Cancer Prevalence Across Vertebrates

Zachary Compton[1,2], Valerie Harris[1], Walker Mellon[1], Shawn Rupp[1], Diego Mallo[1], Stefania E. Kapsetaki[1], Mallory Wilmot[3], Ryan Kennington[3], Kathleen Noble[3], Cristina Baciu[1,4], Lucia Ramirez[5], Ashley Peraza[1], Brian Martins[1], Sushil Sudhakar[1], Selin Aksoy[1], Gabriella Furukawa[3], Orsolya Vincze[6,7], Mathieu Giraudeau[8], Elizabeth G. Duke[9,10], Simon Spiro[11], Edmund Flach[11]*, Hannah Davidson[9,10], Ashley Zehnder[10], Trevor A. Graham[12], Brigid Troan[9,10,13], Tara M. Harrison[9,10†], Marc Tollis[14†], Joshua D. Schiffman[1,3,15†], Athena Aktipis[1,4†], Lisa M. Abegglen[1,3, 15†], Carlo C. Maley[1,2,16†], Amy M. Boddy[1,17†]

Affiliations:
[1] Arizona Cancer Evolution Center, The Biodesign Institute, Arizona State University, Tempe, AZ
[2] School of Life Sciences, Arizona State University, Tempe, AZ
[3] Department of Pediatrics and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT
[4] Department of Psychology, Arizona State University, Tempe, AZ
[5] Genomic Sciences Graduate Program, North Carolina State University, Raleigh, NC
[6] Institute of Aquatic Ecology, Centre for Ecological Research, 4026 Debrecen, Hungary
[7] Evolutionary Ecology Group, Hungarian Department of Biology and Ecology, Babeş-Bolyai University, 400006 Cluj-Napoca, Romania
[8] LIENSs, UMR 7266 CNRS-La Rochelle Université, La Rochelle, France,
[9] North Carolina State College of Veterinary Medicine, Raleigh, NC
[10] Exotic Species Cancer Research Alliance, North Carolina State College of Veterinary Medicine, Raleigh, NC
[11] Wildlife Health Services, Zoological Society of London, London, UK.
[12] Centre for Evolution and Cancer, Institute of Cancer Research, London, UK
[13] The North Carolina Zoo, Asheboro, NC
[14] School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, AZ
[15]Peel Therapeutics, Inc., Salt Lake City, UT
[16] Biodesign Center for Biocomputing, Security and Society, Arizona State University, Tempe, AZ
[17]University of California Santa Barbara, Santa Barbara, CA

[†] co-senior authors
* retired

Supplementary Drive:
https://drive.google.com/drive/folders/1gBTZ9TDPZZ3eK-bKXG7dg7SmHf8Ypin9?usp=sharing

Abstract
Cancer is pervasive across multicellular species. Are there any patterns that can explain differences in cancer prevalence across species? Using 16,049 necropsy records for 292 species spanning three clades (amphibians, sauropsids and mammals) we found that neoplasia and malignancy prevalence increases with adult weight and decreases with gestation time, contrary to Peto's Paradox. Evolution of cancer susceptibility appears to have undergone sudden shifts followed by stabilizing selection. Outliers for neoplasia prevalence include the common porpoise (<1.3%), the Rodrigues fruit bat (<1.6%) the black-footed penguin (<0.4%), ferrets (63%) and opossums (35%). Discovering why some species have particularly high or low levels of cancer may lead to a better understanding of cancer syndromes and novel strategies for the management and prevention of cancer.

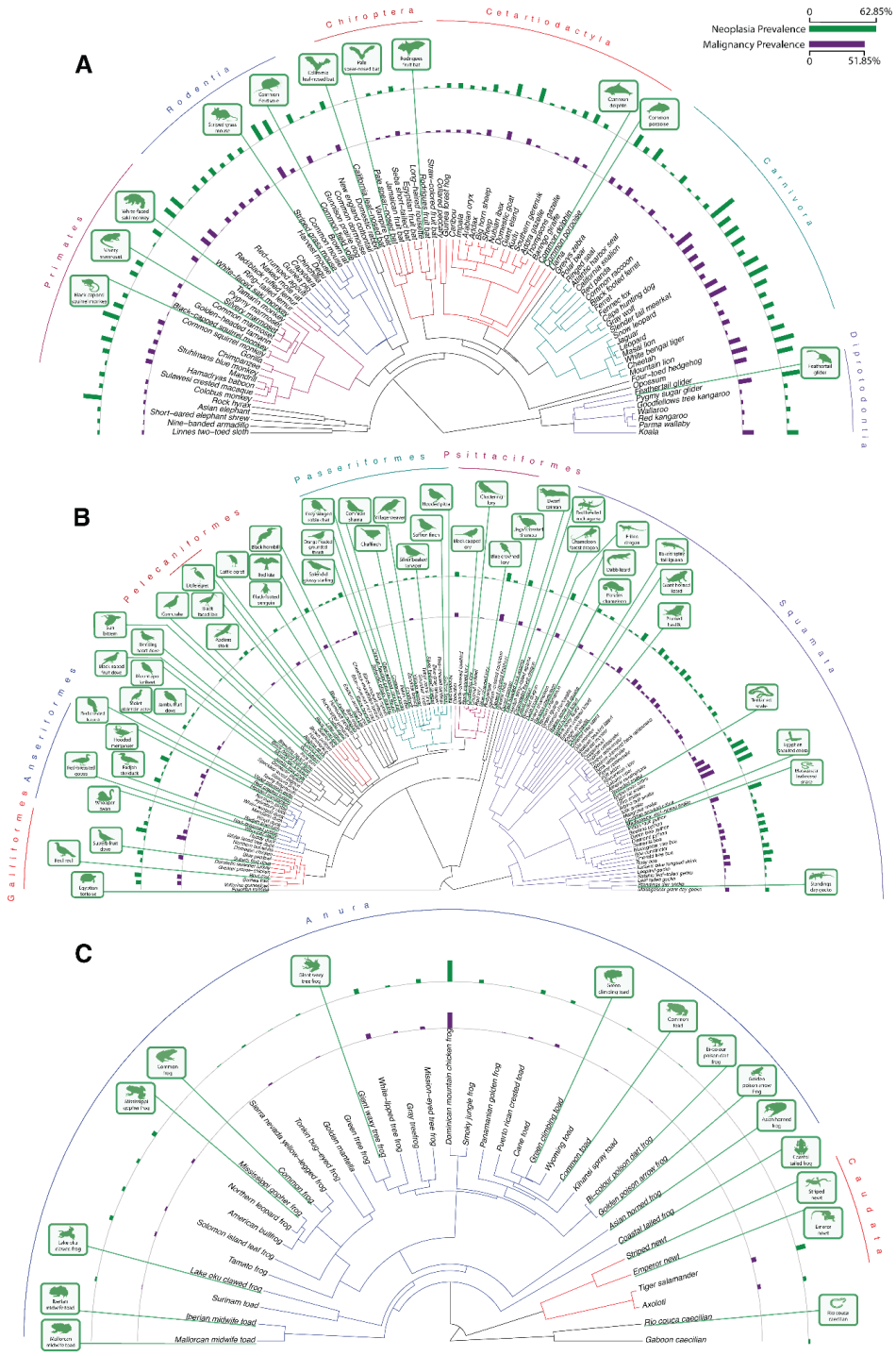
Introduction
Cancer is a ubiquitous problem for multicellular species (*1*) and a leading cause of death in humans (*2*). Every multicellular body is a cooperative cellular system, with cells suppressing replication (*3*), dividing labor (*4*), sharing resources (*5*), regulating cell death (*6*) and taking care of the extracellular environment (*1*). However, cooperative systems are susceptible to cheaters, which emerge as cancers in multicellular organisms (*7*). Because cancer cells can outcompete normal cells with respect to replication, survival, resource use and other cellular behaviors, natural selection within the body can favor cancer cells via somatic evolution.

Cancer has been a strong selective pressure on multicellular organisms and mechanisms for cancer suppression likely co-evolved along with the evolution of multicellularity (*8*, *9*). Despite this persistent selective pressure of cancer, species vary in their investment in cancer defenses across the tree of life. Sir Richard Peto predicted in 1977 that the risk of cancer should scale with the number of cells in an organism and the length of its lifespan (*10*). This prediction is based on the fact that tumors evolve from single cells, partially due to the accumulation of somatic mutations over time (*10*). His observation that cancer risk does not appear to increase with increases in body mass and longevity across species (*10*), a phenomenon known as 'Peto's paradox', launched the field of comparative oncology (*11*).

Early work in comparative oncology found that birds, and to a lesser extent reptiles, develop fewer neoplasms than mammals (*12*–*14*). While single case studies have been reported (*15*), it has been difficult to estimate true neoplasia prevalence in these taxa. In 2015, we published neoplasia prevalence estimates in 37 mammals and reported support for Peto's Paradox, that is, bigger, longer-lived species do not get more cancer (*16*). Follow up studies have supported Peto's Paradox and demonstrated the ubiquity of cancer across mammals (*17*, *18*). The extensive variation in cancer risk across vertebrates provides a unique opportunity to identify species with exceptional cancer resistance that can lead to new discoveries of cancer resistance mechanisms outside the traditional human and murine studies. Additionally, the discovery of cancer vulnerable species could lead to new insights into cancer syndromes as well as provide spontaneous 'natural' animal models of disease that can help us gain a better

understanding of various types of cancer and their treatments. Here we present a large, curated database of tetrapod veterinary necropsy records, including 16,049 individual animals across 292 species of animals, encompassing reptiles, birds, amphibians, and mammals. Because necropsies typically are diagnosed with “neoplasia” which includes both benign and malignant tumors, we developed a terminology dictionary to distinguish benign from malignant neoplasms in the necropsy reports. We calculate and analyze both neoplasia prevalence as well as malignancy (cancer) prevalence. Only a subset of benign neoplasms evolve into cancers over a lifetime, so neoplasia prevalence is always greater than or equal to malignancy prevalence. We also tested for age bias in the animals that died with neoplasias or cancers.

Results


A
B
C
Neoplasia Prevalence
Malignancy Prevalence
0
62.85%
51.85%
Chiroptera
Cetartiodactyla
Rodentia
Carnivora
Primates
Diprotodontia
Passeriformes
Psittaciformes
Pelecaniformes
Anseriformes
Galliformes
Squamata
Anura
Caudata
Common fox
Common porcupine
Common raccoon
Common squirrel monkey
Common wombat
Common brushtail
Common ringtail
Common bottlenose dolphin
Common dolphin
Common gray gibbon
Common marmoset
Common opossum
Common pygmy marmoset
Common squirrel monkey
Common tree shrew
Common vampire bat
Common wallaroo
Common zebra
Common zebra finch
Common zebra dove
Common yellow-throated marten
Common wombat
Common ringtail
Common brush

Variation in Cancer Risk Across Clades

We found evidence of neoplastic disease in necropsies across all analyzed taxonomic clades (Fig.1). Mean prevalence of neoplasia (mean = 9.07%, range = 0% - 62.86%) and malignancy (mean=5.91%, range=0%–40.95%) at death was highest in mammals (mean: Neoplasia= 15.28%, Malignancy = 9.82%; range: Neoplasia = 0% - 62.85%, Malignancy = 0% - 40.95%), followed by sauropsids (mean: Neoplasia=6.40% Malignancy=4.33%; range: Neoplasia= 0% - 39.13%, Malignancy: 0% - 34.78%) and amphibians (mean: Neoplasia = 4.16%; range: Neoplasia = 0% - 45.83%, Malignancy = 0% - 33.33%; Fig. 2), which confirms previous studies (*12*, *14*). Because reptiles are not a monophyletic clade, we have grouped them with birds in the sauropsida clade for the purposes of analysis. Despite a lower mean prevalence for both benign and malignant tumors, sauropsids and amphibians show a wide range of neoplastic disease burden across species. There is a small but highly statistically significant correlation between the prevalence of benign neoplasms and the prevalence of malignant neoplasms across species (r=0.2, p<0.0001, Fig. S64). Supplementary Tables 1 and 2 list the species with the highest and lowest neoplasia and malignancy prevalences, as well as the proportion of neoplasms that are malignant. Among the vertebrates with the highest prevalence of neoplasia (Suppl. Table 8), 63% of ferrets died with a neoplasm (45% of which was lymphoma), 56% of opposums died with a neoplasm (46% of which was in the lung), and 45% of hedgehogs died with a neoplasm (42% of which was in the alimentry tract).

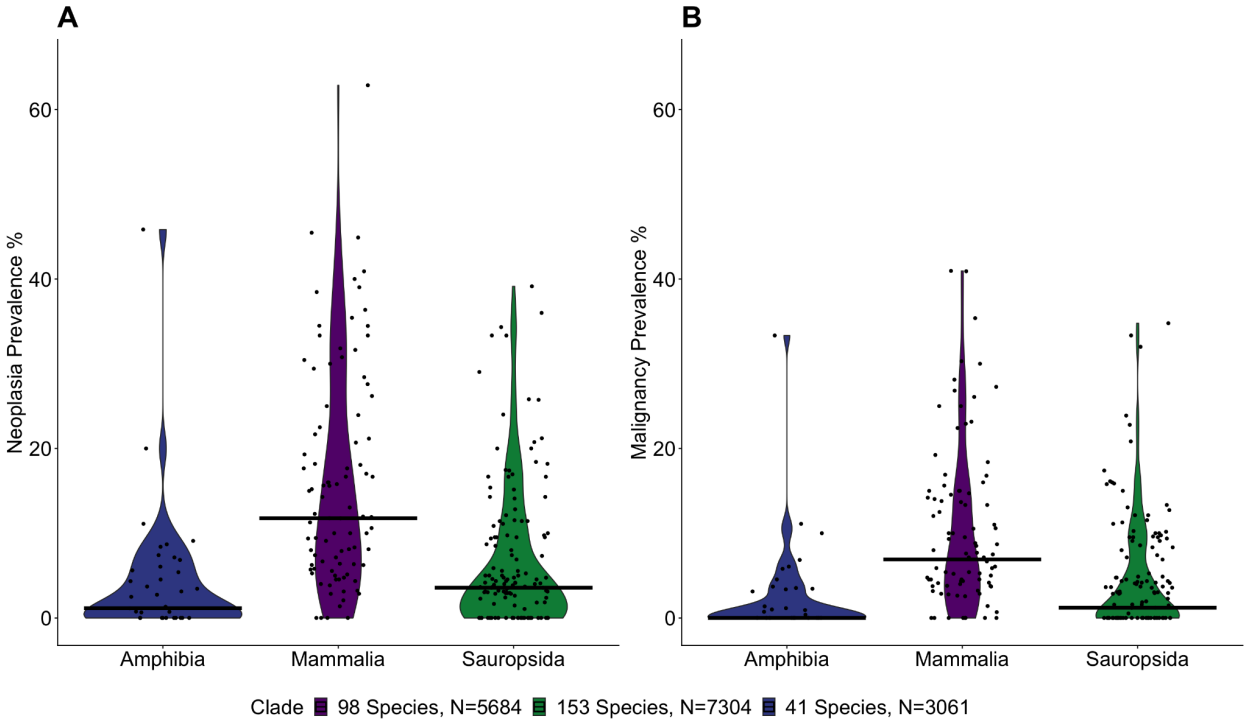


Figure 2. Distributions of of **A.** neoplasia (Kruskal-Wallace test: p = 2.906 x 10^{-12}) and **B.** malignancy (Kruskal-Wallace test: p = 6.519 x 10^{-11}) prevalences are different across three clades, Amphibia, Mammalia, and Sauropsida (Reptilia and Aves). Dots show the estimated species neoplasia prevalence and bars show the median for the clade. Neoplasia and

malignancy prevalence for species were calculated by the proportion of the reported lesions among the total number of necropsies for that species.

Life History Analyses of Neoplasia Prevalence

Evolutionary life history theory provides a framework for understanding the tradeoffs governing species' survival and reproduction (*19*, *20*). Life history theory can be used to explain how species level traits shape organismal cancer risk based on trade-offs between investment in somatic maintenance (e.g., cancer suppression) and reproduction or growth. Several smaller studies have shown that specific life history traits can serve as prognostic indicators of neoplasia prevalence in animals managed under human care (*17*, *21*). We tested for relationships between life history factors and neoplasia or malignancy prevalence, controlling for phylogenetic relatedness, and weighting species data points by the number of necropsies in our dataset. In contrast to previous studies (*16*–*18*), we found an increase in neoplasia prevalence with increases in body mass (2.1% neoplasia per Log_{10}g, $p = 0.007$; 0.65% malignancy per Log_{10}g, $p = 0.287$) and maximum longevity (0.01% neoplasia per Log_{10}g, $p = 0.02$; 0.0047% malignancy per Log_{10}g, $p = 0.276$), not supporting Peto's Paradox (Fig. 3). Animals with longer gestation times also get fewer malignancies (-5.56% malignancies with Log_{10} months, $p = 0.02$; Fig. 3C).

A multivariate model containing all significant predictors of neoplasia or malignancy (adult weight, maximum longevity, and gestation time) shows that both adult body weight (2.9% neoplasia per Log_{10}g, $p = 0.01$) and gestation time (-18.6% neoplasia per month, $p = 0.0001$) provide independent information for estimating neoplasia prevalence. Because gestation time and adult weight are correlated ($r = 0.50$, $p = 2.2 \times 10^{-16}$), but have the opposite relationship to neoplasia and malignancy prevalence, we tested the two-variable model and found that when controlling for adult weight (3.8% neoplasia per Log_{10}g, $p = 0.0005$), gestation time is also a significant predictor of neoplasia prevalence (-15.8% neoplasia per Log_{10} months, $p = 0.001$; $R^2 = 0.27$), and vice versa. When controlling for gestation time, adult weight predicts malignancy prevalence, and when controlling for adult weight, gestation time also predicts malignancy prevalence (0.68% malignancies per Log_{10}g, $p = 0.0008$; -6.61% malignancies per month of gestation, $p = 0.0417$; $R^2 = 0.198$). Adult weight and gestation time were still statistically significant (adjusted $p<0.05$) predictors of neoplasia and malignancy prevalence after a 10% false discovery rate correction for multiple testing.

We found no evidence of a relationship between litter or clutch size and neoplasia prevalence (Figs. S21, S22). However, when we restrict the analysis to mammals, litter size is positively associated with both neoplasia and malignancy prevalence (neoplasia: $p = 0.02$, $R^2=0.55$; malignancy: $p = 0.03$, $R^2 = 0.2$; Suppl. Figs. 26 & 27), supporting our earlier analysis of 37 mammals from the San Diego Zoo (*17*). We also found that time to sexual maturity, growth rate and basal metabolic rates (which were only available for mammals) were not significant predictors of neoplasia or malignancy prevalence (Figs. S11, S12, S13, S14, S17, and S18). In addition to calculating the prevalence of neoplasms and malignancies, we also calculated the proportion of neoplasms that were malignant, which is a measure of the likelihood that a benign neoplasm transforms into a malignant one. We found no statistically significant relationships

between any of those life history factors and the proportion of neoplasms that were malignant (Figs. S11, S13, and S17).

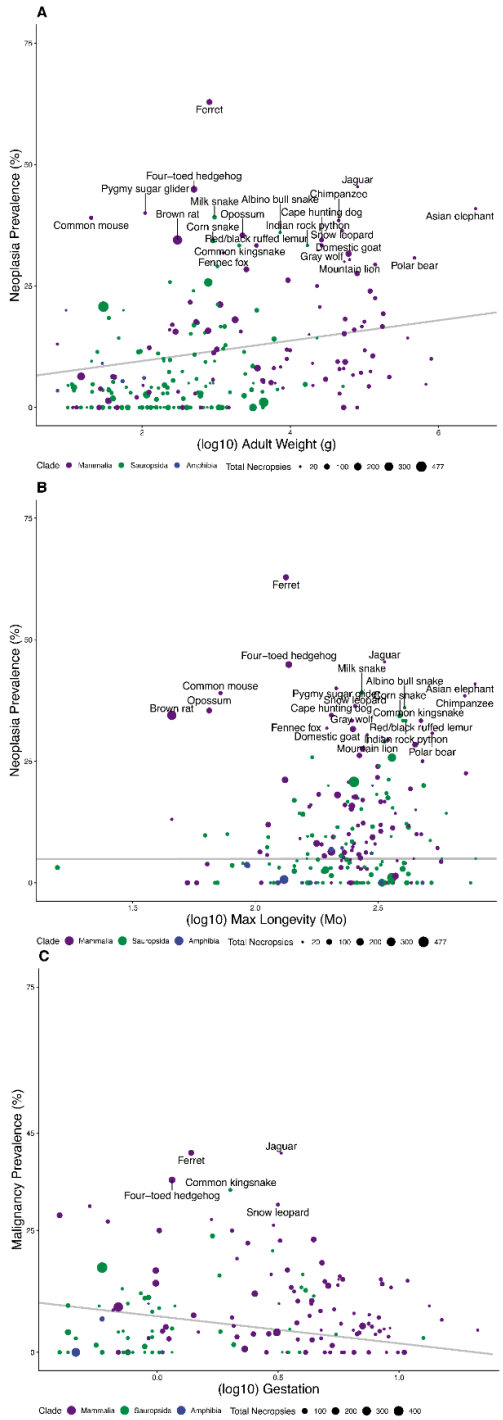


Figure 3. Significant life history factors associated with neoplasia and malignancy prevalence. **A.** Larger organisms have a higher neoplasia prevalence than smaller organisms (2.1% neoplasia per Log_{10}g adult body mass, p = 0.007, R^2 = 0.18, λ = 0.46). **B.** Longer lived

organisms also have more neoplasia (0.01% neoplasia per Log_{10} month lifespan, p = 0.02, R^2 = 0.16, λ = 0.34). **C.** Organisms with longer gestation times have a lower malignancy prevalence (-5.65% malignancies per Log_{10} months, *p* = 0.02, R^2 = 0.01, λ = 0.41). When controlling for adult body mass, organisms with longer gestation times also have fewer neoplasias at death (-5.30% neoplasia per Log_{10} months, *p* = 0.1).

In vitro DNA damage response

We tested primary fibroblasts from 15 species for their response to DNA damage (Fig. 4; Suppl. Figs. 45-60). Mammalian cells have two primary mechanisms to address DNA damage: cell cycle arrest to allow for DNA repair and apoptosis (*16*, *23*). In a previous study, we documented that low prevalence of death from cancer in elephants was correlated with enhanced DNA damage response (*16*). Therefore, we hypothesized that DNA damage response would predict malignancy. Figure 4 shows the predicted trend for both neoplasia and malignancy prevalence, though they are not statistically significant. We also analyzed responses to lower doses of ionizing radiation and increasing doses of a chemotherapeutic drug (doxorubicin). No association with response and neoplasia or malignancy was observed (S45 - S60)

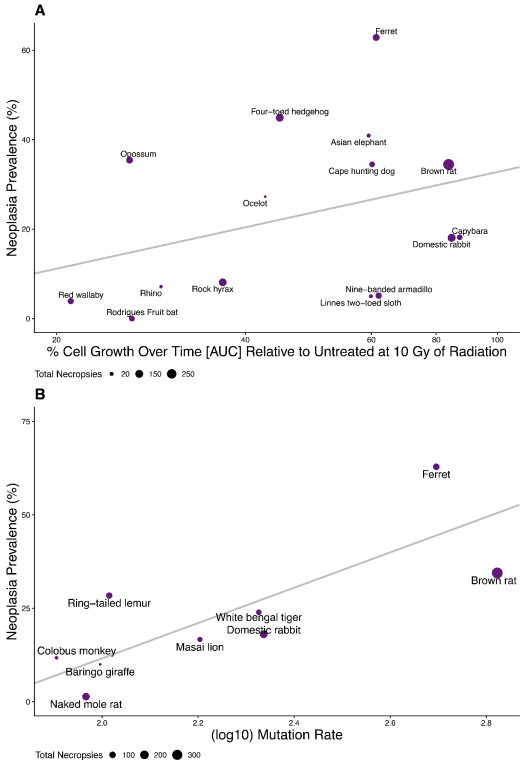


Figure 4. A. % Cell Growth Over Time [AUC] Relative to Untreated at 10 Gy of Radiation (plotted and analyzed on a Log_{10} scale) as a predictor of neoplasia prevalence in species' fibroblast cell lines. (30.89% neoplasia per Log_{10} Cell Count Area Change, *p* = 0.22, R^2 = 0.011, λ = 6.6 x 10^{-5}) **B.** Log_{10} Mean Mutation Rate as a predictor of neoplasia prevalence (47.26% Single Base Substitution per Genome per Year, *p* = 0.0059, R^2 = 0.96, λ = 1.00).

Age as a Cancer Risk Factor in Animals

Age is the single biggest risk factor for the development of cancer in humans (*24*). Most mechanisms of somatic maintenance, including immune cell surveillance, DNA damage response, and telomere shortening, decrease in efficacy as we age (*25–29*). To test if observed neoplasms in animals under human care may be due to the animals living beyond their natural lifespans, we plotted the age of the animals with neoplasia at death, compared to the animals that died without neoplasias, scaled by their average lifespan (Fig. 5). The vast majority of animal deaths with neoplasia diagnoses occur before the average lifespan in most animals. Only amphibians seem to be developing more neoplasias as they live past their normal lifespan under human care(Fig. 5C). The distribution of tumor diagnoses across lifespan in these three clades also demonstrates that cancer is not limited to a disease solely of extended lifespan, and in sauropsids, neoplasia is not particularly a disease of old age (Fig. 5B).

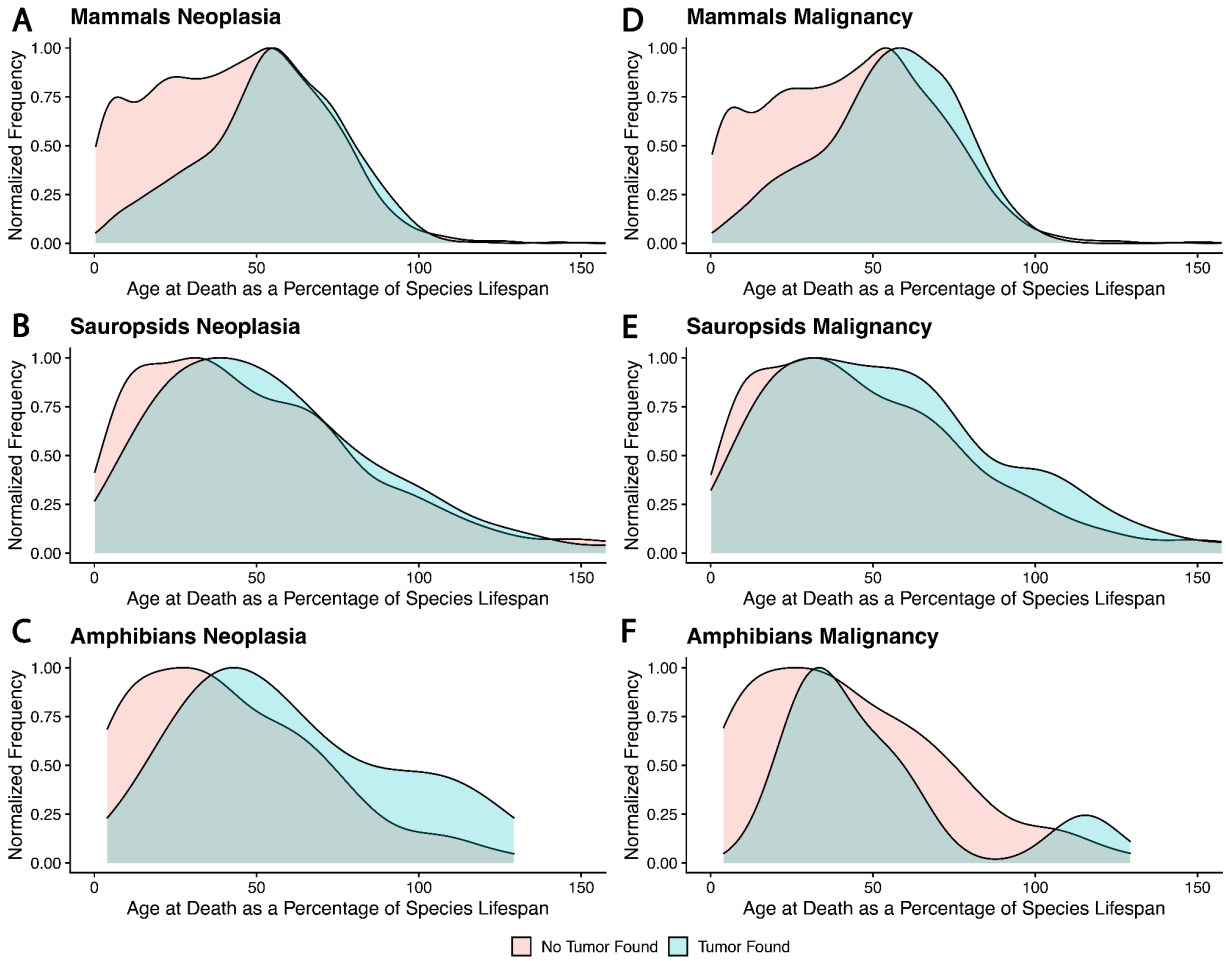


Figure 5. The density distribution and corresponding locally weighted regression (LOESS) of ages at death in animals with neoplasia versus non-neoplasia, adjusted for each species' lifespan as specified in PanTHERIA. While the distributions of ages at death are different between necropsies showing neoplasia versus those that don't (Two Sample Kolmogorov-Smirnov Test: Mammals: D=0.11, p =1.81 x 10^{-6}; Sauropsids: D= 0.18, p = 4.48 x

10^{-8}; Amphibians: D=0.5, p = 0.011), we found few neoplasias that could be explained by an organism living an extraordinarily long time in captivity, except in amphibians.

Phenotypic Models of Cancer Risk

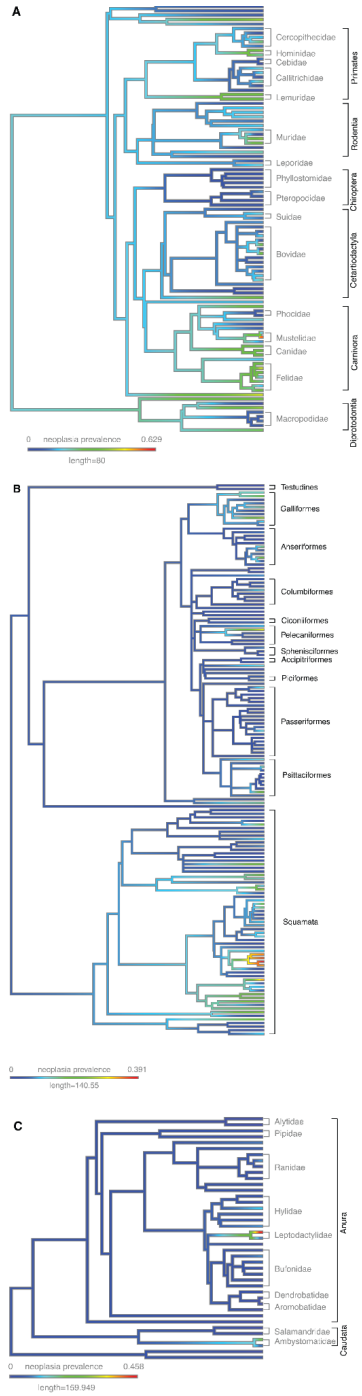


Figure 6. Cladogram depiction of cancer incidence within **A.** Mammals, **B.** Sauropsids, and **C.** Amphibians. Cladograms with the species labels at each tip can be found in Suppl. Fig. 63. Heat map coloration indicates relative prevalence of cancer within each branch, illustrating the diversity of neoplastic disease amongst closely related species. The scale is the same for each panel so that the differences between the clades are apparent.

Evolution of cancer suppression and susceptibility
Comparative phylogenetics provides a wealth of computational tools to model species' trait evolution across a phylogeny (*30*). To explore how cancer susceptibility evolved across the tree of life (Fig. 6), we fit three of the most common phenotype evolution models (Ornstein-Uhlenbeck, Brownian Motion, and Early Burst) to neoplasia prevalence as a continuous trait. We found that a model of stabilizing selection on neoplasia prevalence (Ornstein-Uhlenbeck) fits the distribution of neoplasia prevalence the best (Supp. Tab. 7). Malignancy prevalence evolution is also best explained by the Ornstein-Uhlenbeck model of sudden shifts followed by stasis in the phenotype.

Methods

Analysis of Veterinary Necropsy Records
We collected necropsy records with permission from 99 zoological institutions, aquariums, and other facilities that house animals under managed care. Gross necropsies had all been conducted by veterinary pathologists who specialize in nondomestic species and neoplasia was identified histologically by board-certified pathologists. Cases where suspect neoplasms were not examined histologically were excluded from the dataset. We used a terminology dictionary (Supplemental table 3) to distinguish benign from malignant neoplasms based on the diagnoses in the necropsy reports. We excluded neonatal records to reduce bias from high levels of neonate and infant mortality that is common in many species. Because only common names were recorded for most records, we developed a tool, kestrel, to translate common names into scientific names which is available at https://pkg.go.dev/github.com/icwells/kestrel.

All of the institutions that provided prior approval for the use of their data in these analyses are Association of Zoos and Aquariums (AZA) accredited. AZA accreditation encourages the institution to perform a necropsy on all animals that die under their care to determine cause of death and to monitor morbidity and mortality of each species. Furthermore, each institution had IACUC approval with the Exotic Species Cancer Research Alliance (ESCRA) and the Arizona Cancer and Evolution Center (ACE) for the use of their deceased animal's records of animals with neoplasia for use in this study. Previous analyses included both necropsies for animals diagnosed with neoplasia and animals that were still alive (*18*). In this study, we restricted our analyses to only necropsies - for both cancer and non-cancer diagnosed animals, because alive animals may harbor undetected cancer or might be eventually diagnosed with cancer, thus skewing estimates in cancer prevalence.

The data for our analyses are available in supplemental file 1.

Comparative Phylogenetic Methods in Comparative Oncology
Interspecies comparisons must account for the shared ancestry and the constraint of natural selection on species' traits before a determinant of any correlations can be made. For the life history models of neoplasia and malignancy prevalence, the R programming (*31*) packages "phytools" (*32*), "ape" (*33*), and "caper" (*34*) were all used for phylogenetic comparisons and the handling of phylogenetic data. To accomplish this we wrote the function *pglsSEyPagel* which is

built upon phytool's *pglsSEy* (phylogenetic generalized least squares for uncertainty in Y). pglsSEyPagel expands upon the pglsSEy function by adding the estimate of Pagel's lambda (*35*) to the regression, rather than assuming it is fixed at 1 (i.e., Brownian motion). .

Testing for relationships with life history factors

We extracted data for maximum lifespan, adult body weight, basal metabolic rate, gestation length, litter size, time to sexual maturity, and growth rate from PanTHERIA (*36*). We used a weighted phylogenetic regression to control for non-independence of phenotypes (e.g. neoplasia prevalence) in closely related species. We report the phylogenetic signal, lambda, for each analysis, along with the p-value and R^2. A single phylogenetic tree encompassing the three clades was collected from timetree.org. We pruned the tree to the 292 species in our data set using the setdiff and keep.tip/drop.tip functions in the APE R package. Estimates for neoplasia and malignancy prevalence are more accurate in species with more necropsies. To address the differences in number of necropsies, and to limit the noise from prevalence estimates based on few individuals, we weighted the species data points by the square-root of the number of necropsies records we have. Our R code for all analyses and figures included in this manuscript is freely available at https://github.com/zacharycompton/cancerAcrossVertebrates.git . In addition, we only analyzed species for which we had at least 20 necropsy results (previous studies had used 10 (*16*) or 20 (*17*, *18*) for the lower bound number of individuals). The main *pglsSEyPagel* analyses were done with all species together, including mammals, sauropsids and amphibians. In the analyses of litter size and gestation time, we also tested for a relationship with neoplasia prevalence in mammals alone. We carried out a total of 28 *pglsSEyPagel* analyses. To control for multiple testing, we used a false discovery rate (FDR) of 10%.

DNA Damage Sensitivity Assays

Established, primary cells from mammals were obtained from San Diego Zoo Wildlife Alliance (Capybara, Linne's Two Toed Sloth, Red Necked Wallaby, Rock Hyrax, Rodrigues Fruit Bat, Six Banded Armadillo, Southern White Rhino, and Virginia Opossum) or generated at Huntsman Cancer institute from tissues collected from African Pygmy Hedgehog, Domestic Rabbit, Leopard, Asian Elephant, and Cape hunting dog, Brown rat (Cell Applications) and Normal Human Dermal Fibroblasts (Lonza) were commercially available. Lonza (Normal Human Adult Dermal Fibroblasts). Detailed information on culture conditions, primary donor demographics, and passage numbers can be found in the supplementary information. Cells were seeded in 96-well plates at 2,000 cells per well in cell growth media and allowed to adhere overnight. The following day, doxorubicin was added at one of four concentrations (0μM [DMSO vehicle control], 0.11μM, 0.33μM, and 1μM). Each condition was tested in triplicate in three separate experiments. Cell proliferation and apoptosis were measured by real-time fluorescence microscopy (IncyCyte, Sartorius) at two-hour intervals for three days. Apoptosis was measured using a fluorescent cell death marker, Annexin V Dye (Sartorius). Images were processed and analyzed using IncuCyte software. The number of dead or dying cells were identified by counting Annexin V positive cells. In addition, cell count overtime was calculated using IncuCyte cell-by-cell software. To measure response to radiation-induced DNA damage, cells were

irradiated with one of four doses: 0Gy, 0.4Gy, 2Gy, and 10 Gy. Radiation dose was delivered using an RS-2000 X-Ray Irradiator (Radsource). Cells were then seeded in 96-well plates in cell growth media containing Annexin V Dye (Sartorius). Cells were imaged by real-time fluorescence microscopy (IncuCyte, Sartorius) at two-hour intervals for five days. We observed little to no apoptosis of irradiated fibroblasts, as has been previously reported for this cell type (reference). We estimated cell cycle arrest by normalizing the cell count of irradiated cells to untreated cells by dividing the area under the curve (AUC) of cell count over-time for treated cells by the AUC of cell count over time for the untreated (UT) cells. We converted that number into a percentage that represents the percent of cell proliferation relative to untreated cells. We then tested if this normalized amount of cell growth was predictive of neoplasia prevalence using the phylogenetically controlled *pglsSEyPagel* regression (Fig. 4).

Discussion

We estimated cancer prevalence across a wide range of tetrapod species that includes mammals, amphibians, reptiles and birds. Importantly, and contrary to previous studies, our analyses highlight limitations to Peto's paradox, by showing that large animals do tend to get somewhat more neoplasms, and malignancies when controlling for gestation time, compared with smaller animals. This is particularly apparent when we control for the fact that animals with longer gestation times tend to get both fewer neoplasias and fewer cancers. However, large animals only get slightly more cancer than small animals. Whether or not they get as much cancer as one would expect from their body size and longevity depends on the model one uses to predict cancer prevalence as a function of body size (*37*, *38*). We hypothesize that animals with a longer gestation time than would be expected for their body size, may be investing more resources to control cell proliferation, and thereby reducing their vulnerability to cancer, compared to animals with shorter than expected gestation times.

Cancer prevalence across species varies greatly. Here we have used the largest collection of species to date, and expanded our analyses beyond mammals (*16*–*18*), to test for patterns in cancer prevalence. We only include species with at least 20 necropsies (median 35), compared with 10 individuals per species in our original study (*16*), and weighted species more in our regression analyses if their cancer prevalence estimate is more accurate because it is based on more necropsies. This revealed adult weight and gestation time as significant predictors of neoplasia and malignancy prevalence. The fact that neoplasia prevalence seems to evolve by sudden shifts followed by stabilizing selection (the Ornstein-Uhlenbeck model of phenotypic evolution) is consistent with life history theory predictions that investment in somatic maintenance should be under selection in specific ecological conditions, rather than drifting neutrally consistent with random Brownian motion. Some of the variation in cancer prevalence is still noise, due to estimating cancer prevalence from tens of individuals. However, much of that variation comes from the vast diversity of species across amphibians, reptiles, birds and mammals. We have explained only a small portion (~20%) of the variation in species vulnerability and suppression of cancer. There is clearly more to be discovered.

Peto's Paradox is based on the expectation that large, long-lived animals should get more cancer because they have more cells that exist for a longer amount of time, increasing the likelihood that cancer will arise. Although adult body weight is positively correlated with both neoplasia and malignancy prevalence, partially resolving Peto's paradox, the effect size is much larger for neoplasia (3.8% neoplasia per $Log_{10}g$) than for malignancies (0.68% malignancies per $Log_{10}g$). There may be several explanations for this. The simplest being that malignancies are less common than neoplasias, which include both benign and malignant neoplasms. This reduces the statistical power and the expected size of the effects. However, the blunted relationship between body size and malignancy prevalence may also be due to natural selection acting to evolve mechanisms to suppress malignant transformation. Cancer suppression mechanisms are likely to have been under stronger selection among these larger and longer-lived organisms because it was more critical to suppress cancer for longer in order for these organisms to successfully reproduce. Thus, we might expect a relatively constant cancer rate across species with more cancer suppression mechanisms in large, long-lived organisms (*7*, *16*, *39–43*), and fewer in small, short-lived organisms.

Further, there are at least four transitions in neoplastic progression that natural selection might alter to increase the survivability of cancer in a species: 1) initiation of a neoplasm, 2) transformation of that neoplasm into malignancy (*i.e.*, invasion through the basement membrane), 3) metastasis, and 4) death caused by the cancer. Our data bear on the first two transitions. Specifically, we quantify the prevalence of neoplasms in a species, the prevalence of malignant neoplasms, and the proportion of neoplasms detected that are malignant. However, the selective pressure of cancer is ultimately due to its effects on mortality, and so quantifying the prevalence of cancer as a cause of death would be more relevant for evolutionary studies of comparative oncology (*18*).

The inclusion of cross-species functional assays highlighted in Fig. 4 demonstrates above all that there is tremendous variation in the cellular responses to radiation induced DNA damage. This is the first time functional assays to measure DNA damage response have been paired with species' cancer prevalence. While response to DNA damage was not a significant predictor of neoplasia or malignancy prevalence at 4 or 10 grays, the trend follows our hypothesis that sensitivity to DNA damage may be one mechanism of cancer suppression (*16*). The variation observed in our DNA damage response assays suggests that many species may use other mechanisms of cancer suppression (such as immune surveillance or other forms of DNA damage and somatic mutation), thereby obscuring a simple relationship between our measurements and neoplasia or malignancy prevalence.

Insights from comparative oncology for human cancers

Species with a high prevalence of particular cancers may help to generate targeted studies to elucidate the biological basis of those cancers, help draw informative parallels to particular cancer syndromes in humans, and serve as more realistic models for studying those cancers. For instance 46% of the malignant tumors diagnosed in the opossum were lung adenocarcinomas (Suppl. Tab. 8), which is a leading cause of human cancer mortality in the

United States (*44*). Hedgehogs may hold insights for colorectal cancer, the third leading cause of cancer mortality in the US, and ferrets may help us understand lymphoma. These spontaneous cancers may be more similar to human cancers than genetically manipulated mice, though that remains to be tested.

There is an exciting opportunity to discover the mechanisms for suppressing cancers in species with few to no observed neoplasms, or those that seem to prevent neoplasms from progressing to malignancy (Fig. 1). For example, the paucity of neoplasms in dolphins and porpoises may be due to a legacy of once having had large, long-lived cetacean ancestors that were under strong selection to suppress cancer (*39*). Our earlier analysis of cancer gene evolution in cetaceans found evidence of positive selection in a large number of tumor suppressor genes and proto-oncogenes (*39*).

We previously found that animals that live longer than would be expected for their body size, like bats, tend to have more copies of tumor suppressor genes (*45*). In support of these observations, we find 9 bats, with an average lifespan of 16 years, have low neoplasia prevalence. We had hoped to discover species that are able to prevent malignant transformation by finding species that get a fair number of benign neoplasms, but few to no malignant neoplasms. The common paradigm in understanding the evolution of cancer suppression emphasizes the importance of protecting against tumor initiation. However, mechanisms that suppress malignant transformation could be similarly important in maintaining an organism's fitness. Unfortunately, only a few species in our dataset fit that description. The species with the lowest proportion of malignant to benign neoplasms was the common squirrel, with only 12% of their tumors being malignant.

Challenges for Comparative Oncology

There are a number of potential sources of bias in comparative oncology data. The protection against predation that zoological institutions offer fast life history animals may be extending their lifespan, and thereby unmasking a vulnerability to cancer at an age that they would never attain in the wild. However, Fig. 5 shows that the neoplasms were diagnosed prior to average lifespan in most cases, suggesting the extended lifespan due to managed care is not a large factor in these data. In fact, one surprise was that neoplasms appear in birds and reptiles at relatively young ages, although birds are known to be prone to virally induced cancers (*46*).

Our data results from the combination of the intrinsic cancer susceptibility of a species with the effect of the artificial conditions of managed populations, which is sometimes called an evolutionary mismatch (*47*). These animals were generally protected from predators, provided with veterinary care, had different diets and exercise from their wild counterparts, many lived in an urban environment, and interacted with different species and microbes than a free-ranging animal. It is striking to us that four of the species with the lowest prevalence of neoplasias in our dataset, the gray squirrel, the common dormouse, the striped grass mouse, and the common field vole are all from wild, urban populations. These necropsies come from the London Zoo which has a policy of performing a necropsy on any animal they recover that dies on its

grounds, not just the animals under its care. This is a hint that cancer may well be less common in the wild.

If the “gross” cause of death was obvious for an animal, an institution may not have submitted the animal’s samples for histopathology, and would not be included in our data collection. Similarly, if a particular type of neoplasia is difficult to detect in a necropsy (including some leukemias and intracerebral tumors), or was only present at a microscopic level, it may have been undercounted.

The functional assays were limited to fibroblast cell lines for the species for which we could obtain samples. Limited sample availability precluded the ability to control for biological factors such as age, which we expect to influence DNA damage sensitivity.

The Future of Comparative Oncology

In the future, it will be important to collect additional data to validate our discoveries of species with particularly low and high cancer prevalence, such as those highlighted in Fig. 1. Several life history traits, such as basal metabolic rate, may explain cancer risk but with BMR measured in only a few species, we lacked statistical power to detect a relationship with neoplasia or malignancy prevalence. Here we have dramatically expanded the amount of data on cancer prevalence in non-human animals, but this must continue to be built upon if we are to match the robustness seen in human cancer epidemiology. In particular, much could be learned from analyzing the age-incidence curves of cancer (*18*, *48*), but that would require significantly more individuals for each species.

One of the most important holes in comparative oncology is cancer data on wild animals. Gathering data from free-ranging populations is challenging, as it is difficult to detect cancer due to the animals decaying or being eaten before they can be necropsied. Additionally, accurate age estimates are much more difficult to obtain in wild populations compared to those managed by humans. However, wild animal populations would greatly enhance the field of comparative oncology by validating species that have low cancer prevalence and testing for evolutionary mismatches for animals in captivity.

Conclusion

Cancer is a problem of multicellularity (*1*). While we found a relationship between both body mass and gestation time with cancer prevalence, we are just beginning to discover patterns of cancer susceptibility and cancer defenses across species. It is likely that evolution has developed a variety of mechanisms for preventing cancer. The discovery of particular species with extremely low neoplasia prevalence provides opportunities for elucidating cancer suppression mechanisms that are compatible with life and reproductive success. Investigation of species with extreme vulnerability to a particular cancer may also help us understand those cancers as well as human syndromes that predispose to those cancers. We hope to learn from nature how to better prevent cancer in both humans and non-human animals.

Acknowledgements

We would like to thank our dear colleague Drury Reavill who passed away prior to the preparation of this manuscript. We are forever grateful to her contribution to the field of veterinary science and her dedication to the health and survival of all animals. Many thanks to our continued collaboration with Michael Garner and Northwest Zoo Pathology who provided the annotated veterinarian pathology database. Liam Revell, the developer of the R software package “phytools”, was extremely helpful in providing input for the application of his methods in novel ways. We would also like to thank the zoos and aquaria who cared for these animals and gave us permission to analyze their data. These discoveries could not have been made without them. Among these contributing institutions we would like to specifically thank these zoos: Akron Zoo, Atlanta Zoo, Audubon Nature Institute, Bergen County Zoo, Birmingham Zoo, Buffalo Zoo, Capron Park Zoo, Central Florida Zoo, Dallas Zoo, El Paso Zoo, Elmwood Park Zoo, Fort Worth Zoo, Gladys Porter Zoo, Greensboro Science Center, Henry Doorly Zoo, Utah’s Hogle Zoo, Jacksonville Zoo, John Ball Zoo, Los Angeles Zoo, Louisville Zoo, Mesker Park Zoo, Miami Zoo, Oakland Zoo, Oklahoma City Zoo, Philadelphia Zoo, Phoenix Zoo, Pueblo Zoo, San Antonio Zoo, Santa Ana Zoo, Santa Barbara Zoo, Sedgwick County Zoo, Seneca Park Zoo, The Brevard Zoo, The Detroit Zoo, The Oregon Zoo, and Toledo Zoo.

This work was supported in part by NIH grants U54 CA217376, U2C CA233254, P01 CA91955, and R01 CA140657 as well as CDMRP Breast Cancer Research Program Award BC132057 and the Arizona Biomedical Research Commission grant ADHS18-198847. The findings, opinions and recommendations expressed here are those of the authors and not necessarily those of the universities where the research was performed or the National Institutes of Health.

Author Contributions

AMB, HD, TMH, DM, MG and VH collected the data. SR, ZC, JL, DM and VH curated the data. ZC, SR, WM, AP, BM, EM, SA, and CB carried out the analyses. MG, BT, SS provided veterinary pathology support and TMH provided veterinary supervision. BT created and provided the data dictionary for distinguishing benign from malignant disease. RK, KN and MW generated fibroblast cell lines and functional data generated with them, supervised by LA and JS. ZC, CCM and AA wrote the manuscript. CCM, AMB, AA, MT, TAG, LA, JS supervised the research. All authors edited the manuscript.

References

1. C. A. Aktipis, A. M. Boddy, G. Jansen, U. Hibner, M. E. Hochberg, C. C. Maley, G. S. Wilkinson, Cancer across the tree of life: cooperation and cheating in multicellularity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370** (2015), doi:10.1098/rstb.2014.0219.

2. F. B. Ahmad, R. N. Anderson, The Leading Causes of Death in the US for 2020. *JAMA*. **325**, 1829–1830 (2021).

3. J. W. Pepper, K. Sprouffske, C. C. Maley, Animal cell differentiation patterns suppress somatic evolution. *PLoS Comput. Biol.* **3**, e250 (2007).

4. D. L. Kirk, A twelve-step program for evolving multicellularity and a division of labor. *Bioessays*. **27**, 299–310 (2005).

5. A. M. Nedelcu, The evolution of multicellularity and cancer: views and paradigms. *Biochem. Soc. Trans.* **48**, 1505–1518 (2020).

6. G. Evan, T. Littlewood, A matter of life and cell death. *Science*. **281**, 1317–1322 (1998).

7. A. Aktipis, *The Cheating Cell* (Princeton University Press, 2020).

8. J. DeGregori, Evolved tumor suppression: why are we so good at not getting cancer? *Cancer Res.* **71**, 3739–3744 (2011).

9. T. Domazet-Loso, D. Tautz, Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. *BMC Biol.* **8**, 66 (2010).

10. R. Peto, "Epidemiology, multistage models, and short-term mutagenicity tests" in *The Origins of Human Cancer, Cold Spring Harbor Conferences on Cell Proliferation vol. 4*, H. H. Hiatt, J. D. Watson, J. A. Winsten, Eds. (Cold Spring Harbor Laboratory, 1977), vol. 4, pp. 1403–1428.

11. L. Nunney, C. C. Maley, M. Breen, M. E. Hochberg, J. D. Schiffman, Peto's paradox and the promise of comparative oncology. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370** (2015), doi:10.1098/rstb.2014.0177.

12. M. Effron, L. Griner, K. Benirschke, Nature and Rate of Neoplasia Found in Captive Wild Mammals, Birds, and Reptiles at Necropsy. *J. Natl. Cancer Inst.* **59**, 185–198 (1977).

13. E. G. Duke, S. H. Harrison, A. Moresco, T. Trout, B. V. Troan, M. M. Garner, M. Smith, S. Smith, T. M. Harrison, A Multi-Institutional Collaboration to Understand Neoplasia, Treatment and Survival of Snakes. *Animals (Basel)*. **12** (2022), doi:10.3390/ani12030258.

14. C. V. Kitsoulis, A. D. Baxevanis, T. J. Abatzopoulos, The occurrence of cancer in vertebrates: a mini review. *J. Biol. Res.* . **27**, 9 (2020).

15. T. Madsen, A. Arnal, M. Vittecoq, F. Bernex, J. Abadie, S. Labrut, D. Garcia, D. Faugère, K. Lemberger, C. Beckmann, B. Roche, F. Thomas, B. Ujvari, "Chapter 2 - Cancer Prevalence and Etiology in Wild and Captive Animals" in *Ecology and Evolution of Cancer*, B. Ujvari, B. Roche, F. Thomas, Eds. (Academic Press, 2017), pp. 11–46.

16. L. M. Abegglen, A. F. Caulin, A. Chan, K. Lee, R. Robinson, M. S. Campbell, W. K. Kiso, D. L. Schmitt, P. J. Waddell, S. Bhaskara, S. T. Jensen, C. C. Maley, J. D. Schiffman, Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. *JAMA*. **314**, 1850–1860 (2015).

17. A. M. Boddy, L. M. Abegglen, A. P. Pessier, J. D. Schiffman, C. C. Maley, C. Witte, Lifetime cancer prevalence and life history traits in mammals. *Evol Med Public Health* (2020), doi:10.1093/emph/eoaa015.

18. O. Vincze, F. Colchero, J.-F. Lemaître, D. A. Conde, S. Pavard, M. Bieuville, A. O. Urrutia, B. Ujvari, A. M. Boddy, C. C. Maley, F. Thomas, M. Giraudeau, Cancer risk across mammals. *Nature*. **601**, 263–267 (2022).

19. K. Lika, S. A. L. M. Kooijman, Life history implications of allocation to growth versus reproduction in dynamic energy budgets. *Bull. Math. Biol.* **65**, 809–834 (2003).

20. J. Bielby, G. M. Mace, O. R. P. Bininda-Emonds, M. Cardillo, J. L. Gittleman, K. E. Jones, C. D. L. Orme, A. Purvis, The fast-slow continuum in mammalian life history: an empirical reevaluation. *Am. Nat.* **169**, 748–757 (2007).

21. H. Kokko, M. E. Hochberg, Towards cancer-aware life-history modelling. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370** (2015), doi:10.1098/rstb.2014.0234.

22. S. N. Austad, K. E. Fischer, Mammalian aging, metabolism, and ecology: evidence from the bats and marsupials. *J. Gerontol.* **46**, B47–53 (1991).

23. H. Tian, Z. Gao, H. Z. Li, B. F. Zhang, G. Wang, Q. Zhang, DNA damage response–a double-edged sword in cancer prevention and cancer therapy. *Cancer Lett.* (2015) (available at https://www.sciencedirect.com/science/article/pii/S0304383514007964?casa_token=okGx7LzNOucAAAAA:asBPZzeyl1l5YoYnIX6cF-ZkMZBGC2J4oAi2v5skylpHHwQki3e1EToseGGXu9QY9mt5pyTsswY).

24. M. C. White, D. M. Holman, J. E. Boehm, L. A. Peipins, M. Grossman, S. J. Henley, Age and cancer risk: a potentially modifiable relationship. *Am. J. Prev. Med.* **46**, S7–15 (2014).

25. J. Campisi, Senescent cells, tumor suppression, and organismal aging: good citizens, bad neighbors. *Cell*. **120**, 513–522 (2005).

26. J. Campisi, Aging, tumor suppression and cancer: high wire-act! *Mech. Ageing Dev.* **126**, 51–58 (2005).

27. P. Hasty, J. Campisi, J. Hoeijmakers, H. van Steeg, J. Vijg, Aging and genome maintenance: lessons from the mouse? *Science*. **299**, 1355–1359 (2003).

28. G. A. Garinis, G. T. J. van der Horst, J. Vijg, J. H. J. Hoeijmakers, DNA damage and ageing: new-age ideas for an age-old problem. *Nat. Cell Biol.* **10**, 1241–1247 (2008).

29. T. B. Kirkwood, S. N. Austad, Why do we age? *Nature*. **408**, 233–238 (2000).

30. J. Felsenstein, Phylogenies and the Comparative Method. *Am. Nat.* **125**, 1–15 (1985).

31. R. R. Development Core Team, R: A language and environment for statistical computing (2011)

32. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other

things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

33. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. **20**, 289–290 (2004).

34. D. Orme, R. Freckleton, G. Thomas, T. Petzoldt, S. Fritz, N. Isaac, Others, The caper package: comparative analysis of phylogenetics and evolution in R. *R package version*. **5**, 1–36 (2013).

35. M. Pagel, Inferring the historical patterns of biological evolution. *Nature*. **401**, 877–884 (1999).

36. K. E. Jones, J. Bielby, M. Cardillo, S. A. Fritz, J. O'Dell, C. D. L. Orme, K. Safi, W. Sechrest, E. H. Boakes, C. Carbone, Others, PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184. *Ecology*. **90**, 2648–2648 (2009).

37. P. Calabrese, D. Shibata, A simple algebraic cancer equation: calculating how cancers may arise with normal mutation rates. *BMC Cancer*. **10**, 3 (2010).

38. A. F. Caulin, T. A. Graham, L. S. Wang, C. C. Maley, Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370** (2015), doi:10.1098/rstb.2014.0222.

39. M. Tollis, J. Robbins, A. E. Webb, L. F. K. Kuderna, A. F. Caulin, J. D. Garcia, M. Bèrubè, N. Pourmand, T. Marques-Bonet, M. J. O'Connell, P. J. Palsbøll, C. C. Maley, Return to the Sea, Get Huge, Beat Cancer: An Analysis of Cetacean Genomes Including an Assembly for the Humpback Whale (Megaptera novaeangliae). *Mol. Biol. Evol.* **36**, 1746–1763 (2019).

40. M. Sulak, L. Fong, K. Mika, S. Chigurupati, L. Yon, N. P. Mongan, R. D. Emes, V. J. Lynch, TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *Elife*. **5** (2016), doi:10.7554/eLife.11994.

41. J. M. Vazquez, M. Sulak, S. Chigurupati, V. J. Lynch, A Zombie LIF Gene in Elephants Is Upregulated by TP53 to Induce Apoptosis in Response to DNA Damage. *Cell Rep.* **24**, 1765–1776 (2018).

42. S. Glaberman, S. E. Bulls, J. M. Vazquez, Y. Chiari, V. J. Lynch, Concurrent Evolution of Antiaging Gene Duplications and Cellular Phenotypes in Long-Lived Turtles. *Genome Biol. Evol.* **13** (2021), doi:10.1093/gbe/evab244.

43. J. M. Vazquez, V. J. Lynch, Pervasive duplication of tumor suppressors in Afrotherians during the evolution of large bodies and reduced cancer risk. *Elife*. **10** (2021), doi:10.7554/eLife.65041.

44. F. Islami, E. M. Ward, H. Sung, K. A. Cronin, F. K. L. Tangka, R. L. Sherman, J. Zhao, R. N. Anderson, S. J. Henley, K. R. Yabroff, A. Jemal, V. B. Benard, Annual Report to the Nation on the Status of Cancer, Part 1: National Cancer Statistics. *J. Natl. Cancer Inst.* **113**, 1648–1669 (2021).

45. M. Tollis, A. K. Schneider-Utaka, C. C. Maley, The Evolution of Human Cancer Gene Duplications across Mammals. *Mol. Biol. Evol.* **37**, 2875–2886 (2020).

46. J. W. Beard, AVIAN VIRUS GROWTHS AND THEIR ETIOLOGIC AGENTS. *Adv. Cancer Res.* **7**, 1–127 (1963).

47. A. I. Schulte-Hostedde, G. F. Mastromonaco, Integrating evolution in the management of captive zoo populations. *Evol. Appl.* **8**, 413–422 (2015).

48. G. S. Watson, Age incidence curves for cancer. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 1341–1342 (1977).

31. R Core Team (2021). R: A language and environment for statistical

computing. R Foundation for Statistical Computing, Vienna, Austria. URL

https://www.R-project.org/.