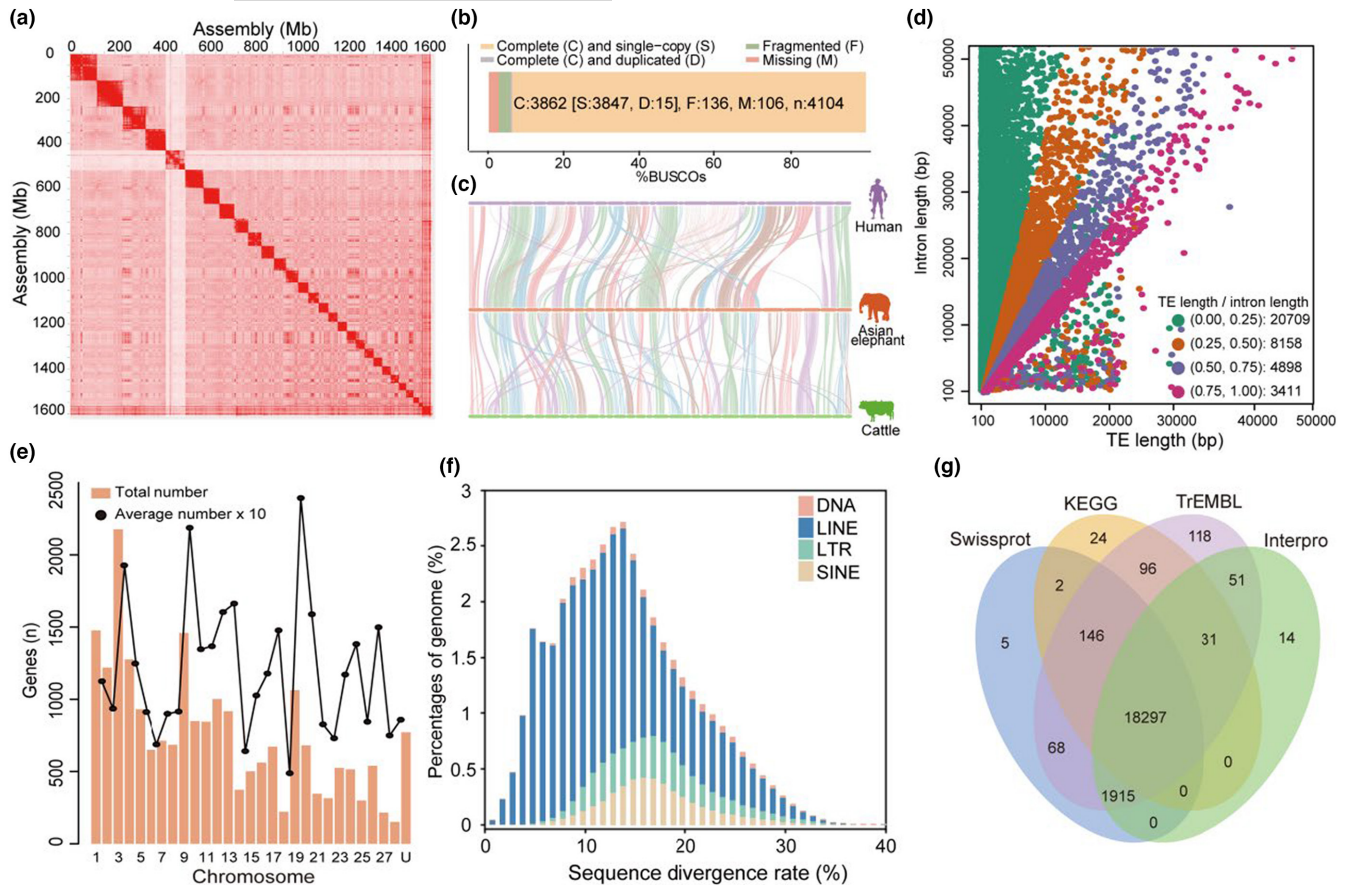


FIGURE 1 Asian elephant genome assembly and annotation. (a) Genome-wide heatmap of Hi-C interactions among 28 chromosomes of Asian elephant. (b) Quality evaluation of the assembled genome in BUSCO software with “mammalia_odb9” dataset. (c) Chromosome synteny among cattle, human, and Asian elephant. The transposon elements (TE) and predicted genes of the Asian elephant genome: (d) Relationship between the intron length and TE length. (e) The number of predicted genes in each chromosome and solid lines denote 10 times of the average number of genes in each numbered chromosome along 10 Mb windows with a step size was 2 Mb. (f) Divergence distribution of repeat elements in the Asian elephant genome. (g) The Venn diagram of functional annotation in the predicted gene set using four protein databases (Swissprot, KEGG, TrEMBL, and Interpro).

In addition, p38MAPK activates cellular responses by regulating various targets, including the prototypical tumor suppressor p53 (Bulavin et al., 1999; Bykov et al., 2018; Hager & Gu, 2014; Mello & Attardi, 2018). Furthermore, 16 genes were enriched in Epstein-Barr virus infection (p value=0.02), of which is the first human tumor virus (Gujer et al., 2015). Other PSGs are also of interesting; for instance, Myosin-6 (*MYO6*, likelihood ratio test (LRT) p value= 6.67×10^{-11}) is significantly upregulated in prostate and breast cancer (Duan et al., 2020; Zhang et al., 2016). Tyrosine kinase 2 (*DOK2*, LRT p value <0.01) is a well-known tumor suppressor gene (Pei et al., 2021) and regulated hematopoietic progenitor cell growth and development (Coppin et al., 2015; Gugasyan et al., 2002). *DOK2* is also associated with colorectal cancer (Wen et al., 2015), lung cancer (Berger et al., 2013; Chen et al., 2019), and renal cancer (Kuřma et al., 2019; Solarek et al., 2019). Numerous reports showed that the downregulation of *DOK2* is associated with cancer development (Berger et al., 2010; Coppin et al., 2015; Ohsugi, 2017).

One of the interesting PSGs in elephants is *CDR2L* (LRT p value= 1.70×10^{-10}) (Figure 3a), which encodes cerebral

degenerated-related protein 2-like protein. Actually, both *CDR2L* and *CDR2* are tumor expression antigens, and targets of Yo-antibodies (Kråkenes et al., 2019). They are associated with a disorder called paraneoplastic cerebellar degradation (PCD), a condition in which the immune system indiscriminately targets tumor antigens, leading to acute cerebellar degradation. *CDR2L* protein appeared to be the main target of Yo-antibodies in PCD-tumors and has been implicated in ovarian and breast cancers (Eichler et al., 2013; Raspotnig et al., 2017, 2022).

We identified one potential positively selected site of *CDR2L* gene in the ancestral lineage leading to Asian elephant and African bush elephant; the residue at position 167 of *CDR2L* gene has a Serine to Arginine substitution in two elephants. This substitution is unique to African bush elephant and Asian elephant while the residue of other mammals with whole-genome resource available is Serine (Figure S7).

We further performed transwell assay to validate the potential ability of residues at positively selected sites (position 167 of the *CDR2L* gene has a Serine to Arginine substitution in two elephants) of



TABLE 1 Statistics of stLFR and Hi-C sequencing for the Asian elephant genome assembly.

Features	stLFR		Hi-C	
	Contigs	Scaffolds	Contigs	Scaffolds
Total number	89,150	58,642	89,150	57,037
Total length	3,132,914,974	3,226,711,689	3,132,914,974	3,227,514,189
Gap N	/	93,796,715	/	94,599,215
Average length	35,142.06	55,023.91	/	/
N50	200,987	40,156,848	200,987	118,153,099
N90	45,879	4,887,711	45,879	73,467,117
Maximum length	1,518,750	163,308,399	1,518,750	237,448,774
Minimum length	48	102	48	102
GC content	40.82	40.82	/	/

TABLE 2 Summary of Asian elephant genome assembly generated in this study compared with existing Asian elephant genome assembly (Tollis et al., 2021).

Features	Our genome data		Tollis et al. (2021)	
	Contigs	Scaffolds	Contigs	Scaffolds
Assembly length	3.13 Gb	3.23 Gb	2.98 Gb	3.13 Gb
Longest	1.52 Mb	163.31 Mb	731 kb	14.6 Mb
Number	89,150	58,642	90,662	6954
N50	201.0 kb	40.16 Mb	79.8 kb	2.77 Mb
L50	4602	24	10,736	336
Percent genome in gaps	0.00	2.90	0.09	4.88
BUSCO results	C: 93.7% [D: 0.4%] F: 3.3%, M: 2.6%, n: 4104		C: 91.5% [D: 0.4%] F: 5.7%, M: 2.8%, n: 4104	

Abbreviations: BUSCO, Benchmarking Universal Single Copy Orthologs; C, complete; D, duplicated; F, fragmented; M, missing.

CDR2L gene (Figure 3b,c). We expressed elephant *CDR2L* gene in A549 cells and measured the in vitro migration ability. The migration assays showed that the number of A549 cells expressing elephant *CDR2L* (Ema-*CDR2L*) that migrated through the transwell polycarbonate filter was significantly lower than that of cells expressing the PEGFP control. Then, we replaced the residue at position 167 (Arginine) in elephant *CDR2L* with Serine and found that the number of A549 cells in mutant group (Ema-*CDR2L*-mut) is higher than elephant group (Ema-*CDR2L*) but still lower than PEGFP control group ($p < 0.001$) (Figure 3d,e, Figure S8a), suggesting that residue replacement at position 167 of elephant *CDR2L* represents a potential ability to cancer resistance.

In parallel, we generated the plasmid of mouse *CDR2L* (Mus-*CDR2L*, position 167 of the *CDR2L* gene is Serine) and plasmid of mouse *CDR2L* mutant (Mus-*CDR2L*-mut, position 167 of the *CDR2L* gene has a replacement from Serine to Arginine) for further assessment. Our results showed that the migration of A549 cells significantly reduced when we expressed the mouse *CDR2L* (Mus-*CDR2L*) compared with PEGFP control, and the migrated number of A549 cells is less than those in mouse group (Mus-*CDR2L*) when we expressed the mutant mouse group (Mus-*CDR2L*-mut) ($p < 0.001$) (Figure 3e,f, Figure S10a).

Then, we also generated the plasmid of human *CDR2L* (Human-*CDR2L*, position 167 of the *CDR2L* gene is Serine) and plasmid of

human *CDR2L* mutant (Human-*CDR2L*-mut, position 167 of the *CDR2L* gene has a substitution from Serine to Arginine) for additional validation. Our transwell results showed that the migrated number of A549-Human-*CDR2L* cells was lower than that of PEGFP control cells, although this difference was not statistically significant. Besides, the migrated number of A549 cells in mutant group (Human-*CDR2L*-mut) was significantly lower than that of cells expressing PEGFP control and human *CDR2L* (Human-*CDR2L*) ($p < 0.001$) (Figure S11a,b). Altogether, these findings suggest that the substitution of *CDR2L* in elephants can attenuate A549 cells migration.

Furthermore, we performed in vitro tumor sphere formation assay as a surrogate to evaluate the capacity of tumor inhibition in the substitutions of *CDR2L*. We found that the number of tumor sphere formation was significantly reduced when we expressed the elephant *CDR2L* (Ema-*CDR2L*) compared with the control group ($p < 0.05$) (Figure S12a,b). Besides, we also expressed the mutant elephant *CDR2L* (position 167 Arginine was replaced with Serine) in A549 cells to evaluate the ability of tumor inhibition. This produced more tumor spheres than the control group (A549) and the elephant *CDR2L* group (Ema-*CDR2L*). Similarly, we found that the number of tumor spheres was significantly reduced when expressing mutant mouse (position 167 Serine was replaced with Arginine) compared

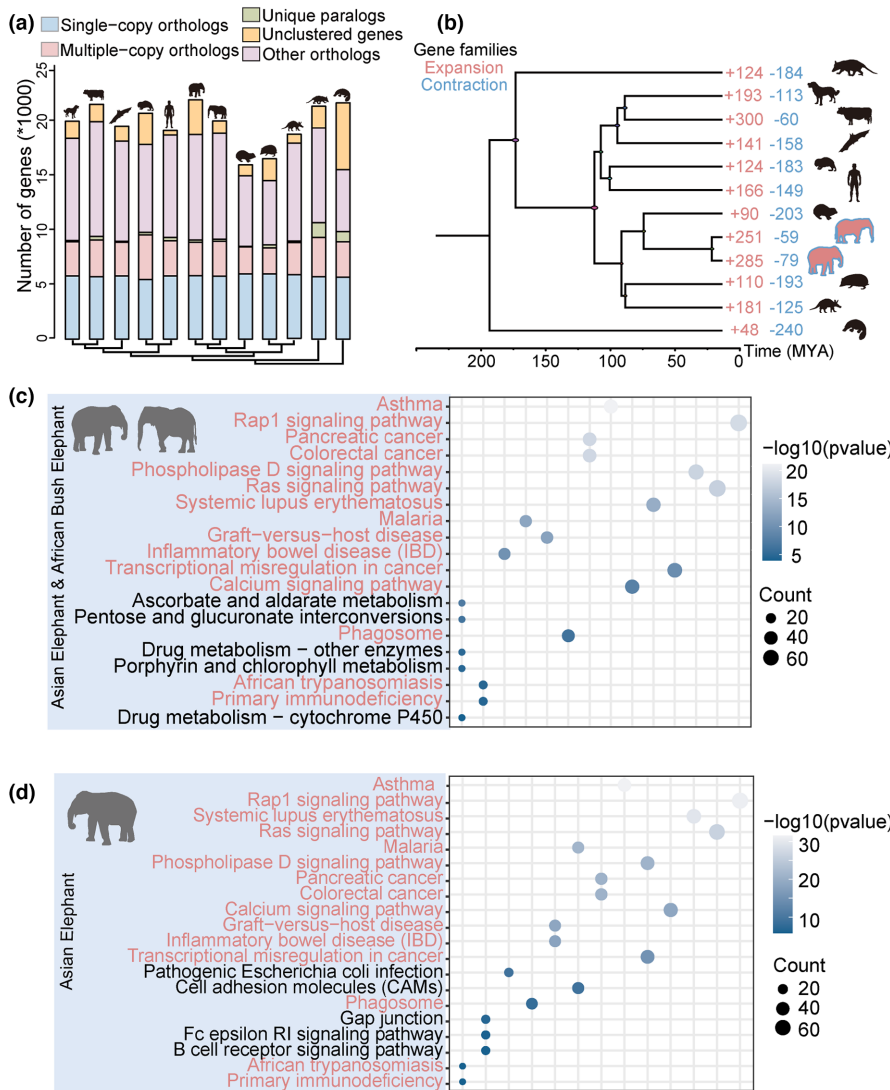


FIGURE 2 Gene family evolution and enrichment of elephants. (a) The number of homolog genes in the 12 mammals analyzed in this study. (b) Phylogenetic tree and divergence time estimated for the Asian elephant and other mammals. The number of significantly expanded (red) and contracted (blue) gene families is designated on each branch. Top 20 KEGG pathways most significantly enriched with expanded gene families in Asian elephant and African bush elephant (c) and Asian elephant only (d).

to expressing mouse CDR2L (Mus-CDR2L) and control group (A549) ($p < 0.05$) (Figure S12a,b). Altogether, tumor sphere formation assays suggested that the substitution of CDR2L in elephants enhanced the tumor inhibition ability.

3.4 | Positively selected genes in four long-lived mammals

To explore genes under positive selection across distant long-lived taxa, we enlarged our test of foreground branch to include other two long-lived mammals: naked mole rat and greater horseshoe bat (Figure 4a). The selective pressure analysis found signatures of positive selection on 132 genes in four long-lived mammals (Table S21). Among them, 79.5% (105 genes) are also under positive selection in the ancestral branch of elephants (Figure S8). These genes are significantly enriched in basal cell carcinoma pathway (p value=0.007), other types of O-glycan biosynthesis pathway (p value=0.02), gap junction (p value=0.02), and adherens junction (p value=0.03) (Figure S6c,d, Table S22). Basal cell carcinoma (BCC)

is one of the most common skin malignancies worldwide (Niculet et al., 2022). The development of BCC is the result of a complex interplays between environmental, phenotypic, and genetic factors, with ultraviolet radiation considered to be the main risk factor (Dika et al., 2020). Previous studies have shown that UV exposure leads to deleterious effects such as skin aging and cancer through generating cellular reactive oxygen species and DNA damage (Yu & Lee, 2017). Several PSGs may involve in tumorigenesis (Table S21). For instance, COP9 signalosome subunit 3 (COPS3, LRT p value= 1.11×10^{-6}) is closely associated with tumor development (Both et al., 2016; van Dartel & Hulsebos, 2004) and knockdown of COPS3 significantly downregulated MEK signaling, reducing metastasis of osteosarcoma cells (Zhang et al., 2018). Guanosine monophosphate reductase (GMPR, LRT p value= 0.05×10^{-2}) is another positively selected gene detected in long-lived mammals, which is a potential tumor suppressor that inhibits the regulatory pathway in tumor cells (Wawrzyniak et al., 2013). We have identified the selective signal in metal responsive transcription factor 1 (MTF1, LRT p value= 9.85×10^{-5}), which is upregulated in malignant ovarian cancer and might contribute to ovarian tumor metastasis.

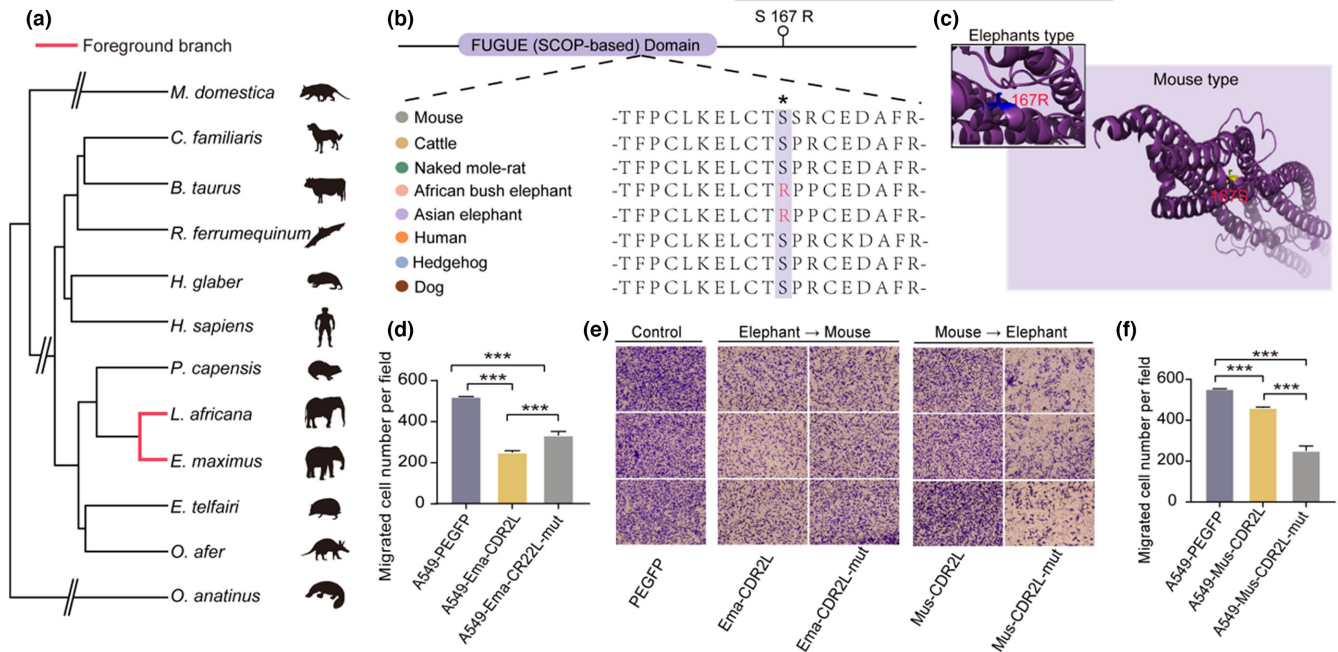


FIGURE 3 Positively selected genes in Asian elephant and African bush elephant. (a) Phylogenetic tree of 12 mammals in comparative analysis with red line represent the foreground branch (Asian elephant and African bush elephant). Specific mutations of elephants in the CDR2L gene. (b) The CDR2L protein sequence of multiple species (indicated with different colors) was aligned. (c) 3D structure simulation of elephants' type compared with that of short-lived mammals. Functional cellular assays of CDR2L gene in short and long-lived mammals' group. (d) Number of migrated cells of A549-PEGFP (control), A549-Ema-CDR2L (Asian elephant type), and A549-Ema-CDR2L-mut (mouse type) group in the A549 cells. (e) Migration ability of A549-PEGFP (control), A549-Ema-CDR2L (Asian elephant type), A549-Ema-CDR2L-mut (mouse type), A549-Mus-CDR2L (mouse type), and A549-Mus-CDR2L-mut (Asian elephant type) group in the A549 cells. (f) Number of migrated cells of A549-PEGFP (control), A549-Mus-CDR2L (mouse type), and A549-Mus-CDR2L-mut (Asian elephant type) group in the A549 cells. **p* < 0.05, ***p* < 0.01, ****p* < 0.001, ns: Non-significant.

Over recent years, there has been growing interest in genome-wide sequencing and comparison of long-lived mammals (Keane et al., 2015; Kim et al., 2011; Seim et al., 2013). Genome sequencing found 45 genes were under positive selection in the naked mole rat (Kim et al., 2011). Interestingly, two of them (*RBM4* and *RBM28*) were detected to be positively selected in four long-lived mammals by this study. Moreover, signatures of positive selection were also detected from other RNA-binding motif proteins (RBM), that is, *RBM14* (LRT *p* value = 1.87×10^{-9}), *RBM27* (LRT *p* value = 0.01×10^{-2}), and *RBM33* (LRT *p* value = 1.14×10^{-10}). The RNA-binding motif proteins are a class of important intracellular proteins, and are associated with the occurrence and development of cancers (Jögi et al., 2009; Zhang et al., 2014). For example, *RBM4* inhibits the apoptosis of breast cancer cells by upregulating the expression of IR-B and MCL-1S (Lin et al., 2014) and *RBM33* participates in promoting proliferation in various cancers (Li et al., 2021). Therefore, RBM protein family might serve as the common targets for antagonizing cancer in long-lived mammals. Vicens and Posada (2018) have assessed the selective pressures on 430 cancer related genes and showed evidence of positive selection on *CSF3R* (Vicens & Posada, 2018), which is the receptor for colony-stimulating factor 3. Mutations in *CSF3R* are common in patients with chronic neutrophilic leukemia or atypical chronic myeloid leukemia (Maxson et al., 2013), and this gene has been also detected under positive selection in four long-lived

mammals (LRT *p* value = 5.00×10^{-4}). Seim et al. (2013) have examined the positively selected genes in the longest-lived bat species, Brandt's bat and found myotilin (*MYOT*) was under positive selection. Our analyses evidenced the signature of selection of this gene (LRT *p* value = 1.67×10^{-8}) (Table S21). *MYOT* belongs to a small protein family of immunoglobulin (Ig) domain-containing proteins in the Z-line associated with the actin cytoskeleton (Otey et al., 2009), where *MYOT* is usually expressed in the heart and involved in muscular dystrophy. Previous observations have shown myotilin to be significantly elevated with age (Han et al., 2022).

3.5 | Functional assay of common substitutions in *ALDH6A1*

One of the PSGs that attracts our interest is *ALDH6A1* (LRT *p* value = 8.80×10^{-5}), which encodes mitochondrial methylmalonate semialdehyde dehydrogenase protein (Figure 4b,c). Members of the aldehyde dehydrogenase superfamily (ALDHs) are a group of oxidizing enzymes that function in diverse cellular activities including aldehyde oxidation, detoxification, and antioxidants (Jackson et al., 2011; Marchitti et al., 2008; Shortall et al., 2021; Vasiliou & Nebert, 2005). Deficiencies or mutations in ALDH have been implicated in various forms of cancer and metabolic disorders (Lu

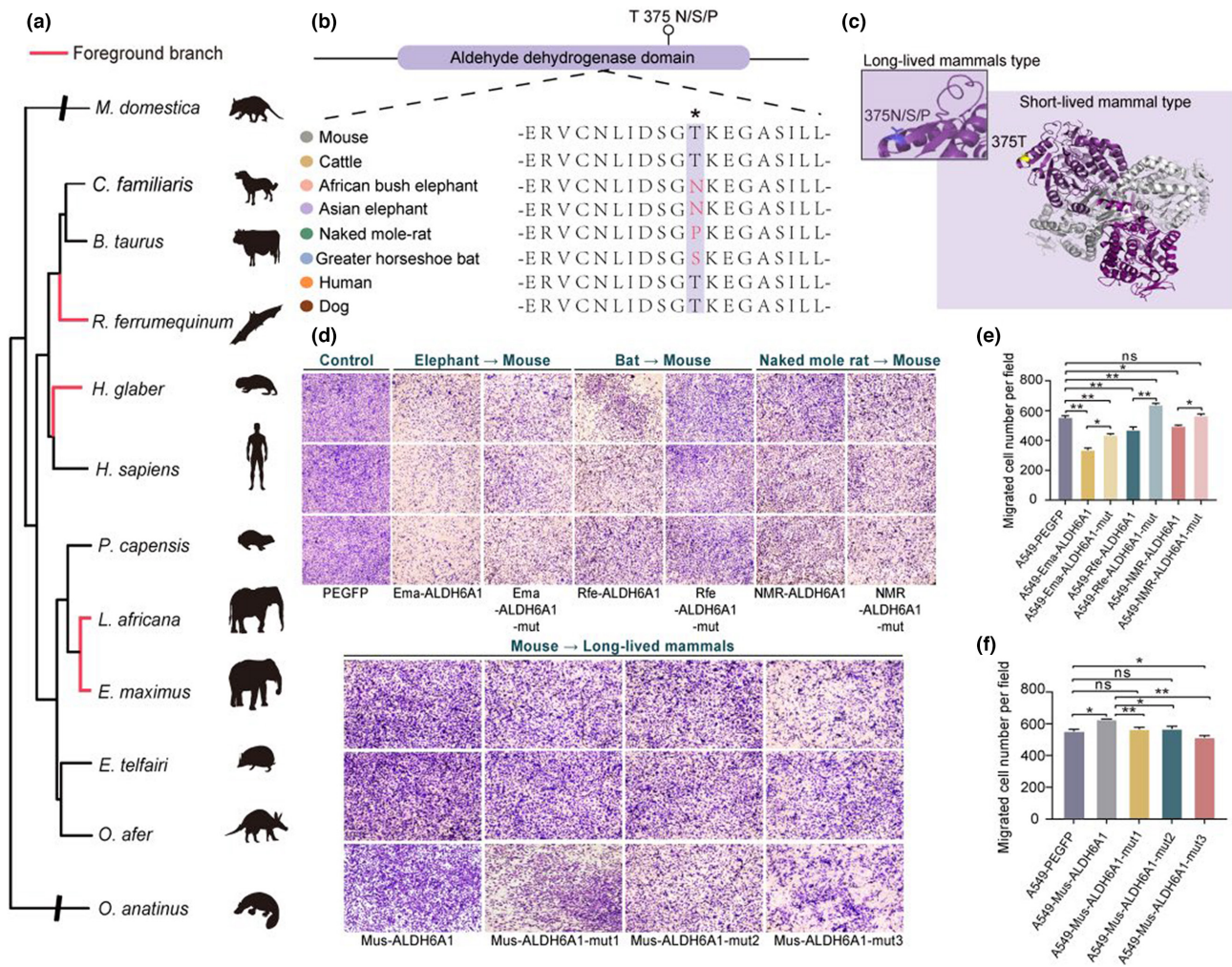


FIGURE 4 Positively selected genes in long-lived mammals. (a) Phylogenetic tree of 12 mammals in comparative analysis with red line representing the foreground branch (Asian elephant, African bush elephant, greater horseshoe bat, and naked mole rat). Specific mutations of long-lived mammals in the *ALDH6A1* gene including one mutation in the Aldehyde dehydrogenase domain. (b) The *ALDH6A1* protein sequence of multiple species (indicated with different colors) was aligned. (c) 3D structure simulation of long-lived mammal's type compared with that of short-lived mammals. Functional cellular assays of *ALDH6A1* gene in short and long-lived mammals. (d) Migration ability of in the A549-PEGFP (control), A549-Ema-*ALDH6A1* (Asian elephant type) and A549-Ema-*ALDH6A1*-mut (mouse type), A549-Rfe-*ALDH6A1* (greater horseshoe bat type) and A549-Rfe-*ALDH6A1*-mut (mouse type), A549-NMR-*ALDH6A1* (naked mole rat type), A549-NMR-*ALDH6A1*-mut (mouse type), A549-Mus-*ALDH6A1* (mouse type), A549-Mus-*ALDH6A1*-mut1 (Asian elephant type), A549-Mus-*ALDH6A1*-mut2 (greater horseshoe bat type) and A549-Mus-*ALDH6A1*-mut3 (naked mole rat type) group in the A549 cells. (e) Number of migrated cells in the A549-PEGFP (control), A549-Ema-*ALDH6A1* (Asian elephant type) and A549-Ema-*ALDH6A1*-mut (mouse type), A549-Rfe-*ALDH6A1* (greater horseshoe bat type) and A549-Rfe-*ALDH6A1*-mut (mouse type), A549-NMR-*ALDH6A1* (naked mole rat type) and A549-NMR-*ALDH6A1*-mut (mouse type) group in the A549 cells. (f) Number of migrated cells migration ability in the A549-PEGFP (control), A549-Mus-*ALDH6A1* (mouse type), A549-Mus-*ALDH6A1*-mut1 (Asian elephant type), A549-Mus-*ALDH6A1*-mut2 (greater horseshoe bat type) and A549-Mus-*ALDH6A1*-mut3 (naked mole rat type) group in the A549 cells. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns: Non-significant.

et al., 2020; Xu et al., 2015; Yu et al., 2010). *ALDH6A1* is reported to be regulated in several processes of cancer, including hepatocellular carcinoma and prostate cancer, and diabetes (Cho et al., 2018; Dharuri et al., 2014; Lu et al., 2020). Specifically, *ALDH6A1* is regulated by a well-known suppressor transcription factor HNF4A, which suppressed tumorigenic capability in clear renal cell carcinoma (Lu et al., 2020). Another study showed that inhibition of *ALDH6A1* may be strongly associated with abnormal proliferation of liver cancer cells (Shin et al., 2020).

We detected one positively selected site of *ALDH6A1* in four long-lived mammals, and we found that the residue of the positively selected site is different in long-lived mammals. The residue of elephants, greater horseshoe bat, and naked mole rat are respectively Asparagine (N), Serine (S), and Proline (P) while the residue of most mammals is Threonine (T) (Figure S9). This residue is specific in elephants compared to other mammals, while the residue of greater horseshoe bat is also unique when compared to other bats with the Threonine (T) residue. We also found guinea



pig also have Asparagine at the site, which is identical to naked mole rate and might be related to the close phylogenetic relationship between these two species.

We then generated plasmids that express *ALDH6A1* of four long-lived mammals (Asian elephant and African bush elephant: Ema-*ALDH6A1*, greater horseshoe bat: Rfe-*ALDH6A1*, naked mole rat: NMR-*ALDH6A1*) and mouse (Mus-*ALDH6A1*). As we expected, the A549 cells that with overexpressed long-lived mammals *ALDH6A1* showed inhibited migration compared to A549 cells transfected with the PEGFP control and mouse *ALDH6A1* (Figure 4d,e, Figure S10b). This not only evidenced that *ALDH6A1* overexpression could inhibit growth and migration of cancer cells (Cai et al., 2022), but also indicated the *ALDH6A1* of long-lived mammals show stronger inhibition of cellular migration of cancer cells. However, there are some subtle differences in migration ability among cells expressed four long-lived mammals *ALDH6A1*. In particular, the migration of A549 cells that expressed elephants *ALDH6A1* is less than A549 cells expressed greater horseshoe bat and naked mole rat *ALDH6A1*, suggesting the elephant *ALDH6A1* has the highest inhibition effects. Next, we created three mutants in *ALDH6A1* of long-lived mammals for additional validation, that is, Ema-*ALDH6A1*-mut, Rfe-*ALDH6A1*-mut, and NMR-*ALDH6A1*-mut, by replacing residues under positive selection in long-lived mammals into mouse residue. Expectedly, we observed a higher cell migration in A549 cells transfected with these mutants, providing additional validation on the role of positively selected sites in long-lived mammals. In vice versa, three plasmids' (Mus-*ALDH6A1*-mut1, Mus-*ALDH6A1*-mut2, Mus-*ALDH6A1*-mut3) mutants were also generated using mouse *ALDH6A1* plasmid by substituting residue at the site of selection to residue of the long-lived mammals. Interestingly, A549 cells transfected with these mutants have shown suppressed cell migration than the Mus-*ALDH6A1* group (Figure 4d,f, Figure S10b).

Besides, we also generated plasmids that express *ALDH6A1* of human (A549-Human-*ALDH6A1*) for the transwell assays. As we expected, the A549 cells with overexpressing human *ALDH6A1* showed inhibited migration compared to A549 cells transfected with the PEGFP control (A549-PEGFP). Furthermore, three plasmids' (Human-*ALDH6A1*-mut1, Human-*ALDH6A1*-mut2, Human-*ALDH6A1*-mut3) mutants were also generated using human *ALDH6A1* plasmid by substituting residue at the site of selection to residue of the long-lived mammals. The transwell results showed that A549 cells transfected with these mutants have shown suppressed cell migration than the Human-*ALDH6A1* group. However, there are also some subtle differences in migration ability among cells expressed three mutants. In particular, the migration of A549 cells that expressed Human-*ALDH6A1*-mut1 (Asian elephant type) is less than A549 cells that expressed Human-*ALDH6A1*-mut2 (greater horseshoe bat type) and Human-*ALDH6A1*-mut3 (naked mole rat type) (Figure S11a,c), suggesting the elephant *ALDH6A1* have the higher inhibition effects which is consisted with the transwell results of A549 cells overexpressing long-lived mammals *ALDH6A1*.

Taken together, our results reveal that the common residues at positively selected sites in *ALDH6A1* of long-lived mammals offer an enhanced function in resisting cancer progression.

4 | DISCUSSION

One of the major restrictions on the evolution of large body sizes across mammals is the high risk of cancer incidence due to the hypothesis that the increase in cell division may bring an increment cancer risk. However, there is no correlation between large body size and the risk of cancer development. Therefore, large mammals might have evolved enhanced cancer resistance mechanisms to suppress the cancer incidence and extend their lifespans as well (Omotoso et al., 2021). For example, elephants are the mammals with large body size (~5500 kg); however, recent studies show that elephants are resistant to cancer, with an estimated cancer mortality rate of 4.81%, compared to 11%–25% for humans (Abegglen et al., 2015). At the same time, multiple copies of TP53, most of them are pseudocopies, were found in elephant genome (Abegglen et al., 2015; Sulak et al., 2016). Subsequent study further found refunctionalizing of another pseudogene, LIF, may mediate the cancer resistance and body size in elephants (Vazquez et al., 2018). Given the long generation time of elephants, these results implied that the expanded pseudocopies of cell cycle check-point genes are key responses to cancer resistance in elephants. In this study, by performing gene family analysis in the ancestral lineage of elephants, we found genes involved in several other pathways, such as Ras-associated protein-1 (Rap1) signaling pathway and base excision repair pathway, were also expanded in elephant genomes. This could relate to cancer resistance because base excision repair pathway is the predominant repair pathway in mammalian cells and is a critical process of genome maintenance, with cancer, premature aging and metabolic disorders, were observed in animals lacking base excision repair function (Mostoslavsky et al., 2006; Vartanian et al., 2006). In addition, several positively selected genes, for example, *CDR2L*, might also involve cancer resistance in elephants. These analyses and results would indicate the large body size and low tumor incidence in elephant is more complex than solely on contributions from pseudogenes.

Another interesting question is whether such cancer resistance in diverse long-lived lineages could be achieved by common approach, that is, "convergent" evolution. In this study, we screened positive selection genes and examined the cellular effect of common substitutions in *ALDH6A1* of long-lived mammals, hence provide a piece of evidence that cancer-resistant substitutions can be utilized by long-lived mammals. However, this does not mean that tumor resistance is convergently evolved in long-lived mammals as it is unknown what are the common benefits brought by such "convergent" phenotype. Moreover, the results could be challenged by including more long-lived or short-lived species. This is also understandable as there is no consistent definition for "long-lived," or strong correlation between longevity and tumor incidence. Nevertheless, the analyses and results presented here could serve as a preliminary survey of common substitutions response to cancer resistance in long-lived animals.

AUTHOR CONTRIBUTIONS

X.Z. conceived the study and designed the project. X.L. and P.W. performed experiments, completed analysis, and wrote the manuscript;



J.D., Z.L., and M.L., prepared the cell cultures; Q.P. implement the data analysis and generated structural figures; X.L., P.W., Q.P., G.L., O.O., W.L., and X.Z. discussed the results and revised this manuscript; all authors contributed to data interpretation.

ACKNOWLEDGMENTS

We thank Fangyi Zhou, Mingwei Bao, and Yisi Hu for supporting sample collection and preparation. This project was funded by grants from the National Natural Science Foundation of China (82050002 and 32270437), Beijing Natural Sciences Foundation (JQ19022), and the National Key Research and Development Projects of the Ministry of Science and Technology of China (2021YFC2301300).

CONFLICT OF INTEREST STATEMENT

The authors declare no competing financial interests.

DATA AVAILABILITY STATEMENT

Genome and RNA sequencing data from this study have been deposited in BioProject at NCBI (PRJCA013758), and the Genome Sequence Archive in National Genomics Data Center, China, National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (CRA009178).

ORCID

Xuanjing Li  <https://orcid.org/0000-0002-1277-4922>

Xuming Zhou  <https://orcid.org/0000-0002-1100-6294>

REFERENCES

- Abegglen, L. M., Caulin, A. F., Chan, A., Lee, K., Robinson, R., Campbell, M. S., Kiso, W. K., Schmitt, D. L., Waddell, P. J., Bhaskara, S., Jensen, S. T., Maley, C. C., & Schiffman, J. D. (2015). Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *JAMA*, *314*(17), 1850–1860.
- Alemayehu, M., Dragan, M., Pape, C., Siddiqui, I., Sacks, D. B., Di Guglielmo, G. M., Babwah, A. V., & Bhattacharya, M. (2013). β -Arrestin2 regulates lysophosphatidic acid-induced human breast tumor cell migration and invasion via Rap1 and IQGAP1. *PLoS One*, *8*(2), e56174.
- Alexa, A., & Rahnenf hrer, J. (2009). Gene set enrichment analysis with topGO. *Bioconductor Improv*, *27*, 1–26.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, T. S. S., & Eppig, J. T. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29.
- Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, *6*, 1–6.
- Bao, Z., & Eddy, S. R. (2002). Automated De novo identification of repeat sequence families in sequenced genomes. *Genome Research*, *12*(8), 1269–1276.
- Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573–580.
- Berger, A. H., Chen, M., Morotti, A., Janas, J. A., Niki, M., Bronson, R. T., Taylor, B. S., Ladanyi, M., van Aelst, L., & Politi, K. J. P. O. (2013). DOK2 inhibits EGFR-mutated lung adenocarcinoma. *PLoS One*, *8*(11), e79526.
- Berger, A. H., Niki, M., Morotti, A., Taylor, B. S., Socci, N. D., Viale, A., Brennan, C., Szoke, J., Motoi, N., Rothman, P. B., Teruya-Feldstein, J., Gerald, W. L., Ladanyi, M., & Pandolfi, P. P. (2010). Identification of DOK genes as lung tumor suppressors. *Nature Genetics*, *42*(3), 216–223.
- Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Research*, *14*(5), 988–995.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., & Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, *31*(1), 365–370.
- Both, J., Wu, T., ten Asbroek, L. M. A., Baas, F., & Hulsebos, T. J. M. (2016). Oncogenic properties of candidate oncogenes in chromosome region 17p11. 2p12 in human osteosarcoma. *Cytogenetic and Genome Research*, *150*(1), 52–59.
- Buffenstein, R. (2008). Negligible senescence in the longest living rodent, the naked mole-rat: Insights from a successfully aging species. *Journal of Comparative Physiology B*, *178*, 439–445.
- Buffenstein, R., & Jarvis, J. U. M. (2002). The naked mole rat—a new record for the oldest living rodent. *Science of Aging Knowledge Environment*, *2002*(21), pe7.
- Bulavin, D. V., Saito, S., Hollander, M. C., Sakaguchi, K., Anderson, C. W., Appella, E., & Fornace-Jr, A. J. (1999). Phosphorylation of human p53 by p38 kinase coordinates N-terminal phosphorylation and apoptosis in response to UV radiation. *The EMBO Journal*, *18*(23), 6845–6854.
- Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., & Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, *31*(12), 1119–1125.
- Bykov, V. J., Eriksson, S. E., Bianchi, J., & Wiman, K. G. (2018). Targeting mutant p53 for efficient cancer therapy. *Nature Reviews Cancer*, *18*(2), 89–102.
- Cai, Y., Zeng, R., Peng, J., Liu, W., He, Q., Xu, Z., & Bai, N. (2022). The downregulated drug-metabolism related ALDH6A1 serves as predictor for prognosis and therapeutic immune response in gastric cancer. *Aging*, *14*(17), 7038–7051.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552.
- Caulin, A. F., & Maley, C. C. (2011). Peto's paradox: Evolution's prescription for cancer prevention. *Trends in Ecology & Evolution*, *26*(4), 175–182.
- Chang, L., & Karin, M. (2001). Mammalian MAP kinase signalling cascades. *Nature*, *410*(6824), 37–40.
- Chen, M., Zhang, J., Berger, A. H., Diolombi, M. S., Ng, C., Fung, J., Bronson, R. T., Castillo-Martin, M., Thin, T. H., Cordon-Cardo, C., Plevin, R., & Pandolfi, P. P. (2019). Compound haploinsufficiency of Dok2 and Dusp4 promotes lung tumorigenesis. *The Journal of Clinical Investigation*, *129*(1), 215–222.
- Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, *5*(1), 4–10.
- Chi, P., Allis, C. D., & Wang, G. G. (2010). Covalent histone modifications—Miswritten, misinterpreted and mis-erased in human cancers. *Nature Reviews Cancer*, *10*(7), 457–469.
- Cho, S. Y., Kang, S., Kim, D. S., Na, H. J., Kim, Y. J., Choi, Y. D., & Cho, N. H. (2018). HSP27, ALDH6A1 and prohibitin act as a trio-biomarker to predict survival in late metastatic prostate cancer. *Anticancer Research*, *38*(11), 6551–6560.
- Cloos, P. A., Christensen, J., Agger, K., & Helin, K. J. (2008). Erasing the methyl mark: Histone demethylases at the center of cellular differentiation and disease. *Genes & Development*, *22*(9), 1115–1140.
- Coppin, E., Gelsi-Boyer, V., Morelli, X., Cervera, N., Murati, A., Pandolfi, P. P., Birnbaum, D., & Nunès, J. J. (2015). Mutational analysis of the DOK2 haploinsufficient tumor suppressor gene in chronic myelomonocytic leukemia (CMML). *Leukemia*, *29*(2), 500–502.



- Corre, I., Paris, F., & Huot, J. J. (2017). The p38 pathway, a major pleiotropic cascade that transduces stress and metastatic signals in endothelial cells. *Oncotarget*, *8*(33), 55684.
- De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics*, *22*(10), 1269–1271.
- de Magalhães, J. P., Costa, J., & Toussaint, O. (2005). HAGR: The human ageing genomic resources. *Nucleic Acids Research*, *33*, D537–D543.
- Dharuri, H., t'Hoën, P. A. C., Van Klinken, J. B., Henneman, P., Laros, J. F., Lips, M. A., Bouazzaoui, F. E., Van Ommen, G.-J. B., Janssen, I., Van Ramshorst, B., Van Wagenveld, B. A., Pijl, H., Van Dijk, K. W., & Van Harmelen, V. (2014). Downregulation of the acetyl-CoA metabolic network in adipose tissue of obese diabetic individuals and recovery after weight loss. *Diabetologia*, *57*, 2384–2392.
- Dika, E., Scarfi, F., Ferracin, M., Broseghini, E., Marcelli, E., Bortolani, B., Campione, E., Riefolo, M., Ricci, C., & Lambertini, M. (2020). Basal cell carcinoma: A comprehensive review. *International Journal of Molecular Sciences*, *21*(15), 5572.
- Duan, Q., Xu, M., Wu, M., Zhang, X., Gan, M., & Jiang, H. (2020). Long noncoding RNA UCA1 promotes cell growth, migration, and invasion by targeting miR-143-3p in oral squamous cell carcinoma. *Cancer Medicine*, *9*(9), 3115–3129.
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). De novo assembly of the *Aedes aegypti* genome using hi-C yields chromosome-length scaffolds. *Science*, *356*(6333), 92–95.
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. *Cell Systems*, *3*(1), 95–98.
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.
- Eichler, T. W., Totland, C., Haugen, M., Qvale, T. H., Mazengia, K., Storstein, A., Haukanes, B. I., & Vedeler, C. A. (2013). CDR2L antibodies: A new player in paraneoplastic cerebellar degeneration. *PLoS One*, *8*(6), e66002.
- Elsik, C. G., Mackey, A. J., Reese, J. T., Milshina, N. V., Roos, D. S., & Weinstock, G. M. (2007). Creating a honey bee consensus gene set. *Genome Biology*, *8*(1), 1–8.
- Gomes, N. M., Ryder, O. A., Houck, M. L., Charter, S. J., Walker, W., Forsyth, N. R., Austad, S. N., Venditti, C., Pagel, M., Shay, J. W., & Wright, W. E. (2011). Comparative biology of mammalian telomeres: Hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell*, *10*(5), 761–768.
- Gorbunova, V., & Seluanov, A. (2009). Coevolution of telomerase activity and body mass in mammals: From mice to beavers. *Mechanisms of Ageing and Development*, *130*(1–2), 3–9.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652.
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., & Bateman, A. (2005). Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Research*, *33*, D121–D124.
- Guevara-Aguirre, J., Balasubramanian, P., Guevara-Aguirre, M., Wei, M., Madia, F., Cheng, C. W., Hwang, D., Martin-Montalvo, A., Saavedra, J., Ingles, S., de Cabo, R., Cohen, P., & Longo, V. D. (2011). Growth hormone receptor deficiency is associated with a major reduction in pro-aging signaling, cancer, and diabetes in humans. *Science Translational Medicine*, *3*(70), 70ra13.
- Gugasyan, R., Quilici, C., Stacey, T. T. I., Grail, D., Verhagen, A. M., Roberts, A., Kitamura, T., Dunn, A. R., & Lock, P. (2002). Dok-related protein negatively regulates T cell development via its RasGTPase-activating protein and Nck docking sites. *The Journal of Cell Biology*, *158*(1), 115–125.
- Gujer, C., Chatterjee, B., Landtwing, V., Raykova, A., McHugh, D., & Münz, C. (2015). Animal models of Epstein Barr virus infection. *Current Opinion in Virology*, *13*, 6–10.
- Hager, K. M., & Gu, W. (2014). Understanding the non-canonical pathways involved in p53-mediated tumor suppression. *Carcinogenesis*, *35*(4), 740–746.
- Han, Y., Wennersten, S. A., Wright, J., Ludwig, R., Lau, E., & Lam, M. P. (2022). Proteogenomic characterization of age and sex interactions in cardiac gene expression. *bioRxiv*.
- Huang, Z., Jebb, D., & Teeling, E. C. (2016). Blood miRNomes and transcriptomes reveal novel longevity mechanisms in the long-lived bat, *Myotis myotis*. *BMC Genomics*, *17*(1), 1–15.
- Huang, Z., Whelan, C. V., Foley, N. M., Jebb, D., Touzalin, F., Petit, E. J., Puechmaillie, S. J., & Teeling, E. C. (2019). Longitudinal comparative transcriptomics reveals unique mechanisms underlying extended healthspan in bats. *Nature Ecology & Evolution*, *3*(7), 1110–1120.
- Ikeno, Y., Hubbard, G. B., Lee, S., Cortez, L. A., Lew, C. M., Webb, C. R., Berryman, D. E., List, E. O., Kopchick, J. J., & Bartke, A. (2009). Reduced incidence and delayed occurrence of fatal neoplastic diseases in growth hormone receptor/binding protein knockout mice. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, *64*(5), 522–529.
- Jackson, B., Brocker, C., Thompson, D. C., Black, W., Vasiliou, K., Nebert, D. W., & Vasiliou, V. (2011). Update on the aldehyde dehydrogenase gene (ALDH) superfamily. *Human Genomics*, *5*, 1–21.
- Jögi, A., Brennan, D. J., Rydén, L., Magnusson, K., Fernö, M., Stål, O., Borgquist, S., Uhlen, M., Landberg, G., Pålman, S., Pontén, F., & Jirström, K. (2009). Nuclear expression of the RNA-binding protein RBM3 is associated with an improved clinical outcome in breast cancer. *Modern Pathology*, *22*(12), 1564–1574.
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, *36*, W5–W9.
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30.
- Keane, M., Semeiks, J., Webb, A. E., Li, Y. I., Quesada, V., Craig, T., Madsen, L. B., van Dam, S., Brawand, D., Marques, P. I., Michalak, P., Kang, L., Bhak, J., Yim, H.-S., Grishin, N. V., Nielsen, N. H., Heide-Jørgensen, M. P., Oziolor, E. M., Matson, C. W., ... de Magalhães, J. P. (2015). Insights into the evolution of longevity from the bowhead whale genome. *Cell Reports*, *10*(1), 112–122.
- Kim, E. B., Fang, X., Fushan, A. A., Huang, Z., Lobanov, A. V., Han, L., Marino, S. M., Sun, X., Turanov, A. A., Yang, P., Yim, S. H., Zhao, X., Kasaikina, M. V., Stoletzki, N., Peng, C., Polak, P., Xiong, Z., Kiezun, A., Zhu, Y., ... Gladyshev, V. N. (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature*, *479*(7372), 223–227.
- Kim, T. D., Oh, S., Shin, S., & Janknecht, R. (2012). Regulation of tumor suppressor p53 and HCT116 cell physiology by histone demethylase JMJD2D/KDM4D. *PLoS One*, *7*(4), e34618.
- Kosakovsky Pong, S. L., Poon, A. F. Y., Velazquez, R., Weaver, S., Hepler, N. L., Murrell, B., Shank, S. D., Magalis, B. R., Bouvier, D., Nekrutenko, A., Wisotsky, S., Spielman, S. J., Frost, S. D. W., & Muse, S. V. (2019). HyPhy 2.5—A customizable platform for evolutionary hypothesis testing using phylogenies. *Molecular Biology and Evolution*, *37*(1), 295–299.
- Krâkenes, T., Herdlevaer, I., Raspotnig, M., Haugen, M., Schubert, M., & Vedeler, C. A. (2019). CDR2L is the major Yo antibody target in



- paraneoplastic cerebellar degeneration. *Annals of Neurology*, 86(2), 316–321.
- Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, 34(7), 1812–1819.
- Kužma, M., Killinger, Z., Jackuliak, P., Vaňuga, P., Hans, D., Binkley, N., & Payer, J. (2019). Pathophysiology of growth hormone secretion disorders and their impact on bone microstructure as measured by trabecular bone score. *Physiological Research*, 68, S121–S129.
- Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., & Ussery, D. W. (2007). RNAMmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35(9), 3100–3108.
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), 1–10.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13(9), 2178–2189.
- Li, Z., Guo, Q., Zhang, J., Fu, Z., Wang, Y., Wang, T., & Tang, J. (2021). The RNA-binding motif protein family in cancer: Friend or foe? *Frontiers in Oncology*, 11, 757135.
- Lin, J. C., Lin, C. Y., Tarn, W. Y., & Li, F. Y. (2014). Elevated SRPK1 lessens apoptosis in breast cancer cells through RBM4-regulated splicing events. *RNA*, 20(10), 1621–1631.
- Lin, K. T., Yeh, Y. M., Chuang, C. M., Yang, S. Y., Chang, J.-W., Sun, S.-P., Wang, Y.-S., Chao, K.-C., & Wang, L.-H. (2015). Glucocorticoids mediate induction of microRNA-708 to suppress ovarian cancer metastasis through targeting Rap1B. *Nature Communications*, 6(1), 1–13.
- Liu, Y. Y., Lee, C. H., Dedaj, R., Zhao, H., Mrabat, H., Sheidlin, A., Syrkinina, O., Huang, P. M., Garg, H. G., Hales, C. A., & Quinn, D. A. (2008). High-molecular-weight hyaluronan—a possible new treatment for sepsis-induced lung injury: A preclinical study in mechanically ventilated rats. *Critical Care*, 12, 1–11.
- Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25(1), 955–964.
- Lu, J., Chen, Z., Zhao, H., Dong, H., Zhu, L., Zhang, Y., Wang, J., Zhu, H., Cui, Q., Qi, C., Wang, S., Chen, S., & Shao, J. (2020). ABAT and ALDH6A1, regulated by transcription factor HNF4A, suppress tumorigenic capability in clear cell renal cell carcinoma. *Journal of Translational Medicine*, 18, 1–15.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., ... Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1(1), 1–18.
- Lyle, K. S., Raaijmakers, J. H., Bruinsma, W., Bos, J. L., & de Rooij, J. (2008). cAMP-induced Epac-Rap activation inhibits epithelial cell migration by modulating focal adhesion and leading edge dynamics. *Cellular Signalling*, 20(6), 1104–1116.
- Marchitti, S. A., Brocker, C., Stagos, D., & Vasiliou, V. (2008). Non-P450 aldehyde oxidizing enzymes: The aldehyde dehydrogenase superfamily. *Expert Opinion on Drug Metabolism & Toxicology*, 4(6), 697–720.
- Maxson, J. E., Gotlib, J., Pollyea, D. A., Fleischman, A. G., Agarwal, A., Eide, C. A., Bottomly, D., Wilmot, B., McWeeney, S. K., Tognon, C. E., Pond, J. B., Collins, R. H., Goueli, B., Oh, S. T., Deininger, M. W., Chang, B. H., Loriaux, M. M., Druker, B. J., & Tyner, J. W. (2013). Oncogenic CSF3R mutations in chronic neutrophilic leukemia and atypical CML. *New England Journal of Medicine*, 368(19), 1781–1790.
- Mello, S. S., & Attardi, L. D. (2018). Deciphering p53 signaling in tumor suppression. *Current Opinion in Cell Biology*, 51, 65–72.
- Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H.-Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, L., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., ... Finn, R. D. (2019). InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, 47, D351–D360.
- Morgan, C. C., Mc Cartney, A. M., Donoghue, M. T., Loughran, N. B., Spillane, C., Teeling, E. C., & O'Connell, M. J. (2013). Molecular adaptation of telomere associated genes in mammals. *BMC Evolutionary Biology*, 13(1), 1–13.
- Mostoslavsky, R., Chua, K. F., Lombard, D. B., Pang, W. W., Fischer, M. R., Gellon, L., Liu, P., Mostoslavsky, G., Franco, S., & Murphy, M. M. (2006). Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell*, 124(2), 315–329.
- Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29(22), 2933–2935.
- Niculescu, E., Craescu, M., Rebegea, L., Bobeica, C., Nastase, F., Lupasteanu, G., Stan, D. J., Chioncel, V., Anghel, L., Lungu, M., & Tatu, A. L. (2022). Basal cell carcinoma: Comprehensive clinical and histopathological aspects, novel imaging tools and therapeutic approaches. *Experimental and Therapeutic Medicine*, 23(1), 1–8.
- Ohsugi, T. (2017). Effects of expressing human T-cell leukemia virus type 1 (HTLV-I) oncoprotein tax on DOK1, DOK2 and DOK3 gene expression in mice. *Journal of Veterinary Medical Science*, 79(5), 935–938.
- Omotoso, O., Gladyshev, V. N., & Zhou, X. (2021). Lifespan extension in long-lived vertebrates rooted in ecological adaptation. *Frontiers in Cell and Developmental Biology*, 9, 704966.
- Otey, C. A., Dixon, R., Stack, C., & Goicoechea, S. M. (2009). Cytoplasmic Ig-domain proteins: Cytoskeletal regulators with a role in human disease. *Cell Motility and the Cytoskeleton*, 66(8), 618–634.
- Padariya, M., Jooste, M. L., Hupp, T., Fähræus, R., Vojtesek, B., Vollrath, F., Kalathiya, U., & Karakostis, K. (2022). The elephant evolved p53 isoforms that escape mdm2-mediated repression and cancer. *Molecular Biology and Evolution*, 39(7), msac149.
- Pei, S., Li, R., Meng, Y., Xi, S., Wang, Q., Yang, X., Peng, X., & Cai, J. (2021). Introduction to DOK2 and its potential role in cancer. *Physiological Research*, 70(5), 671.
- Peto, R., Roe, F. J., Lee, P. N., Levy, L., & Clack, J. (1975). Cancer and ageing in mice and men. *British Journal of Cancer*, 32(4), 411–426.
- Podlutzky, A. J., Khritankov, A. M., Ovodov, N. D., & Austad, S. N. (2005). A new field record for bat longevity. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 60(11), 1366–1368.
- Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics*, 21(1), 351–358.
- Raspotnig, M., Haugen, M., Thorsteinsdottir, M., Stefansson, I., Salvessen, H. B., Storstein, A., & Vedeler, C. A. (2017). Cerebellar degeneration-related proteins 2 and 2-like are present in ovarian cancer in patients with and without Yo antibodies. *Cancer Immunology, Immunotherapy*, 66, 1463–1471.
- Raspotnig, M., Kråkenes, T., Herdlevær, I., Haugen, M., & Vedeler, C. (2022). Expression of cerebellar degeneration-related proteins CDR2 and CDR2L in human and rat brain tissue. *Journal of Neuroimmunology*, 362, 577766.
- Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Functammasan, A., & Kim, J. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592(7856), 737–746.
- Seim, I., Fang, X., Xiong, Z., Lobanov, A. V., Huang, Z., Ma, S., Feng, Y., Turanov, A. A., Zhu, Y., & Lenaz, T. L. (2013). Genome analysis reveals



- insights into physiology and longevity of the Brandt's bat *Myotis brandtii*. *Nature Communications*, 4(1), 1–8.
- Seluanov, A., Hine, C., Azpurua, J., Feigenson, M., Bozzella, M., Mao, Z., Catania, K. C., & Gorbunova, V. (2009). Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *Proceedings of the National Academy of Sciences of the United States of America*, 106(46), 19352–19357.
- Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, 16(1), 1–11.
- Shah, S., Brock, E. J., Jackson, R. M., Ji, K., Boerner, J. L., Sloane, B. F., & Mattingly, R. R. (2018). Downregulation of Rap1Gap: A switch from DCIS to invasive breast carcinoma via ERK/MAPK activation. *Neoplasia*, 20(9), 951–963.
- Shin, H., Cha, H. J., Lee, M. J., Na, K., Park, D., Kim, C.-Y., Han, D. H., Kim, H., & Paik, Y.-K. (2020). Identification of ALDH6A1 as a potential molecular signature in hepatocellular carcinoma via quantitative profiling of the mitochondrial proteome. *Journal of Proteome Research*, 19(4), 1684–1695.
- Shortall, K., Djeghader, A., Magner, E., & Soulimane, T. (2021). Insights into aldehyde dehydrogenase enzymes: A structural perspective. *Frontiers in Molecular Biosciences*, 8, 659550.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212.
- Solarek, W., Koper, M., Lewicki, S., Szczylik, C., & Czarnecka, A. M. (2019). Insulin and insulin-like growth factors act as renal cell cancer intratumoral regulators. *Journal of Cell Communication and Signaling*, 13, 381–394.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research*, 34, W435–W439.
- Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N. P., Emes, R. D., & Lynch, V. J. (2016). TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *eLife*, 5, e11994.
- Tian, X., Azpurua, J., Hine, C., Vaidya, A., Myakishev-Rempel, M., Ablaeva, J., Mao, Z., Nevo, E., Gorbunova, V., & Seluanov, A. (2013). High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature*, 499(7458), 346–349.
- Tollis, M., Ferris, E., Campbell, M. S., Harris, V. K., Rupp, S. M., Harrison, T. M., Kiso, W. K., Schmitt, D. L., Garner, M. M., Aktipis, C. A., Maley, C. C., Boddy, A. M., Yandell, M., Gregg, C., Schiffman, J. D., & Abegglen, L. M. (2021). Elephant genomes reveal accelerated evolution in mechanisms underlying disease defenses. *Molecular Biology and Evolution*, 38(9), 3606–3620.
- Vallés, A. M., Beuvin, M., & Boyer, B. (2004). Activation of Rac1 by paxillin-Crk-DOCK180 signaling complex is antagonized by Rap1 in migrating NBT-II cells. *Journal of Biological Chemistry*, 279(43), 44490–44496.
- van Dartel, M., & Hulsebos, T. J. M. (2004). Amplification and overexpression of genes in 17p11. 2~ p12 in osteosarcoma. *Cancer Genetics and Cytogenetics*, 153(1), 77–80.
- Vartanian, V., Lowell, B., Minko, I. G., Wood, T. G., Ceci, J. D., George, S., Ballinger, S. W., Corless, C. L., McCullough, A. K., & Lloyd, R. S. (2006). The metabolic syndrome resulting from a knockout of the NEIL1 DNA glycosylase. *Proceedings of the National Academy of Sciences of the United States of America*, 103(6), 1864–1869.
- Vasilou, V., & Nebert, D. W. (2005). Analysis and update of the human aldehyde dehydrogenase (ALDH) gene family. *Human Genomics*, 2(2), 1–6.
- Vazquez, J. M., Sulak, M., Chigurupati, S., & Lynch, V. J. (2018). A zombie LIF gene in elephants is upregulated by TP53 to induce apoptosis in response to DNA damage. *Cell Reports*, 24(7), 1765–1776.
- Venkat, A., Hahn, M. W., & Thornton, J. W. (2018). Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nature Ecology & Evolution*, 2(8), 1280–1288.
- Vicens, A., & Posada, D. (2018). Selective pressures on human cancer genes along the evolution of mammals. *Genes*, 9(12), 582.
- Wang, O., Chin, R., Cheng, X., Wu, M. K. Y., Mao, Q., Tang, J., Sun, Y., Anderson, E., Lam, H. K., Chen, D., Zhou, Y., Wang, L., Fan, F., Zou, Y., Xie, Y., Zhang, R. Y., Drmanac, S., Nguyen, D., Xu, C., ... Peters, B. A. (2019). Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Research*, 29(5), 798–808.
- Wawrzyniak, J. A., Bianchi-Smiraglia, A., Bshara, W., Mannava, S., Ackroyd, J., Bagati, A., Omilian, A. R., Im, M., Fedtsova, N., & Miecznikowski, J. C. (2013). A purine nucleotide biosynthesis enzyme guanosine monophosphate reductase is a suppressor of melanoma invasion. *Cell Reports*, 5(2), 493–507.
- Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Research*, 27(5), 757–767.
- Wen, X., Zhou, M., Guo, Y., Zhu, Y., Li, H., Zhang, L., Yu, L., Wang, X., & Peng, X. (2015). Expression and significance of DOK2 in colorectal cancer. *Oncology Letters*, 9(1), 241–244.
- Wisotsky, S. R., Kosakovsky Pond, S. L., Shank, S. D., & Muse, S. V. (2020). Synonymous site-to-site substitution rate variation dramatically inflates false positive rates of selection analyses: Ignore at your own peril. *Molecular Biology and Evolution*, 37(8), 2430–2439.
- Xu, X., Chai, S., Wang, P., Zhang, C., Yang, Y., Yang, Y., & Wang, K. (2015). Aldehyde dehydrogenases and cancer stem cells. *Cancer Letters*, 369(1), 50–57.
- Xu, Z., & Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, 35, W265–W268.
- Yang, Y., Li, M., Yan, Y., Zhang, J., Sun, K., Qu, J. K., Wang, J. S., & Duan, X. Y. (2015). Expression of RAP1B is associated with poor prognosis and promotes an aggressive phenotype in gastric cancer. *Oncology Reports*, 34(5), 2385–2394.
- Yang, Y., Zhang, J., Yan, Y., Cai, H., Li, M., Sun, K., Wang, J., Liu, X., Wang, J., & Duan, X. (2017). Low expression of Rap1GAP is associated with epithelial-mesenchymal transition (EMT) and poor prognosis in gastric cancer. *Oncotarget*, 8(5), 8057.
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591.
- Yang, Z., & Dos Reis, M. (2010). Statistical properties of the branch-site test of positive selection. *Molecular Biology and Evolution*, 28(3), 1217–1228.
- Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *OmicS: A Journal of Integrative Biology*, 16(5), 284–287.
- Yu, H.-S., Oyama, T., Isse, T., Kitagawa, K., Tanaka, M., & Kawamoto, T. (2010). Formation of acetaldehyde-derived DNA adducts due to alcohol exposure. *Chemico-Biological Interactions*, 188(3), 367–375.
- Yu, S.-L., & Lee, S.-K. (2017). Ultraviolet radiation: DNA damage, repair, and human disorders. *Molecular & Cellular Toxicology*, 13, 21–28.
- Zhang, F., Yan, T., Guo, W., Sun, K., Wang, S., Bao, X., Liu, K., Zheng, B., Zhang, H., & Ren, T. (2018). Novel oncogene COPS3 interacts with Beclin1 and Raf-1 to regulate metastasis of osteosarcoma through autophagy. *Journal of Experimental & Clinical Cancer Research*, 37(1), 1–12.
- Zhang, J., Xu, E., Ren, C., Yan, W., Zhang, M., Chen, M., Cardiff, R. D., Imai, D. M., Wisner, E., & Chen, X. (2014). Mice deficient in Rbm38, a target of the p53 family, are susceptible to accelerated aging and spontaneous tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 111(52), 18637–18642.



- Zhang, X., Huang, Z., Hu, Y., & Liu, L. (2016). Knockdown of myosin 6 inhibits proliferation of oral squamous cell carcinoma cells. *Journal of Oral Pathology & Medicine*, 45(10), 740–745.
- Zhao, Y., Qiao, S., Hou, X., Tian, H., Deng, S., Ye, K., Nie, Y., Chen, X., Yan, H., & Tian, W. (2019). Bioengineered tumor microenvironments with naked mole rats high-molecular-weight hyaluronan induces apoptosis in breast cancer cells. *Oncogene*, 38(22), 4297–4309.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Li, X., Wang, P., Pan, Q., Liu, G., Liu, W., Omotoso, O., Du, J., Li, Z., Yu, Y., Huang, Y., Zhu, P., Li, M., & Zhou, X. (2023). Chromosome-level Asian elephant genome assembly and comparative genomics of long-lived mammals reveal the common substitutions for cancer resistance. *Aging Cell*, 22, e13917. <https://doi.org/10.1111/ace1.13917>