

Cancer risk across mammals

<https://doi.org/10.1038/s41586-021-04224-5>

Received: 26 April 2021

Accepted: 9 November 2021

Published online: 22 December 2021

Open access

 Check for updates

Orsolya Vincze^{1,2,3,4,✉}, Fernando Colchero^{5,6,7}, Jean-Francois Lemaître⁸, Dalia A. Conde^{6,7,9}, Samuel Pavard¹⁰, Margaux Bieuville¹⁰, Araxi O. Urrutia^{11,12}, Beata Ujvari¹³, Amy M. Boddy¹⁴, Carlo C. Maley¹⁵, Frédéric Thomas¹ & Mathieu Giraudeau^{1,2}

Cancer is a ubiquitous disease of metazoans, predicted to disproportionately affect larger, long-lived organisms owing to their greater number of cell divisions, and thus increased probability of somatic mutations^{1,2}. While elevated cancer risk with larger body size and/or longevity has been documented within species^{3–5}, Peto's paradox indicates the apparent lack of such an association among taxa⁶. Yet, unequivocal empirical evidence for Peto's paradox is lacking, stemming from the difficulty of estimating cancer risk in non-model species. Here we build and analyse a database on cancer-related mortality using data on adult zoo mammals (110,148 individuals, 191 species) and map age-controlled cancer mortality to the mammalian tree of life. We demonstrate the universality and high frequency of oncogenic phenomena in mammals and reveal substantial differences in cancer mortality across major mammalian orders. We show that the phylogenetic distribution of cancer mortality is associated with diet, with carnivorous mammals (especially mammal-consuming ones) facing the highest cancer-related mortality. Moreover, we provide unequivocal evidence for the body size and longevity components of Peto's paradox by showing that cancer mortality risk is largely independent of both body mass and adult life expectancy across species. These results highlight the key role of life-history evolution in shaping cancer resistance and provide major advancements in the quest for natural anticancer defences.

Complex multicellular organisms are built of millions to quadrillions of cells, ultimately all being derived from a single cell, the zygote. During the course of the organisms' lifetime and owing to various mutational processes, cell lineages tend to accumulate mutations^{7,8}. While the majority of mutations are harmless, some enable cells to escape cell cycle control, to grow and proliferate uncontrollably, resulting in cancer^{9,10}. Cancer is a multistage process, where a set of mutations is required for both initiation and malignant progression. Given that every cell division carries a risk of generating mutations, organisms with large bodies (composed of more cells) and extended longevity (with a longer time to accumulate mutations) should be more likely to develop cancer^{1,6,11,12}. Indeed, within humans^{3,11} and dogs⁴, larger individuals are more likely to develop cancer than smaller ones. Similarly, increasing age is one of the most potent carcinogenic factors in species in which cancer aetiology is well studied. Yet, while current evidence suggests that large body size and extended longevity result in increased cancer risk within species, this relationship may not hold across taxa¹³.

Limited data available so far indicate that vertebrates do not face clear size-dependent cancer risks despite their size and longevity varying by orders of magnitude. This poses a logical challenge, first

formulated by Sir Richard Peto^{6,14}. He noted that although mice have approximately 1,000 times fewer cells and >30 times shorter lifespans than humans, their risk of carcinogenesis is not markedly different (coined as Peto's paradox)⁵. Peto's paradox is an evolutionary conundrum that has puzzled the scientific community and has led to lively debate regarding the evolution of anticancer mechanisms. It is often postulated that natural selection on large size or extended longevity is inherently inseparable from the evolution of anticancer defences. Knowledge gained from investigating Peto's paradox might thus largely contribute to our knowledge on natural anticancer mechanisms that could potentially be harnessed for medical use. Further, understanding cross-species variation of cancer vulnerability is an important next step in animal health and welfare. While a few studies aimed to establish cross-species variation in cancer risk^{15,16}, most estimates and analyses have considerable limitations. These include small cross- or within-species sample sizes^{1,6,17}, lacking information on the age distribution of cancer^{15–17}, data heterogeneity (for example, biases due to domestication¹⁷ or combining data from multiple taxa^{17,18}) or lack of control for phylogenetic relatedness among species¹⁷. Moreover, the effect of longevity was generally tested using the much-debated metric

¹CREEC/CANECEV, MIVEGEC (CREES), University of Montpellier, CNRS, IRD, Montpellier, France. ²Littoral, Environnement et Sociétés (LIENSs), UMR 7266 CNRS-La Rochelle Université, La Rochelle, France. ³Institute of Aquatic Ecology, Centre for Ecological Research, Debrecen, Hungary. ⁴Evolutionary Ecology Group, Hungarian Department of Biology and Ecology, Babeş-Bolyai University, Cluj-Napoca, Romania. ⁵Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark. ⁶Interdisciplinary Centre on Population Dynamics, University of Southern Denmark, Odense, Denmark. ⁷Species360 Conservation Science Alliance, Bloomington, MN, USA. ⁸Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, Université Lyon 1; CNRS, UMR5558, Villeurbanne, France. ⁹Department of Biology, University of Southern Denmark, Odense, Denmark. ¹⁰Eco-Anthropologie (EA), Muséum National d'Histoire Naturelle, CNRS, Université de Paris, Musée de l'Homme, Paris, France. ¹¹Instituto de Ecología, UNAM, Mexico City, Mexico. ¹²Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, UK. ¹³Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, Victoria, Australia. ¹⁴Department of Anthropology, University of California Santa Barbara, Santa Barbara, CA, USA. ¹⁵Arizona Cancer Evolution Center, Biodesign Institute and School of Life Sciences, Arizona State University, Tempe, AZ, USA. ✉e-mail: vincze.orsolya@ecolres.hu

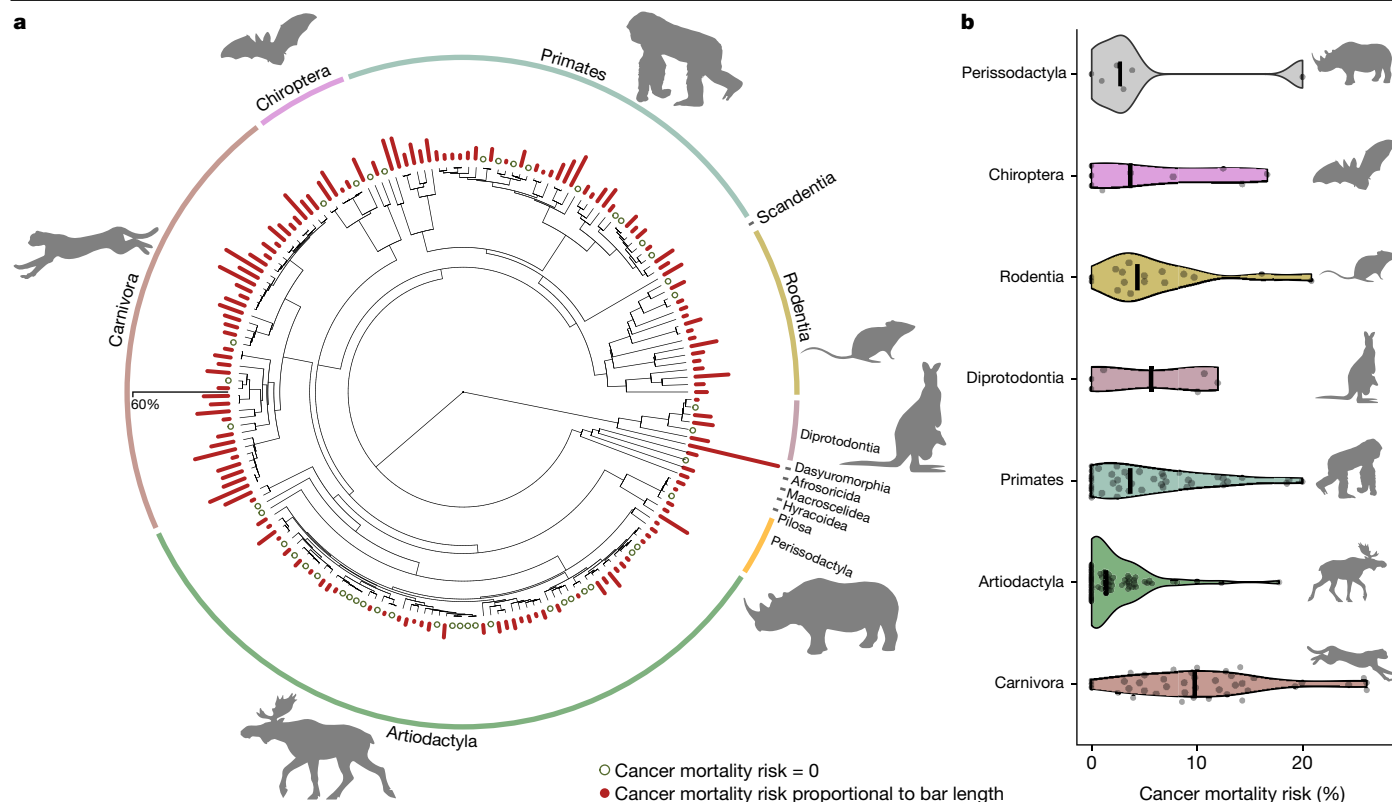


Fig. 1 | Distribution of cancer mortality risk across the mammalian phylogeny. a, CMR in various mammals (scale to bar plots is provided on the left of the graph). **b**, Violin plots indicating order differences in CMR in orders

with a minimum of two species represented. Solid black lines represent order medians. Animal silhouettes used to visually represent mammalian orders were downloaded from PhyloPic (<http://www.phylopic.org>).

of maximum reported lifespan^{15–17}. Nonetheless, cancer prevalence (the parameter exclusively considered by earlier studies^{15–17}) is expected to correlate with life expectancy (the average time lived by individuals in the population of interest), not maximum lifespan potential (that very few individuals achieve), making these analyses inherently flawed.

To characterize cancer incidence in a homogeneous sample and across a wide taxonomic range, here we used the Zoological Information Management System (ZIMS), managed by Species360 (a non-profit organization custodian of zoo and aquarium data)¹⁹. We assembled information on 110,148 adult non-domesticated mammals distributed over 191 species, including data on their age, sex, dead/alive status and postmortem pathological records for 11,840 individuals. Cancer is registered in this database only for deceased animals and only if the inspecting veterinary pathologist considered it to be a factor that contributed to the individual's death. First, to characterize species longevities, we used survival modelling ($n = 110,148$) and calculated species-specific adult life expectancies, representing average adult longevity in our sample^{20,21}. Second, to estimate cancer mortality risk, we used two metrics, both estimating the proportion of individuals dying of cancer. We first calculated a simple measure of cancer mortality risk (hereafter CMR; that is, the ratio between the number of cancer-related deaths and the total number of individuals whose post-mortem pathological records were entered in the database, $n = 11,840$), a measure adopted by earlier comparative studies^{15,17}. This measure relies solely on dead individuals, ignoring the incomplete records of live animals, potentially introducing bias in cancer mortality estimates. Therefore, we also calculated the cumulative incidence of cancer mortality (hereafter ICM), a metric of cancer mortality risk eliminating potential biases due to disregarding left-truncation (that is, cancer before individuals enter the study) and right-censoring (individuals alive, thus with unknown fate at data extraction). Using these two metrics, we explored the phylogenetic distribution of cancer-related

mortality across mammals. We then investigated Peto's paradox and tested whether cancer mortality risk is associated with body size or the mean number of years lived by adults in the explored populations (that is, adult life expectancy).

Cancer across the mammalian phylogeny

CMR was highly variable among species, ranging from 0% (in 47 species out of 191) to 57.14% in the kowari (*Dasyuroides byrnei*). CMR exceeded 10% in 41 species (21.5% of all species inspected), indicating that the oncogenic process is a prevailing source of mortality of many mammalian species distributed along the phylogeny, at least in managed populations (Fig. 1). ICM showed strong consistency with CMR (Pearson correlation test, $r = 0.89$, $t = 25.14$, $df = 170$, $P < 0.0001$). Nonetheless, all models were performed using both metrics to test for consistency in the results.

Cross-species variation in cancer risk showed strong phylogenetic signal²² (CMR: $n = 191$, $\lambda = 0.87$, $P < 0.0001$; ICM: $n = 172$, $\lambda = 0.69$, $P < 0.0001$). To explore this, we compared cancer risk among mammalian orders represented by at least two species using linear regressions. Results indicated that the phylogenetic signal was mostly driven by cancer mortality risk in Carnivora, which was significantly higher than in Primates or in Artiodactyla (Extended Data Table 1, Fig. 1 and Extended Data Fig. 1). Both cancer mortality risk metrics indicated that Artiodactyla is the least cancer-prone mammalian order, despite the frequency of large-bodied species in this group (Extended Data Table 1).

High cancer risk in managed populations of Carnivora has previously been reported^{23,24}. Possible explanations include the use of hormonal contraception (for example, progestins) and pregnancy postponement in zoo carnivores, both being significant risk factors for certain cancers in humans as well as non-domestic felids^{24–26}. Nonetheless, if contraception was the key factor driving elevated cancer risk

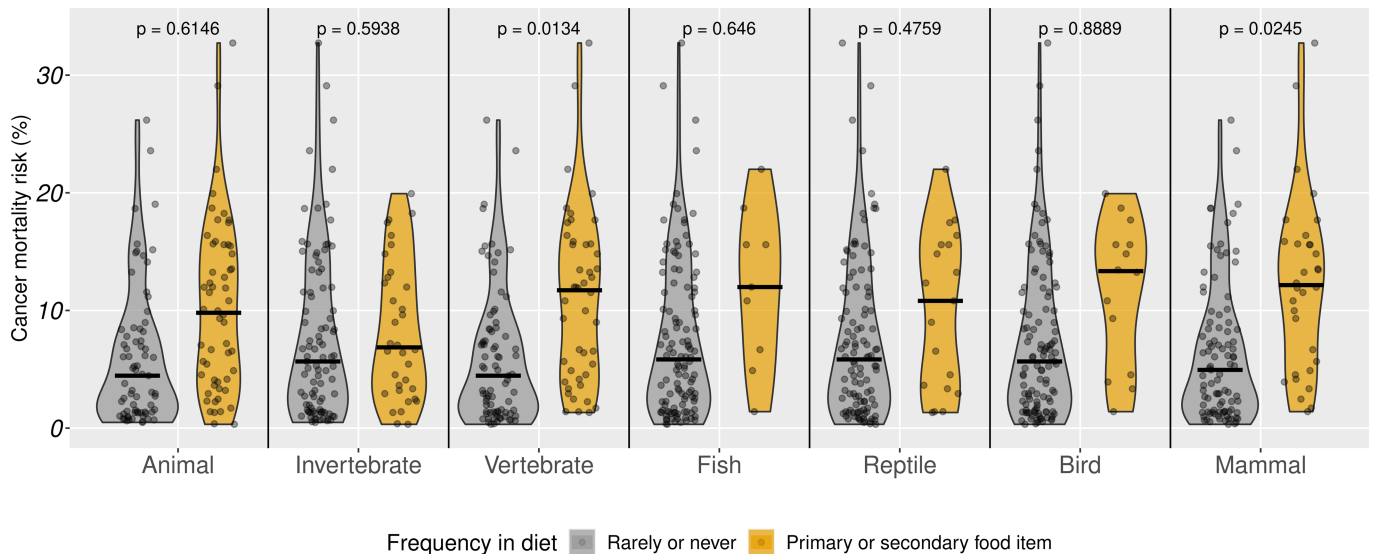


Fig. 2 | Cancer mortality risk in mammals as a function of animal content in diet. Violin plots show CMR as a function of seven diet items, each coded as rarely/never occurring in the diet or representing the primary/secondary food

item of the species. Medians are marked with solid black lines. *P* values indicate pairwise differences as indicated by models presented in Extended Data Table 2 that also control for body mass, life expectancy and phylogeny.

in Carnivora, a significant sex bias in cancer risk would be expected in this group, because hormonal contraception is usually administered to females. To test for sex bias in cancer risk across Carnivora, we estimated sex-specific CMR and ICM (only species with a minimum of ten males and ten females with available postmortem pathological records: $n = 36$ and $n = 30$ species, respectively). Pairwise comparison between sexes revealed no sex bias in either measure of cancer mortality risk (phylogenetic paired *t*-tests, CMR: $t = 0.52$, $df = 33$, $P = 0.6061$; ICM: $t = -0.6815$, $df = 27$, $P = 0.5014$) (ref. ²³). Therefore, the generally high cancer risk in Carnivora is unlikely to be driven solely by the carcinogenic effects of reproductive management in zoo populations.

A high-fat, low-fibre diet, a known risk factor for carcinogenesis, has also been suggested to explain the elevated cancer risk in Carnivora^{26,27}. Moreover, carnivores are on the top of the food chain, exposing them to bio-magnified effects of carcinogenic compounds²⁸, such as pollutants²⁶. Importantly, the consumption of raw meat can also expose carnivores to pathogens that can drive oncogenic transformation²⁹. For instance, in humans it was estimated that 10–20% of all cancers are of viral origin³⁰. While this figure is unknown in any other animal species²⁹, it is arguable that raw meat consumption might exacerbate the spread of carcinogenic pathogens³¹. Exploring the association between diet and cancer risk could help to disentangle the influences of these risk factors.

To explore the link between carnivorous diet and cancer risk, we collected data on the species' natural diet (that is, consumption of animals, including invertebrates or vertebrates, and specifically of fish, reptiles, birds and mammals) from the literature³². Phylogenetic generalized least-squares (PGLS) regressions controlling for differences in longevity and body mass run separately for each diet item (Fig. 2, Extended Data Figs. 2 and 3 and Extended Data Table 2) indicated that species with animal-based diets have comparable cancer mortality risks (both CMR and ICM) to species that rarely or never consume animals. Nonetheless, consumption of vertebrate but not invertebrate prey was associated with increased cancer mortality risk. Specifically, mammals frequently consuming mammalian prey had significantly higher cancer mortality risk compared to mammals that rarely or never consume other mammals. Similar differences could not be detected in the case of fish, reptile or bird prey frequencies. These results indicate that a carnivorous diet has significant costs in terms of heightened oncogenic predisposition across mammals, particularly for

diets high in mammalian prey. The nonsignificant association between cancer mortality risk and diet content of invertebrate, fish, reptile or bird indicates bio-magnification as a less likely source of elevated cancer risk among Carnivora. Nonetheless, the limited number of species primarily consuming these preys in our sample does not allow for definitive conclusions regarding this hypothesis. By contrast, the result that mammals consuming other mammals appear to have the highest cancer risk of all diet categories is consistent with a pathogenic origin of elevated cancer mortality risk among Carnivora. Host jumping of pathogens is most likely to occur in the case of phylogenetic proximity between the reservoir prey and the predator species³³, making a mammal-to-mammal transmission the most likely host jump scenario. These results suggest that pathogen-driven oncogenesis might have a considerable role in shaping cancer mortality risk in mammals and urges the search for pathogens in various cancer types, while considering the notorious difficulty of proving oncogenic properties of pathogens³⁰. Alternatively, high cancer risk in carnivorous animals might be related to their low microbiome diversity³⁴, limited physical exercise under human care or other aspects of their physiology. Nonetheless, a lack of bias in these results, caused by potential alterations in the diet of housed carnivores, should be confirmed by studying natural populations. Importantly, these results probably reflect a complex, maybe indirect evolutionary link between diet and cancer vulnerability; therefore, the effect of meat consumption on cancer risk should be interpreted with caution.

Test of Peto's paradox

Owing to the large number of zero cancer mortality risk estimates and thus non-Gaussian distributions, cancer risks were analysed using zero-inflated phylogenetic models (Methods), as a function of sample size, body mass and life expectancy. The probability of detecting at least one individual with cancer in a species increased steeply with increasing number of individuals with available postmortem pathological records (Extended Data Table 3). In fact, cancer was detected in at least one individual in almost all species with more than 82 individual pathological records available (Extended Data Figs. 4 and 5). Exceptions were the blackbuck (*Antelope cervicapra*) and the Patagonian mara (*Dolichotis patagonum*), where no cancer was detected despite postmortem pathological records being available for 196 and 213 individuals, respectively.

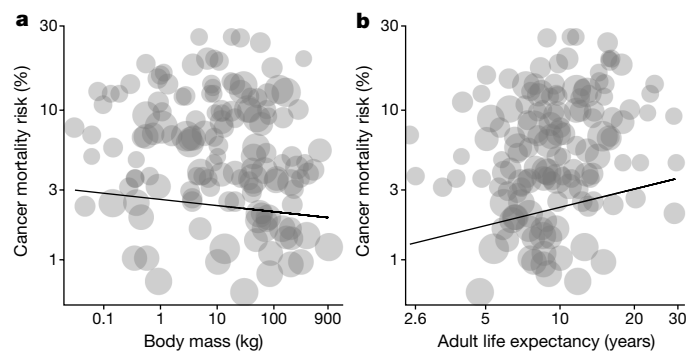


Fig. 3 | Association between cancer mortality risk and body mass or adult life expectancy across mammals. a, b, Non-zero CMR plotted against body mass (**a**) or adult life expectancies (**b**). Slopes were obtained from the PGLS model presented in Extended Data Table 3a. Points are proportional to the log number of individuals with available postmortem pathological records.

This highlights the quasi-universality of oncogenic phenomena across mammals, illustrating that with adequate sampling, cancer is likely to be detected in all mammals. Our results further emphasize the fact that some members of the order Artiodactyla, besides rodents, are particularly cancer-resistant. Rodents have long been subject to scrutiny in the search for natural cancer resistance mechanisms³⁵, owing to notoriously low cancer incidence in some species³⁶. Nonetheless, cancer mortality risk in our dataset was lowest among ruminants, complying with rare cancer case reporting in this taxonomic group^{15,37}. This indicates that other mammalian groups, especially Artiodactyla, might serve as informative model organisms in cancer research.

The probability of detecting cancer (CMR: $n = 188$; ICM: $n = 141$) (Extended Data Table 3 and Extended Data Figs. 4 and 5), as well as non-zero cancer mortality risk (CMR: $n = 141$; ICM: $n = 128$) (Extended Data Table 3 and Fig. 3), tended to decrease with larger body masses and to increase with longer life expectancies. These effects were not significant, and were consistent between the two cancer mortality metrics. These associations were also largely independent of each other, and were similar in single-predictor models (Supplementary Table 3). The effect of body mass on cancer risk was even slightly negative, but the models indicated only a 2.8–2.9% decrease in cancer risk for a doubling of the body weight (for example, CMR changes from 3.82% (1 kg) to 3.71% (2 kg), or from 2.86% (1,024 kg) to 2.78% (2,048 kg), respectively, predictions obtained from the model presented in Extended Data Table 3, with an arbitrary life expectancy of 27 years, Extended Data Fig. 6). Additionally, body mass accounted for only 0.78% of the cross-species variance in CMR (that is, partial coefficient of determination). Similarly, cancer mortality risk only minimally and nonsignificantly increased with higher life expectancy, indicating a 24.7–25.2% increase in cancer risk for a doubling of the adult life expectancy (for example, CMR changes from 0.89% (1 year) to 1.18% (2 years), or from 2.80% (16 years) to 3.72% (32 years), respectively, predictions obtained from the model presented in Extended Data Table 3 with an arbitrary body mass of 10 kg, Extended Data Fig. 6). Adult life expectancy accounted for only 2.94% of the variance observed in CMR (that is, partial coefficient of determination). Overall, these results provide the largest-scale and most robust support to the body size and life expectancy components of Peto's paradox in mammals. They suggest that lifespan extension and larger body size jointly evolved with better anticancer mechanisms across mammals.

Since the first indication of species differences in cancer predisposition, an intense search has been conducted to identify mechanisms explaining cancer resistance in certain species, mostly rodents³⁵, and very large animals¹⁷. Although these studies demonstrated key species-specific anticancer mechanisms^{17,28,35}, a considerable gap remains in our knowledge on the taxonomic and phylogenetic diversity

and distribution of tumour-suppressing mechanisms. Our results provide a solid foundation for future studies scrutinizing these questions, by providing information on the generality and frequency of oncogenic phenomena across the mammalian phylogeny. We also highlight the exceptional resources provided by zoos for studies of cancer in wildlife^{35,36}.

Our study indicates that death due to oncogenic phenomena is frequent and taxonomically widespread in mammals. In some species more than 20–40% of the managed adult population die of cancer-related pathologies. This estimate is staggering, especially knowing that cancer incidences estimated here are conservative (Methods). This observation urges the extensive exploration of cancer in wildlife, especially in the context of recent environmental perturbations³⁸, as serious threats to animal welfare²⁹.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-04224-5>.

- Leroi, A. M., Koufopanou, V. & Burt, A. Cancer selection. *Nat. Rev. Cancer* **3**, 226–231 (2003).
- Armitage, P. & Doll, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **8**, 1–12 (1954).
- Wirén, S. et al. Pooled cohort study on height and risk of cancer and cancer death. *Cancer Causes Control* **25**, 151–159 (2014).
- Fleming, J. M., Creevy, K. E. & Promislow, D. E. L. Mortality in North American dogs from 1984 to 2004: an investigation into age-, size-, and breed-related causes of death. *J. Vet. Intern. Med.* **25**, 187–198 (2011).
- Nunney, L. Lineage selection and the evolution of multistage carcinogenesis. *Proc. R. Soc. B* **266**, 493–498 (1999).
- Peto, R. in *Origins of Human Cancer* Vol 45 (eds. Hiatt, H. et al.) 1403–1428 (Cold Spring Harbor Laboratory, 1977).
- Couzin-Frankel, J. The bad luck of cancer. *Science* **347**, 12 (2015).
- Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **58**, 235–263 (2017).
- Nunney, L. The real war on cancer: the evolutionary dynamics of cancer suppression. *Evol. Appl.* **6**, 11–19 (2013).
- Ujvari, B., Roche, B. & Thomas, F. *Ecology and Evolution of Cancer* (Academic, 2017).
- Nunney, L. Size matters: height, cell number and a person's risk of cancer. *Proc. R. Soc. B* **285**, 20181743 (2018).
- Caulin, A. F. & Maley, C. C. Peto's paradox: evolution's prescription for cancer prevention. *Trends Ecol. Evol.* **26**, 175–182 (2011).
- Nunney, L., Maley, C. C., Breen, M., Hochberg, M. E. & Schiffman, J. D. Peto's paradox and the promise of comparative oncology. *Philos. Trans. R. Soc. B* **370**, 20140177 (2015).
- Peto, R. Epidemiology, multistage models, and short-term mutagenicity tests. *Int. J. Epidemiol.* **45**, 621–637 (2016).
- Boddy, A. M. et al. Lifetime cancer prevalence and life history traits in mammals. *Evol. Med. Public Health* **2020**, 187–195 (2020).
- Møller, A. P., Erritzøe, J. & Soler, J. J. Life history, immunity, Peto's paradox and tumours in birds. *J. Evol. Biol.* **30**, 960–967 (2017).
- Abegglen, L. M. et al. Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *JAMA* **314**, 1850–1860 (2015).
- Tollis, M. et al. Elephant genomes reveal accelerated evolution in mechanisms underlying disease defenses. *Mol. Biol. Evol.* **38**, 3606–3620 (2021).
- Conde, D. A. et al. Data gaps and opportunities for comparative and conservation biology. *Proc. Natl Acad. Sci. USA* **116**, 9658–9664 (2019).
- Ronget, V. & Gaillard, J. M. Assessing ageing patterns for comparative analyses of mortality curves: going beyond the use of maximum longevity. *Funct. Ecol.* **34**, 65–75 (2020).
- Tidière, M. et al. Comparative analyses of longevity and senescence reveal variable survival benefits of living in zoos across mammals. *Sci. Rep.* **6**, 36361 (2016).
- Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
- Moresco, A. et al. Taxonomic distribution of neoplasia among non-domestic felid species under managed care. *Animals* **10**, 2376 (2020).
- Moresco, A. *The Pro-carcinogenic Effects of Progesterone on Carnivore Target Tissues* (2009).
- Harrenstien, L. A. et al. Mammary cancer in captive wild felids and risk factors for its development: a retrospective study of the clinical behavior of 31 cases. *J. Zoo Wildl. Med.* **27**, 468–476 (1996).
- Munson, L. & Moresco, A. Comparative pathology of mammary gland cancers in domestic and wild animals. *Breast Disease* **28**, 7–21 (2007).
- Chao, A. et al. Meat consumption and risk of colorectal cancer. *JAMA* **293**, 172–182 (2005).

28. Kelly, B. C., Ikononou, M. G., Blair, J. D., Morin, A. E. & Gobas, F. A. P. C. Food web-specific biomagnification of persistent organic pollutants. *Science* **317**, 236–239 (2007).
29. Pesavento, P. A., Agnew, D., Keel, M. K. & Woolard, K. D. Cancer in wildlife: patterns of emergence. *Nat. Rev. Cancer* **18**, 646–661 (2018).
30. Bogolyubova, A. V. Human oncogenic viruses: old facts and new hypotheses. *Mol. Biol.* **53**, 767–775 (2019).
31. Khatami, A. et al. Bovine leukemia virus (BLV) and risk of breast cancer: a systematic review and meta-analysis of case-control studies. *Infect. Agents Cancer* **15**, 48 (2020).
32. Kissling, W. D. et al. Establishing macroecological trait datasets: digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. *Ecol. Evol.* **4**, 2913–2930 (2014).
33. Olival, K. J. et al. Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
34. Ley, R. E. et al. Evolution of mammals and their gut microbes. *Science* **320**, 1647–1651 (2008).
35. Seluanov, A., Gladyshev, V. N., Vijg, J. & Gorbunova, V. Mechanisms of cancer resistance in long-lived mammals. *Nat. Rev. Cancer* **18**, 433–441 (2018).
36. Herrera-Álvarez, S., Karlsson, E., Ryder, O. A., Lindblad-Toh, K. & Crawford, A. J. How to make a rodent giant: genomic basis and tradeoffs of gigantism in the capybara, the world's largest rodent. *Mol. Biol. Evol.* **38**, 1715–1730 (2021).
37. Albuquerque, T. A. F., Drummond do Val, L., Doherty, A. & de Magalhães, J. P. From humans to hydra: patterns of cancer across the tree of life. *Biol. Rev.* **93**, 1715–1734 (2018).
38. Giraudeau, M., Sepp, T., Ujvari, B., Ewald, P. W. & Thomas, F. Human activities might influence oncogenic processes in wild animal populations. *Nat. Ecol. Evol.* **2**, 1065–1070 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Methods

Documenting cancer in wild animals is extremely challenging in most cases owing to the lack of information on the age of individuals, the difficulty retrieving the bodies for necropsy and the likelihood of cancer negatively influencing survival before cancer itself could be detected. Although data on cancer incidence from wild populations would be indispensable to describe natural incidences of malignancies, such data, especially with corresponding ages and demographic histories, are unfortunately still far from our reach. Therefore, to estimate cancer mortality risk, we used data provided by Species360 and the Zoological Information Management System (ZIMS, Data Use Approval Number 73836), an international non-profit organization that maintains a real-time and centralized database of animals under human care (regrouping information from over 1,200 zoos worldwide). Although we recognize that the interpretation of data gathered on zoo animals requires caution, owing to strong human control on the diet, health, mortality factors, environment or standard biological functions of the animals, zoos provide exceptionally high data resolution on the demography and cause of death for a wide range of species. Here we rely on the high probability of body retrieval of deceased zoo animals and the necropsy routinely performed on most of them (unless found in an advanced stage of decomposition), aiming to identify the most likely pathology causing the death of the animal. These examinations are likely to reveal most solid tumours, but (although possible) benign tumours, liquid tumours (for example, leukaemia) or early-stage cancers are unlikely to be recorded here, either owing to their diagnostic difficulty or their perceived limited role in contributing to the death of the animal.

Specifically, here we use the husbandry module of ZIMS, providing information on birth, death, sex and pre-defined categories of pathological findings, including neoplasia (by definition tumours that contributed to the death, albeit with no option to specify the cancer type or other details). No statistical methods were used to predetermine sample size, but to minimize bias caused by potential temporal heterogeneity in data management practices and necropsy record-keeping¹⁵, here we focused on individuals alive or born after 1 January 2010 (data extraction: 30 May 2020). This sample was then used to characterize species-specific life expectancies and cancer incidence, but only after the exclusion of data that did not fulfil a series of criteria, to ensure the highest and most homogeneous data quality possible. First, cancer is an age-related disease that rarely occurs in juveniles, and pediatric cancers are usually medically distinct from adults' cancers. As such, infant mortality differences observed across species can significantly confound cancer incidence estimates. Therefore, we gathered sex-specific or species-specific (wherever the former was not available) ages at sexual maturity and we considered individuals for analyses only if they reached maturity before or during our sampling period. For individuals of unknown sex (about 12% of all individuals in the raw data extract), the maximum of ages at sexual maturity of males and females was used as an inclusion age threshold. Sex-specific age at sexual maturity for each species was obtained from Conde et al.¹⁹ or from published literature resources (see data sources in <https://github.com/OrsolyaVincze/VinczeEtal2021Nature/blob/main/SupplementaryData.xls>). Second, given that age is a key predictor of cancer emergence, we considered only individuals for which date of birth was recorded precisely or within a narrow (maximum 30 days) time interval. Third, we considered only species in which postmortem pathological records were available for at least 20 adult individuals, irrespective of the cause of death (for example, infection, accident, geriatric disease and so on). Nonetheless, models presented were performed with increased thresholds of 40, 60, 80 and 100 individuals to check for consistency in the results (Supplementary Table 2). Fourth, given that the process of domestication is widely regarded as a major contributing factor to inbreeding depression and higher incidence of cancer³⁹, we excluded all species

that were subject to domestication as well as their wild ancestors (taxa excluded owing to being subject to domestication are listed in Supplementary Table 1). Following these restrictions, data extraction on age and cause of death resulted in information for 110,148 (62,556 live and 47,592 dead) individuals ($n = 191$ species). For calculation of ICM, we included only species in which survival is correctly estimated until old ages (that is, data allowing the estimation of age-specific survival until the age at which only 10% of individuals are surviving, $n = 172$ species). While these restrictions removed multiple sources of bias in our cancer mortality risk estimates, we cannot exclude the possibility that some species (for example, more charismatic ones) are subject to more frequent or more detailed necropsies. Nonetheless, our statistical approach, especially the complete case analysis, is largely insensitive to such biases, as individuals not having available postmortem diagnostic records are considered censored (see below). Also, while the depth of necropsy might vary slightly among species, neoplasia that had a significant contribution to the death of the animals (the focus of our study) are generally detected even at gross necropsies. Additionally, large species are considered of key importance for zoos, also reflected by the fact that the proportion of dead individuals with postmortem pathological records is larger in larger species (Pearson's correlation: $r = 0.24$, $t = 3.35$, $df = 189$, $P = 0.001$). Accordingly, if charisma had a role in cancer detection, we would expect a larger cancer risk in large mammals, opposite to the (nonsignificant) negative body mass effect in our models. Consequently, we believe that charisma is unlikely to represent a major source of bias in our analysis.

Estimation of adult life expectancy

As we have no reason to believe that censored individuals would not have the same prospect of survival as those who continue to be followed, we estimate adult life expectancy from age-specific survival estimated using the Kaplan–Meier procedure (using the `survfit` function in the R package `survival`⁴⁰). Individuals older than their age at sexual maturity on 1 January 2010 were left-truncated at their age at this date; individuals reaching sexual maturity after this date were left-truncated at their age at sexual maturity. Individuals still alive at the time of data extraction were considered right-censored (samples per species varied from 42 to 5,816 individuals), while known fate individuals were assigned as dead ($n = 47,592$), irrespective of whether their cause of death was specified or not.

Estimation of ICM

ICM was calculated using a competing risk approach, based on the cumulative hazard of cancer-related deaths and survival probability of the species under human care. First, age-specific survival S_x at age x , was estimated from KM analysis as above. However, here we performed a complete-case analysis, using only 11,840 individuals for which the cause of death was specified together with right-censored survivors. Complete-case analysis assumes that missingness in the cause of failure is random, but we had no reason to believe that this was not the case in our dataset. Postmortem examinations are routinely carried out on most recovered bodies in zoos, and once examinations are performed the results are equally likely to be entered in the database irrespective of the pathologies identified. ICM estimates were thus based on $n = 74,396$ individuals, $n = 179$ species. Second, the cancer mortality hazard h_x^c was estimated using a KMx1 analysis where only deaths by cancer were incorporated as a death event. ICM is then such that

$$\text{ICM} = \sum_{x=\alpha}^{\infty} S_x h_x^c$$

where α is the age at sexual maturity. The only difference with classic estimation is that we extracted h_x^c (and S_x) for each time unit with discrete jumps (and falls) at event times at age t and with $h_{t \leq x < t+1}^c = 0$ (S_x constant) between these events; instead of estimating discrete

hazard $h_t^c = d_t^c/n_t$ on these time intervals (where d_t^c is the number of deaths by cancer within the interval and n_t is the number of survivors at the beginning of the interval). We chose this method to reflect true variation in the data for the interspecific comparison where species differ greatly in the number of events and time interval between these (sometimes a third of the organism's adult lifespan), a situation rarely met when comparing human groups.

Covariates and statistical analyses

For each species, we obtained sex-specific adult body mass data from Species360's ZIMS (see <https://github.com/OrsolyaVincze/VinczeEtal2021Nature/blob/main/SupplementaryData.xls>). Species-specific body mass was calculated as the average of all body mass measurements recorded in the ZIMS database in adults, while species-specific values were obtained by averaging the body masses of males and females. These were calculated only in species for which there were at least 100 adult body mass records; otherwise, body mass was taken from the literature and database review by Conde et al.¹⁹. We verified that there was a one-to-one correspondence in the body mass information for species with records in both datasets.

Diet information was obtained from a global diet dataset for terrestrial mammals³², providing information on diet composition at four hierarchical levels of food items (never consumed, occasionally consumed, secondary food item, primary food item). We collected information on animal content in diet, as well as subcategories of this diet class, namely invertebrate or vertebrate consumption, as well as specifically fish, reptile, bird and mammal consumption. Given that most food items had few species in the intermediate levels (occasional consumption and secondary food item), we re-categorized the diet variables at two levels: never/rarely consumed or representing the primary/secondary food item of the species. The effect of diet was tested in PGLS regressions using only species with non-zero cancer mortality risks. Models were run separately for each food item that were entered to a base model including body mass and adult life expectancy as covariates. Results are shown in Extended Data Table 2.

To account for statistical non-independence due to phylogenetic relationships, we obtained a sample of 1,000, equally plausible phylogenetic trees, from the posterior distribution published by Upham et al.⁴¹, covering 5,911 species. We then obtained a rooted consensus tree using the sumtrees Python library⁴². Two species recently raised to species level were manually added to the tree as sister taxa of the species it was recently separated from (that is, *Cervus canadensis* to *Cervus elaphus* and *Gazella marica* to *Gazella subgutturosa*). Phylogenetic signal of cancer risk was assessed using the function `phylosig` from the R package `phytools`²². Partial coefficients of determinations were calculated using the function `R2.pred` from the R package `rr2` (ref.⁴³), based on models presented in Extended Data Table 3.

Models of cancer risk testing Peto's paradox were performed using zero-inflated logistic models, which allow us to make inferences on the probability of detecting at least one cancer case in the species and, given that cancer was detected, inferences on the CMR or ICM. Therefore, the first part of this consisted of a phylogenetic binomial regression (using the function `binaryPGLMM`, in the R package `ape`⁴⁴), where the dependent variable explained the presence of zeros and non-zeros in CMR or ICM. This model contained the log number of deceased individuals with available postmortem pathological records as an explanatory variable, due to the higher probability of detecting cancer with a higher number of dead individuals inspected. Additionally, the model contained body mass and adult life expectancy as covariates. The second part of the model consisted of a PGLS regression that investigated variance only in non-zero cancer risks. ICM and CMR were logit-transformed in all PGLS models as recommended when analysing proportions⁴⁵. These models were weighted by log number of deceased individuals with available postmortem pathological records, as the precision of cancer mortality risk estimates is expected to increase with the number

of dead individuals subject to postmortem examination, but it is not expected to explain bias in the estimation of the dependent variable in any particular direction (as in the case of the binomial models). These models also contained body mass and adult life expectancy as explanatory variables. Given the expected additive effect of body mass and longevity, the interaction between body mass and longevity metrics was also tested in all four models (binomial and logistic regressions for CMR and ICM), but these interactions did not increase model fit in any case and are therefore not presented. Both models were controlled for phylogenetic relatedness among species, where the scaling parameter of phylogenetic dependence (that is, s^2 /Pagel's λ in PGLMMs and PGLSs respectively) was set to the most appropriate values assessed by likelihood ratio statistics in each model separately. PGLS models in which Pagel's λ converged to negative values were refitted with Pagel's λ fixed at 0. Three species (*Lagurus lagurus*, *Cricetus cricetus* and *Dasyuroides byrnei*) had been removed from the latter models, due to their high leverage caused by their very low adult life expectancy compared to the rest of the species and therefore concerns of strong influence of these points over model fit. Nonetheless, all models were performed using the entire dataset, and the results were highly consistent with and without the exclusions (Supplementary Table 4 and Extended Data Fig. 7).

Order differences in cancer incidence were tested using standard linear regressions, built using only taxonomic orders in which at least two species had their cancer incidence estimated. The model contained CMR or ICM (non-transformed) as dependent variables and order as the sole explanatory factor. Pairwise order differences were assessed using the R package `emmeans`⁴⁶. All analysis were performed in the R Statistical and Programming Environment, version 4.0.4 (ref.⁴⁷). Cancer mortality risks were transformed to percentages in figures and in the analysis performed on order differences (Extended Data Table 1), for easier interpretation. Models presented in Extended Data Tables 2 and 3 and Supplementary Tables 2–4 are based on probabilities.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data used for the analysis presented in the paper are available at <https://github.com/OrsolyaVincze/VinczeEtal2021Nature/blob/main/SupplementaryData.xls>. Raw data used to estimate cancer risk (Species360 Data Use Approval Number 73836) cannot be publicly shared, as Species360 is the custodian (not the owner) of their members' data. Raw data are accessible through Research Request applications (form available at https://docs.google.com/forms/d/1znoy62VEkDlHAp_ORfEvF7Zsx03g4W5AlppJHqo3_WQ/viewform?edit_requested=true&pli=1). Research Requests are reviewed by both the Species360 Research Committee and their Board of Trustees every four months. The Board of Trustees makes the final decision on data sharing, based on recommendations by the Research Committee. Once Species360 grants access to data, they are intended only for and restricted to use in the project they were approved for and for a single publication. The researcher cannot use them for other projects, publications and/or purposes, nor can the researcher share the data with third parties. For any other inquiries, all of the details for the submission of research requests to Species360 can be found at https://conservation.species360.org/wp-content/uploads/2020/08/Species360-Sharing-Data-v3-3_komprimeret.pdf. Any email communications should be directed to support@species360.org.

Code availability

Data and R code needed to reproduce the analysis are publicly available at <https://github.com/OrsolyaVincze/VinczeEtal2021Nature>.

39. Thomas, F. et al. Rare and unique adaptations to cancer in domesticated species: an untapped resource? *Evol. Appl.* 12920 (2020).
40. Therneau, T. M. & Lumley, T. Package 'survival'. *CRAN* (2014).
41. Upham, N. S., Esselstyn, J. A. & Jetz, W. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol.* **17**, e3000494 (2019).
42. Sukumaran, J. & Holder, M. T. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26**, 1569–1571 (2010).
43. Ives, A. R. R²s for correlated data: phylogenetic models, LMMs, and GLMMs. *Syst. Biol.* **68**, 234–251 (2018).
44. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
45. Warton, D. I. & Hui, F. K. C. The arcsine is asinine: the analysis of proportions in ecology. *Ecology* **92**, 3–10 (2011).
46. Lenth R. V. emmeans: estimated marginal means, aka least-squares means, <https://cran.r-project.org/package=emmeans> (2021).
47. R Core Team. *R: A Language and Environment for Statistical Computing*, <http://www.R-project.org/> (R Foundation for Statistical Computing, 2021).

Acknowledgements We are grateful to P. Bustamante and T. Székely for constructive criticism on an earlier version of the manuscript, and R. Thompson and A. Teare for invaluable explanations of ZIMS data. We are grateful to more than 1,200 zoo and aquarium members of Species360 that record data in ZIMS, making this study possible. O.V. was financed by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the New National Excellence Programme of the Hungarian Ministry of Innovation and Technology. F.T. is

supported by the MAVA Foundation, the ANR TRANSCAN (ANR-18-CE35-0009) and a CNRS International Associated Laboratory Grant. C.C.M. and A.M.B. were supported in part by NIH grant U54 CA217376. C.C.M. was also in part supported by NIH grant U2C CA233254, as well as CDMRP Breast Cancer Research Program Award BC132057 and the Arizona Biomedical Research Commission grant ADHS18-198847. D.A.C. was financed by the Species360 CSA sponsors: Copenhagen Zoo, Wildlife Reserves of Singapore and the World Association of Zoos and Aquariums. 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 United States National Institutes of Health.

Author contributions O.V. and M.G. contributed to the study conception; O.V., F.C., M.B., M.G., D.A.C., J.-F.L. and S.P. contributed to data collection and interpretation. O.V. performed statistical analysis with significant contribution from F.C. and S.P.; O.V. led the writing of the manuscript; all authors contributed to the interpretation of the results and writing of the manuscript. All authors read and approved the final manuscript.

Competing interests The authors declare no competing interests.

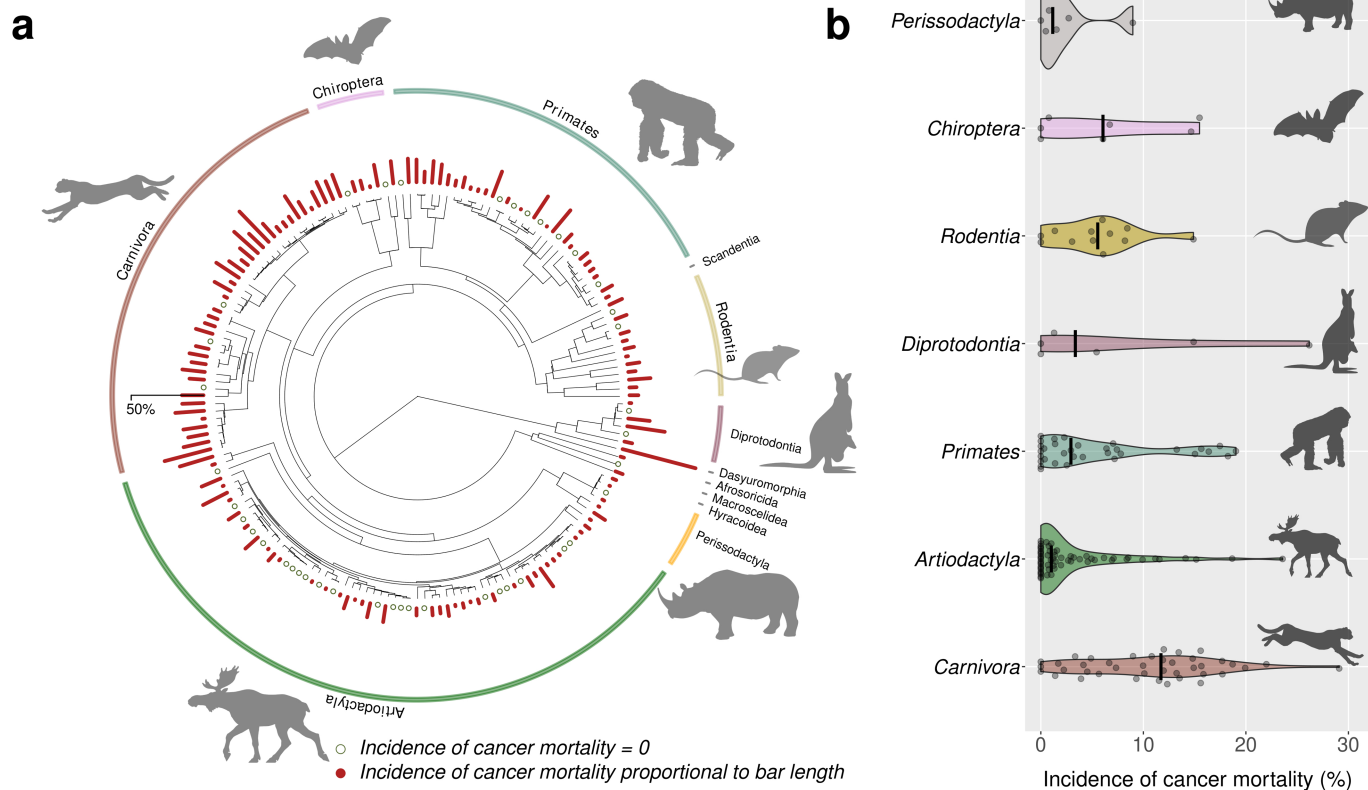
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-04224-5>.

Correspondence and requests for materials should be addressed to Orsolya Vincze.

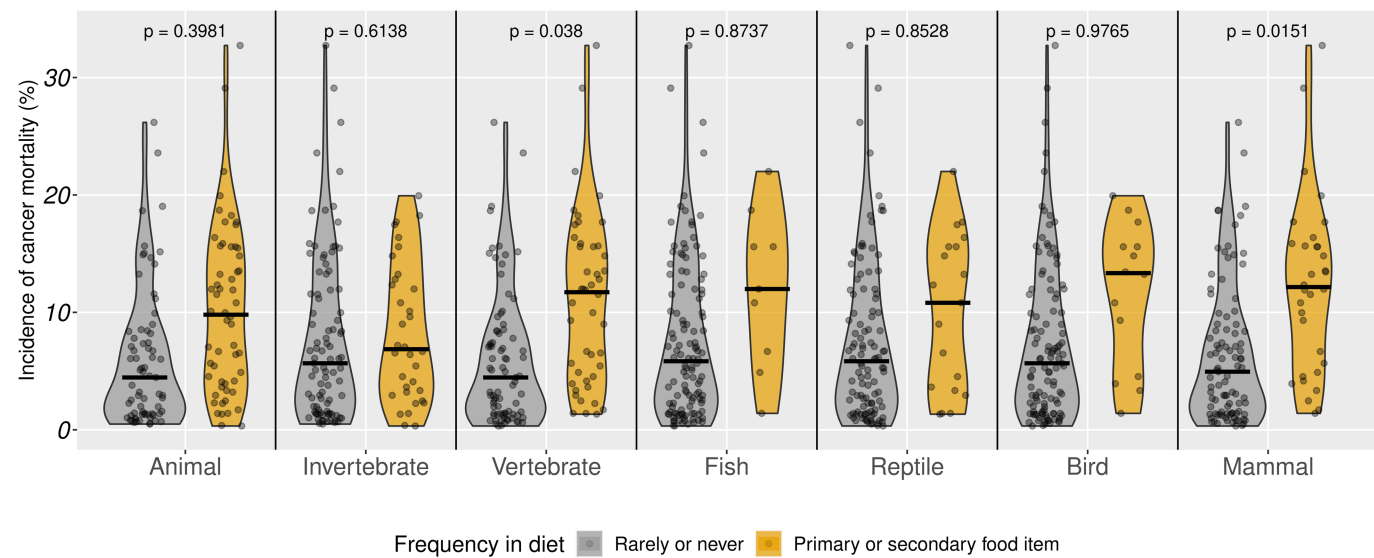
Peer review information *Nature* thanks Oliver Ryder and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



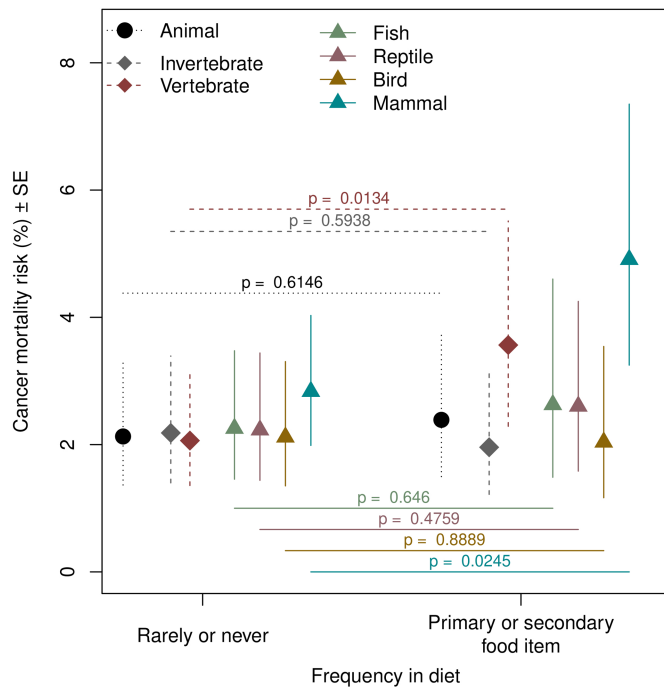
Extended Data Fig. 1 | Phylogenetic distribution and order differences of ICM. a, Phylogenetic distribution of ICM (%). b, Violin plots indicating order differences in ICM (%) in orders with minimum two species assessed. Solid

black lines indicate order medians. Animal silhouettes used to visually represent mammalian orders were downloaded from PhyloPic (<http://www.phylopic.org>).

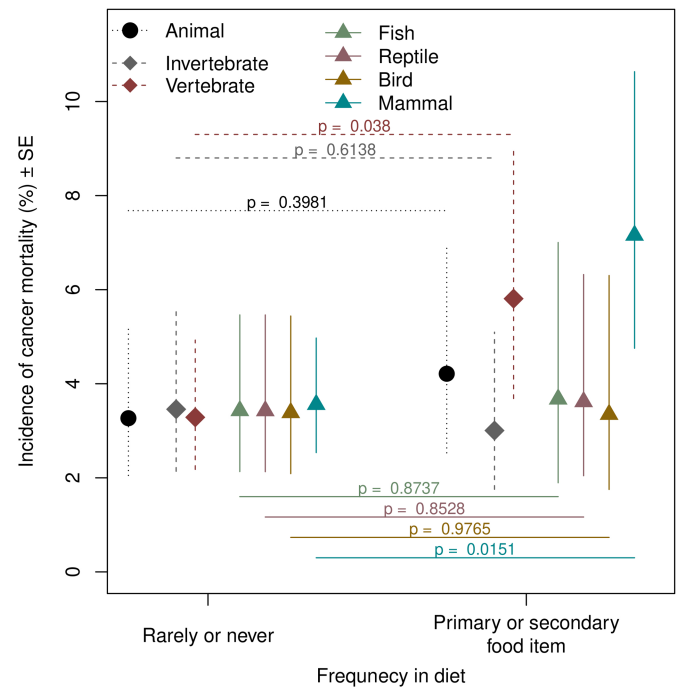


Extended Data Fig. 2 | ICM in function of diet animal content. Violin plots indicate ICM in function of diet animal content. Diet is characterized by variables on three taxonomic levels: I. animal content (including any vertebrate or invertebrate prey); II. invertebrate or vertebrate prey; and III. within vertebrates fish, reptile, bird and mammal preys. Each variable is coded as a two-level factor: rarely/never occurring in diet and representing the primary/

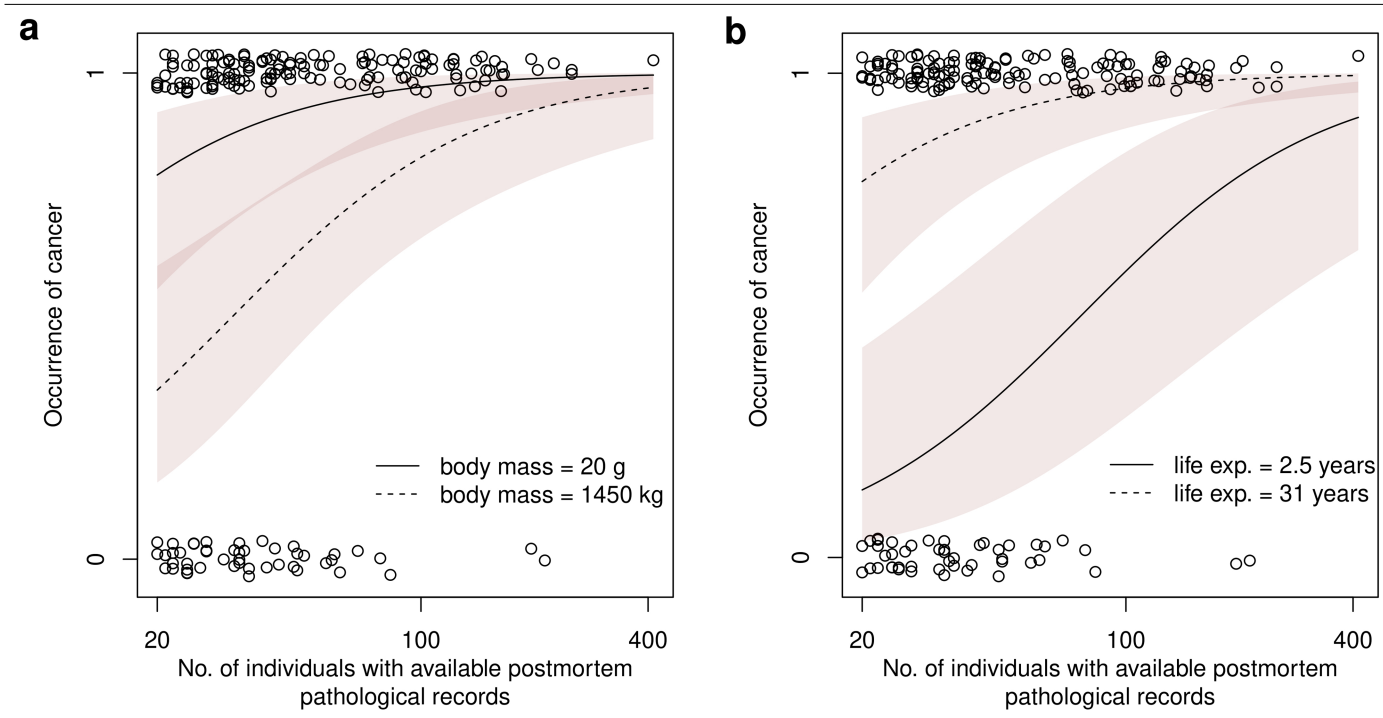
secondary food item of the species. Plots indicate the range and distribution of cancer risks in each category, where the width of each curve corresponds with the approximate frequency of data points in each region. Medians are marked with solid black lines. P values indicate pairwise differences as indicated by models presented in Extended Data Table 2 that also control for body mass, life expectancy and phylogeny.



Extended Data Fig. 3 | Predicted cancer mortality risk in function of animal content in diet. Values represent estimated marginal means of **a**, CMR or **b**, ICM in function of diet animal content, based on models presented in Extended Data Table 2. Diet is characterized by variables on three taxonomic levels: I. animal content (including any vertebrate or invertebrate prey); II. invertebrate

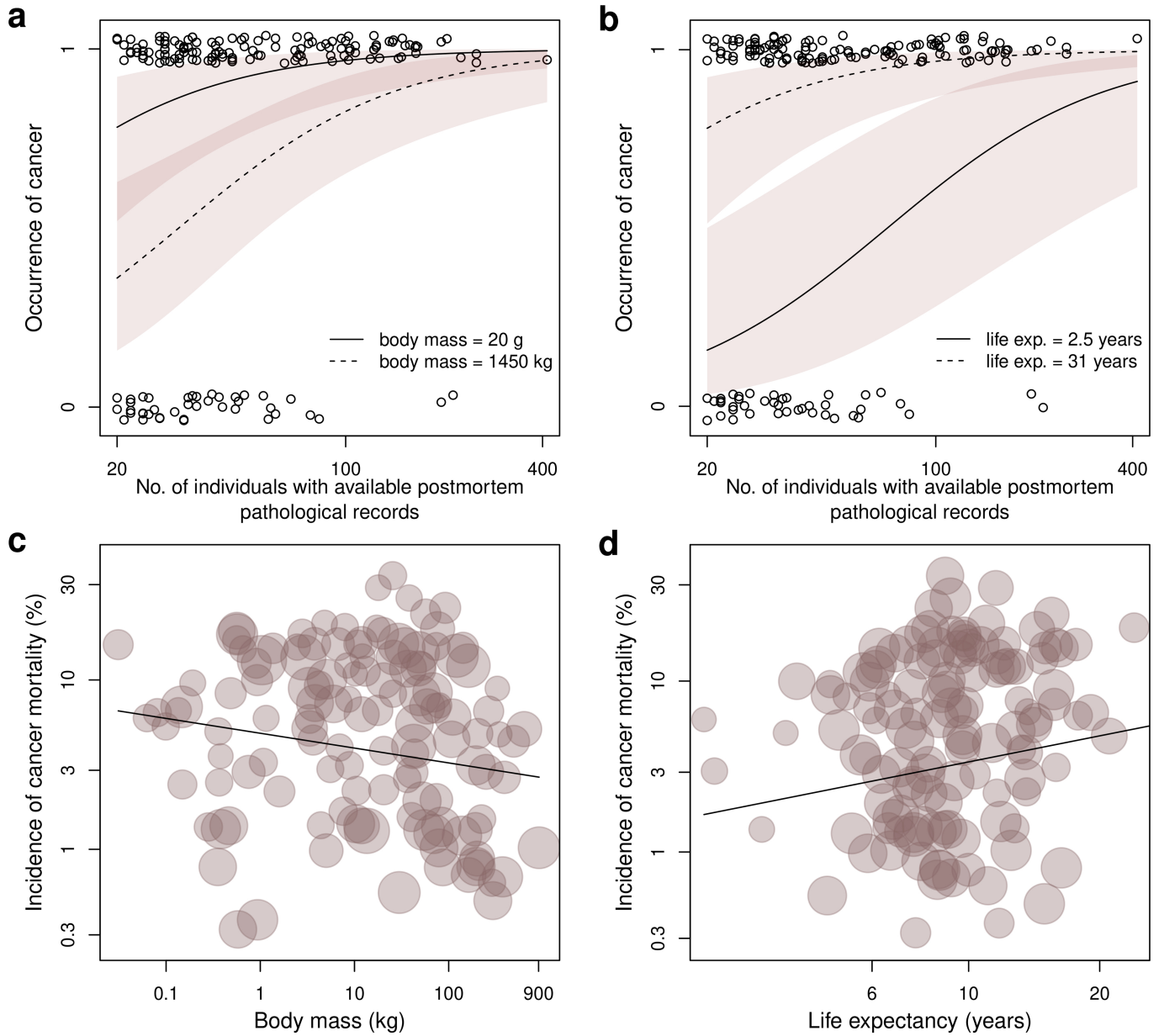


or vertebrate prey; and III. within vertebrates fish, reptile, bird and mammal preys. Each variable is coded as a two-level factor: rarely/never occurring in diet and representing the primary/secondary food item of the species. P-values shown were obtained from models presented in Extended Data Table 2.



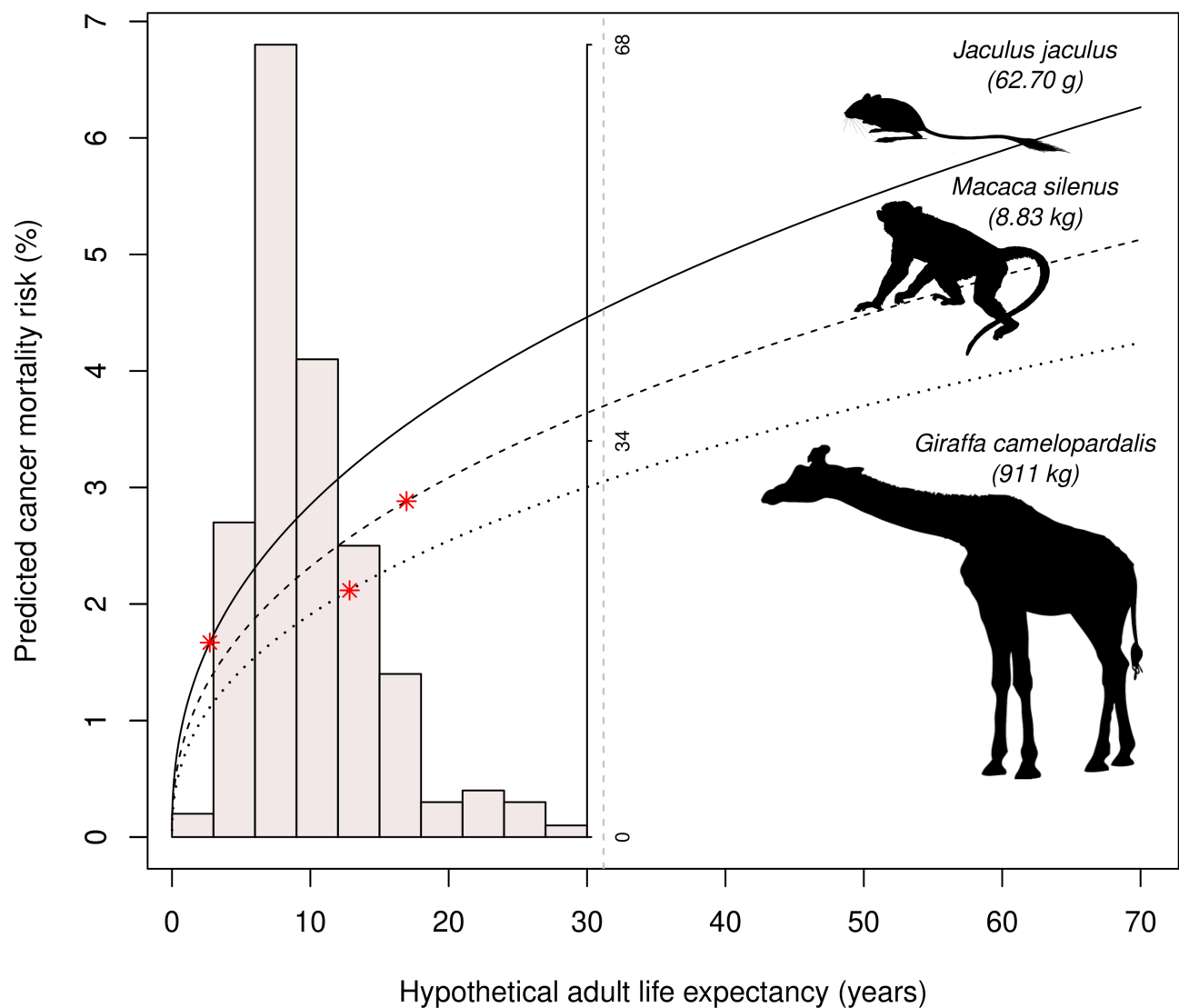
Extended Data Fig. 4 | Association between occurrence of non-zero CMR and a, body mass or b, adult life expectancy across species. Occurrence of cancer in each species is plotted in function of the number of deceased individuals for which cause of death was known. Predictions were obtained for two scenarios, one for low and one for high **a**, body masses and **b**, adult life

expectancies respectively. Random noise was added to cancer occurrence to facilitate the visualization of overlapping points. Predictions and associated 95% confidence intervals were obtained from a binomial GLM without phylogenetic control.



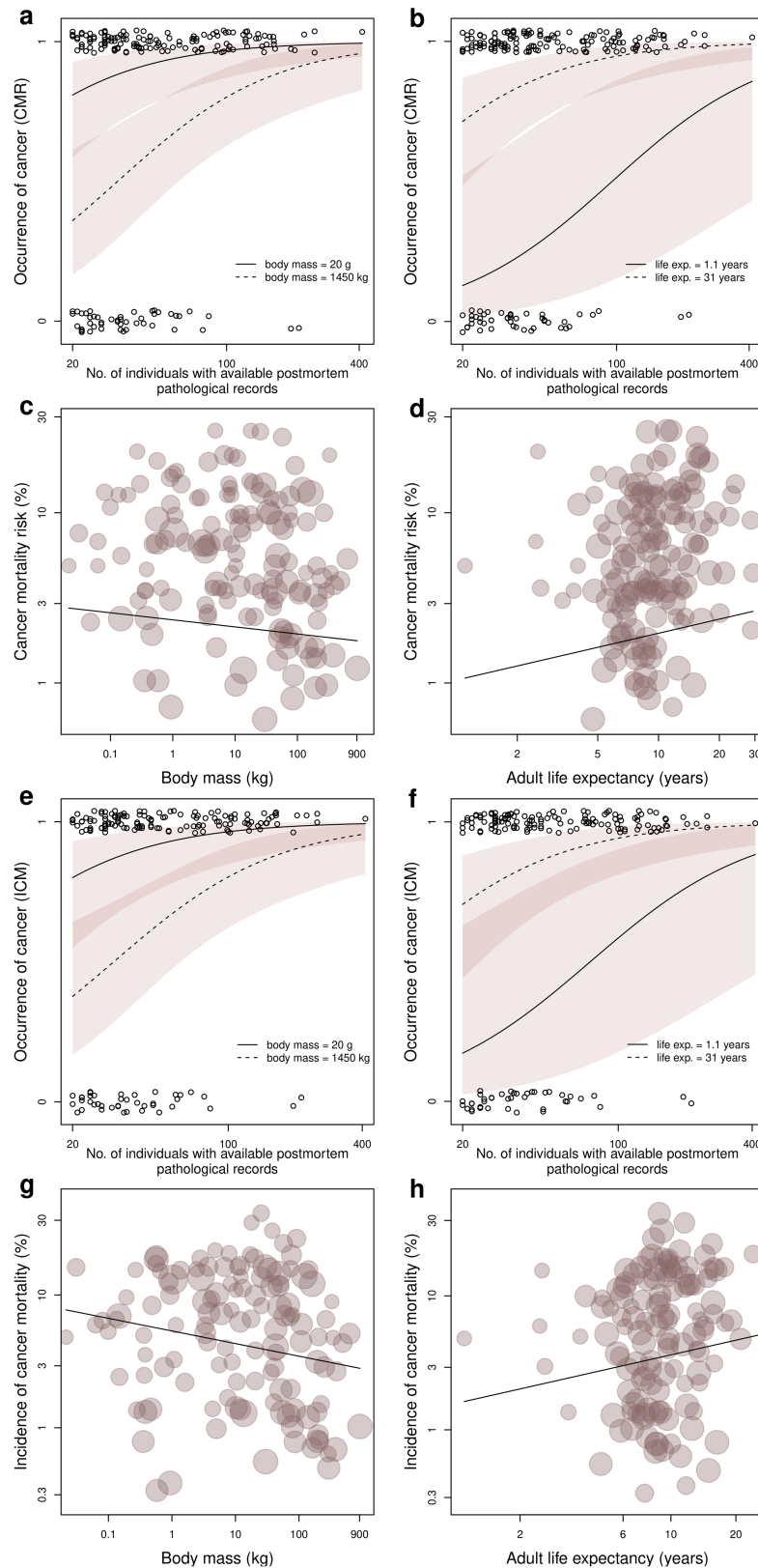
Extended Data Fig. 5 | Association between ICM and a, c, body mass or b, d, adult life expectancy. Occurrence of cancer is plotted in function of the number of deceased individuals for which cause of death was known and predictions were obtained for two scenarios, one for low and high **a**, body masses and **b**, adult life expectancies respectively. Random noise was added to cancer occurrence to facilitate the visualization of overlapping points.

Predictions and associated 95% confidence intervals were obtained from a binomial GLM without phylogenetic control. Non-zero cancer mortality risks were plotted against **c**, body mass or **d**, adult life expectancies. Slopes were obtained from the PGLS model presented in Extended Data Table 3b. Points are proportional to the log number of individuals with known cause of death.



Extended Data Fig. 6 | Predicted CMR at various hypothetical life expectancies for small, medium, and large-bodied species. CMR was predicted based on the logistic model presented in Extended Data Table 3a for a series of hypothetical adult life expectancies, ranging from one to 70 years (x-axis). Predictions were obtained for three body masses, corresponding to a small, medium and a large bodied mammal in our dataset. True life

expectancies of the three species are marked with red stars. Distribution of life expectancies across the species set of the model is shown by a histogram. Vertical dashed line marks maximum life-expectancy in our dataset. *Jaculus jaculus* silhouette by Maija Karala. The other animal silhouettes used to visually represent mammalian orders were downloaded from PhyloPic (<http://www.phylopic.org>).



Extended Data Fig. 7 | Association between cancer mortality risk and body mass as well as adult life expectancy across 191 mammal species. These analysis include the three species excluded in some models due to their high leverages. Plots show **a-d**, CMR or **e-h**, ICM. Occurrence of cancer in each species is plotted in function of the number of individuals with post-mortem pathological records. Predictions were obtained for two scenarios: for small or large **a,e**, body masses and low or high **b,f**, adult life expectancies respectively.

Random noise was added to cancer occurrence to facilitate the visualization of overlapping points. Predictions and associated 95% confidence intervals were obtained from a binomial GLM without phylogenetic control. Non-zero cancer mortality risks were plotted against **c,g**, body mass or **d,h**, adult life expectancies. Slopes were obtained from the PGLS model presented in Supplementary Table 4. Points are proportional to the log number of individuals with known cause of death.

Article

Extended Data Table 1 | Order differences in cancer mortality risk

order	a. CMR			b. ICM		
	Average %	95% CI	No. species	Average %	95% CI	No. species
Artiodactyla	2.31	0.99 - 3.64	65	3.18	1.64 - 4.73	62
Carnivora	10.64	8.98 - 12.31	41	11.17	9.27 - 13.07	41
Chiroptera	6.21	2.65 - 9.77	9	6.24	1.65 - 10.84	7
Diprotodontia	5.66	1.3 - 10.02	6	7.97	3.01 - 12.93	6
Perissodactyla	5.04	0.68 - 9.4	6	2.42	-2.54 - 7.38	6
Primates	5.63	3.98 - 7.28	42	5.55	3.44 - 7.67	33
Rodentia	6.08	3.41 - 8.75	16	5.40	1.89 - 8.91	12
contrast	c. Order differences in CMR			d. Order differences in ICM		
	estimate	t.ratio	p.value	estimate	t.ratio	p.value
<i>Artiodactyla-Carnivora</i>	-8.33	-7.72	<0.0001	-7.99	-6.45	<0.0001
Artiodactyla-Chiroptera	-3.90	-2.02	0.4034	-3.06	-1.25	0.8748
Artiodactyla-Diprotodontia	-3.35	-1.45	0.7732	-4.78	-1.82	0.5379
Artiodactyla-Perissodactyla	-2.72	-1.18	0.9009	0.77	0.29	0.9999
<i>Artiodactyla-Primates</i>	-3.32	-3.10	0.0362	-2.37	-1.79	0.5599
Artiodactyla-Rodentia	-3.76	-2.49	0.1685	-2.21	-1.14	0.9143
Carnivora-Chiroptera	4.43	2.23	0.2871	4.93	1.96	0.4451
Carnivora-Diprotodontia	4.98	2.10	0.3543	3.20	1.19	0.8967
Carnivora-Perissodactyla	5.61	2.37	0.2179	8.75	3.25	0.0230
<i>Carnivora-Primates</i>	5.01	4.22	0.0008	5.62	3.90	0.0026
Carnivora-Rodentia	4.57	2.86	0.0691	5.77	2.86	0.0705
Chiroptera-Diprotodontia	0.55	0.19	1.0000	-1.73	-0.50	0.9988
Chiroptera-Perissodactyla	1.17	0.41	0.9996	3.82	1.12	0.9222
Chiroptera-Primates	0.58	0.29	0.9999	0.69	0.27	1.0000
Chiroptera-Rodentia	0.13	0.06	1.0000	0.84	0.29	1.0000
Diprotodontia-Perissodactyla	0.63	0.20	1.0000	5.55	1.56	0.7066
Diprotodontia-Primates	0.03	0.01	1.0000	2.42	0.88	0.9744
Diprotodontia-Rodentia	-0.41	-0.16	1.0000	2.57	0.83	0.9809
Perissodactyla-Primates	-0.59	-0.25	1.0000	-3.13	-1.15	0.9122
Perissodactyla-Rodentia	-1.04	-0.40	0.9997	-2.98	-0.97	0.9600
Primates-Rodentia	-0.45	-0.28	1.0000	0.15	0.07	1.0000

measured as **a.** CMR or **b.** ICM. Only orders represented by a minimum of two species were included in these analyses. Models were constructed using R function "lm" using cancer mortality risks as a dependent variables and order as a sole predictor. Values presented are estimated marginal means and their associated 95% confidence intervals. The number of species for which data was available in each order is also shown. Post-hoc test of order differences in **c.** CMR or **d.** ICM across mammalian orders are shown.

Extended Data Table 2 | Association between cancer risk and diet animal content

a. CMR				b. ICM			
	β (SE)	t-value	p-value		β (SE)	t-value	p-value
Base model				Base model			
Intercept	-7.13 (1.83)	-3.90	0.0002	Intercept	-7.15 (2.55)	-2.81	0.0058
Log body mass	-0.04 (0.05)	-0.87	0.3850	Log body mass	-0.09 (0.06)	-1.48	0.1415
Log life expectancy	0.42 (0.23)	1.83	0.0698	Log life expectancy	0.51 (0.32)	1.59	0.1153
<i>ModelStats</i>	$\lambda = 0.53$	<i>AIC</i> = 367.42	<i>n</i> = 140	<i>ModelStats</i>	$\lambda = 0.36$	<i>AIC</i> = 401.49	<i>n</i> = 127
Animal				Animal			
Intercept	-7.15 (1.83)	-3.91	0.0001	Intercept	-6.81 (2.55)	-2.67	0.0085
Log body mass	-0.04 (0.05)	-0.70	0.4869	Log body mass	-0.07 (0.07)	-1.11	0.2673
Log life expectancy	0.42 (0.23)	1.82	0.0714	Log life expectancy	0.46 (0.33)	1.40	0.1653
Animal	0.12 (0.23)	0.50	0.6146	Animal	0.26 (0.31)	0.85	0.3981
<i>ModelStats</i>	$\lambda = 0.53$	<i>AIC</i> = 370.24	<i>n</i> = 140	<i>ModelStats</i>	$\lambda = 0.34$	<i>AIC</i> = 403.3	<i>n</i> = 127
Invertebrate				Invertebrate			
Intercept	-7.12 (1.84)	-3.88	0.0002	Intercept	-7.29 (2.57)	-2.84	0.0054
Log body mass	-0.05 (0.05)	-0.99	0.3237	Log body mass	-0.11 (0.07)	-1.55	0.1235
Log life expectancy	0.43 (0.23)	1.84	0.0682	Log life expectancy	0.54 (0.33)	1.64	0.1038
Invertebrate	-0.11 (0.21)	-0.53	0.5938	Invertebrate	-0.14 (0.29)	-0.51	0.6138
<i>ModelStats</i>	$\lambda = 0.54$	<i>AIC</i> = 370.45	<i>n</i> = 140	<i>ModelStats</i>	$\lambda = 0.37$	<i>AIC</i> = 403.9	<i>n</i> = 127
Vertebrate				Vertebrate			
Intercept	-7.1 (1.77)	-4.01	0.0001	Intercept	-6.65 (2.42)	-2.75	0.0069
Log body mass	-0.02 (0.05)	-0.49	0.6232	Log body mass	-0.07 (0.06)	-1.11	0.2691
Log life expectancy	0.41 (0.22)	1.80	0.0735	Log life expectancy	0.44 (0.31)	1.41	0.1616
Vertebrate	0.56 (0.22)	2.51	0.0134	Vertebrate	0.60 (0.28)	2.10	0.0380
<i>ModelStats</i>	$\lambda = 0.5$	<i>AIC</i> = 364.45	<i>n</i> = 140	<i>ModelStats</i>	$\lambda = 0.27$	<i>AIC</i> = 400.29	<i>n</i> = 127
Fish				Fish			
Intercept	-6.94 (1.86)	-3.72	0.0003	Intercept	-7.05 (2.62)	-2.69	0.0081
Log body mass	-0.04 (0.05)	-0.90	0.3721	Log body mass	-0.09 (0.06)	-1.48	0.1421
Log life expectancy	0.41 (0.24)	1.73	0.0864	Log life expectancy	0.50 (0.33)	1.51	0.1329
Fish	0.16 (0.34)	0.46	0.6460	Fish	0.07 (0.45)	0.16	0.8737
<i>ModelStats</i>	$\lambda = 0.52$	<i>AIC</i> = 369.54	<i>n</i> = 140	<i>ModelStats</i>	$\lambda = 0.36$	<i>AIC</i> = 403.22	<i>n</i> = 127
Reptile				Reptile			
Intercept	-7.10 (1.83)	-3.89	0.0002	Intercept	-7.13 (2.55)	-2.79	0.0060
Log body mass	-0.04 (0.05)	-0.70	0.4822	Log body mass	-0.09 (0.07)	-1.38	0.1692
Log life expectancy	0.42 (0.23)	1.81	0.0720	Log life expectancy	0.51 (0.33)	1.57	0.1185
Reptile	0.16 (0.22)	0.71	0.4759	Reptile	0.06 (0.31)	0.19	0.8528
<i>ModelStats</i>	$\lambda = 0.53$	<i>AIC</i> = 370.09	<i>n</i> = 140	<i>ModelStats</i>	$\lambda = 0.36$	<i>AIC</i> = 403.99	<i>n</i> = 127
Bird				Bird			
Intercept	-7.17 (1.86)	-3.86	0.0002	Intercept	-7.16 (2.57)	-2.78	0.0062
Log body mass	-0.04 (0.05)	-0.87	0.3873	Log body mass	-0.09 (0.07)	-1.45	0.1500
Log life expectancy	0.43 (0.23)	1.82	0.0713	Log life expectancy	0.51 (0.33)	1.57	0.1183
Bird	-0.04 (0.28)	-0.14	0.8889	Bird	-0.01 (0.38)	-0.03	0.9765
<i>ModelStats</i>	$\lambda = 0.54$	<i>AIC</i> = 370.13	<i>n</i> = 140	<i>ModelStats</i>	$\lambda = 0.37$	<i>AIC</i> = 403.57	<i>n</i> = 127
Mammal				Mammal			
Intercept	-6.92 (1.70)	-4.08	0.0001	Intercept	-6.73 (2.31)	-2.92	0.0042
Log body mass	-0.06 (0.04)	-1.41	0.1622	Log body mass	-0.10 (0.05)	-1.89	0.0615
Log life expectancy	0.44 (0.21)	2.06	0.0418	Log life expectancy	0.47 (0.29)	1.62	0.1081
Mammal	0.57 (0.25)	2.27	0.0245	Mammal	0.74 (0.30)	2.47	0.0151
<i>ModelStats</i>	$\lambda = 0.36$	<i>AIC</i> = 366.49	<i>n</i> = 140	<i>ModelStats</i>	$\lambda = 0.17$	<i>AIC</i> = 400.28	<i>n</i> = 127

Models explain variance in **a.** CMR or **b.** ICM in function of body mass, life expectancy and diet animal content. Diet animal content is characterized by variables on three taxonomic level: 1. animal content (including any vertebrate or invertebrate prey); 2. invertebrate or vertebrate prey; and 3. within vertebrates fish, reptiles, birds and mammals. Each diet item is coded as a two-level factor: rarely/never occurring in diet and representing the primary/secondary food item of the species. Each diet variable is added one by one to a base model containing the two significant predictors of cancer risk: body mass and life expectancy. Akaike's Information Criteria (AIC) are directly comparable among models with the same independent variable. All models are PGLS regressions.

Extended Data Table 3 | Results of phylogenetic models exploring variation in of cancer mortality risk

	a. CMR			b. ICM		
	β (SE)	z/t-value	p-value	β (SE)	z/t-value	p-value
Probability of detecting cancer						
Intercept	-11.39 (5.07)	-2.25	0.0247	-10.06 (5.44)	-1.85	0.0644
Log no. individuals with post-mortem pathological record	1.52 (0.39)	3.93	0.0001	1.46 (0.39)	3.72	0.0002
Log body mass	-0.05 (0.12)	-0.43	0.6650	-0.07 (0.13)	-0.53	0.5980
Log life expectancy	0.88 (0.57)	1.54	0.1237	0.73 (0.63)	1.17	0.2414
Model Stats.	$P_{s2} = 0.0006$	$s2 = 1.88$	$n = 188$	$P_{s2} = 0.0065$	$s2 = 1.71$	$n = 169$
Weighted logistic reg.						
Intercept	-7.11 (1.82)	-3.92	0.0001	-7.27 (2.54)	-2.87	0.0049
Log body mass	-0.04 (0.05)	-0.89	0.3754	-0.09 (0.06)	-1.44	0.1534
Log life expectancy	0.42 (0.23)	1.83	0.0688	0.53 (0.32)	1.63	0.1049
Model Stats.	$AIC = 368.98$	$\lambda = 0.53$	$n = 141$	$AIC = 403.93$	$\lambda = 0.37$	$n = 128$

Results are presented for both **a**, CMR (%) and **b**, ICM (%). Binomial phylogenetic GLMMs are presented first and phylogenetic GLSs exploring variation in non-zero cancer mortality risks are presented second. For each model sample size (n) and phylogenetic inertia (s2/ λ) are presented at the bottom of the results.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No software or code was used to collect the data. All raw data was obtained from Species360 (a non-profit organization centralizing zoo data), through a research request. Raw data can be obtained directly from the organization (see <https://www.species360.org/serving-conservation/research-data-request/>)

Data analysis All data analysis and graphical presentation was performed using R Statistical and Programming Environment, version 4.0.4. We estimated species-specific adult life expectancies from age-specific survival, estimated using a Kaplan-Meier procedure (using the "survfit" function in R package "survival" v. 3.2-11). Models of cancer risk were performed using zero-inflated logistic models. The first part of this model consisted of a phylogenetic binomial regression (using the function binaryPGLMM, in R package "ape" v. 5.5), the second part of the model consisted of a phylogenetic generalised least squares regression (PGLS, built using the function gls in R package "nlme" v. 3.1-153). Phylogenetic signal of cancer risk was assessed using the function 'phylosig' from R package 'phytools' v. 0.7-80. All packages and functions are described and fully referenced in the methods. Consensus phylogenetic tree was obtained using the 'sumtrees' library (v. 4.4.0) in python (v. 2.7.16).

All codes used to generate the presented results are available on <https://github.com/OrsolyaVincze/VinczeEtal2021Nature>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data required for the replication of the analysis are results are given as Source Data. Raw data are in the possession of Species360 and are available through research request from the organization (<https://www.species360.org/serving-conservation/research-data-request/>).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We worked with the widest database and largest sample size available at Species360. Species were considered for statistical analysis if minimum 20 individuals had postmortem pathological records. Models were however run by increasing this sample size to 40, 60, 80 and 100 to confirm the robustness of the results. Results of this sensitivity analysis is given in Supplementary Table 2.
Data exclusions	We excluded all species that were subject to domestication, because the process of domestication is widely regarded as a major contributing factor to inbreeding depression and higher incidence of cancer. We list in the species excluded due to this reason in Supplementary Table 1 and provide references in the manuscript summarizing the potential biases caused by domestication. Due to their high leverages (unusually low life expectancies) three species (<i>Lagurus lagurus</i> , <i>Cricetus cricetus</i> , and <i>Dasyuroides byrnei</i>) were eliminated from the main analysis presented in the manuscript. The three species are listed in the Methods section, and the reason for their exclusion is explained. All analysis were run however also including these three species to ensure that the results are robust and not affected by the above exclusion. Results of the latter sensitivity analysis are presented in Supplementary Table 4 and Extended Data Figure 7.
Replication	To ensure the replicability of the results, we estimated two different metrics of cancer mortality risk, both estimating the proportion of individuals dying of cancer. All models are presented using both metrics as dependent variables. Moreover, to ensure the replicability of the results we performed sensitivity analysis using different minimum threshold of number of individuals with available postmortem pathological records. Results of these analyses are presented in Supplementary Table 2. Where collinearity was plausible, models were re-run with single predictors to ensure that effects are consistent, as shown in Supplementary Table 3.
Randomization	Initial species pool used here, represented the widest available husbandry database of Species360 at the time of data extraction. Exclusion included species subject to domestication (as shown in Supplementary Table 1) and three species whose leverage was concerning (all analysis were repeated including species, as presented in Supplementary Table 4). Diet categories were pre-defined by an independent published work (Kissling et al. 2014 Ecol.Evol.) and were adopted without modifications.
Blinding	Diet categories were pre-defined by an independent published work (Kissling et al. 2014 Ecol.Evol.) and were adopted without modifications. No other group allocation was necessary.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging