Why Do Whales Exist?

Cancer Resistance in Cetaceans

Matthew Speight

Brasenose College

A thesis submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

Michaelmas Term 2021

Abstract

Cancer is the inevitable fate for all multicellular organisms should they live long enough. However, if all mammalian cells have approximately equivalent probabilities of cancer-associated gene mutation then, all else being equal, the number of cells susceptible to transformation should scale with body size and longevity, implying larger, longer-lived mammals such as whales should be disproportionately more prone to cancer than smaller mammals, such as mice, simply due to cell number alone. Empirical evidence to date does not support this prediction, and the discrepancy between theory and observation is known as Peto's paradox.

The mere existence of whales and other large mammals demands the evolution of a suite of cancer suppressive mechanisms in order to resolve the paradox; the establishment and understanding of which being the focus of this research.

First, cancer incidence rates across the animal kingdom are collected and assessed, quantitatively revealing the veracity of Peto's paradox and the relationship between incidence risk and scaling basal metabolic rate. Next, computational models are developed to test the hypothesis that intrinsic metabolic scaling may itself be sufficient to explain Peto's paradox. Results reveal that though incorporating scaling life-history traits into the narrative helps, they alone aren't sufficient to explain cancer suppression in large mammals. Instead, genetic control is established as playing a crucial role. Finally, comparative analysis of the available cetacean genomes and transcriptomes uncovers several gene duplication events in key cancer-associated pathways, alongside lineage-specific and size-dependent selection in gene families associated with cancer resistance. This work demonstrates though there exists no singular explanation for Peto's paradox, the disparate mechanisms that underlie the trajectory towards cancer resistance in independent lineages are all variations on a theme - the understanding of which may elucidate insight towards the development of future targets for human cancer therapies.

Acknowledgements

The last five years damn near killed me. My DPhil often the least of my worries.
The great trials and tribulations we go through however teach us the essence of joy, and the real value of others. I could not have braved the storm and made it to shore without the help and support of all those below.
First, I'd like to thank my supervisor, Professor Sebastian Shimeld, for his advice, tremendous patience, kindness and understanding, and for convincing me not to throw in the towel when the skies were grey. I owe a great deal to the Zoological Society of London and The British and Irish Association of Zoos and Aquariums, without whom this research could not have been pursued.
I'd like to thank the Dept. of Zoology, and assorted members of the Shimeld-Holland 'EvoDevo' consortium, for putting up with me; and my previous project supervisors: Prof. Michael Bonsall, Dr. James Rosindell, Dr. Samraat Pawar, and Prof. Ian Tomlinson for their assistance getting this project off the ground. I also gratefully acknowledge funding received by the Biotechnology and Biological Sciences Research Council (BBSRC) and University of Oxford, with which I kept a roof over my head and whales on my mind, and support from the Disability Advisory Service, and Dr. Jean Debarros, who helped push me over the finishing line.
Brasenose College has been more than just a fancy dining hall and place to pick up my post. It's been a cherished community, and, more, my *home*. Many thanks to Owen Lewis, my college advisor, Julia Baldwin, the college chaplain, and members of the wider Welfare Team for their support. I want to thank Professor Elspeth Garman in particular, for her wisdom and guidance – not only as Tutor for Graduate Studies, but as a mentor and source of inspiration.
Speaking of Brasenostrils - Oh, my friends! Attempting to put into mere words what you have meant to me is an exercise in futility. Luisa Kapp, Callum Brodie, Leila Al-Izzi, Ariane Laurent-Smith, Benjamin Singer, Angus Fisk, Eric Haney, Thom and Laura Diment, Ben Cariss, Manal Bougazzoul, Will Brady - it has been a privilege being a part of your lives. I would not be here today without you. And I look forward to being there with you tomorrow.
People can enter and exit our lives for only fleeting moments, but leave lasting memories. From my time here in Oxford, I'm indebted to Alexandra, Franziska, Miriam, Ciaran, Sam, Jo, Alice, Kiki, Bryony, Jake, Tom, Ellie, Sjef, Victoria, and the Silwood '15 gang, for being remarkable individuals; Chris too, as an overdue catalyst for change. Though I've only met them through their words, I nonetheless thank Dawkins for introducing me to 'real' biology, Wittgenstein for teaching me how to think, and Baudrillard for showing me what's real.
Last, but not least, I'm grateful for Bincho, my axolotl, who was always there by my side, even if only because he had no say in the matter.

And, well, thank *you*, whomever and wherever you are, for reading.

Matthew Speight, 5th November 2021

Supervisor: Prof. Sebastian Shimeld
Dept. of Zoology, University of Oxford

Table of Contents

Chapter 4: Modelling Cancer Resistance in Mammals

Figures

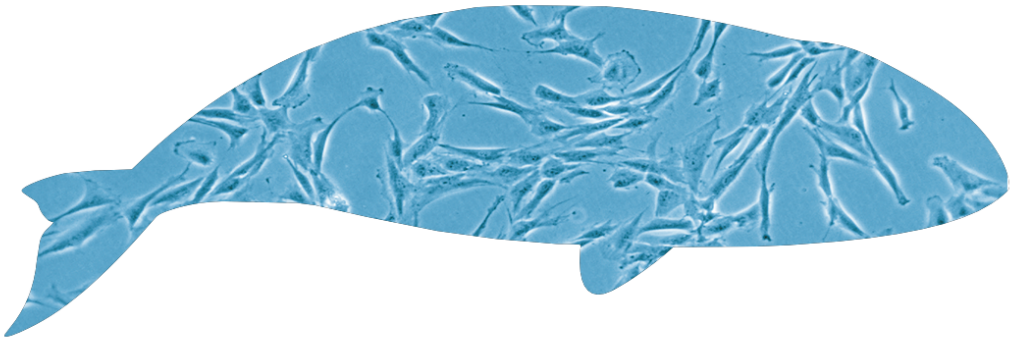
Supplementary Information

Supplementary materials, including all data, code and output (where available) is accessible online, unless otherwise stated, at the following repository: https://bit.ly/3GUaY3B

See Readme.txt file for details.

Chapter 1

Introduction



Introduction

1. Multicellularity and the Evolution of Cancer

The evolution of multicellularity represents a momentous transition in the history of life on Earth. A formless biological world of unicellular life was transformed into one where, as Darwin put it, "endless forms most beautiful and most wonderful... are being evolved". This transition however was subject to the fitness of previously independent unicellular organisms being balanced by, and indeed being superseded by, the fitness of the newly emerged multicellular individual. Individual cells needed to behave altruistically and cooperate, and, further, collectively re-organise the distribution of labour and specialise for the purposes of either survival (somatic cells) or reproduction (germ-line cells); only by then doing so being able to realise a greater synergistic genetic fitness. Such a system only works when a combination of mechanisms can effectively suppress the occurrence of somatic mutation (i.e. suppressing cellular variation) and decrease, if not cancel out, any selective advantage somatic mutation may provide (i.e. suppressing cellular-level selection) (Nedelcu, 2020; Trigos *et al*., 2017; Bussey *et al*., 2017). Conceptually, a multicellular organism therefore represents a system of cellular self-control; each cell navigating a behavioural balance between the intrinsic 'unicellular desire' to pursue accelerating self-replication, and the 'multicellular desire' to achieve greater gene-level fitness by figuratively applying the brakes.

Cancer is the resulting outcome from the breakdown of this self-control. If a cell can avoid or negate the mechanisms otherwise keeping itself from successfully pursuing a 'selfish' or atavistic 'unicellular' agenda, the cell can abandon multicellular cooperation and realise the cellular-level fitness potentials afforded by natural selection. Cells, and their descendants, that 'cheat' the system in this way, via inheritable genetic and epigenetic alteration, are free to self-proliferate, compete, and progressively adapt to the pressures of their tissue microenvironment; a process which, without intervention, can result in a growing heterogenous population of cells known as a tumour (Vogelstein & Kinzler, 2004). Tumours can be detrimental to organism-level fitness to varying degrees; indeed, malignant tumours almost always, given enough time, ultimately prove fatal. Malignant tumours, or cancers, which are so incredibly costly to their host organism, are differentiated from other, less severe tumour types by several phenotypic traits, or hallmarks (Hanahan & Weinberg, 2000 & 2011). Here, seven characteristics are used to define cancer, as per Fouad & Aanei (2017):

Cancer cells are those which form a heterogenous competing population, under selective pressure for:

i) Acquired growth and proliferative advantages. Cancer cells can grow and replicate in the absence of growth- and replication-promoting signals, or be insensitive to signals intended to inhibit growth and replication, or both.

ii) An altered stress response; cells better adapted to bypass or ameliorate the pressures of apoptosis, senescence and other crisis responses.

iii) Sustained promotion of environmental vascularisation, or angiogenesis, thereby ensuring a supply of oxygen and other nutrients.

iv) Successful invasion of surrounding tissue (cancer invasion), and the seeding of distant sites to form secondary and tertiary tumours (metastasis).

v) Metabolic rewiring; the promotion of nutrient uptake, and dysregulation or other alteration to key metabolic pathways.

vi) Immune modulation, and evasion of the immune response.

vii) Existing and actively engaging with a dynamic heterogenous microenvironment that promotes continued tumour growth.

The biological hallmarks that define cancer arise within a cell lineage gradually, a process known as carcinogenesis, via the accumulation of contributory genetic and epigenetic changes and instability (Colotta *et al.*, 2009; Knudson, 2001). This progression is often defined by multi-stage models of carcinogenesis (Nunney, 1999). Importantly, these alterations happen in both the nuclear and mitochondrial genome, and coincide with, and are in part caused by, the transition of cellular energy production from oxidative phosphorylation to substrate-level phosphorylation (Pavlova & Thompson, 2016; Koppenol, Bounds & Dang, 2011).

2. Causes of Cancer

Cancer-promoting genetic and epigenetic changes are either inherited or acquired endogenously in a multitude of ways. Most genetic changes are via mutation, though non-mutational changes, such as chromosomal aberrations (either numerical, such as aneuploidy - the presence of an abnormal number of chromosomes - or structural, such as segment deletions or translocations), may also contribute. Genetic mutation, or acute alteration to the nucleotide sequence of DNA via DNA damage is considered a primary factor that provides fitness variation among somatic cells. Damage can be caused by exposure to either exogenous or environmental agents (chemical mutagens, radiation, viral action etc.), or endogenous agents (reactive oxygen species produced by mitochondria), or via a deficiency in DNA repair mechanisms. Reduced efficacy of DNA repair itself can be due to prior direct genetic mutations in associated genes (e.g. MGMT, a protein involved in mis-match and other error repair during DNA replication) or via epigenetic alteration, wherein expression of genes associated with repair is reduced. As with DNA damage, gene regulatory epigenetic markers - including DNA methylation profiles, non-coding RNA sequences, histone covalent modifications etc. - are also subject to cancer-promoting alteration via exogenous and endogenous agents, and likewise accumulate at an increasing rate with prior alteration. Both genetic and epigenetic change can facilitate a cancer-causing cellular phenotype by either promoting or inhibiting expression in particular sets of genes.

Proto-oncogenes, for example, are genes typically associated with the promotion of cell growth, and the inhibition of differentiation and cell death. If overexpressed or otherwise mutated or dysregulated, proto-oncogenes are said to become oncogenes. Oncogenes act as dominant mutants; only a single proto-oncogene activation required for a oncogenic cancer-promoting phenotype to manifest.

Given that the detrimental consequences of non-cooperating cells arising, and evolving into cancer, on organismal fitness are substantial, organism-level selective pressure acts to reduce the risk of carcinogenesis. This has led to the evolution, or co-adoption, of several mechanisms in multicellular organisms that act together to prevent genetic and epigenetic mutation in somatic cells, limit the replicative ability of cells, and, further, effectively reduce the selective advantage of a lineage, or otherwise expunge it entirely, should a cell initiate a non-cooperative behavioural trajectory towards cancer.

3. Mechanisms to Suppress Carcinogenesis

Suppressing Somatic Variation: Mechanisms that prevent somatic mutations from occurring and propagating are known as tumour suppressive mechanisms, and, by categorisation, provide either 'caretaker', 'gatekeeper', or 'landscaper' functions. 'Caretaker' genes encode products that stabilise the genome and prevent mutations arising, for example, by sensing and repairing DNA damage. 'Gatekeeper' genes encode a system of checks and balances that monitor cell growth, division and death. If a mutation arises and isn't corrected, for example, 'gatekeepers' can induce early senescence or programmed cell death (apoptosis) to prevent unwarranted replication and propagation of the abnormal genotype. 'Landscaper' genes, in contrast, are not associated with intracellular control, but do encode products that determine the composition of the cell membrane, or mediate intercellular and cell-to-extracellular communication. 'Landscapers' therefore create environments that control cell growth by, for example, moderating growth factor signals or access to nutrients.

The genes that collectively encode these three tumour-suppressive functions are known as tumour-suppressor genes (TSGs). Unlike proto-oncogenes, TSGs are typically recessive as mutants; both alleles of a given TSG need to be affected before a phenotypic effect is manifested (known as the 'two-hit' hypothesis). Further, the impact of mutation or dysregulation in a TSG will be different, depending largely upon which class it belongs to. For example, loss-of-function of a 'caretaker' TSG will not result in the promotion of a cancer-phenotype if 'gatekeeper' TSGs remain unaffected; any increase in genetic instability due to the inactive 'caretaker' will still trigger a 'gatekeeper'-associated apoptotic response and the caretaker mutant will not replicate. In contrast, loss-of-function of a 'gatekeeper' TSG, in a cell with normally functioning 'caretaker' TSGs, will promote a lineage of more proliferative cells, given the function of 'caretakers' is not perfect.

Reducing Selective Advantage: In animals, additional traits and mechanisms reduce the selective advantage of somatic mutations, by capping or otherwise limiting their replicative potential. Replicative senescence, for instance, is a fundamental feature of somatic cells that sets a limit on the number of divisions a lineage can undertake, governed by telomere length. During DNA replication, telomeres - repetitive non-coding regions of DNA at the ends of chromosomes - shorten, eventually, after some number of replication cycles (known as the Hayflick limit), becoming too short and triggering an apoptotic response.

Tissue architecture and replicative dynamics also help maintain tissue integrity. Most somatic cells are compartmentalised, derived from the asymmetric division of a small pool of multipotent stem cells Upon asymmetric division, one daughter receives a disproportionately high proportion of cellular waste and damage, and becomes rapidly differentiated and senescent following only a small number of undifferentiated intermediaries (Neumuller & Knoblich, 2009; Campisi, 1997); the other daughter remains a multipotent stem cell, effectively rejuvenated and replicatively quiescent - accumulation of mutation in cells with relatively long lifespans therefore taking a long time. Compartmentalisation likewise provides physical extracellular barriers further impeding and localising proliferative cell growth. Finally, immune policing can actively detect and remove abnormally proliferative cells.

It is important to note that, though many mechanisms contribute to tumour suppression, this does not necessarily mean they are evolved adaptive responses to the risk of cancer; many, including 'caretaker' genes, apoptosis, and immune response play (and are primarily selected for) indispensable roles unrelated to cancer (e.g. apoptosis in development). Further, the effectiveness of these mechanisms decreases with age, particularly once the organism post-reproductive, due, in part, to diminishing selection strength. For this reason, and that genetic, epigenetic, and metabolic alterations promoting a tumorigenic phenotype need to accumulate over time (i.e. multi-stage models of carcinogenesis), cancer is a phenomenon strongly associated with increasing age.

Beyond temporal variation in selective pressure, cancer both shapes, and is shaped by, varying selective pressures acting at multiple levels within and between different multicellular lineages.

4. Peto's Paradox

Cancer, or cancer-like, phenotypes are known to arise in most major multicellular lineages, their frequency and fitness effect varying considerably (Albuquerque *et al.*, 2018; Ewald & Ewald, 2015). For example, while tumours develop frequently in plants, given the immobile architectural nature of plant cells and tissue, and the non-centralised modularity of plant organs, tumours can neither metastasise nor otherwise cause fatal harm, thus the effect on overall fitness is relatively low (Doonan & Sablowski, 2010). In contrast, cancer is common in animals, its impact often severe, though there remains considerable diversity across taxa. Many animal groups display little to no observed cancer (e.g. nematodes, rotifers) (Aktipis *et al.*, 2015); in other groups, such as mammals, prevalence

of cancer can vary substantially, even within single orders - in rodents, for example, lifetime cancer incidence is ~50% in mice (*Mus musculus*) (Andervont & Dunn, 1962), whilst there have only been three reports of neoplasia present in larger capybara (*Hydrochoerus hydrochaeris*) (Herrera-Álvarez *et al*., 2018).

Richard Peto (1975), Richard Doll (1971), and John Cairns (1975) were first to note the diversity in mammalian cancer frequency in the 1970s, and the discrepancy it posed against the understood multistage models of cancer progression: simply put, if all mammalian cells have approximately equivalent probabilities of cancer-associated genetic, epigenetic, and metabolic alteration then, all else being equal, the number of cells susceptible to transformation should scale with body size and lifespan. Large, long-lived mammals, such as whales and elephants, should be disproportionately more prone to cancer than relatively smaller mammals, such as mice, simply due to time and cell number alone. Whales, as currently exist on Earth, simply should not exist, given predicted rates of cancer would mean too few survive to reproductive age. Despite evidence that the relationship between cancer incidence and body size holds within a given species (Fleming, Creevey & Promislow, 2011; Green *et al*., 2011), no empirical evidence supports this prediction between species (Peto, 2015). This discrepancy between prediction and observation came to be known as Peto's paradox.

Assuming no additional mechanisms of tumour suppression in large mammals, previous estimations, such as the multi-stage algebraic model used by Caulin & Maley (2011) suggest that 100% of whales should succumb to colorectal cancer by age 90 (P = 1). This is without even considering other causes of mortality, including other types of cancer, and stands in clear contradiction to observed lifespans that, in some instances, run into centuries - wild bowhead whales (*Balaena mysticetus*) are recorded living up to 211 years (George *et al*., 1999). Furthermore, pathological observations to date suggest not only do larger, long-lived mammals not have a disproportionately higher risk of cancer, but may indeed have lower observed cancer incidence than smaller mammals. A review of a worldwide database of 15,000+ elephants suggested only 18 of 616 (~3%) deceased elephants had cancer (Dang, 2015; Caulin *et al*., 2015); this in contrast to the observed ~35% incidence of cancer, attributable to all causes, in humans (Smittenaar *et al.*, 2016), and upwards of ~50% of cancer-associated deaths in a study of a natural murine population (Pompei *et a*l., 2001; Andervont & Dunn, 1962) (see Fig.1).

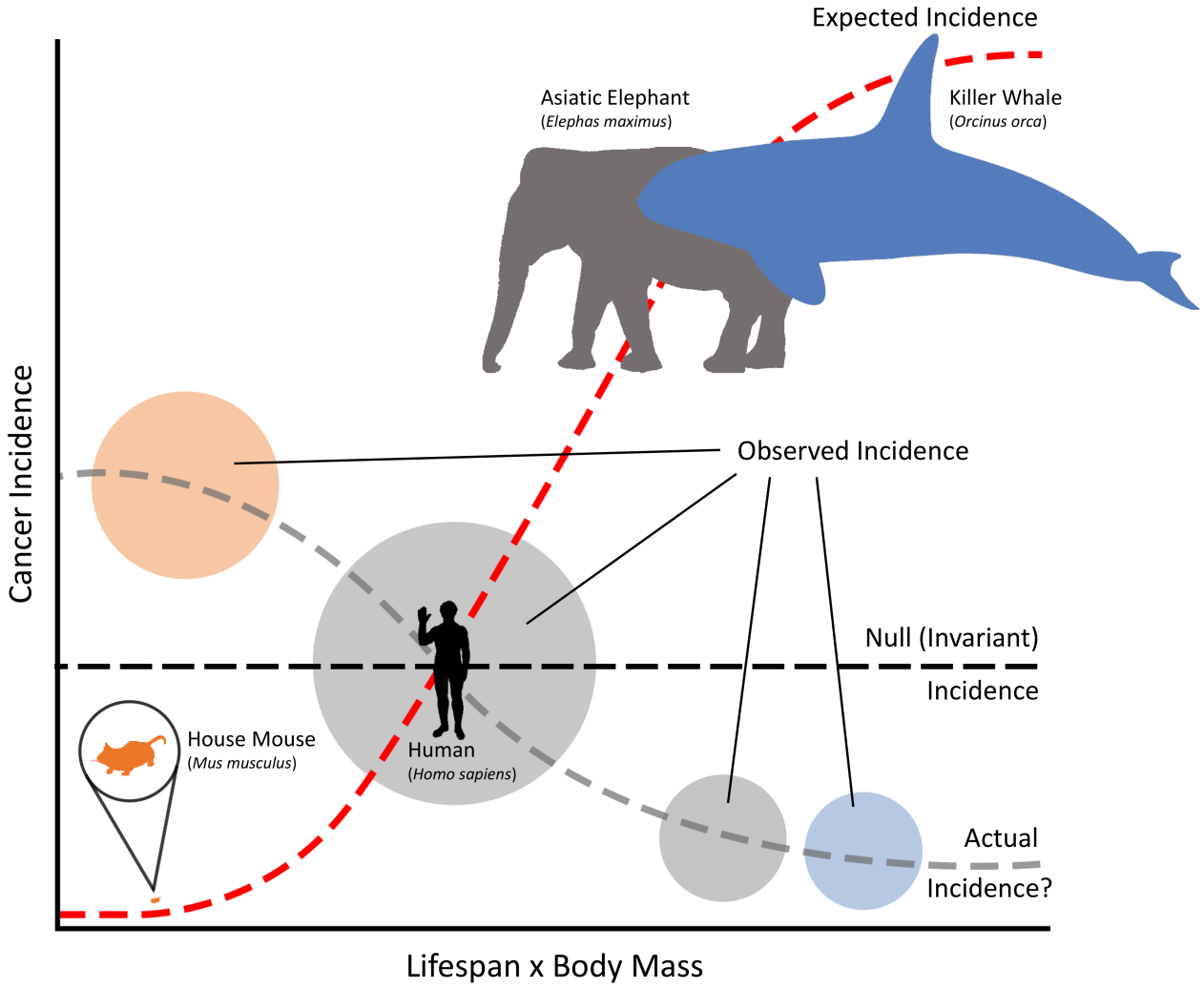


Fig. 1: Illustrating Peto's paradox: simple model estimates (red dashed line) suggest cancer incidence approaches ~100% in large mammals as mass and longevity increases, however observation to date suggests no such relationship between cancer risk and body size and lifespan - real cancer incidence being either invariant (black dashed line) or, based on scattered data, perhaps even a negative relationship (grey dashed line). This project aims to elucidate a more accurate relationship between cancer and size and longevity, and uncover the mechanisms which explain the difference between expected and observed incidence rates. Illustrations taken from Phylopic (Keesey, 2020). Figure based on original by Tollis, Boddy & Maley (2017).

The mere existence of whales supports the general and obvious evolutionary explanation to this paradox; that, given the higher selective pressure to suppress cancer risk, large, long-lived mammals have evolved greater cancer resistance mechanisms than smaller, shorter-lived mammals. What remains puzzling, and more remarkable, however are the biochemical mechanisms that produce this resistance, and, more, the relative degree to which they must be effective. Firstly, from a body size perspective, the x3,000 difference in body mass between a 20g mouse and an 80kg human suggests

a ~20g mouse-sized piece of human tissue is x3,000 more cancer-proof than a mouse-sized piece of mouse tissue (Peto, 2015). Secondly, and more impressively, from an age perspective, it has long been recognised that a power-law relationship exists between cancer incidence and age, to approximately the sixth power (Peto, 1976; Armitage & Doll, 1954):

$$P = C \times t^6 \tag{1}$$

Eq. 1: A constant of proportionality (C) determines susceptibility to cancer incidence (P) if multiplied by age (t), in years, to the sixth power (the exponent varies between 4-8 depending on cancer type; following a distribution curve peaking ~6) (Peto, 2015). This power-law formula suggests that doubling the time available for cancer to arise not just doubles the risk, but increases risk by six twofold increases, or a 64-fold increase in risk. For a human living approximately 80 years to have equivalent lifetime cancer risk compared to a mouse living 2.5 years, the constant C must be ~10^{10} times smaller.

Consequently, for a human being to have comparable lifetime cancer incidence to a mouse, combining the billion-fold factor reflecting difference in lifespan with the 3000-fold factor reflecting differences in body size, gram-for-gram a human being must have ~3 trillion times more effective cancer suppression than a mouse. As Peto himself puts it:

"*Presumably some concomitant of our evolved ability to grow big and to live for threescore years and ten is involved.*" (1977)

"*By what biological mechanisms does evolution produce trillion-fold differences between species in this constant of proportionality, which are essential for large body size and long lifespan to exist?*" (2015)

Given that humans have seemingly evolved more effective cancer suppression than mice, it follows large mammals such as whales and elephants have evolved more effective cancer suppression than humans.

5. The Promise of Comparative Oncology

Despite several decades of sustained research, cancer remains a major contributor to human mortality. As of 2016, in the UK the projected lifetime risk of a cancer diagnosis is ~50%, and though mortality rates are falling, by ~9% for all combined cancer types since 2009, it nonetheless remains that ~28% of all deaths per year are attributable to cancer (Smittenaar *et al.*, 2016; Cancer Research UK, 2020). Most cancer research to date has focused on the mechanisms of tumour growth, with a significant proportion of this research attempting to identify treatment methods limiting tumour growth and proliferation following diagnosis. With a contemporary understanding, now viewing cancer as a phenomenon through an ecological and evolutionary lens, we now understand the limitations of this previous research approach - many treatment options simply aggressively select for increasingly drug-resistant cancers, with high rates of recurrence and remission. Attention is increasingly moving away towards cancer prediction and prevention, an approach to which the auspices of comparative oncology are well suited to contribute.

Cancer is an unavoidable disease, or essential 'cost', of multicellularity. However, because of this, life on Earth has been evolving mechanisms to suppress cancer ever since the origins of multicellularity approximately ~1.5 billion years ago. Recognising that Peto's paradox is resolved by adaptive evolution, we can apply evolutionary and genetic theory, leveraging the similarities and differences between taxa and their responses to cancer, in order to not only better understand cancer as a pervasive, fundamental constraint on the evolution of multicellular life, but also in the hope of applying and clinically translating our understanding to develop effective preventative and treatment strategies for human and veterinary medicine.

6. Overview of Thesis

At the onset of this project, limited research efforts had focused on resolving and understanding Peto's paradox. Since, an explosion of interest and research has started to successfully address many of its questions - from uncovering several cancer-suppressing copies of the tumour-suppressor gene *TP53* in elephants (Abegglen *et al.*, 2015), to analysis of the capybara genome uncovering evidence of gigantism-related trade-off and cancer resistance (Herrera-Álvarez *et al.*, 2020). This investigation attempts to contribute to this ongoing body of research and understand the underlying evolutionary

basis and mechanisms for Peto's paradox, via three approaches; life-history trade-off analysis; computational modelling; and comparative genomics.

In the following chapter (Chapter 2), I review the hypotheses and results to date associated with resolving Peto's paradox, and present the case for investigating mammalian taxa, and particularly cetaceans as a case-study taxon, for studying Peto's paradox.

Despite discussion over decades, the relationship between body size, longevity and other life history traits, and cancer has yet to be widely and quantitatively established. In Chapter 3, I collate data from zoological and veterinary institutions across the UK, and beyond, to present a dataset of cancer incidence across the animal kingdom, the analysis of which illuminates several relationships, notably between cancer incidence and mass-specific metabolic rate.

The metabolic rate hypothesis suggests cellular metabolic rate and, subsequently, oxidative stress, decreases with increasing body size and longevity, providing a cancer protective mechanism. In Chapter 4, in exploring a metabolic explanation for Peto's paradox, I develop and present three models incorporating scaling metabolic rate, and associated biology at different levels, and *in silico* demonstrate how metabolic scaling can account for only some of the variance in cancer incidence across mammals.

Low mass-specific basal metabolic rates alone remain insufficient to explain low cancer incidence in large cetaceans. In Chapter 5, I undertake comparative genomic analyses of the 27 available cetacean genomes and transcriptomes, present evidence for duplication events, size-associated accelerated selection, and convergent evolution in several tumour-suppressor genes, and provide case-study hypotheses on the functional significance of several key genes.

Peto's paradox is a multivariate phenomenon, and though different taxa have invested in different solutions, there remain common trajectories. Chapter 6 synthesises and critically evaluates what I have uncovered with respect to the literature at large, and offers hypotheses and perspectives on future directions.

References

- Abegglen, L.M., Caulin, A.F., Chan, A., Lee, K., Robinson, R., Campbell, M.S., Kiso, W.K., Schmitt, D.L., Waddell, P.J., Bhaskara, S., Jensen, S.T., Maley, C.C. & Schiffman, J.D. (2015) Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. *JAMA*. 314 (**17**), 1,850-1,860

- Aktipis, C.A., Boddy, A.M., Jansen, G., Hibner, U., Hochberg, M.E., Maley, C.C. & Wilkinson, G.S. (2015) Cancer across the tree of life: cooperation and cheating in multicellularity. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 370 (**1673**), e20140219

- Albuquerque, T.A.F., Drummond do Val, L., Doherty, A. & Magalhães, J.P. (2018) From humans to hydra: patterns of cancer across the tree of life. *Biological Reviews*. 93 (**3**), 1,715-1,734

- Andervont, H.B. & Dunn, T.B. (1962) Occurrence of Tumours in Wild House Mice. *Journal of the National Cancer Institute*. 28 (**5**), 1,153-1,163

- Armitage, P. & Doll, R. (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer*. 8 (**1**), 1-12

- Bussey, K.J., Cisneros, L.H., Lineweaverm C.H. & Davies, P.C. (2017) Ancestral gene regulatory networks drive cancer. *PNAS*. 114 (**24**), 6,160-6,162

- Cairns, J. (1975) Mutation selection and the natural history of cancer. *Nature*. 255, 197-200

- Campisi, J. (1997) Aging and cancer: the double-edged sword of replicative senescence. *Journal of American Geriatrics Society*. 45 (**4**), 482-488

- Cancer Research UK (2020) Cancer Statistics. [online] Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk [Accessed: 10/06/20]

- Caulin, A.F., Graham, T.A., Wang, L.S. & Maley, C.C. (2015) Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 370 (1,673), 20140222

- Caulin, A.F. & Maley, C.C. (2011) Peto's Paradox: evolution's prescription for cancer prevention. *Trends in Ecology & Evolution*. 26 (**4**), 175-182

- Colotta, F., Allavena, P., Sica, A., Garlanda, C. & Mantovani, A. (2009) Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis*. 30 (**7**), 1,073-1,081

- Dang, C.V. (2015) A metabolic perspective of Peto's paradox and cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 370 (**1673**), e20140223

- Doll, R. (1971) The Age Distribution of Cancer: Implications for Models of Carcinogenesis. *Journal of the Royal Statistical Society: Series A (General)*. 134 (**2**), 133-166

- Doonan, J.H. & Sablowski, R. (2010) Walls around tumours - why plants do not develop cancer. Nature Reviews Cancer. 10, 794-802

- Ewald, P.W. & Ewald, H.A.S. (2015) Infection and cancer in multicellular organisms. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 370 (1673), e20140224

– Fleming, J.M., Creevy, K.E. & Promislow, D.E.L. (2011) Mortality in north american dogs from 1984 to 2004: an investigation into age-, size-, and breed-related causes of death. *Journal of Veterinary Internal Medicine*. 25 (**2**), 187-198

– Fouad, Y.A. & Aanei, C. (2017) Revisiting the hallmarks of cancer. *American Journal of Cancer Research*. 7 (**5**), 1,016-1,036

– George, J., Bada, J.L., Zeh, J. & Scott, L. (1999) Age and growth estimates of bowhead whales (*Balaena mysticetus*) via aspartic acid racemization. *Canadian Journal of Zoology*. 77 (**4**), 571-580

– Green, J., Cairns, B.J., Casabonne, D., Wright, L.F., Reeves, G., Beral, V. & Million Women Study collaborators (2011) Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *The Lancet: Oncology*. 12 (**8**), 785-794

– Hanahan, D. & Weinberg, R.A. (2000) The Hallmarks of Cancer. *Cell*. 100 (**1**), 57-70

– Hanahan, D. & Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*. 144 (**5**), 646-674

– Herrera-Álvarez, S., Karlsson, E., Ryder, O.A., Lindblad-Toh, K. & Crawford, A.J. (2018) How to make a rodent giant: Genomic basis and tradeoffs of gigantism in the capybara, the world's largest rodent. *BioRxiv*.

– Keesey, M.T. (2020) Phylopic. [online] Available at: https://www.phylopic.org [Accessed: 01/09/2020]

– Knudson, A.G. (2001) Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*. 1, 157-162

– Koppenol, W.H., Bounds, P.L. & Dang, C.V. (2011) Otto Warburg's contributions to current concepts of cancer metabolism. *Nature Reviews Cancer*. 11, 325-337

– Nedelcu, A.M. (2020) The evolution of multicellularity and cancer: views and paradigms. *Biochemical Society Transactions*. 48 (**4**), 1,505-1,518

– Neumuller, R.A. & Knoblich, J.A. (2009) Dividing cellular asymmetry: asymmetric cell division and its implications for stem cells and cancer. *Genes & Development*. 23 (**23**), 2,675-2,699

– Nunney, L. (1999) Lineage selection and the evolution of multistage carcinogenesis. *Proceedings of the Royal Society B: Biological Sciences*. 266 (**1418**), 493-498

– Pavlova, N.N. & Thompson, C.B. (2016) The emerging hallmarks of cancer metabolism. *Cell Metabolism*. 23 (**1**), 27-47

– Peto, R., Roe, F.J., Lee, P.N. & Clack, J. (1975) Cancer and ageing in mice and men. *British Journal of Cancer*. 32, 411-426

– Peto, R. (1977) Epidemiology, multi-stage models and short-term mutagenicity tests. In: Hiatt, H.H., Watson, J.D. & Winsten, J.A. (eds.) *Origins of Human Cancer*. Cold Spring Harbor Laboratory, New York, NY. pp.1,403-1,428

– Peto, R. (2015) Quantitative implications of the approximate irrelevance of mammalian body size and lifespan to lifelong cancer risk. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 370 (**1673**), e20150198

– Pompei, F., Polkanov, M. & Wilson, R. (2001) Age distribution of cancer in mice: the incidence turnover at old age. *Toxicology & Industrial Health*. (**17**), 7-16

– Smittenaar, C.R., Peterson, K.A., Stewart, K. & Moitt, N. (2016) Cancer incidence and mortality projections in the UK until 2035. *British Journal of Cancer*. 115 (**9**), 1,147-1,155

– Tollis, M., Boddy, A.M. & Maley, C.C. (2017) Peto's Paradox: how has evolution solved the problem of cancer prevention?. BMC Biology. 15 (1), 60

– Trigos, A.S., Pearson, R.B., Papenfuss, A.T. & Goode, D.L. (2017) Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. *PNAS*. 114 (**24**), 6,406-6,411

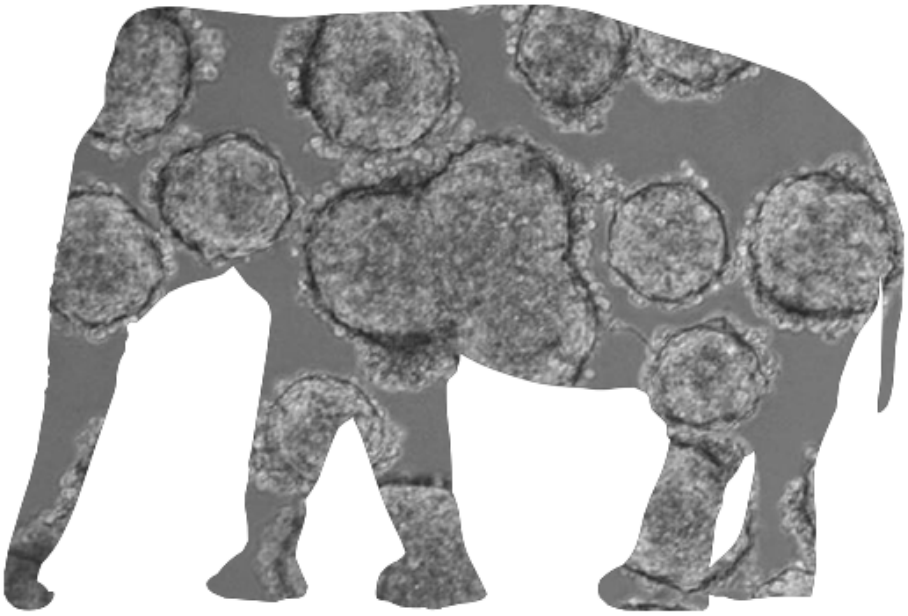
– Vogelstein, B. & Kinzler, K.W. (2004) Cancer genes and the pathways they control. *Nature Medicine*. 10, 789-799

Bibliography

– Trigos, A.S., Pearson, R.B., Papenfuss, A.T. & Goode, D.L. (2018) How the evolution of multicellularity set the stage for cancer. *British Journal of Cancer*. 118, 145-152

Chapter 2

Resolving Peto's Paradox



Resolving Peto's Paradox

Peto's paradox is likely a multifaceted phenomenon, and many alternative, co-compatible and complementary hypotheses have been proposed to resolve it. Further, larger, longer-lived organisms have evolved multiple times independently in several lineages, and one should expect, unless tumour suppression can be explained by only a few shared characteristics, a multitude of lineage-specific solutions. In many ways, this is indeed a major promise from understanding Peto's paradox; the more evolved solutions there may be, the more avenues of therapeutic investigation that could result in meaningful advances in medical and veterinary cancer prevention and treatment. Here I consider hypothesised solutions and, where available, evidence that supports them.

1. Growing Large and Living Long

1.1. Body Size Evolution

Despite the risk of cancer acting as a constraining factor on evolutionary trajectories towards larger body size, several proposed adaptive advantages may be gained from increasing body size, thereby providing pressure to evolve resistance mechanisms in order to accommodate. Suggestions proposed have included increased tolerance to environmental extremes, reduced mortality (e.g. via predation), enhanced predation success, greater metabolic efficiency, among other advantages (Brown & Silby, 2006; Hone & Benton, 2005; Peters, 1986). Large body sizes have after all evolved multiple times independently in several lineages through evolutionary time, including in 10 out of 11 mammal orders as detailed by Baker *et al.* (2015); and if indeed driven by active selection, as opposed to neutral or random processes, could explain the existence of the Cope-Depéret rule.

The Cope-Depéret rule postulates that population lineages actively trend towards an increase in body size over evolutionary time (Cope, 1896). Studies based on extant and palaeontological data over several decades have been found to support (Hone *et al.*, 2005; Hone & Benton, 2007), reject (Monroe & Bokma, 2010; Moen, 2006), or otherwise cast conceptual doubt on the validity of adaptive trends towards gigantism. More recently however, germane phylogenetic statistical and macroevolutionary approaches corroborate the Cope-Depéret rule, providing evidence for a consistent signal for directional change in size (Baker *et al.*, 2015; Heim *et al.*, 2015). Further, and interestingly, large-bodied species express inherently faster rates of evolution above certain thresholds, suggestively owing to the

relaxation of, as per the authors, some unspecified size-associated constraint.

In marine mammals, Clauset (2013) accurately predicted the evolution and body size distribution of extant cetaceans using a simple macroevolutionary trade-off model; between short-term evolutionary advantages, attained via increasing size, against long-term extinction risks (given low reproductive rates etc. in larger mammals), and constrained by environmental thermoregulatory limits, thereby further validating the notion that consistent adaptive (non-passive) micro-evolutionary pressure results in larger body sizes, and can accurately predict observed body size.

Given therefore that many different lineages - from sauropod dinosaurs weighing ~65 tonnes, to 22-tonne indricothere rhinoceroses, to ~150 tonne modern whales - have independently faced the trade-off between adaptively-selected large body size and cancer risk during the course of evolution, there have likely been many different pathways by which cancer resistance has evolved.

1.2. Increased Longevity and Post-Reproductive Lifespan

Similarly, increase in longevity presents itself widely and independently across the animal kingdom. Life-history trade-off theories, with emphasis on reproductive success, account for much of this variance and the rate of senescence and ageing (see Section 4.1). The contemporary paradigm posits biological processes are optimised for early-life function but become harmful when they continue to run long-term, and the process of natural selection on maintaining these processes is maximised during pre-reproductive development and progressively declines following sexual maturation, reaching zero at age of last reproduction (Maklakov & Chapman, 2019). The default trend therefore being towards shorter lifespan, constraints on early life reproductive need and success are typical factors that drive pressure promoting slower reproductive senescence and longer lifespan in animals - to, for example, accommodate slow growth and development rates in fewer numbers of larger offspring, and/or lengthier periods of rearing of more altricial offspring.

In a rare number of cases, this associated pressure has driven longevity beyond reproductive lifespan (alongside accelerated reproductive senescence) (Ellis *et al.*, 2018a) - typically in social species including humans (i.e. menopause), elephants (Chapman *et al.*, 2019) and cetaceans (Ellis *et al.*, 2018b), Photopoulou *et al.*, 2017). Numerous non-adaptive (Nichols, Zecherle & Arbuckle, 2016; Cohen, 2007) and adaptive hypotheses (Takahashi, Singh & Stone, 2016; Pavard, Metcalf & Heyer, 2008) have been presented to explain post-reproductive life and associated phenomena such as menopause.

Of note, the close-kin social cultures of many post-reproductive (and likewise large and long-lived) species such as elephants and cetaceans are matriarchal - the eldest female members are dominant, their activities supporting grandoffspring promoting inclusive kin fitness, and therefore selection for long post-reproductive life (known as the 'grandmother hypothesis'; Williams, 1957). Nattrass *et al*. (2019) (and previously Foster *et al*., 2012) quantifiably demonstrated the influence of the grandmother effect on fitness in killer whales (*Orcinus orca*), which have the longest post-reproductive lifespans of all non-human species. Killer whales have likewise recently and evolutionarily rapidly considerably increased in body size (McGowen *et al*., 2020). Perhaps selective pressure promoting post-reproductive longevity can partially provide and explain trajectories towards greater cancer resistance in some species, and therefore (or in co-incident) large body size. Further, given the female-specific nature of post-reproductive selection (if adaptive), perhaps cancer rates in post-reproductive species are disproportionately lower in females compared to males.

As with body size, the trade-off between pressure driving longevity and cancer risk has been approached multiple times independently, to varying degrees and different reasons, and we can expect no single pathway to cancer suppression.

As discussed in the previous chapter, pathways to cancer resistance can largely be categorised into mechanisms that; i) suppress somatic mutation and variation (2.1); and ii) suppress selective advantage of cancer-promoting mutants (2.2).

2. Suppressing Somatic Mutation and Variation

2.1. Lower Somatic Mutation Rates: Error Rates during DNA Replication

An assumption that facilitates the conceptualisation of Peto's paradox is that somatic mutation rates per cell generation are relatively similar across taxa. Is this really the case?

During the process of DNA replication, double-helix DNA is separated and each single-strand is used as a template upon which DNA polymerase enzymes bind, matching nucleotides to generate a complementary strand (Alberts *et al*., 2008). DNA polymerase activity is not intrinsically perfect, and errors in the form of incorrect nucleotide bases may be introduced. These errors may be corrected via proof-reading mechanisms inherent to polymerase activity, or otherwise via a sophisticated network of DNA damage-response (DDR) systems including mis-match repair (MMR), base-excision repair (BER), nucleotide-excision repair (NER) amongst others, mediated by other enzymes. These DDR mechanisms likewise deal with DNA damage occurring during all stages of the cell cycle. The rate at which errors are generated and persist post-repair is very low; proteins in the polymerase family produce errors per base *in vitro* between 10^{-1} and 10^{-6}, and as low as 10^{-9} for MMR proteins - human DNA polymerase δ and ε, for example, has a single base error rate of 4.4×10^{-5} (Korona, LeCompte & Pursell, 2011; for review, see: Bebenek & Ziuzia-Graczyk, 2018). The degree and significance to which these error rates differ between organisms, beyond at the highest taxonomic levels (e.g. eukaryotes vs. prokaryotes), is unclear (Chintalapati & Moorjani, 2020; Matsuda *et al*., 2001); it may be the case larger, longer-lived animals may have lower error, and therefore intrinsic mutation, rates via selection acting upon, for example, polymerase activity and proof-reading ability. Further, DDR repair mechanisms may likewise be more effective at dealing with DNA damage.

In mice and men, mutation rates estimated via specific genomic loci following radiation-exposure suggest comparative rates of persistent mutation (i.e. following repair mechanisms), at least within a cell line. Analysis following genome-sequencing of HRPT-null mutant cells in human and mouse kidney adenocarcinomas likewise suggests incident mutation rates per cell division are broadly comparable (Leroi, Koufopanou & Burt, 2003). More recently and in contrast, by sequentially sequencing somatic and germline cells, Milholland *et al.* (2017) were the first to conclusively determine the true spontaneous rate of mutation rates to be significantly higher in mouse cell lines (4.4×10^{-7}) compared to human (2.8×10^{-7}). However, further evidence is lacking and, beyond murine models, little to no

investigation has been done into intrinsic error and mutation rates across taxa at lower hierarchical levels (i.e. order; mammals). Further investigation is needed before considering this hypothesis. Cross-taxa comparisons of DNA damage repair systems to exogenously-induced DNA damage however have shown greater efficacy in longer-lived species (MacRae *et al.*, 2015; Hart *et al.*, 1979).

2.2. Tumour-Suppressor Gene Redundancy

Given tumour suppressor genes (TSGs) are recessive mutants, in order for a more cancer-like phenotype to express itself, both alleles of a given TSG must be inactivated - summarised by Knudson's 'two-hit' hypothesis (Knudson, 1971). It therefore follows that, given a rate-limiting step in a multi-stage model of carcinogenesis is the number of TSG 'hits' required, the more copies of a TSG allele that require inactivation, the longer it will take for the requisite number of inactivation events to accumulate. It might be expected therefore to find a greater incidence of TSG copies in the genomes of larger, long-lived mammals.

Indeed, Abegglen *et al.* (2015) and Sulak *et al.* (2016) determined the tumour-suppressor *TP53* had been duplicated as many as ~19 times in elephants, largely via segmental duplications, with several of the retrogenes demonstrably being transcribed and translated in elephant tissue. Further investigation into genomes across the larger elephantine family, including mammoths and mastodons, revealed via molecular phylogenetic dating methods the timing of the *TP53RTG* expansion took place coincident with the evolution of the large body size, as determined by palaeontological evidence.

The TSG *LIF* has likewise been duplicated up to ~11 times within afrotherian species (inc. elephants, manatee and hyrax), though none were demonstrably expressed with exception to a single *LIF6* in elephant cells, and only substantially under induced genotoxic stress (Vazquez *et al.*, 2018). Further, *FBXO31*, involved in the regulation of *TP53* degradation, has suggestively been duplicated ~63 and ~57 times respectively in the little brown bat (*Myotis lucifugus*) and Brandt's bat (*Myotis brandtii*) (Caulin *et al.*, 2015), and a single putative duplication of *PCNA*, similarly involved in DNA damage repair, was noted in the original analysis of the bowhead whale (*Balaena mysticetus*) genome (Keane *et al.*, 2015).

In addition to the number of TSGs in the genome, greater redundancy can also manifest itself via expression profiles. Clinically relevant expression profiles of many TSGs are known to be tissue specific (e.g. *SMAD4* and *STK11* are relatively highly expressed in the gastrointestinal tract, compared to

elsewhere; Muir & Nunney, 2015), with some TSGs being categorically inactive as a normal phenotype in many tissues. Instead of duplicating TSGs, larger, longer-lived mammals may therefore have simply altered the expression profiles of existing TSGs; i.e. to be more ubiquitous in tissue types presenting particular risk, via epigenetic or other expression pathway alterations.

If greater TSG redundancy can result in reduced cancer risk, the question then follows why don't most multicellular organisms simply maximise TSG redundancy?

Experimental evidence demonstrates overexpression of *TP53* in transgenic mice, though demonstrably more cancer resistant, results in major life history trade-offs, including slower pre- and post-natal growth and reduced body size (Rodier, Campisi & Bhaumik, 2007; Maier *et al*., 2004), accelerated senescence and a shorter lifespan (Maier *et al*., 2004; Tyner *et al*., 2002), and reduced fertility - to a particularly high and costly degree in males (Maier *et al*., 2004; Allemand *et al*., 1999). Seemingly, as much as increased TSG copy number does enhance tumour suppression, the associated fitness cost is seemingly prohibitive.

Evidence provided by elephant genome analysis indicates that *TP53*RTG expansion in the elephant family likely evolved via a birth and death process; while new copies arose via duplication, only a few were maintained whilst others become non-functional and were lost. Under this model, selection acts to maintain only a minimum number of functional copies. This firstly implies there's at least no adaptive advantage to retaining duplicates and that total TSG copy may be a selectively neutral process, and, secondly, because of this, analysis of any genome must consider any duplicate TSGs found may simply be driven by higher rates of duplication and low rates of loss, rather than adaptive expansion.

As for the fitness cost, though the exact function of the expressed *TP53* retrogenes remains unclear, evidence suggests retrogene expression - via a transposable element derived from a promoter evolutionarily younger than the duplicate retrogenes - occurred only after several mutations altering the retrogene sequences arose. Further, a prevailing hypothesis suggests expressed elephant *TP53*RTGs are indirectly functional: in translating *TP3*RTG protein domains that bind with *TP53* ubiquitination proteins, the *TP53*RTG 'decoys' reduce negative inhibition of *TP53* via ubiquitination, thereby elevating the persistent pool of 'canonical' *TP53* in the cell. *TP53*RTGs therefore do not act in an equivalent manner to fitness-cost inducing *TP53* duplicates in *in vivo* mouse studies.

In short, direct duplication and expression of TSGs retaining original function incurs too high a fitness cost on individual growth and reproduction; in support of this, observed genomic TSG expansion presents evidence of selective forces acting to minimise copy number, and any expressed function of copies is altered substantially from the canonical TSG.

2.3. Elimination of Proto-Oncogenes

As much as increasing tumour-suppressor redundancy can reduce cancer risk under a multistage model, so too could be the elimination of proto-oncogenes. As proto-oncogenes are dominant mutants, requiring only a single 'activation' to express a cancer-promoting phenotype, the fewer proto-oncogene alleles available to 'activation' in this way, the lower the risk of carcinogenic progression. Experiments have demonstrated nullification of proto-oncogenic alleles does indeed reduce cancer risk; Ise *et al.* (2000) developed murine *HRAS* null mutants which, following induction, produced six times fewer papillomas compared to control. Similarly, transgenic murine lines with an inactivated *Grb2* allele took twice as long to develop mammary tumours compared to control (Cheng *et al.*, 1998).

Unlike mechanisms to produce redundant tumour-suppressors however, the evolutionary removal of a proto-oncogenic gene or pathway may be more constrained, given their highly conserved, and often functionally critical nature (Bi *et al.*, 2018; Lodish *et al.*, 2007). From an investigational perspective, null oncogenic alleles are likewise more difficult to determine; it's relatively easy to spot a duplicated gene, it's more difficult to robustly determine mono-allelic loss-of-function, especially if epigenetically modulated.

A considerable number of oncogenes active in cancerous tissue have been demonstrated to be viral in origin. Collectively known as oncoviruses, cancer-promoting viruses come in a variety of forms, either a DNA genome (e.g. adenoviruses, and HPV - the human papillomavirus; Whyte *et al.*, 1988; Scheffner *et al.*, 1990) or RNA genome (e.g. hepatitis C virus; Perz & Bell, 2006), and during their processes of replication can insert viral oncogenic genes into the host cell. These can either act as oncogenes directly, or otherwise enhance the influence of activated oncogenes already present in the genome (Zheng, 2010; Choi *et al.*, 2015). Alternatively, viruses can influence carcinogenesis indirectly, for example, by inducing tumour-promoting tissue microenvironments via chronic non-specific inflammation (e.g. as demonstrated in HCV-induced liver carcinoma; Axley *et al.*, 2017), or

via insertional mutagenesis (e.g. viral genes inserted within, and therefore inactivating, host tumour-suppressing genes).

Over longer periods of time, genes retrovirally incorporated into a host genome can persist, leaving descendants known as endogenous retroviruses (ERVs) (Belshaw *et al.*, 2004). ERVs are genomically abundant (5-8% of the human genome are ERVs or ERV-derived), though the overwhelming majority of ERV sequences are inactive, many can be actively expressed, and indeed be manifestly oncogenic (e.g. *HERV-K* in human melanoma and breast cancers; Downey *et al.*, 2014).

With viral oncogenes in mind, it may be the case that larger, longer-lived organisms have more effective means to immunologically or otherwise police viral oncogenic insertion and expression. Indeed, Katzourakis *et al.* (2014) demonstrated the number of ERVs present in mammalian genomes is negatively correlated with body size; one can therefore infer larger mammals are either more directly or indirectly effective at limiting retroviral insertion and integration, or are under pressure to inactivate and remove potentially oncogenically-active ERVs, or both. Genetic variation in immunological genes in humans is likewise associated with variation in viral-associated cancer incidence (Klein, 2009).

3. Suppressing Somatic Advantage and Proliferation

Once somatic mutation arises, numerous secondary response mechanisms exist to prevent mutated cells presenting cellular-level selective advantages from persisting and proliferating.

3.1. DNA Damage Response (DDR): Arrested Growth and Apoptosis

The sensitivity and efficacy of DNA damage responses, in particular the apoptotic and cell cycle arrest strategies, may be increased in larger, longer-lived organisms. If cells are better able to detect and respond to DNA damage via, for example, lower damage thresholds required to induce apoptosis, it's less likely mutations will be carried through the cell cycle and passed on via replication to daughter cells. Indeed, both the functional expression of several duplicated *TP53* retrogenes, and the over-expression of *LIF6*, in elephant cells has been experimentally associated with increased sensitivity to genotoxic stress and a lower associated threshold to inducing an apoptotic response, compared to human cells (Vazquez *et al.*, 2018; Abegglen *et al.*, 2015). Likewise, the mis-match repair and associated apoptotic response in human cell cultures exposed to DNA-damaging chemical agents results

in a larger proportion of cell death than the comparative response in mouse cell cultures (Martin & Chang, 2018; Humbert *et al.*, 1999).

Increased apoptotic sensitivity is likely to be constrained however, given the need to maintain stem cell pools of a given size throughout lifespan (Tyner *et al.*, 2002), and age-dependent resistance to apoptosis is well reported *in vitro* and *in vivo* (Park, 2017; Suh *et al.*, 2002; Yeo *et al.*, 2000). Rates of cell cycle arrest are similarly likely to be constrained, owing to the need for efficient tissue damage repair, and to maintain growth rates.

3.2. Shorter Telomeres

Larger, longer-lived organisms may present altered replicative senescence via shorter somatic telomeres. Telomerase activity retains, to some extent, telomere length in stem cell populations; increased regenerative capacity of telomerase could mean stem cells in larger, longer-lived organisms could effectively retain relatively shorter telomeres and both remain viable in long stem cell lineages, and present a means to ensure short lineage lengths in daughter stem cells, thereby capping proliferation potential. Alternatively, telomeres in differentiated daughters may lose telomere length at a higher rate.

In contrast to the hypothesis, evidence to date suggests a dichotomous relationship between body size and telomere length; smaller, short-lived mammals have the longest telomeres: for example, >50Kb in several shrew species, compared to ~9Kb in bowhead whales (*Balaena mysticetus*) (Gomes *et al.*, 2011). There is however a positive relationship between longevity and shortening rate. Human telomeres shorten at a rate of approximately ~70bp per year (Canela *et al.*, 2007), compared to ~7,000bp per year in mice (Vera *et al.*, 2012). In a cross-taxa comparison performed by Whittermore *et al.* (2019), telomere shortening rate was determined to correlate with lifespan, but not body size; Asiatic elephants (*Elephas maximus*) present a shortening rate of ~109bp per year.

3.3. Increased Immunocompetence

Variation in immunological efficacy is associated with viral-associated cancer incidence (see section 2.3), and may also have an influence on non-viral cancer progression. Though cancer, at first glance, may seem immunologically 'invisible', given immune responses typically require a 'non-self' target, the process that establishes self-tolerance requires self-exposure, and any proteins that are not exposed to the immune system triggers an immune response. In the context of cancer, these can include

proteins normally produced only in extremely small quantities but highly expressed in cancer (e.g. tyrosinase, with elevated expression in melanomas; Vargas, Sittadjody & Thangasamy, 2017), those usually sequestered from the immune system (e.g. alphafetoprotein, normally produced only during early embryogenesis, prior to a developed immune system; Ertle *et al*., 2013), and, most commonly, proteins whose structure is modified due to mutation. These unrecognised proteins, known as tumour antigens, are immunogenic and induce a response (Pandolfi *et al*., 2011).

Typically, cancer cell types present early during tumour development are most immunogenic in this way, and are eliminated via the immune system. This response however provides selective pressure for cancer cells to adaptively evolve immune-evading mechanisms; said lineages emerge and reach equilibrium within the tumour population, and tumours (and their metastatic 'daughters') therefore become increasingly non-immunogenic with time (a process called "immunoediting") (Dunn, Old & Schreiber, 2004).

Further, immune responses such as local chronic inflammation can produce microenvironments favourable to cancer progression (e.g. by promoting angiogenesis and lymphangiogenesis) (Coussens & Werb, 2002), or indeed induce it outright: inflammatory signalling in response to, for example, microbial infection, leads to, amongst other things, local suppression and promotion of apoptotic and proliferative processes respectively which further lead to increased genomic destabilisation (Colotta *et al*., 2009).

Larger, longer-lived organisms may have evolved means to better suppress the immunoediting process, or present lower tissue inflammatory potential, thereby reducing cancer risk (Mantovani *et al*., 2008). Evidence from several human cancers suggests some types of inflammation can also be cancer-inhibitory as opposed to promotional, determined by levels of the transcription factor NF-kB successfully altering the behaviour of macrophages, at least in certain tissue types (Saccani *et al*., 2006; Hagemann *et al*., 2008); overexpression of NF-kB in experimental skin tissue models inhibits invasive epidermal neoplasia (Dajee *et al*., 2003) (though the opposite effect manifests in other tissue types). By similar mechanisms, larger, longer-lived organisms may more reliably trigger tumour-inhibiting inflammatory microenvironments.

3.4. Altered Tissue Architecture and Dynamics

In addition to the influence of chronic inflammation on tissue microenvironments, it is also increasingly recognised broader tissue architecture and homeostasis plays a major role in the development of malignant phenotypes (Soto & Sonnenschein, 2011; Komarova & Cheng, 2006; Nelson & Bissell, 2006). Accordingly, even if cell lines accrue sufficient mutations to initiate carcinogenesis, malignant cells won't develop without a permissive ecological niche, which is dependent on tissue-type, organ-type, and species.

Most tissues are comprised of fundamental units of genetically-isolated hierarchical lineages, or 'compartments', derived from a small pool of multipotent stem cells. Stem cells give rise to lines of progenitor cells, which themselves produce differentiated non-dividing cells associated with mature tissue functions. As demonstrated by Rodriguez-Brenes, Wodarz & Komarova (2013), the replicative potential of a cell line is influenced by the characteristics of the 'compartments' it derives from, both physical and behavioural, such as self-renewal capability, division rates, and number of intermediary progenitors; modelled characteristics optimising limits on proliferative capability of cell lineages were found to be those exhibited in most tissue types.

A foundational assumption of Peto's paradox is that larger organisms have (linearly) more dividing cells relative to smaller organisms. However, as above, tissues are organised in such a way that most cells are differentiated, and differentiated cells are non-replicative evolutionary 'dead ends'; proliferative potential therefore depends almost solely on the number and dynamics of stem cells. Increasing stem cell number is not necessary to increase the number of differentiated cells and thereby increase body size - the same outcome can be achieved by, for example, increasing the number of precursor intermediaries. Peto's paradox likewise assumes equivalent cell size across taxa. If larger organisms increase body size through combination of enlarged cell size and altered compartment characteristics, the isometric scaling assumption of Peto's paradox would be violated, perhaps explaining some of the disparity between observed and predicted cancer incidence.

Cell size is however constrained by biophysical limitations, such as diffusion potential, and evidence from morphological comparison suggests the sizes of a majority of cell types in whales are not significantly larger than human equivalents, exceptions being (as expected) nerve cells, fat cells, and erythrocytes (the latter being ~110, ~170 and ~215um^2 in mice, humans, and whales, respectively) (Hosokawa & Sekino, 2016; Hoogduijn *et al.*, 2013).

3.5. Increased Sensitivity to Contact Inhibition

Another microenvironmental factor influencing cancer progression is contact inhibition (Stoker, 1967). Contact inhibition is a fundamental property of somatic cells, whereby cells cease proliferation once they occupy the space available to them in their microenvironment, determined by mechanical signals caused by, for example, cell shape deformation or pushing/pulling forces of the extracellular matrix (Pavel *et al*., 2018; Gérard & Goldbeter, 2014). Hypersensitivity to contact inhibition is associated with cancer resistance in naked mole rats (*Heterocephalus glaber*): mole rat cells, in culture, cease dividing at lower densities compared to human and mouse cells, due to early activation of the *P16*, and subsequent activation of the *P27* pathways, thereby providing redundancy in the mechanism of contact inhibition (Seluanov *et al*., 2009). Similar mechanisms may play a role in other cancer resistant taxa.

---

4. Life-History Traits Conferring Tumour-Suppression

4.1. Life History Trade-Offs

Evolutionary life history theory states, given resources are finite and allocation thereof must be carefully considered, the evolution and diversity of phenotypes is determined by the result of several trade-offs: organisms simply cannot maximise all competing fitness-relevant traits, such as growth, maintenance, and reproduction, simultaneously (Stearns, 1992). These trade-offs apply to all entities subject to natural selection, including all living organisms and, indeed, neoplastic cells (Aktipis *et al*., 2014).

The most fundamental life-history trade-offs are those between proliferation and survival, governed by offspring production timing (early vs. late), offspring number (many vs. few), and offspring quality (more vs. less parental and developmental investment) (Stearns, 1989). Despite all organisms having unique natural histories, all life history strategies lie along the axes defined by the above trade-offs. Further, broadly-speaking, these strategies developed by living entities undergoing natural selection can be categorised into ‘fast life’ and ‘slow life’ strategies, each describing sets of generalised characteristics (albeit not exclusively nor inclusively) (Bielby *et al*., 2007; Gaillard *et al*., 2005).

Further, life-history trade-offs selected at the level of the organism fundamentally impact the envi-

ronmental and ecological constraints at the cellular-level. With cancer progression described as a process of cellular-level fitness selection for 'fast life' (proliferative) cellular strategies - in opposition to organism-fitness pressure for 'slow life' (survival) tissue and cellular strategies - selective pressure acting at the organism-level for 'fast life' strategies can predispose cellular strategies to one of proliferation (and vice-versa). As such, in attempting to understand mechanisms of cancer suppression across taxa, it is important to consider life-history traits intrinsic to the natural history of host organisms that may, adaptively or not, via multi-level selection contribute to cancer inhibition. This section will consider the influence of various life-history traits that present selective pressures or predispositions for cancer progression.

4.2. Placentation

Metastatic cancers present several phenotypes, or hallmarks, in common with the process of placentation. The placenta is an organ that has evolved independently multiple times in diverse taxa, to varying degrees of development (Roberts, Green & Schulz, 2016; Blackburn, 2015). As an interface between maternal and fetal tissue, it anchors the embryo to the uterus and mediates, amongst other things, fetal nutrient uptake, gas exchange, waste elimination, and generally provides an environment optimised to promote fetal development, including immune-sequestration and tolerance. In mammals, the organ develops following implantation of a fetal blastocyst - a cellular mass in the early stages of embryogenesis - into the maternal endometrium, or epithelial lining of the uterus. Shortly thereafter the outer cell layer of the blastocyst becomes the trophoblast, a body of cells (trophoblasts) which invade maternal tissue and develop into the various layers of the outer placenta (Burton & Fowden, 2015; Maltepe & Fisher, 2015). Like cancer cells, trophoblasts rapidly proliferate, engage in host tissue invasion, induce vessel formation (angiogenesis), and promote an environment increasingly protected from the immune system.

Given the behavioural similarities, it has been hypothesised the evolution of invasive placentation in mammals also established a set of molecular pathways that could be co-opted by cells under cancer-promoting selective pressure (Costanzo *et al*., 2018; Ferretti *et al*., 2007; Murray & Lessey, 1999). Further, given placentation has a high early life fitness benefit, and most cancers only impart late life fitness costs, the consequence of metastatic cancer arising from co-option of placenta-promoting pathways would be tolerable; placenta invasiveness would be selected for and selection against the antagonistic pleiotropic effect (cancer) would be low.

The morphology and tissue organisation, alongside mechanisms of implantation and invasion, of the placenta is highly variable across mammals (Enders & Carter, 2004); even trophoblast cells - the fundamental building block units of the placenta - can bear little morphological and molecular resemblance across taxa (Selwood & Johnson, 2006; Blackburn, Taylor & Padykula, 1988). Broadly speaking, placentas in different species are categorised based on the degree of attachment and invasiveness (Renfree, 1982) (see Fig. 1):

i) Epitheliochorial placentas are the least invasive and present the highest number of maternal tissue layers separating the fetus from maternal blood.

ii) Endotheliochorical placentas are partially invasive, as only a single layer - the endothelial wall of maternal blood vessels and associated connective tissue - separates the fetus from maternal blood.

iii) Haemochorial placentas are the most invasive and most diffuse. Fetal tissue is in direct contact with maternal blood.

Trophoblast-derived cells behave increasingly similarly to metastatic cancer cells as the degree of attachment and invasiveness of maternal tissue, and level of immunosuppression, increases. For example, the villi that make up the structure attaching the placenta to maternal tissue are comprised of several cell types; uniquely in haemochorial placentas, these include extravillous trophoblast cells (EVTs) that, like aggressive cancer cells, excrete exosomes to prepare for tissue invasion, convert from an epithelial to mesenchymal cell type, and then directly invade maternal arteries, replacing maternal endothelial cells lining the vessel wall with themselves (Menkhorst *et al*., 2016).

In humans (a haemochorial species), several molecular mechanisms driving feto-maternal tolerance are directly reactivated in cancer (notably aggressive metastatic epithelial-derived adenocarcinomas; Rousseaux *et al*., 2013), suggesting straightforward parallelism between feto-maternal tolerance and metastatic progression. Of note, *PLAC1* - a gene normally only expressed during embryogenesis (Rawn & Cross, 2008) - is found to be re-expressed in prostate, breast and ovarian cancers (Nejadmoghaddam *et al*., 2017). Beyond behavioural and functional aspects, similarities are also noted at the epigenomic level: placentation is associated with widespread hypo- and hypermethylation profiles at particular loci and genes otherwise only observed in tumour cells (Costanzo *et al*., 2018; Smith *et al*., 2017; Gama-Sosa *et al*. 1983).

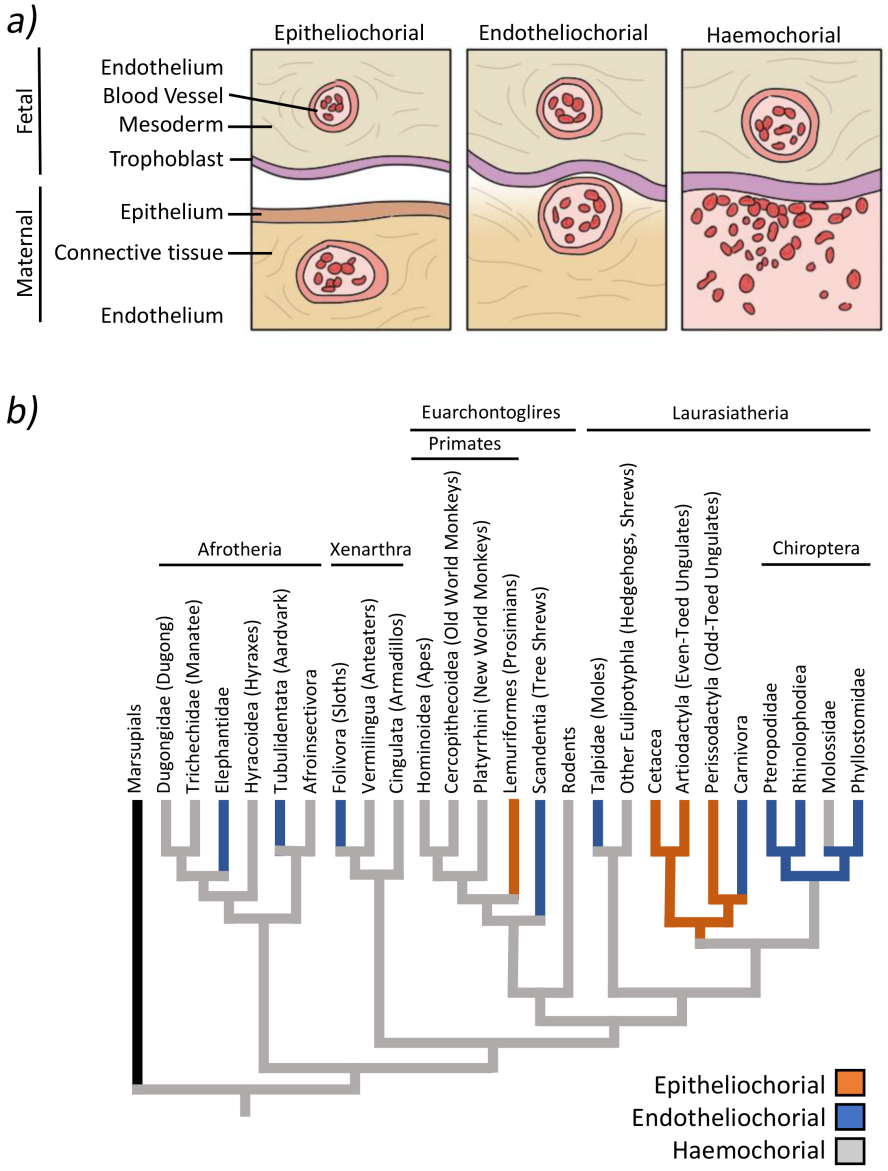


Fig. 1: a) Placentas are classified according to the number and kinds of cell layers that separates the maternal bloodstream from fetal tissue, with increasing invasiveness from epi- to endo- to haemochorial placentation. b) Within placental mammals, haemochoriality is the ancestral trait, with degrees of lesser invasiveness having evolved as a derived phenotype several times independently. In marsupials (not shown), the opposite is true; epitheliochorality being ancestral, with various lineages independently developing an increase in invasiveness. Cladogram based on Roberts, Green & Schulz, 2016.

Given the diversity of placentation and degree of invasiveness across mammals, and animal taxa at large, cross-taxa comparison may present evidence for a positive pleiotropic relationship between evolutionary trajectories towards increasing invasiveness and increased cancer incidence. Further, antagonistic pleiotropic effects may act as a factor constraining invasiveness in some lineages, or otherwise provide selective pressure to mitigate the antagonistic effect and therefore cancer incidence.

4.3. Exogenous Exposure: Diet and the Environment

Diet has been hypothesised to play a role in inter-specific variation in cancer incidence and life span. Though diet is an established risk factor for several types of cancer in humans, there remains little to no evidence suggesting a singular diet reduces risk from all types of cancer neither inter- nor intra-specifically (Steck & Murphy; 2020). The exception here being marine species occupying high trophic positions in their respective ecosystems and, via their diet, being exposed to high levels of carcinogenic chemical pollutants - but this is a recent phenomenon and totally divorced from natural cancer incidence. In a similar vein, anthropogenic environmental change is associated with increased cancer incidence in some, particularly urban, wildlife populations owing to increased carcinogenic and radiation exposure (Sepp *et al*., 2019; Sweet *et al*., 2012), but isn't of consideration here. Though exposure to UV-radiation and other forms of natural environmental carcinogenic agents does vary between species (e.g. between nocturnal and diurnal animals), and has been suggested as a contributory factor in Peto's paradox, it is unlikely to be of significance given the diversity in body size, longevity and cancer incidence within given environments - though, variation in taxon-specific responses to exposure may be of interest (see 5.2;p38).

4.4. Metabolic Rate: Lower Production of Endogenous Genotoxic Agents

Basal metabolic rate (BMR) is defined as the minimal metabolic cost of living and endothermy for inactive, resting animals in a steady, post-absorptive state (i.e. no digestion nor anabolic build-up, or growth, of tissues), typically determined by rate of oxygen consumption (ml O_2 h^{-1}) and subsequent heat production (kJ^{-h}). BMR is directly proportional to rate of oxidative metabolism at a cellular level and, in mammals, BMR is further correlated with body mass, described by allometric scaling laws, with mass-specific metabolic rate (msBMR) scaling negatively with increase in body size.

Cellular metabolism in the mitochondria generates toxic by-products, collectively known as reactive oxygen species (ROS), which damage cellular membranes and contribute considerably towards genotoxic damage. Given rates of somatic mutations resulting in oncogene activation and TSG inactivation are therefore tied directly to rates of net ROS production, which in turn is associated with general cellular metabolic rate, it is hypothesised therefore that species with low measured msBMR (i.e. larger, longer-lived species) have inherent, non-adaptive protection from cancer due to their lower metabolic rates. Kliemann *et al*. (2018) determined intraspecific rates of colorectal cancer are associated with BMR in humans, and experimental alteration of metabolic rate via caloric restriction in mouse models

results in higher longevity and lower cancer incidence (Meynet & Ricci, 2014), though no analysis has yet been undertaken between mammalian species.

Though the underlying association between metabolic rate and cancer incidence is complex, several key pathways are thought to be important. The mTOR regulatory pathway, for example, is essential for anabolic metabolism, with a reduction in mTOR signalling associated with decrease in cellular energy expenditure (Papadopoli *et al*., 2019). Observed over-activation contributes significantly towards cancer initiation and development in several human cancers (Xu, Liu & Wei, 2015), and experimental mTOR knock-out results in decreased rates of spontaneous neoplasia in mice (Wu *et al*., 2013). It's presumed decreased cellular energetic demand via mTOR inhibition also decreases demand on general mitochondrial output, suggesting a mechanistic link between mTOR, and similar pathways, with ROS production, and subsequent longevity and rates of tumorigenesis. Comparative experimental or genomic analysis of key metabolic pathways across species may provide further insight - large mammals, for example, may present molecular alterations to disproportionately inhibit mTOR compared to smaller, shorter-lived species.

Alongside the mutagenic impact of mitochondrial ROS production, ROS play a further role in intracellular signalling, aberrations from typical levels thereby influencing other cells in the microenvironment, and potential immune response. Oxidative stress resulting from altered cancer metabolism can further alter the ability of cancer cells to handle ROS - for example, by degrading the mitochondrial membrane allowing increased ROS diffusion into the cytosol - thereby instigating potential positive feedback loops, facilitating further mutations and more rapid neoplastic evolution.

4.5. Metabolic Rate: Low Growth and Replication Rates

Another fundamental assumption made by Peto's paradox is that cell replication rates are equivalent across species, however rates of replication and cell cycle progression can differ substantially between different species, scaling proportionally to basal metabolic rate. Incorporating the influence of metabolically-scaling reproductive rate into models of interspecific cancer progression may explain Peto's paradox.

4.6. Body Size Effect and Hypertumours

An alternative hypothesis to explain Peto's paradox is that larger, longer-lived organisms do indeed more frequently produce tumours compared to smaller, shorter-lived organisms, but given the physiological considerations of organismal size and growth rate, tumours grow too slowly and remain too small relative to body size for longer periods of time, such that their fitness cost is disproportionately low and they persist chronically as benign growths. Nagy, Victor & Cropper (2007) further suggested natural selection within large, slow-growing tumours may produce cell lineages that explore aggressive 'parasitic' niches within the tumour ecosystem; incapable of supporting themselves in a hypoxic microenvironment, they take advantage of angiogenic vessel-producing cells and necrotise other competing cells, essentially 'hyperparasitising' the tumour and slowing growth of the overall mass. There is no observational nor experimental evidence to date supporting the existence of hypertumours, though no experimental investigation (e.g. via serial passage of cancer through several individuals) has been done. Pathological records likewise don't support the suggestion that tumours are frequent albeit benign in larger mammals (Martineau *et al*., 2002; Bossart *et al*., 1996).

5. Tumour-Suppression in the Animal Kingdom

5.1. On Mole-Rats

Despite being relatively small mammals, both naked mole rats (*Heterocephalus glaber*) and relatively distantly related blind mole rats (*Spalax* spp.) prove remarkable exceptions to the otherwise highly cancer-prone taxon they belong to - rodents. Both are subterranean, living in hypercapnic and hypoxic burrows, unusually long-lived (up to ~32 and ~21 years, respectively) (AnAge, 2020; Edrey *et al*., 2011), and highly cancer resistant; neoplasia is not reported in blind mole rats (Manov *et al*., 2013), and only six tumours have ever been reported in naked mole rats (Delaney *et al*., 2016; Delaney *et al*., 2013); and then only in colonies living in non-natural conditions, exposed to light, and greater temperature ranges and different atmospheric gas levels (i.e. ~21% oxygen, compared to ~2-9%) compared to wild colonies. Neither exhibit replicative senescence, a trait otherwise associated with cancer-promotion, and compensate for this by utilising different mechanisms (for full review, see: Seluanov *et al*., 2018; and Lagunas-Rangel & Chávez-Valencia, 2017.

Naked mole-rats seemingly rely largely on early-acting anti-hyperplastic tumour suppressive mechanisms. Naked mole-rat cells produce and secrete large quantities of uniquely high molecular mass hyaluronan (HMM-HA) (Tian *et al*., 2013), a carbohydrate polymer that constitutes a major component of the extracellular matrix (ECM) in mammals. HA, which is comparatively much shorter in other taxa, is usually associated with inflammation and a cancer-promotive phenotype; the larger HMM-HA in naked mole-rats however has anti-proliferative, anti-inflammatory and anti-metastatic properties (Toole, 2004). In addition, HMM-HA acts as an antioxidant, reducing reactive oxygen species (ROS)-induced DNA and protein damage, and, further, is directly involved in triggering early contact inhibition (as above in 3.5) via interaction with CD44 surface receptors, which in turn activate the *P16* pathway (Seluanov *et al*., 2009).

Naked mole-rats additionally exhibit a more stable epigenome, given the propensity of their cells to resist induced pluripotent stem cell (iPSC) reprogramming (Tan *et al*., 2017). Further, via mechanisms still unclear, inactivation of several key TSGs (e.g. *TP53* and *RB1*) results in paradoxical higher apoptotic response, suggesting mole-rat cells can 'sense' when key tumour-suppressive genes are inactivated and redirect pathways to compensate for the loss-of-function (Seluanov *et al*., 2009).

Additional tumour-suppressive mechanisms include high expression of several DNA-damage response

genes (equivalent to human levels, in contrast to more closely related and size equivalent mice), ribosomes that present high-fidelity protein synthesis (Azpurua *et al*., 2013), and Nrf2-mediated alteration to regulatory pathways involving antioxidants, detoxicants and other cytoprotective molecules (Lewis *et al*., 2015).

Blind mole-rats are particularly tolerant of hypoxic conditions (Shams, Avivi & Nevo, 2004), and their anti-cancer resistance stems partly from the need to tolerate hypoxia. Hypoxia typically induces apoptosis; in order to avoid this, blind mole-rats have an altered *TP53* sequence, inhibiting its apoptotic potential. This ought to increase cancer susceptibility, but blind mole-rats compensate via a unique phenomenon known as 'concerted cell death' (Gorbunova *et al*., 2012). They allow cells to proliferate, at least until a certain proliferative hyperplastic threshold is reached, at which point an IFNβ-mediated response triggers 'mass apoptosis' of the entire cell lineage. Should cancer arise, it would too be subject to the same concerted cell death mechanism; figuratively akin to suppressing an office fire (cancer) by deliberately lighting multiple separate fires elsewhere in the building (rapidly proliferating hyperplasia) in order to trigger the building-wide sprinkler system. Further, blind mole-rats differentially express several genes associated with DNA damage repair, considered adaptive to maintaining genome integrity under hypoxic conditions (Malik *et al*., 2016).

In addition, blind mole-rats also produce heavy weight hyaluronan (HMM-HA) (Tian *et al*., 2013), however the response differs from naked mole-rats: instead of triggering early contact inhibition mechanisms (which blind mole-rats lack), HMM-HA's role is to primarily neutralise reactive oxygen species, and contribute to a cancer-resistant microenvironment (Manov *et al*., 2013). Further, heparinase in blind mole-rats is alternatively spliced, compared to other species, in such a way as to inhibit extracellular matrix degradation which, together with HMM-HA, promotes a more robustly structured ECM and tissue microenvironment that restricts tumour growth and metastasis (Nasser *et al*., 2009).

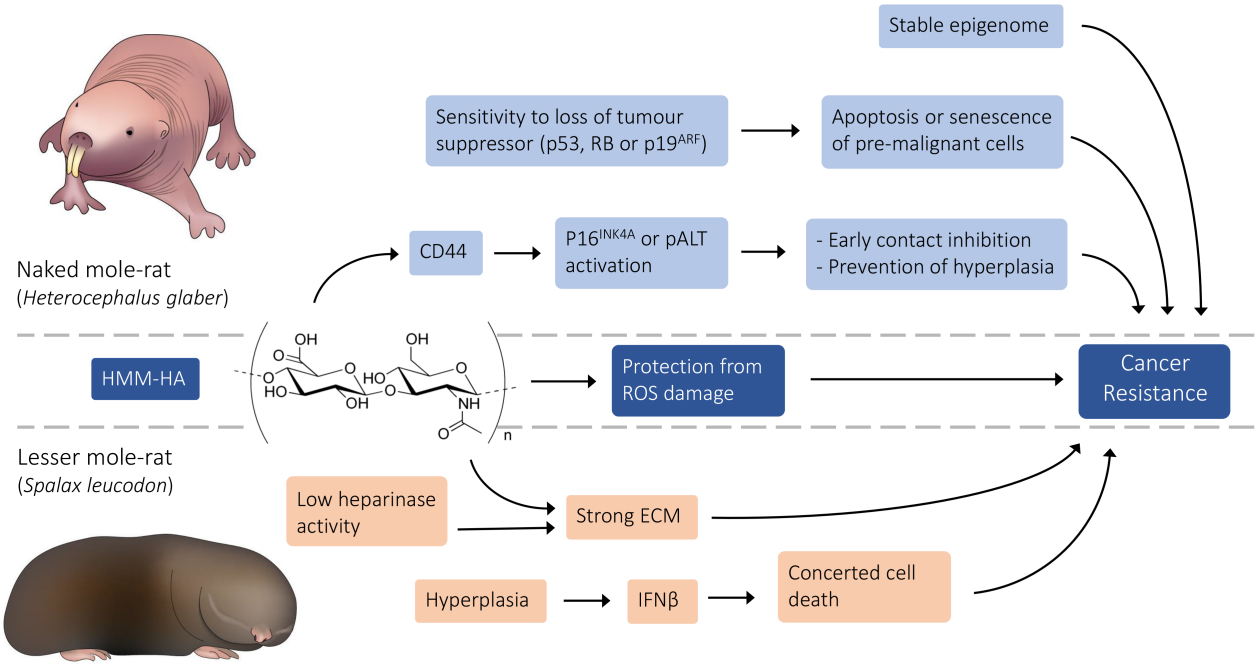


Fig. 2: Both mole-rat species produce large quantities of high molecular mass hyaluronan (HMM-HA), which acts as an antioxidant, reducing ROS-induced DNA damage. Beyond this however, no further cancer suppressing mechanisms are shared, including any further roles of HMM-HA in resistance pathways. Naked mole-rat (*Heterocephalus glaber*; light blue) HMM-HA interacts with CD44 receptors to trigger early contact inhibition via p16 activation or via pALT in order to arrest cell development. In addition, a stable epigenome resists reprogramming associated with cell transformation, and present a unique ability to detect the loss of a single tumour-suppressor and trigger apoptosis. Blind-mole rats (*Spalax* spp.), by contrast, use HMM-HA, along with alternative heparanase variants and expression, to promote robust extracellular matrix (ECM), presenting a tissue microenvironment resistant to tumour growth and metastasis. In addition, any hyperplastic growth which does arise is rapidly removed via concerted cell death, modulated by the interferon-β(INFβ) pathway. Figure modified from Seluanov *et al.*, 2018.

Both naked- and blind mole-rats exhibit multiple independent and coordinated tumour-suppressive mechanisms, and therefore demonstrate solutions to Peto's paradox - even within a single species - could be diverse (see Fig.2 summary). Likewise, both mole-rats exhibit some convergence, or at least co-adoption, of similar mechanisms to solve the same problem; in both mole-rats, HMM-HA likely evolved first as an adaptation to their subterranean lifestyles, conferring a stronger and more flexible skin that constantly rubs against tunnel walls - HMM-HA then later providing a foundation, in both instances, on which cancer-inhibiting phenotypes were developed, albeit in different ways (see Fig.2) (Seluanov *et al.*, 2018).

Though it provides interesting case studies, mole-rat cancer suppression is tangential within the wider context of Peto's paradox - a phenomenon associated with size and longevity - given that mole-rat evolved resistance is due to unique ecological and environmental drivers. To investigate Peto's paradox itself, we need to investigate cancer incidence and suppressive mechanisms beyond singular species examples; we require an approach based upon multiple independent contrasts of species with varying lifespans and body sizes, ideally whilst minimising other influencing factors such as specialisation to particular ecologies or environments, and phylogenetic history. In short, we need a clade of relatively closely-related species which are broadly similar, expressing as large a degree in body size and longevity as possible.

The order Rodentia might be a good clade for consideration. From an experimental point of view, the house mouse (*Mus musculus*) is amongst the most well-studied species on Earth, and murine (and broader rodent) molecular models and experimental methods are well established. Rodents likewise exhibit a relatively high degree of variance in size and longevity; from the African pygmy mouse (*Mus minutoides*) weighing 3-12g and living ~2 years, to capybara (*Hydrochoerus hydrochaeris*) at ~48kg and living ~15 years (Weigl, 2005). On the other hand, at project onset, very few rodent whole genome sequences had been published, notably lacking species from a wide range of sizes including the capybara (whose draft sequence was only available in 2018; Herrera-Álvarez *et al*., 2020). Further, given the desire to understand suppressive mechanisms that may eventually be of therapeutic usefulness in human biology, mechanisms conferring relative resistance amongst species smaller and living shorter lives than us are less likely to be directly informative. In contrast, the clade Cetacea, containing whales and dolphins, is potentially more promising.

5.2. Why Whales?

Cetaceans are marine mammals constituting the infraorder Cetacea, a sub-division of the even-toed ungulate order Artiodactyla. There are 89 recognised species of extant cetaceans (though this number is continuously revised based on new genomic evidence and interpretation), divided between two parvorders: Odontoceti, or the 'toothed whales' (~70 species, including dolphins, porpoises and beaked whales), and the Mysticeti, or 'baleen whales' (~14 species) (McGowen *et al*., 2020).

Cetaceans originated from a clade of terrestrial even-toed ungulates ~50 million years ago (Gatesy & O'Leary, 2001). A rapid transformation and radiation of the Archaeoceti ("ancient whales") known as the 'first cetacean radiation' took place 45-53 million years ago, resulting in a total (and dramatic) adaption to oceanic life (Thewissen & Williams, 2002). Modern Odontoceti and Mysticeti lineages split approximately ~36 million years ago during the 'second cetacean radiation' (Steeman *et al*., 2009); toothed whales thereafter acquiring echolocation and a highly specialised inner ear, while baleen whales lost their teeth and developed novel keratinous material for filter-feeding (Gatesy & McGowen, 2013).

All cetaceans are carnivorous and predatory; toothed whales engage in active hunting of macroscopic prey (fish, cephalopods etc.), whilst baleen whales are obligate filter feeders, straining small prey items (copepods, zooplankton, krill etc.) from the water. All cetaceans have a thick layer of blubber under their skin, a relatively metabolic-inactive fatty tissue used, amongst other things, to store energy, and can constitute up to 35-40% of a cetacean's body weight (Nishiwaki, 1950; Lockyer, 1976). Although they must surface to breathe, cetaceans are adapted to dive to great depths, often for long durations, and are physiologically adapted to tolerate hypoxia, hypercapnia and other diving-associated stresses (Tian *et al*., 2016; Noren & Williams, 2000). Cetaceans are highly intelligent (Chinea, 2017; Marino, 2007), often living in complex and highly communicative social groups, and can present long post-reproductive socially active lives (Pavelka *et al*., 2018).

Extant cetaceans exhibit a wide range of body sizes and lifespans, from the pig-sized vaquita (*Phocoena sinus*) at ~43kg, living up to ~21 years, to the blue whale (*Balaenoptera musculus*), the largest animal to have ever existed, weighing ~136,000kg and living up to ~110 years. Body size expansion has taken place independently multiple times within several cetacean clades, allowing for analytical contrast at multiple taxonomic levels and across different evolutionary time frames (See Fig.3).

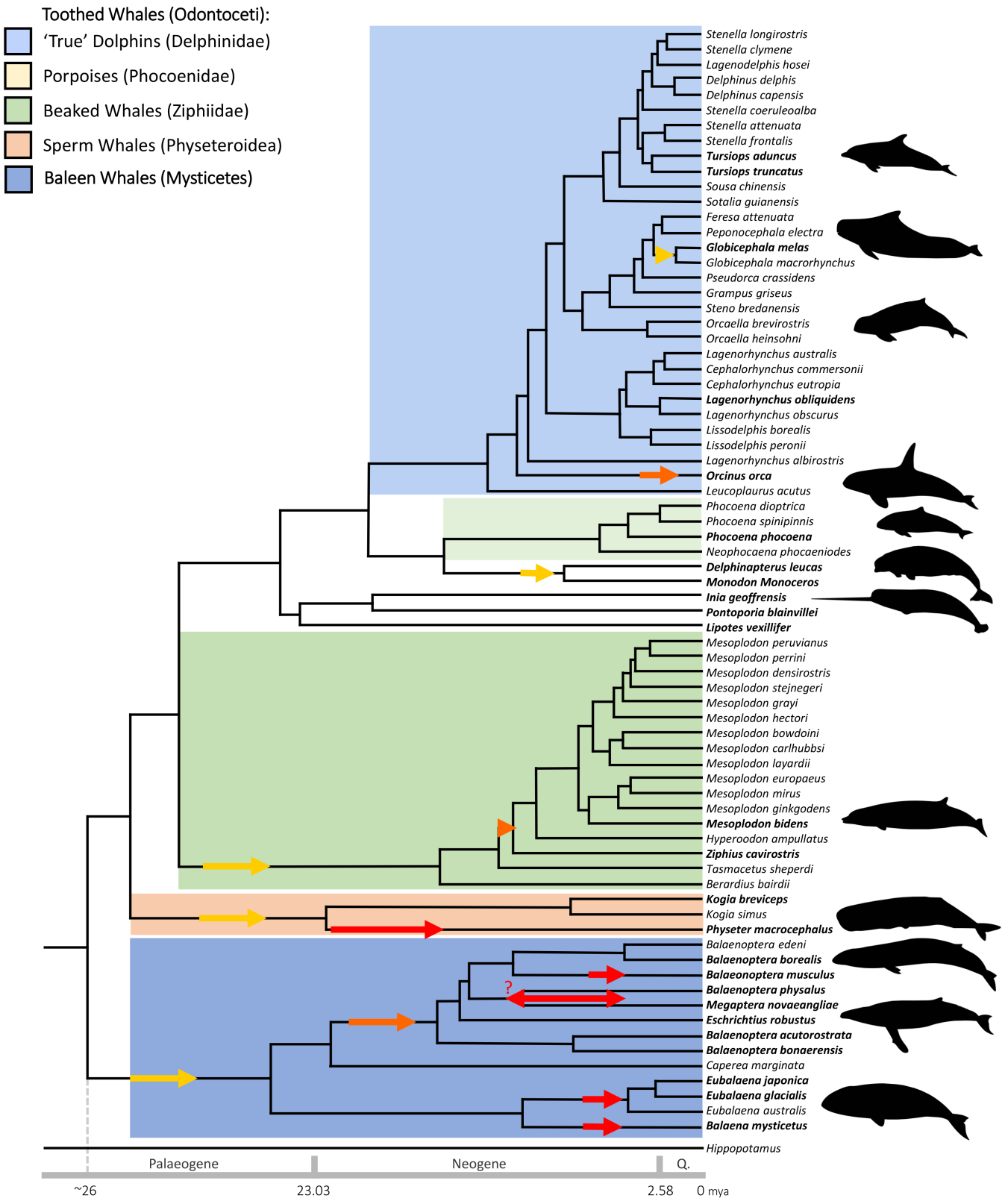


Fig. 3: Phylogenetic time tree for 73 cetacean species, based on comprehensive comparative genomic dataset by McGowen *et al.* (2020). Independent evolution of body size increase is overlain, given by yellow, orange and red arrows respectively reflecting approximate one, two and three orders of magnitude in increase from ancestral condition, based on palaeontological evidence (Slater, Goldbogen & Pyenson, 2017), though some uncertainty remains, particularly whether gigantism in extant fin

whales (*Balaenoptera physalus*) and humpback whales (*Megaptera novaeangliae*) occured before or after their most recent common ancestor. Gigantism in baleen whales is thought to be recent (~5mya), linked to Plio-Pleistocene changes in ocean ecosystem dynamics. The independent trajectory towards gigantism in at least 7 lineages presents opportunity for comparative investigation to uncover genetic and molecular changes that must have evolved to suppress cancer, and whether there are similar or diverse mechanisms. Species highlighted in bold have published whole genome data at time of writing. Illustrations taken from PhyloPic (Keesey, 2020).

The evolutionary and morphological history of cetaceans is increasingly well studied and understood, with palaeontological data providing independent and robust evidence on body size expansion within the clade. Further, at study onset, whole genome sequences were assembled and publicly available for 25 cetacean species representing diverse body sizes and lifespans from across the taxon (see Chapter 5). Hippopotamus (*Hippopotamus amphibious*) and other ungulate genomes were likewise available to present appropriate outgroups. Ample data was therefore available for comparative genomic analyses to explore molecular changes associated with cancer against increases in body size and longevity.

Unlike rodents, robust experimental methods utilising cetacean material is comparatively lacking, though wild-type cell lines from several species have been established (Burkard *et al*., 2015; Yajing *et al*., 2018; Yu *et al*., 2005) - at least one commercially available - and associated culture techniques are well documented, providing some promise of experimental work if desirable.

At project onset, few studies had been undertaken on cancer resistance in cetaceans. Martinez-Levasseur *et al*. (2013) explored genotoxic stress associated with solar ultraviolet radiation exposure in three cetacean species, demonstrating differences in physiological and molecular responses to UV-stress between sperm whales (*Physeter macrocephalus*; a toothed whale) and baleen whales. Further, papers associated directly with the publication of cetacean genomes have noted several genetic changes associated with cancer and longevity; for example, with their assemblage and analysis of the bowhead whale (*Balaena mysticetus*) genome, Keane *et al*. (2015) identified the duplication of the tumour-suppressor *LAMTOR1*. However, no comparative genomic investigation had been undertaken among and between cetaceans, nor a targeted genomic investigation of cancer associated genes and pathways. (Since, and during my own comparative investigation, Tollis *et al*. (2019) undertook a broad analysis of 10 cetacean genomes, uncovering numerous candidate genes of interest; see Chapter 5).

For these reasons, investigating the cetacean clade in-depth for evidence of cancer suppressive mechanisms via comparative genomic and other methods has yet to be undertaken, and may prove insightful into elucidating cancer suppressive mechanisms and resolving Peto's paradox. Establishing cancer incidence rates in cetaceans, as with the rest of the animal kingdom, however must first be undertaken in order to provide context and comparison, as elaborated in the following chapter.

6. References

- Abegglen, L.M., Caulin, A.F., Chan, A., Lee, K., Robinson, R., Campbell, M.S., Kiso, W.K., Schmitt, D.L., Waddell, P.J., Bhaskara, S., Jensen, S.T., Maley, C.C. & Schiffman, J.D. (2015) Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. *JAMA*. 314 (**17**), 1,850-1,860

- Aktipis, C.A., Boddy, A.M., Gatenby, R.A., Brown, J.S. & Maley, C.C. (2014) Life history tradeoffs in cancer evolution. *Nature Reviews Cancer*. 12 (**12**), 883-892

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2008) *Molecular Biology of the Cell* (5th ed.). Garland Science, New York, NY, USA.

- Allemand, I., Anglo, A., Jeantet, A.Y., Cerutti, I. & May, E. (1999) Testicular wild-type *p53* expression in transgenic mice induces spermiogenesis alterations ranging from differentiation defects to apoptosis. *Oncogene*. 18 (**47**), 6,521-6,530

- AnAge: The Animal Ageing and Longevity Database (2020). Accessed online at: `https://genomics.senescence.info/species/entry.php?species=Heterocephalus_glaber`

- Axley, P., Ahmed, Z., Ravi, S. & Singal, A.K. (2017) Hepatitis C Virus and Hepatocellular Carcinoma: A Narrative Review. *Journal of Clinical and Translational Hepatology*. 6 (**1**), 79-84

- Azpurua, J., Ke, Z., Chen, I.X., Zhang, Q., Ermolenko, D.N., Zhang, Z.D., Gorbunova, V. & Seluanov, A. (2013) Naked mole-rat has increased translational fidelity compared with the mouse, as well as a unique 28S ribosomal RNA cleavage. *PNAS*. 110 (**43**), 17,350-17,355

- Baker, J., Meade, A., Pagel, M. & Venditti, C. (2015) Adaptive evolution toward larger size in mammals. *PNAS*. 112 (**15**), 5,093-5,098

- Bebenek, A. & Ziuzia-Graczyk, I. (2018) Fidelity of DNA replication—a matter of proofreading. *Current Genetics*. 64, 985-996

- Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Pačes, J., Burt, A. & Tristem, M. (2004) Long-term reinfection of the human genome by endogenous retroviruses. *PNAS*. 101 (**14**), 4,894-4,899

- Bi, K., Chen, T., He, Z., Gao, Z., Zhao, Y., Fu, Y., Cheng, J., Xie, J. & Jiang, D. (2018) Proto-oncogenes in a eukaryotic unicellular organism play essential roles in plasmodial growth in host cells. *BMC Genomics*. 19 (**1**)

- Bielby, J., Mace, G.M., Bininda-Emonds, O.R.P., Cardillo, M., Gittleman, J.L., Jones, K.E., Orme, C.D.L. & Purvis, A. (2007) The Fast-Slow Continuum in Mammalian Life History: An Empirical Reevaluation. *The American Naturalist*. 169 (**6**), 748-757

- Blackburn, D.G. (2015) Evolution of vertebrate viviparity and specializations for fetal nutrition: a quantitative and qualitative analysis. *Journal of Morphology*. 276 (**8**), 961-990

- Blackburn, D.G., Taylor, J.M. & Padykula, H.A. (1988) Trophoblast Concept as Applied to Therian Mammals. *Journal of Morphology*. 196, 127-136

– Bossart, G.D., Cray, C., Solorzano, J.L., Decker, S.J., Cornell, L.H. & Altman, N.H. (1996) Cutaneous Papillomaviral-like Papillomatosis in a Killer Whale (*Orcinus orca*). *Marine Mammal Science*. 12 (**2**)

– Brown, J.H. & Sibly, R.M. (2006) Life-history evolution under a production constraint. *PNAS*. 103 (**47**), 17,595-17,599

– Burkard, M., Whitworth, D.J., Schirmer, K. & Bengson Nash, S.M. (2015) Establishment of the first humpback whale fibroblast cell lines and their application in chemical risk assessment. *Aquatic Toxicology*. 167, 240-247

– Burton, G.J. & Fowden, A.L. (2015) The placenta: a multifaceted, transient organ. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 370 (**1663**), e20140066

– Canela, A., Klatt, P. & Blasco, M.A. (2007) Telomere length analysis. *Methods in Molecular Biology*. 371, 45-72

– Caulin, A.F., Graham, T.A., Wang, L.S. & Maley, C.C. (2015) Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. *Proceedings of the Royal Society B: Biological Sciences*. 370 (**1673**), e20140222

– Chapman, S.N., Jackson, J., Htut, W., Lummaa, V. & Lahdenperä, M. (2019) Asian elephants exhibit post-reproductive lifespans. *BMC Evolutionary Biology*. 19

– Cheng, A.M., Saxton, T.M., Sakai, R., Kulkarni, S., Mbamalu, G., Vogel, W., Tortorice, C.G., Cardiff, R.D., Cross, J.C., Muller, W.J. & Pawson, T. (1998) Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell*. 95 (**6**), 793-803

– Chinea, A. (2017) Is Cetacean Intelligence Special? New Perspectives on the Debate. *Entropy*. 19 (**10**), 543

– Chintalapati, M. & Moorjani, P. (2020) Evolution of the mutation rate across primates. *Development*. 62, 58-64

– Choi, H.S., Jain, V., Krueger, B., Marshall, V., Kim, C.H., Shisler, J.L., Whitby, D. & Renne, R. (2015) Kaposi's Sarcoma-Associated Herpesvirus (KSHV) Induces the Oncogenic miR-17-92 Cluster and Down-Regulates TGF-β Signaling. *PLoS Pathology*. 11 (**11**), e1005255

– Clauset, A. (2013) How Large Should Whales Be?. *PLoS One*. 8 (**1**)

– Cohen, A.A. (2007) Female post-reproductive lifespan: a general mammalian trait. *Biological Reviews*. 79 (**4**)

– Colotta, F., Allavena, P., Sica, A., Garlanda, C. & Mantovani, A. (2009) Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis*. 30 (**7**), 1,073-1,081

– Cope, E.D. (1896) *The Primary Factors of Organic Evolution*. The Open Court Publishing Company, Chicago, USA.

– Costanzo, V., Bardelli, A., Siena, S. & Abrignani, S. (2018) Exploring the links between cancer and placenta development. *Open Biology*. 8 (**6**), e180081

– Coussens, L.M. & Werb, Z. (2002) Inflammation and cancer. *Nature*. 420 (**6917**), 860-67

– Dajee, M., Lazarov, M., Zhang, J.Y., Cai, T., Green, C.L., Russell, A.J., Marinkovich, M.P., Tao, S., Lin, Q., Kubo, Y. & Khavari, P.A. (2003) NF-κB blockade and oncogenic Ras trigger invasive human epidermal neoplasia. Nature. 421 (6923), 639-643

– Delaney, M.A., Ward, J.M., Walsh, T.F., Chinnadurai, S.K., Kerns, K., Kinsel, M.J. & Treuting, P.M. (2016) Initial Case Reports of Cancer in Naked Mole-rats (*Heterocephalus glaber*). *Veterinary Pathology*. 53 (**3**), 691-696

– Delaney, M.A., Nagy, L., Kinsel, M.J. & Treuting, P.M. (2013) Spontaneous histologic lesions of the adult naked mole rat (*Heterocephalus glaber*): a retrospective survey of lesions in a zoo population. *Veterinary Pathology*. 50 (**4**), 607-621

– Downey, R.F., Sullivan, F.J., Wang-Johanning, F., Ambs, S., Giles, F.J. & Glynn, S.A. (2014) Human endogenous retrovirus K and cancer: Innocent bystander or tumorigenic accomplice?. *International Journal of Cancer*. 137 (**6**), 1,249-1,257

– Dunn, G.P., Old, L.J. & Schreiber, R.D. (2004) The three Es of cancer immunoediting. *Annual Review of Immunology*. 22, 329-360

– Edrey, Y.H., Casper, D., Huchon, D., Mele, J., Gelfond, J.A., Kristan, D.M., Nevo, E. & Buffenstein, R. (2011) Sustained high levels of neuregulin-1 in the longest-lived rodents; a key determinant of rodent longevity. *Aging Cell*. 11 (**2**), 213-222

– Ellis, S., Franks, D.W., Nattrass, S., Cant, M.A., Bradley, D.L., Giles, D., Balcomb, K.C. & Croft, D.P. (2018a) Postreproductive lifespans are rare in mammals. *Ecology and Evolution*. 8 (**5**), 2,482-2,494

– Ellis, S., Franks, D.W., Nattrass, S., Currie, T.E., Cant, M.A., Giles, D., Balcomb, K.C. & Croft, D.P. (2018b) Analyses of ovarian activity reveal repeated evolution of post-reproductive lifespans in toothed whales. *Scientific Reports*. 8

– Enders, A.C. & Carter, A.M. (2004) What can comparative studies of placental structure tell us?–A review. *Placent*a. 25, Supplement, S3-S9

– Ertle, J.M., Heider, D., Wichert, M., Keller, B., Kueper, R., Hilgard, P., Gerken, G. & Schlaak, J.F. (2013) A Combination of α-Fetoprotein and Des-γ-Carboxy Prothrombin Is Superior in Detection of Hepatocellular Carcinoma. *Digestion*. 87 (**2**), 121-131

– Ferretti, C., Bruni, L., Dangles-Marie, V., Pecking, A.P. & Bellet, D. (2007) Molecular circuits shared by placental and cancer cells, and their implications in the proliferative, invasive and migratory capacities of trophoblasts. *Human Reproduction Update*. 13 (**2**), 121-141

– Foster, E.A., Franks, D.W., Mazzi, S., Darden, S.K., Balcomb, K.C., Ford, J.K.B. & Croft, D.P. (2012) Adaptive Prolonged Postreproductive Life Span in Killer Whales. *Science*. 337 (**6100**), 1,212

– Gaillard, J.M., Yoccoz, N.G., Lebreton, J.D., Bonenfant, C., Devillard, S., Loison, A., Pontier, D. & Allaine, D. (2005) Generation Time: A Reliable Metric to Measure Life-History Variation among Mammalian Populations. *The American Naturalist*. 166 (**1**), 124-128

/item Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W. & Ehrlich, M. (1983) The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Research*. 11 (**19**), 6,883-6,894

– Gatesy, J. & O'Leary, M.A. (2001) Deciphering whale origins with molecules and fossils. *Trends in Ecology & Evolution*. 16 (**10**), 562-570

– Gatesy, J. & McGowen, M.R. (2013) A phylogenetic blueprint for a modern whale. *Molecular Phylogenetics and Evolution*. 66 (**2**), 479-506

– Gérard, C. & Goldbeter, A. (2014) The balance between cell cycle arrest and cell proliferation: control by the extracellular matrix and by contact inhibition. *Interface Focus*. 4 (**3**)

– Gomes, N.M.V., Ryder, O.A., Houck, M.L., Charter, S.J., Walker, W., Forsyth, N.R., Austad, S.N., Venditti, C., Pagel, M., Shay, J.W. & Wright, W.E. (2011) Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell*. 10 (**5**), 761-768

– Gorbunova, V., Hine, C., Tian, X., Ablaeva, J., Gudkov, A.V., Nevo, E. & Seluanov, A. (2012) Cancer resistance in the blind mole rat is mediated by concerted necrotic cell death mechanism. *PNAS*. 109 (**47**), 19,392-19,396

– Gulland, F. M., Trupkiewicz, J. G., Spraker, T. R., & Lowenstine, L. J. (1996). Metastatic carcinoma of probable transitional cell origin in 66 free-living California sea lions (*Zalophus californianus*), 1979 to 1994. *Journal of Wildlife Diseases*. 32 (**3**), 250-258

– Hagemann, T., Lawrence, T., McNeish, I., Charles, K.A., Kulbe, H., Thompson, R.G., Robinson S.C. & Balkwill, F.R. (2008) "Re-educating" tumor-associated macrophages by targeting NF-κB. *The Journal of Experimental Medicine*. 205 (**6**), 1,261-1,268

– Hart, R. W., D'Ambrosio, S. M., Ng, K. J., & Modak, S. P. (1979) Longevity, stability and DNA repair. *Mechanisms of Ageing and Development*. 9, 203–223

– Heim, N.A., Knope, M.L., Schaal, E.K., Wang, S.C. & Payne, J.L. (2015) Cope's rule in the evolution of marine animals. *Science*. 347 (**6224**), 867-870

– Herrera-Álvarez, S., Karlsson, E., Ryder, O.A., Lindblad-Toh, K. & Crawford, A.J. (2020) How to Make a Rodent Giant: Genomic Basis and Tradeoffs of Gigantism in the Capybara, the World's Largest Rodent. *Molecular Biology and Evolution*. msaa285

– Hone, D.W.E. & Benton, M.J. (2005) The evolution of large size: how does Cope's Rule work?. *Trends in Ecology & Evolution*. 20 (**1**), 4-6

– Hone, D.W.E., Keesey, T.M., Pisani, D. & Purvis, A. (2005) Macroevolutionary trends in the Dinosauria: Cope's rule. *Journal of Evolutionary Biology*. 18 (**3**), 587-595

– Hone, D.W.E. & Benton, M.J. (2007) Cope's Rule in the Pterosauria, and differing perceptions of Cope's Rule at different taxonomic levels. *Journal of Evolutionary Biology*. 20 (**3**), 1,164-1,170

– Humbert, O., Fiumicino, S., Aquilina, G., Branch, P., Oda, S., Zijno, A., Karran, P. & Bignami, M. (1999) Mismatch repair and differential sensitivity of mouse and human cells to methylating agents. *Carcinogenesis*. 20 (**2**), 205-214

– Ise, K., Nakamura, K., Nakao, K., Shimizu, S., Harada, J., Ichise, T., Miyoshi, J., Gondo, Y., Ishikawa, T., Aiba, A. & Katsuki, M. (2000) Targeted deletion of the H-ras gene decreases tumor formation in mouse skin carcinogenesis. *Oncogene*. 19, 2,951-2,956

– Katzourakis, A., Magiorkinis, G., Lim, A.G., Gupta, S., Belshaw, R. & Gifford, R. (2014) Larger Mammalian Body Size Leads to Lower Retroviral Activity. *PLoS Pathogens*. 10 (**7**), e1004214

– Keane, M., Semeiks, J., Webb, A.E., Li, Y.I., Quesada, V., Craig, T., Madsen, L.B., van Dam, S., Brawand, D., Marques, P.I., Michalak, P., Kang, L., Bhak, J., Yim, H.S., Grishin, N.V., Nielsen, N.H., Heide-Jorgensen, M.P., Oziolor, E.M., Matson, C.W., Church, G.M., Stuart, G.W., Patton, J.C., George, J.C., Suydam, R., Larsen, K., Lopez-Otin, C., O'Connell, M.J., Bickham, J.W., Thomsen, B. & de Magalhaes, J.P. (2015) Insights into the evolution of longevity from the bowhead whale genome. *Cell Reports*. 10, 112-120.

– Keesey, M.T. (2020) Phylopic. [online] Available at: https://www.phylopic.org [Accessed: 01/09/2020]

– Klein, G. (2009) Toward a genetics of cancer resistance. *PNAS*. 106 (**3**), 859-863

– Kliemann, N., Huybrechts, I., Murphy, N. & Gunter, M. (2018) Basal metabolic rate and risk of colorectal cancer in the European prospective investigation into cancer and nutrition. *Revue d'Épidémiologie et de Santé Publique*. 66S5, S252

– Knudson Jr, A.G. (1971) Mutation and Cancer: Statistical Study of Retinoblastoma. *PNAS*. 68 (4), 820-823

– Korona, D.A., LeCompte, K.G. & Pursell, Z.F. (2011) The high fidelity and unique error signature of human DNA polymerase ε. *Nucleic Acids Research*. 39 (**5**), 1,763-1,773

– Komarova, N.L. & Cheng, P. (2006) Epithelial tissue architecture protects against cancer. *Mathematical Biosciences*. 200 (**1**), 90-117

– Lagunas-Rangel, F.A. & Chávez-Valencia, V. (2017) Learning of nature: The curious case of the naked mole rat. *Mechanisms of Ageing and Development*. 164, 76-81

– Leroi, A.M., Koufopanou, V. & Burt, A. (2003) Cancer Selection. *Nature Reviews Cancer*. 3 (**3**), 226-231

– Lewis, K.N., Wason, E., Edrey, Y.H., Kristan, D.M., Nevo, E. & Buffenstein, R. (2015) Regulation of Nrf2 signaling and longevity in naturally long-lived rodents. *PNAS*. 112 (**12**), 3,722-3,727

– Lockyer, C. (1976) Body weights of some species of large whales. *Journal du Conseil / Conseil Permanent International pour l'Exploration de la Mer*. 36 (**3**), 259-273

– Lodish, H. (2007) Proto-Oncogenes and Tumor-Suppressor Genes. In: Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H. & Matsudaira, P. (eds.) *Molecular Cell Biology* (6th ed.). W.H. Freeman and Company, New York, NY, USA.

– MacRae, S.L., Croken, M.M., Calder, R., Aliper, A., Milholland, B., White, R.R., Zhavoronkov, A., Gladyshev, V.N., Seluanov, A., Gorbunova, V., Zhang, Z.D. & Vijg, J. (2015) DNA repair in species with extreme lifespan differences. *Aging*. 7 (**12**), 1,171-1,182

– Maier, B., Gluba, W., Bernier, B., Turner, T., Mohammad, K., Guise, T., Sutherland, A., Thorner, M. & Scrable, H. (2004) Modulation of mammalian life span by the short isoform of *p53*. *Genes & Development*. 18 (**3**), 306-319

– Maklakov, A.A. & Chapman, T. (2019) Evolution of ageing as a tangle of trade-offs: energy versus function. *Proceedings of the Royal Society B: Biological Sciences*. 286 (**1911**), e20191604

– Malik, A., Domankevich, V., Lijuan, H., Xiaodong, F., Korol, A., Avivi, A. & Shams, I. (2016) Genome maintenance and bioenergetics of the long-lived hypoxia-tolerant and cancer-resistant blind mole rat, Spalax: a cross-species analysis of brain transcriptome. *Scientific Reports*. 6, 38624

– Maltepe, E & Fisher, S.J. (2015) Placenta: The Forgotten Organ. *Annual Review of Cell and Developmental Biology*. 31 (**1**), 523-552

– Manov, I., Hirsh, M., Iancu, T.C., Malik, A., Sotnichenko, N., Band, M., Avivi, A. & Shams, I. (2013) Pronounced cancer resistance in a subterranean rodent, the blind mole-rat, *Spalax*: *in vivo* and *in vitro* evidence. *BMC Biology*. 91

– Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. (2008) Cancer-related inflammation. *Nature*. 454, 436-444

– Marino, L. (2007) Cetacean brains: How aquatic are they?. *The Anatomical Record*. 290 (**6**), 694-700

– Martin, L.J. & Chang, Q. (2018) DNA Damage Response and Repair, DNA Methylation, and Cell Death in Human Neurons and Experimental Animal Neurons Are Different. *Journal of Neuropathology & Experimental Neurology*. 77 (**7**), 636-655

– Martineau, D., Lemberger, K., Dallaire, A., Labelle, P., Lipscomb, T.P., Michel, P. & Mikaelian, I. (2002) Cancer in wildlife, a case study: beluga from the St. Lawrence estuary, Québec, Canada. *Environmental Health Perspectives*. 110 (**3**), 285-292

– Martinez-Levasseur, L.M., Birch-Machin, M.A., Bowman, A., Gendron, D., Weatherhead, E., Knell, R.J. & Acevedo-Whitehouse, K. (2013) Whales Use Distinct Strategies to Counteract Solar Ultraviolet Radiation. *Scientific Reports*. 3, 2,386

– Matsuda, T., Bebenek, K., Masutani, C., Rogozin, I.B., Hanaoka, F. & Kunkel, T.A. (2001) Error rate and specificity of human and murine DNA polymerase η. *Journal of Molecular Biology*. 312 (**2**), 335-346

– McGowen, M.R., Tsagkogeorga, G., Álvarez-Carretero, S., dos Reis, M., Struebig, M., Deaville, R., Jepson, P.D., Jarman, S., Polanowski, A., Morin, P.A. & Rossiter, S.J. (2020) Phylogenomic Resolution of the Cetacean Tree of Life Using Target Sequence Capture. *Systematic Biology*. 69 (**3**), 479-501

– Menkhorst, E., Winship, A., Sinderen, M.V. & Dimitriadis, E. (2016) Human extravillous trophoblast invasion: intrinsic and extrinsic regulation. Reproduction, Fertility and Development. 28 (4), 406-415

– Meynet, O. & Ricci, J.E. (2014) Caloric restriction and cancer: molecular mechanisms and clinical implications. *Trends in Molecular Medicine*. 20 (**8**), 419-427

– Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y. & Vijg, J. (2017) Differences between germline and somatic mutation rates in humans and mice. *Nature Communications*. 8, 15183

– Moen, D.S. (2006) Cope's rule in cryptodiran turtles: do the body sizes of extant species reflect a trend of phyletic size increase?. *Journal of Evolutionary Biology*. 19 (**4**), 1,210-1,221

– Monroe, M.J. & Bokma, F. (2010) Little evidence for Cope's rule from Bayesian phylogenetic analysis of extant mammals. *Journal of Evolutionary Biology*. 23 (**9**), 2,017-2,021

– Muir, B. & Nunney, L. (2015) The expression of tumour suppressors and proto-oncogenes in tissues susceptible to their hereditary cancers. *British Journal of Cancer*. 113 (**2**), 345-53

– Murray, M.J. & Lessey, B.A. (1999) Embryo Implantation and Tumor Metastasis: Common Pathways of Invasion and Angiogenesis. *Reproductive Medicine*. 17 (**3**), 275-290

– Nagy, J.D., Victor, E.M. & Cropper, J.H. (2007) Why don't all whales have cancer? A novel hypothesis resolving Peto's paradox. *Integrative & Comparative Biology*. 47 (2), 317-328

– Nasser, N.J., Avivi, A., Shafat, I., Edovitsky, E., Zcharia, E., Ilan, N., Vlodavsky, I. & Nevo, E. (2009) Alternatively spliced Spalax heparanase inhibits extracellular matrix degradation, tumor growth, and metastasis. *PNAS*. 106 (**7**), 2,253-2,258

– Nattrass, S., Croft, D.P., Ellis, S., Cant, M.A., Weiss, M.N., Wright, B.M., Stredulinsky, E., Doniol-Valcroze, T., Ford, J.K.B., Balcomb, K.C. & Franks, D.W. (2019) Postreproductive killer whale grandmothers improve the survival of their grandoffspring. *PNAS*. 116 (**52**), 26,669-26,673

– Nejadmoghaddam, M.R., Zarnani, A.H., Ghahremanzadeh, R., Ghods, R., Mahmoudian, J., Yousefi, M., Nazari, M., Ghahremani, M.H., Abolhasani, M., Anissian, A., Mahmoudi, M. & Dinarvand, R. (2017) Placenta-specific1 (*PLAC1*) is a potential target for antibody-drug conjugate-based prostate cancer immunotherapy. *Scientific Reports*. 7 (**1**), 13,373

– Nelson, C.M. & Bissell, M.J. (2006) Of Extracellular Matrix, Scaffolds, and Signaling: Tissue Architecture Regulates Development, Homeostasis, and Cancer. *Annual Review of Cell and Developmental Biology*. 22, 287-309

– Nichols, H.J., Zecherle, L. & Arbuckle, K. (2016) Patterns of philopatry and longevity contribute to the evolution of post-reproductive lifespan in mammals. *Biology Letters*. 12 (**2**), e20150992

– Nishiwaki, M. (1950) On the Body Weight of Whales [online]. [online] Available at: tknk.io/ZdyG

– Noren, S.R. & Williams, T.M. (2000) Body size and skeletal muscle myoglobin of cetaceans: adaptations for maximizing dive duration. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*. 126 (**2**), 181-191

– Pandolfi, F., Cianci, R., Pagliari, D., Casciano, F., Bagalà,C., Astone, A., Landolfi, R & Barone¡ C. (2011) The Immune Response to Tumors as a Tool toward Immunotherapy. *Clinical and Developmental Immunology*. 2011, e894704

– Papadopoli, D., Boulay, K., Kazak, L., Pollak, M., Mallette, F.A., Topisirovic, I. & Hulea, L. (2019) mTOR as a central regulator of lifespan and aging. *F1000 Research*. 8, 998

– Park, S.C. (2017) Survive or thrive: tradeoff strategies for cellular senescence. *Experimental & Molecular Medicine*. 49 (**6**), e342

– Pavard, S., Metcalf, C.J.E. & Heyer, E. (2008) Senescence of reproduction may explain adaptive menopause in humans: A test of the "mother" hypothesis. *American Journal of Physical Anthropology*. 136 (**2**), 194-203

– Pavel, M., Renna, M., Park, S.J., Menzies, F.M., Ricketts, T., Füllgrabe, J., Ashkenazi, A., Frake, R.A., Lombarte, A.C., Bento, C.F., Franze, K. & Rubinsztein, D.C. (2018) Contact inhibition controls cell survival and proliferation via YAP/TAZ-autophagy axis. *Nature Communications*. 9, 2961

– Pavelka, M.S.M., Brent, L.J.N., Croft, D.P. & Fedigan, L.M. (2018) Post-Fertile Lifespan in Female Primates and Cetaceans. In: Kalbitzer, U & Jack, K.M. (eds.) *Primate Life Histories, Sex Roles, and Adaptibility*. Springer Publishing, New York, NY, USA. pp37-55

– Perz, J.F. & Bell, B.P. (2006) The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *Journal of Hepatology*. 45 (**4**), 529-538

– Peters, R.H. (1986) *The Ecological Implications of Body Size*. Cambridge University Press, Cambridge, UK

– Photopoulou, T., Ferreira, I.M., Best, P.B., Kasuya, T. & Marsh, H. (2017) Evidence for a postreproductive phase in female false killer whales (*Pseudorca crassidens*). *Frontiers in Zoology*. 14, 1-14

– Rawn, S.M. & Cross, J.C. (2008) The evolution, regulation, and function of placenta-specific genes. *Annual Review of Cell and Developmental Biology*. 24, 159-181

– Renfree, M.B. (1982) Implantation and Placentation. In: Austin, C.R. & Short, R.V. (eds.) *Reproduction in Mammals Vol.2: Embryonic and Fetal Development*. Cambridge University Press, Cambridge, UK.

– Roberts, R.M., Green, J.A. & Schulz, L.C. (2016) The Evolution of the Placenta. *Reproduction*. 152 (**5**), 179-189

– Rodier, F., Campisi, J. & Bhaumik, D. (2007) Two faces of p53: aging and tumor suppression. *Nucleic Acids Research*. 35 (**22**), 7,475-7,484

– Rodriguez-Brenes, I.A., Komarova, N.L. & Wodarz, D. (2013) Tumor growth dynamics: insights into evolutionary processes. *Trends in Ecology & Evolution*. 28 (**10**), 597-604

– Rosseaux, S., Debernardi, A., Jacquiau, B., Vitte, A.L., Vesin, A., Nagy-Mignotte, H., Moro-Sibilot, D., Brichon, P.Y., Lantuejoul, S., Hainaut, P., Laffaire, J., de Reyniès, A., Beer, D.G., Timsit, J.F., Brambilla, C., Brambilla, E. & Khochbin, S. (2013) Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. Science Translational Medicine. 5 (186), 186ra66

– Saccani, A., Schioppa, T., Porta, C., Biswas, S.K., Nebuloni, M., Vago, L., Bottazzi, B., Colombo, M.P., Mantovani, A. & Sica, A. (2006) p50 nuclear factor-κB overexpression in tumor-associated macrophages inhibits M1 inflammatory responses and antitumor resistance. *Cancer Research*. 66 (**23**), 11,432-11,440

– Scheffner, M., Werness, B.A., Huibregste, J.M., Levine, A.J. & Howley, P.M. (1990) The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell*. 63 (**6**), 1,129-1,136

– Seluanov, A., Hine, C., Azpurua, J., Feigenson, M., Bozzella, M., Mao, Z., Catania, K.C. & Gorbunova, V. (2009) Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *PNAS*. 106 (**46**), 19,352-19,357

– Seluanov, A., Gladyshev, V.N., Vjig, J. & Gorbunova, V. (2018) Mechanisms of cancer resistance in long-lived mammals. *Nature Reviews Cancer*. 18, 433-441

– Selwood, L. & Johnson, M.H. (2006) Trophoblast and hypoblast in the monotreme, marsupial and eutherian mammal: evolution and origins. *Bioessays*. 28 (**2**)

– Sepp, T., Ujvari, B., Ewald, P.W., THomas, F. & Giraudeau, M. (2019) Urban environment and cancer in wildlife: available evidence and future research avenues. *Proceedings of the Royal Society B: Biological Sciences*. 286 (1,894)

– Shams, I., Avivi, A. & Nevo, E. (2004) Hypoxic stress tolerance of the blind subterranean mole rat: Expression of erythropoietin and hypoxia-inducible factor 1α. *PNAS*. 101 (**26**), 9,698-9,703

– Slater, G.J., Goldbogen, J.A. & Pyenson, N.D. (2017) Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. *Proceedings of the Royal Society B: Biological Sciences*. 284 (**1855**), e20170646

– Smith, Z.D., Shi, J., Gu, H., Donaghey, J., Clement, K., Cacciarelli, D., Gnirke, A., Michor, F. & Meissner, A. (2017) Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature*. 549, 543-547

– Soto, A.M. & Sonnenschein, C. (2011) The tissue organization field theory of cancer: A testable replacement for the somatic mutation theory. *Bioessays*. 35 (**5**), 332-340

– Stearns, S.C. (1989) Trade-offs in life-history evolution. *Functional Ecology*. 3 (**3**), 259-268

– Stearns, S.C. (1992) *The Evolution of Life Histories*. Oxford University Press, Oxford, UK.

– Steck, S.E. & Murphy, E.A. (2020) Dietary patterns and cancer risk. *Nature Reviews Cancer*. 20 (**2**), 125-138

– Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y.W., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sørensen, M.V. & Willersley, E. (2009) Radiation of Extant Cetaceans Driven by Restructuring of the Oceans. *Systematic Biology*. 58 (**6**), 573-585

– Stoker, M.G.P. (1967) Density Dependent Inhibition of Cell Growth in Culture. *Nature*. 215 (**5097**), 171-172

– Suh, Y., Lee, K.A., Kim, W.H., Han, B.G., Vijg, J. & Park, S.C. (2002) Aging alters the apoptotic response to genotoxic stress. *Nature Medicine*. 8, 3–4

– Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N.P., Emes, R.D. & Lynch, V.J. (2016) *TP53* copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *eLife*. e11994

– Sweet, M., Kirkham, N., Bendall, M., Currey, L., Bythell, J. & Heupel, M. (2012) Evidence of Melanoma in Wild Marine Fish Populations. *PLoS One*. 7, e41989

– Takahashi, M., Singh, R.S. & Stone, J. (2016) A Theory for the Origin of Human Menopause. *Frontiers in Genetics*. 7 (222)

– Tan, L., Ke, Z., Tombline, G., Macoretta, N., Hayes, K., Tian, X., Lv, R., Ablaeva, J., Gilbert, M., Bhanu, N.V., Yuan, Z.F., Garcia, B.A., Shi, Y.G., Shi, Y., Seluanov, A. & Gorbunova, V. (2017) Naked Mole Rat Cells Have a Stable Epigenome that Resists iPSC Reprogramming. *Stem Cell Reports*. 9 (**5**), 1,721-1,734

– Thewissen, G.M. & Williams, E.M. (2002) The Early Radiations of Cetacea (Mammalia): Evolutionary Pattern and Developmental Correlations. *Annual Review of Ecology and Systematics*. 33, 73-90

– Tian, X., Azpurua, J., Hine, C., Vaidya, A., Myakishev-Rempel, M., Ablaeva, J., Mao, Z., Nevo, E., Gorbunova, V. & Seluanov, A. (2013) High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature*. 499, 346-349

– Tian, R., Wang, Z., Niu, X., Zhou, K., Xu, S. & Yang, G. (2016) Evolutionary Genetics of Hypoxia Tolerance in Cetaceans during Diving. *Genome Biology and Evolution*. 8 (**3**), 827-839

– Toole, B.P. (2004) Hyaluronan: from extracellular glue to pericellular cue. *Nature Reviews Cancer*. 4, 528-539

– Tollis, M., Robbins, J., Webb, A.E., Kuderna, L.F.K., Caulin, A.F., Garcia, J.D., Bèrubè, M., Pourmand, N., Marques-Bonet, T., O'Connell, M.J., Palsbøll, P.J. & Maley, C.C. (2019) Return to the Sea, Get Huge, Beat Cancer: An Analysis of Cetacean Genomes Including an Assembly for the Humpback Whale (*Megaptera novaeangliae*). *Molecular Biology and Evolution*. 36 (**8**), 1,746-1,763

– Tyner, S.D., Venkatachalam, S., Choi, J., Jones, S., Ghebranious, N., Igelmann, H., Lu, X., Soron, G., Cooper, B., Brayton, C., Park, S.H., Thompson, T., Karsenty, G., Bradley, A. & Donehower, L.A. (2002) *p53* mutant mice that display early ageing-associated phenotypes. *Nature*. 415, 45-53

– Vargas, A.J., Sittajody, S. & Thangasamy, T. (2017) Exploiting Tyrosinase Expression and Activity in Melanocytic Tumors: Quercetin and the Central Role of p53. *Integrative Cancer Therapies*. 10 (**4**), 328-340

– Vazquez, J.M., Sulak, M., Chigurupati, S. & Lynch, V.J. (2018) A Zombie LIF Gene in Elephants Is Upregulated by TP53 to Induce Apoptosis in Response to DNA Damage. *Cell Reports*. 24 (**7**), 1,765-1,776

– Vera, E., de Jesus, B.B., Foronda, M., Flores, J.M. & Blaso, M.A. (2012) The rate of increase of short telomeres predicts longevity in mammals. *Cell Reports*. 2 (**4**), 732-737

– Weigl, R. (2005) *Longevity of Mammals in Captivity; from the Living Collections of the World*. Kleine Senchenberg-Reihe 48, Stuttgart, Germany.

– Whyte, P., Buchkovich, K.J., Horowitz, J.M., Friend, S.H., Raybuck, M., Weinberg, R.A. & Harlow, E. (1988) Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product. *Nature*. 334, 124-129

– Williams, G.C. (1957) Pleiotropy, natural selection and the evolution of senescence. *Evolution*. 11, 398–411

– Wu, J.J., Chen, E.B., Wang, J.J., Cao, L., Narayan, N., Fergusson, M.M., Rovira, I.I., Allen, M., Springer, D.A., Lago, C.U., Zhang, S., DuBois, W., Ward, T., deCabo, R., Gavrilova, O., Mock, B. & Finkel,T. (2013) Increased Mammalian Lifespan and a Segmental and Tissue-Specific Slowing of Aging after Genetic Reduction of mTOR Expression. *Cell Reports*. 4 (**5**), 913-920

– Whittemore, K., Vera, E., Martinez-Nevado, E., Sanpera, C. & Blasco, M.A. (2019) Telomere shortening rate predicts species life span. *PNAS*. 116 (**30**), 15,122-15,127

– Xu, K., Liu, P. & Wei, W. (2015) mTOR signalling in tumorigenesis. *Biochimica et Biophysica Acta*. 1846 (**2**), 638-654

– Yajing, S., Rajput, I.R., Ying, H., Fei, Y., Sanganyado, E., Ping, L., Jingzhen, W. & Wenhua, L. (2018) Establishment and characterisation of pygmy killer whale (*Feresa attenuata*) dermal fibroblast cell line. *PLoS One*. 13 (**3**)

– Yeo, E.J., Hwang, Y.C., Kang, C.M., Choy, H.E. & Park, S.C. (2000) Reduction of UV-induced cell death in the human senescent fibroblasts. *Molecules and Cells*. 10, 415-422

– Yu, J., Kindy, M.S., Ellis, B.C., Baatz, J.E., Peden-Adams, M., Ellingham, T.J., Wolff, D.J., Fair, P.A. & Gattoni-Celli, S. (2005) Establishment of epidermal cell lines derived from the skin of Atlantic bottlenose dolphin (*Tursiops trunactus*). *The Anatomical Record Part A Discoveries in Molecular Cellular and Evolutionary Biology*. 287 (**2**), 1,246-1,255

– Zheng, Z.M. (2010) Viral Oncogenes, Noncoding RNAs, and RNA Splicing in Human Tumor Viruses. *International Journal of Biological Sciences*. 6 (**7**), 730-755

7. Bibliography

– Konarzewski, M. & Ksiażek, A. (2013) Determinants of intra-specific variation in basal metabolic rate. *Journal of Comparative Physiology B*. 183, 27-41

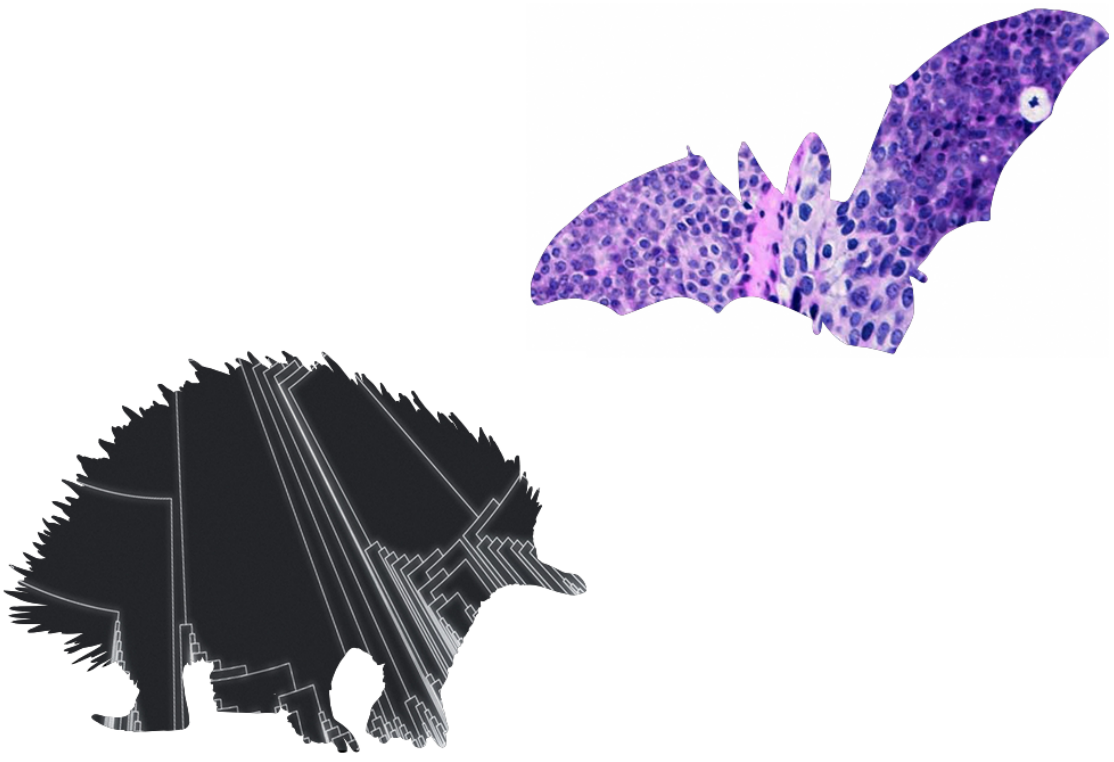
– Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V. E., de Magalhaes, J. P. (2018) Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research*. 46 (D1, D1083-D1090

– Trigos, A.S., Pearson, R.B., Papenfuss, A.T. & Goode, D.L. (2018) How the evolution of multicellularity set the stage for cancer. *British Journal of Cancer*. 118, 145-152

Chapter 3

Cancer in the Animal Kingdom

Quantifying Peto's Paradox



Abstract

Cancer presents itself ubiquitously across the animal kingdom. While elevated cancer incidence with respect to increasing lifespan and body size is well documented within individual species, Peto's paradox asserts the lack of such an association across taxa. Little evidence exists however to either support nor contradict this hypothesis, owing to the difficulty assessing cancer incidence rates in wild animal species.

With data collected from zoological and veterinary institutions across the UK, and beyond, here I present the first large, cross-taxa dataset of incidence rates across the animal kingdom (125,997 entries, representing 543 species), and quantitatively demonstrate the validity of Peto's paradox - with exception to some clade-specific relationships, cancer incidence is not only largely invariant with both mass and lifespan across the animal kingdom, but within mammals longevity is a significant negative predictor of neoplastic risk. Further life history trait analysis additionally demonstrates a significant negative relationship between mass-specific metabolic rate and incidence rates across mammals, suggesting intrinsic life history may passively play a role in cancer prevention. These results quantitatively demonstrate Peto's paradox as a valid phenomenon, and highlight the role life history evolution plays a role in shaping evolutionary trajectories towards cancer resistance.

Introduction

Peto's hypothesis, which predicts an increase in cancer risk with size and longevity, was originally based upon a review of multistage models of cancer progression (Peto, 1975), and recognised as paradoxical with respect to observed incidence based on disparate qualitative accounts. To date, despite being widely accepted, scant quantitative evidence is available to verify the paradox across animal taxa. Cancer incidence rates are well documented in humans and domesticated animals (albeit biased given potential environmental and inheritable influencing factors). Incidence in mice approximates ~40-50% (Pompei, 2001; Andervont & Dunn, 1962), humans and canines ~30% (Smittenaar *et al*., 2016; Dobson *et al*., 2006), horses ~10% (Knottenbelt, Snalune & Patterson-Kane, 2016), and elephants ~3% (Caulin & Maley, 2011; Dang, 2015) which indeed supports the paradox's incidence trajectory with increasing size. However there remains little reliable quantitative evidence supporting this trajectory, nor indeed of cancer incidence rates in general, in wild animal species.

Multispecies analyses of wildlife tumours are seldom reported in the literature, especially at scale and with any associated data for review. Those few publications which are available are typically presented as descriptive lists of tumour types uncovered during post-mortem examinations of deceased captive animals, providing little to no information on incidence probability (Chang, Liu & Jeng, 2012; Yanai *et al*., 2002; Hubbard, Schmidt & Fletcher, 1983).

In contrast, though scattered and narrow in focus, species-specific cancer prevalence and pathology, and its effect on physiology and population, has on occasion been subject to intensive research. Reported by Martineau *et al*. (2002), over the course of 17 years, 129 necropsy examinations were undertaken on stranded beluga (*Delphinapterus leucas*) carcasses from an isolated population in the St. Lawrence estuary, Canada, reporting tumour growth in 27% of examined adults. This particularly high abundance of reported cancer - estimated annually at a rate of 163 out of 100,000 animals and significantly higher than reported for any other cetacean species - was suggestively associated with influence of heavy contamination of the environment by polycyclic aromatic hydrocarbons sourced from local heavy industry, and not representative of expected background rates of carcinogenesis in beluga whales. Similarly, persistent organic pollutants were thought to be associated with high cancer rates in California sea lions (*Zalophus californianus*) and Baltic grey seals (*Halichoerus grypus grypus*): ~26% and ~64% of stranded animals suffering neoplastic growth respectively (though only

sick animals are likely to strand themselves, with actual prevalence likely significantly lower, as reported by Gulland *et al*., 1996 and Bäcklin & Olovsson, 2003). Looking more broadly across all vertebrates, further studies found high cancer prevalence in several fish species, including English sole (*Parophrys vetulus*; ~24%), walleye (*Sander vitreus*; 20-30%) and white sucker (*Catostomus commersonii*; ~59%), the etiology of which, again, heavily associated with high levels of pollution and chemical carcinogens (Coffee *et al*., 2013; Myers *et al*., 1990; Baumann, 1998 & 1984; Bowser *et al*., 1988).

Beyond long-term studies associated with marine industry and ecotoxicology however, there appears to be a dearth of robust investigation into cancer incidence rates in animal species in the wild. A comprehensive literature review of cancer prevalence in wild vertebrates was in part attempted by Madsen *et al*. (2017). Though summarising the results of over 130 individual publications, they were only successful in collating incidence data for 29 species from published datasets, only 11 of which were mammals. Furthermore, the disease etiology of most remained unknown or was otherwise associated with, as above, environmental pollution.

It's likely robust data on cancer incidence in wild populations will remain scarce, given the difficulty in collecting reliable observations. Captive animals, on the other hand, present opportune populations to quantify cancer incidence which can approximate rates expected for species in their selective environment, associated caveats taken into consideration. With few notable exceptions (e.g. most large marine mammals), most animal species exhibit greater average life expectancies in captivity compared to the wild (Tidiere *et al*., 2016). Animals in captivity suffer significantly less from environmentally driven mortality, such as predation, food competition or scarcity, and infectious disease. Living artificially longer lives however, many species eventually suffer from post-reproductive, age-associated morbidity - including, of note to this investigation, cancer. Furthermore, not all species adapt as well to captive life as others, and the pressures of a captive existence for some species, such as stress or improper diet, can encourage morbidity otherwise not present in wild individuals. As such, this must be taken into consideration when interpreting cancer prevalence data gained from captive animals, particularly when concerned with questions involving evolutionary selective pressures which act largely upon pre- and intra-reproductive life.

There are a small number of published reports of cancer vulnerability in captive species, including the naked-mole rat (*Heterocephalus glaber*; Delaney *et al*., 2016 & 2013), blind-mole rats (*Spalax*

sp.; Manov *et al*., 2013), Tasmanian devils (*Sarcophilus harrisii*; McAloose & Newton, 2009), and elephants (Elephantidae; Abegglen *et al*., 2015). Beyond singular published reports, there are few multispecies or cancer-specific analyses of reported incidence. In recent years however a surge of retrospective analyses of species- or clade-specific captive animal morbidity and mortality has been undertaken. The bulk of this material is presented as broad veterinary reference material for captive husbandry and management interests, and thus, the collation and analysis of which, in the interest of establishing cancer incidence rates in lieu of Peto's hypothesis, has been, to date, seemingly overlooked.

In this chapter, sourcing incidence data both from (as above) published retrospective morbidity records and veterinary and necropsy records from zoological institutions, I aim to establish a robust dataset on cancer prevalence across the animal kingdom. Further, given the association between cancer risk and life history traits (as described in Chapter 2.4), analyses will be undertaken to establish relationships (if any) between incidence and life history in order to gain insight into how species across the animal kingdom have adapted given the evolutionary constraint of cancer.

Methods

2.1. Cancer Prevalence across the Animal Kingdom

In order to estimate cancer incidence in wild animal species, veterinary and necropsy records were first compiled from 26 institutions, all members of the British and Irish Association of Zoos and Aquaria (BIAZA), European Association of Zoos and Aquaria (EAZA) and the (North America) Association of Zoos and Aquariums (AZA) between 2016-2019. Collaborating institutions and trusts are listed as follows (some trusts oversee multiple separate institutions):

- Action for The Wild & Colchester Zoo
- Bristol Zoological Society
- Cotswold Wildlife Park Conservation Trust
- Durrell Wildlife Conservation Trust
- FlamingoLand & The Centre for Integrated Research,
- Conservation and Learning (CIRCLE)
- International Zoo Veterinary Group UK (IZVG)
- Marwell Wildlife
- National Zoological Society of Wales
- Rare Species Convervation Centre (RSCC)
- The Aspinall Foundation
- The North of England Zoological Society (NEZS)
- The Royal Zoological Society of Scotland (RZSS)
- White Oak Conservation Trust
- Zoo Leipzig
- Zoological Society of East Anglia (ZSEA)
- Zoological Society of London (ZSL)

The overwhelming majority of this data was collated from the above institutional repositories which used the Zoological information Management Software (ZIMS) database protocol. Access to this ZIMS data was negotiated on an institution-by-institution basis, typically upon agreement to several confidentiality and privacy agreements. As such; i) no data included in the following chapters' databases nor analyses will include identifying information directly linking any dataset or entry to any particular institution / zoological collection; ii) however, given recording methodology between institutions may vary, datasets will nonetheless be tagged with an alphanumeric identifier, as a means

to introduce control over potential statistical uncertainty during data analysis; the relation of the alphanumeric identifier to the institution of origin remaining known only to the author(s). Any additional non-ZIMS data from collaborating institutions was taken either from a small number of paper records held in archive, or shared excel .xlm databases.

All available necropsy reports logged per institution (typically from present back in time until first entry into the ZIMS database, else approx. last 15 years) were collated, with exception to those relating to unhatched or unborn deceased individuals. In addition, veterinary records of individual animals still extant at time of data collection and tagged as involving tumour growth (typically undertaking, or having undertaken successful, treatment) were also reviewed and included.

Information collected from each individual necropsy and veterinary report included species, common name, identified sex, age (in years/months/days), weight at necropsy, and, in relation to cancer, a record of the presence/absence of tumour and, if present, the type of tumour, both with an associated confidence score alongside methodological justification for said score, described as follows:

i) Presence/absence of tumour: recorded as yes or no. If yes, a confidence score of 1-3 describes; 1: an observed cancerous neoplasia or similar, with an appropriate histopathological and/or immunological examination of the tumour for confirmation; 2: an observed cancerous neoplasia assigned as such by a physical examination, without histopathological or immunological examination; and 3: any pathology possibly attributable to hyperplasia, lipoma, edema or other observed abnormal cell or tissue growth, but not confirmed by further physical or pathological examination (i.e. benign non-cancerous growths, or possible albeit unlikely neoplasia)

ii) Type of tumour: recorded as described by necropsy record author. As above, an associated confidence score of 1-3 describes; 1: an acute clinical identification of tumour type, inc. cellular origin, via histopathological and/or immunological examination (e.g. “multilocular biliary cystadenoma”); 2: a more general clinical identification of tumour type and possible cellular origin, via histopathological or physical examination (e.g. “adenocarcinoma”); 3: unknown or no type specificity recorded (e.g. “neoplasia”).

In addition, for each individual necropsy record with an identified tumour, the veterinarian’s notes were also archived, and an associated reference number included in the database, for potential further follow-up (see Table 1).

In addition to necropsy and veterinary data collected by myself, a comprehensive literature search was undertaken to find and include additional data entries if they met a set of inclusion criteria (see SI.3.1 for sources and full details). Of particular note, these included 636 necropsy entries of captive elephants (Elephantidae sp.) from the online Elephant Encyclopaedia Database (Koehl, 2020), and 5,441 and 5,223 entries of beached cetaceans respectively provided by the Cetacean Strandings Investigation Programme (CSIP) (2020) and New Zealand Dept. of Conservation (Duignan *et al*., 2004).

Having collated all necropsy and veterinary reports from all sources into a single database, an automated R script was used to format and process the data (see SI.3.2 for working code and full annotation):

- Broken text, incorrect capitalisation and other text entry-associated errors are found and replaced.
- ZIMS standardised age format is converted into a decimalised year format.
- Using the taxize R package (Chamberlain *et al*., 2020) taxonomic species names are screened for error, and defunct species names are replaced *nomen novum*, under supervision. The following online databases used as reference are: Catalogue of Life, Wikispecies, IT IS, NCBI, EOL, FishBase, Reptile Database, Mammal Species of the World, and BirdLife International (Roskov *et al*., 2019; Parr *et al*., 2014; Froese & Pauly, 2019; Uetz *et al*., 2020; Wilson & Reeder, 2005).
- Higher taxonomic ranks are generated for each species, up to class, using the NCBI database.

The revised and corrected full set of individual necropsy and veterinary records is then ready to be used to generate a species-by-species cancer incidence database for analysis (see Table 2 for example), following input into a second automated script (see SI.3.3). There, the positive and negative observations of tumours are noted per species, and associated upper and lower 95% confidence intervals calculated:

$$Cancer\ \% \pm 1.96 \times SE \tag{1}$$

$$SE = \sqrt{(p(1-p))/n} \tag{2}$$

Eq. 1 & 2: 95% confidence intervals are calculated (a) using the standard error following inclusion of Laplace smoothing. Given no species is wholly immune to cancer (and thereby has a 0% incidence rate), a pseudo-count is therefore added to both positive and negative tumour observation counts to account for this, before calculating the standard error (2); where n is the total number of necropsies (including pseudo-counts) and p is the proportion of positive tumour observation (including pseudo-counts).

Species	Common Name	Sex	Age (Y/M/D)	Weight (g)	Tumour	Tumour Conf.	Tumour Type	Type Conf.	Pathology	Ref.
Ptychochromis oligacanthus	Juba cichlid	M		66	Y	1	Colorectal adenocarcinoma	1	Physical + Histo	27
Ptychochromis oligacanthus	Juba cichlid	M		75.2	N					
Ptilinopus melanospilus	Black-naped fruit-dove	M	1Y 8M 29D	139.5	N					
Ptilinopus melanospilus	Black-naped fruit-dove	M	0Y 3M 23D	100.9	N					
Ptilinopus melanospilus	Black-naped fruit-dove	F	1Y 1M 10D	91.2	N					
Ptilinopus melanospilu	Black-naped fruit-dove	F	1Y 2M 20D	88.5	N					
Pteropus livingstonii	Livingstone's fruit bat	M	7Y 8M 11D	604	Y	1	Osteosarcoma	3	Physical	38
Pteropus livingstonii	Livingstone's fruit bat	F	15Y 6M 11D	784	Y	1	Squamous cell carcinoma (Larynx)	1	Physical + Histo	61
Pteropus livingstonii	Livingstone's fruit bat	M	15Y 11M 12D	856	Y	1	Myelolipoma (Adrenal Gland)	1	Physical + Histo	70
Pteropus livingstonii	Livingstone's fruit bat	M	15Y 7M 22D	917	N					
Pteropus livingstonii	Livingstone's fruit bat	F	21Y 1M 0D	1297	N					
Pteropus livingstonii	Livingstone's fruit bat	F	25Y 8M 5D	579	N					
Pteropus livingstonii	Livingstone's fruit bat	F	0Y 3M 19D	115.3	N					
Pteropus livingstonii	Livingstone's fruit bat	F	0Y 10M 22D	283.1	N					

Table 1: Example necropsy report entries with select data taken from institutional ZIMS database. Species, common name and tumour type are taken as entered by institutional record author(s). Sex taken as M (male), F (female) or U (Unknown). Age as formatted by ZIMS software for ease of input. Weight taken in grams at time of necropsy recording. Tumour confidence and type confidence input based upon previously described scoring system (see p54). Grammatical errors and defunct species names are rectified later, alongside the reformatting of age inputs, by an (supervised) automated script. Reference numbers refer to archived full necropsy reports for any future analysis or interpretation.

Species	Common Name	# Necropsies	# Tumours	Mass (g)	Lifespan (years)	BMR (W)	msBMR (W/g)	% Tumour	Lower 95% CI	Upper 95% CI
Acionyx jubatus	Cheetah	490	100	38446.1	20.5	61.77	0.0016067	20.408	20.372	23.977
Acomys cilicicus	Turkish spiny mouse	32	3	34	4	0.258	0.0075882	9.375	9.267	20.205
Acomys cahirinus	Cairo spiny mouse	16	0	42	5.9	0.258	0.0061429	0	0	10.582
Agalychnis lemur	Lemur leaf frog	25	1	2.7	9.6			4	3.901	13.879
Agapornis nigrigenis	Black-cheeked lovebird	15	0	64.65	8.4	0.63357	0.0098	0	0	11.185
Ailurus fulgens	Red panda	73	5	5740	19	4.898	0.0008533	6.849	6.788	12.989
Alligator mississippiensis	American alligator	116	7	1079	77	0.1539	0.0001426	6.034	5.989	10.57
Alouatta palliata	Mantled howler monkey	18	1	4670	24	11.464	0.0024548	5.556	5.425	18.704
Aotus trivirgatus	Northern night monkey	27	2	914.5	30.1	2.499	0.0027326	7.407	7.296	18.491
Ambystoma mexicanum	Axolotl	43	15	13	17	0.005	0.0003846	34.884	34.744	48.87
Ambystoma tigrinum	Tiger salamander	15	0	3.21	25	0.00196	0.0006106	0	0	11.185
Alces alces	Moose	122	6	325000	25	286.847	0.0008826	4.918	4.877	8.98
Anas meller	Meller's duck	16	0	1100	18	4.068	0.0036982	0	0	10.582
Arctocephalus australis	South American fur seal	24	9	127551	32.1	76.29	0.0005981	37.5	37.313	56.201
Astronotus ocellatus	Tiger oscar	2	1	710	7.8			50	49.51	99
Arctictis binturong	Binturong	80	5	14280	27	12.747	0.0008926	6.25	6.194	11.887

Table 2: Example subset of collated species records. Displayed are actual numbers of tumours and necropsies recorded and included, however the tumour percentage and associated confidence intervals are calculated from the recorded numbers plus a pseudo-count, such that the minimum number of tumours present is 1 (given no species is 100% immune to cancer). Though mean mass and lifespan from recorded individual necropsies can be used, given a substantial amount of data collated came from sources without such data, mass and lifespan are therefore taken from published trait databases (see 2.2).

2.2. Life History Traits

Values for mean adult body mass, maximum longevity and basal metabolic rate (BMR) for most species were taken from the AnAge database (Tacutu *et al.*, 2018), with supplementation from other sources including the PanTHERIA dataset of lifehistory and ecology traits (Jones *et al.*, 2009), the Mammalian Life History database (Ernest, 2003), metabolic rate data collated by McNab (2008) and numerous clade-specific datasets hosted on DRYAD (2019) (see SI.3.1). Maximum longevity is used given the mean lifespan of most species is unknown. Values for supplemental species traits including litter size, weight and sex ratio, gestation length, postnatal growth rate, age- and size-specific mortality schedules, diet and activity cycle were likewise sourced from the above, for use in taxon-specific life history trait analyses. Categorisation of placentation was taken from literature sources (listed in SI.3.3).

2.2.1. Basal Metabolic Rate (BMR)

For species where reliable values for BMR were unavailable or absent, for some analyses as noted, a predictive BMR was estimated and included (and noted) based upon allometric scaling models applied to the complete data set:

$$\log B = \alpha + B_0 \log M \tag{3}$$

$$\log B = \beta_0 + \beta_1 \log M + \beta_2 (\log M)^2 + \epsilon \tag{4}$$

$$x = \beta_1 + 2\beta_2 (\ln M) \tag{5}$$

Eq. 3 & 4: Kleiber (1932) suggested BMR (B) scaled with body mass (M) across mammal species, concluding the relationship be described by a ¾ power law, known as Kleiber's Law (3). More recently, log-linear and quadratic models of metabolic scaling have been proposed (Kolokotrones *et al.*, 2010)(4). When applying Kleiber's Law and log-linear models of increasing higher-order terms to the species dataset with reliable BMR values taken from the literature, the resulting AIC scores agree with the recent literature that the relationship between mass and metabolic rate has convex curvature on a logarithmic scale (AIC = -368 $\rightarrow$ -415.13), correctly predicting the known BMR values for elephants and orca when they were removed from the analysis (see SI:3.4). As such, a quadratic model of basal metabolic rate was applied to the accumulated dataset in order to estimate the BMR values for select species without one (i.e. most cetaceans).

2.3. Statistical Analysis

Accurate statistical analyses of the relationship between cancer incidence and numerous life history traits, when each response variable (i.e. species) is phylogenetically associated to varying degree and therefore non-independent, requires careful consideration (Garland Jr *et al*., 2005). Given cancer incidence is expressed as proportion data derived from counts, beta and (multinomial / non-binary) logistic regression models are most appropriate.

Typically, when undertaking regression analysis of interspecies data, the independent and dependent variables are first univariately tested for phylogenetic signal, to determine whether a need to account for phylogenetic bias is required. However, diagnosing *a priori* alone whether a phylogenetic regression is required can provide misleading results (see Revell, 2010), therefore, when phylogenetic regression is undertaken, a retrospective review based on testing for bias in residual errors from non-phylogenetic regression should be undertaken. The phylolm and sensiPhy packages in R (Ho & Ané, 2014; Paterno *et al*., 2018) were therefore used to estimate phylogenetic bias and undertake appropriate phylogenetically weighted logistic regression and analysis, where appropriate (otherwise glm was used in base R). As no available packages provided for phylogenetic bias when undertaking non-binary beta regression, if required, an R pipeline was developed to enable this type of analysis (based directly on Ives & Garland Jr, 2010; see S.I.3.5).

In order to undertake each phylogenetic regression on any given input of species, phylogenetic trees were automatically generated in each instance using the taxize package in R (Chamberlain *et al*., 2020)), using NCBI databases as reference. In instances where species taxonomy or classification discrepancy was flagged (i.e. NCBI using synonymous species names, or not recognising more recent phylogeny change), that species was noted and removed. In addition, where phylogenetic uncertainty arose due to competing alternative tree hypotheses (i.e. relationships between three or more species uncertain) with different topologies and/or branch lengths, the analysis was repeated using a sample of all relevant trees and the influence of any uncertainty on output regression parameters compared and controlled for, using sensiPhy (Paterno *et al*., 2018).

Furthermore, beyond phylogenetic bias, other bias in the form of sampling uncertainty generated by the variance in sample size between species (i.e. number of necropsy entries) was accounted for via sample size sensitivity analysis using sensiPhy, and species disproportionately skewing output

(determined by significant change in output regression parameter estimates) were removed where appropriate, and noted.

In smaller taxon-specific analyses, where the origin of entries for a given species or subset of species was known, any intraspecific variation due to difference between institutional measurement errors was controlled for. This is the only instance where the alphanumeric identifier w.r.t to a specific zoological institution providing data was used. Otherwise, for the most part, data uncertainty was not controlled for, and taken as given - as a confounding variable.

Results & Discussion

3.1. Global Cancer Incidence Dataset

Approximately ~36,000 individual necropsy and (extant individual) veterinary records on institutional ZIMS databases were manually reviewed, of which 33,412 met the threshold criteria required for entry into the global cancer prevalence dataset. To this, a further 27,399 entries were included from ZIMS global species-specific profiles which met inclusion criteria, alongside 12,024 entries from taxon-specific literature sources (see SI.3.1.). Including data on domesticated species and a representative sample of human lifetime cancer prevalence (n=1000), a total of 125,997 entries representing 543 species were therefore compiled (see Table 3).

Datasets used during the analyses in the following results sections, unless otherwise stated, excluded those species with 15 or fewer individual entries, only non-estimated values for basal metabolic rate (BMR) were used, and species-specific tumour counts excluded individual entries with a tumour confidence score of 3 (therefore noted as non-tumour). For any relationship investigated between cancer prevalence and one or more life history trait(s), unless otherwise stated, model parameter results presented originate from the model with most appropriate term, and least biased and most precise estimates, following a stepwise model-fitting exercise and residual diagnostic plot review.

3.2. Cancer Prevalence across the Animal Kingdom

Placental mammals presented significantly more observed incidence of malignant neoplasia than any other vertebrate taxon, including at every order level with a few notable exceptions (see Fig. 1). Mammals likewise were the only class presenting a significantly different modelled regression trajectory between cancer incidence and mass-specific basal metabolic rate (msBMR) compared to other taxa (Fig. 2). Sensitivity analysis determined Cetacea and Soricidae (whales and shrews resp.), both taxa at the extreme ends of the size, longevity and BMR scales, had a significantly disproportionate effect on model outcome ($p<0.05$); following removal, mammals exhibited a near identical relationship between cancer and mSBMR as other classes, suggesting a universally similar relationship between msBMR and lifetime neoplasia risk in vertebrates.

Taxon	Species		Representation	
	Total	# Necropsies >= 15	% Represent	Total Sp. in Taxon*
Mammal	**522**	**284**	**6.5**	**6,399**
Monotreme	1	1	20	5
Marsupial	40	25	7.6	328?
Afrotheria	13	10	12.5	80-86
Afroinsectiphilia	8	4	8.7	69-75
Paenungulata	5	5	36.4	11
Xenartha	6	5	16.1	31
Rodents	43	32	1.4	2,277
Tree Shrews (Scandentia)	5	3	15	20
Primates	64	51	10.9	470-475
Prosimian	23	19	12.9	148-152
New World Monkeys	17	9	5.6	161
Old World Monkeys	14	14	10.1	138
Apes	10	9	39.1	23-24
“Insectivore” (Eulipotyphla)	2	1	0.22	450?
Chiroptera	19	18	1.82	986-1,200+
Carnivora	64	55	19.7	279-?
Ungulates	94	83	25.4	326
Even-Toed Ungulates	45	45	20.6	220
Odd-Toed Ungulates	8	7	41.8	17
Cetacea	41	31	34.8	∼89
Birds	**173**	**88**	**0.72-0.8**	**9,000-10,000?**
Reptiles	**47**	**21**	**0.18**	**11,136**
Amphibians	**22**	**12**	**0.14**	**8,161**
Fish	**55**	**9**	**0.02**	**33,230-35,000**
Invertebrates	**23**	**0**	**0.0001<**	**?**
Total:	**793**	**542**		

Table 3: 125,997 records were reviewed and collated into a dataset representing 793 species, 542 of which had ≥15 individual necropsy and veterinary records meeting the threshold for analysis inclusion (see Methods 2.1.). With exception to the species-rich rodent (Rodentia), bat (Chiroptera) and ‘insectivore’ (Eulipotyphla) clades, which are disproportionately underrepresented in captive collections (and/or appropriate morbidity study), at least ~10 (±3)% or more of total species within 15 mammal orders (of 26 extant) are represented, including approximately 1/3rd of all cetacean species. Non-mammalian and non-avian vertebrates are comparatively record-poor, lacking sufficiently representative samples for robust life history trait analyses, and no invertebrates met analysis threshold criteria. Total species counts per taxon sourced from Burgin *et al.* (2018); Barroclough *et al.* (2016); Uetz *et al.* (2020); AmphibiaWeb (2020), and; Froese & Pauly (2020).

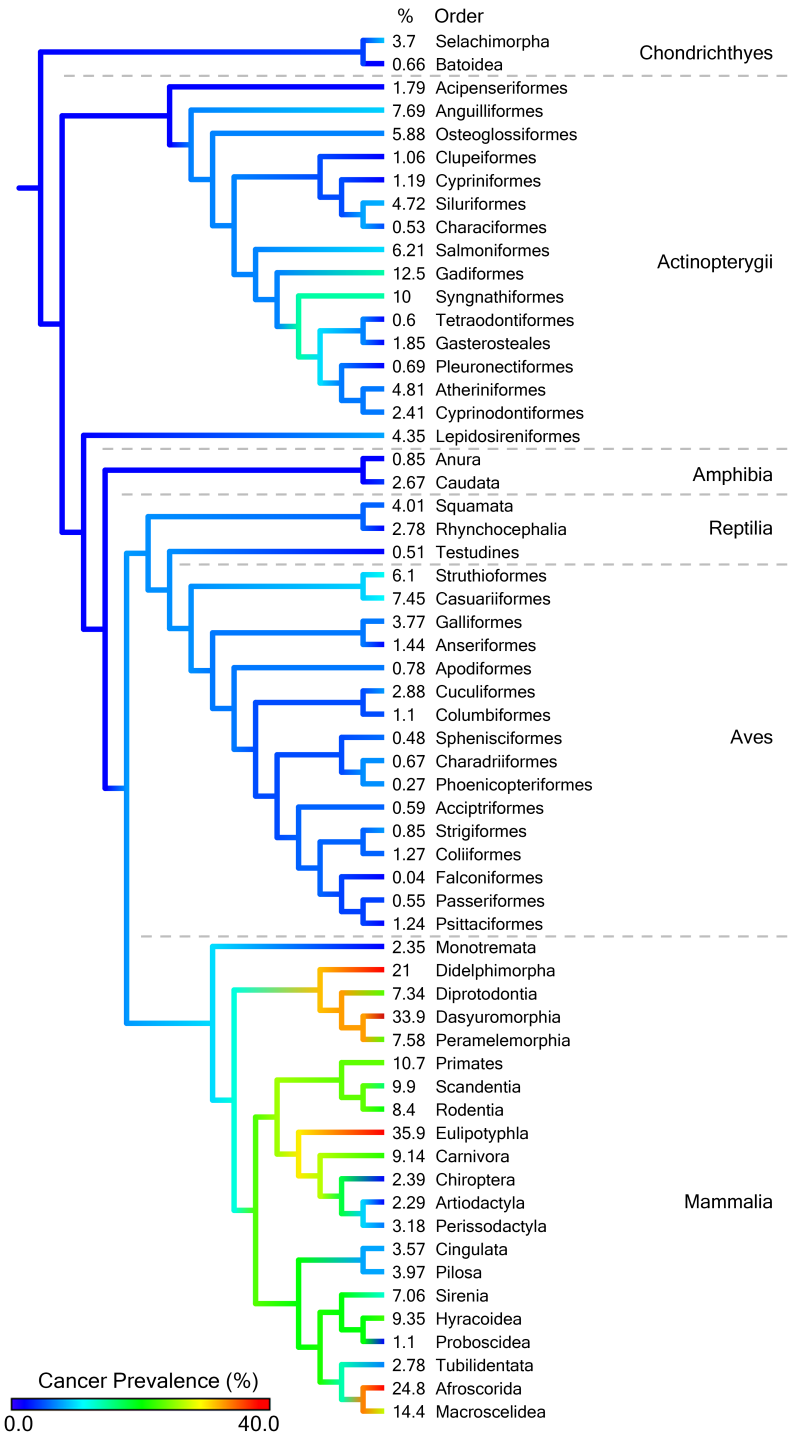


Fig. 1: Collated cancer prevalence across 61 vertebrate orders. Placental mammals as a clade are considerably more prone to cancer than any other vertebrate group, with notable orders including Chiroptera (bats), Artio- and Perissodactyla (even- and odd-toed ungulates, including cetaceans) and Proboscidea (elephants) presenting disproportionately lower incidence compared to sister taxa and estimated ancestral condition. Beyond mammals, higher prevalence is observed in Struthioformes and Casuariiformes (large, flightless ratites), and Syngnathiformes (seahorses and pipefish). Trait gradient mapping calculated via Bayesian MCMC using PhyTools in R (Revell, 2012), utilising NCBI taxonomy database branch length, and colour mapped onto Open Tree of Life cladogram (Rees & Cranson, 2017).

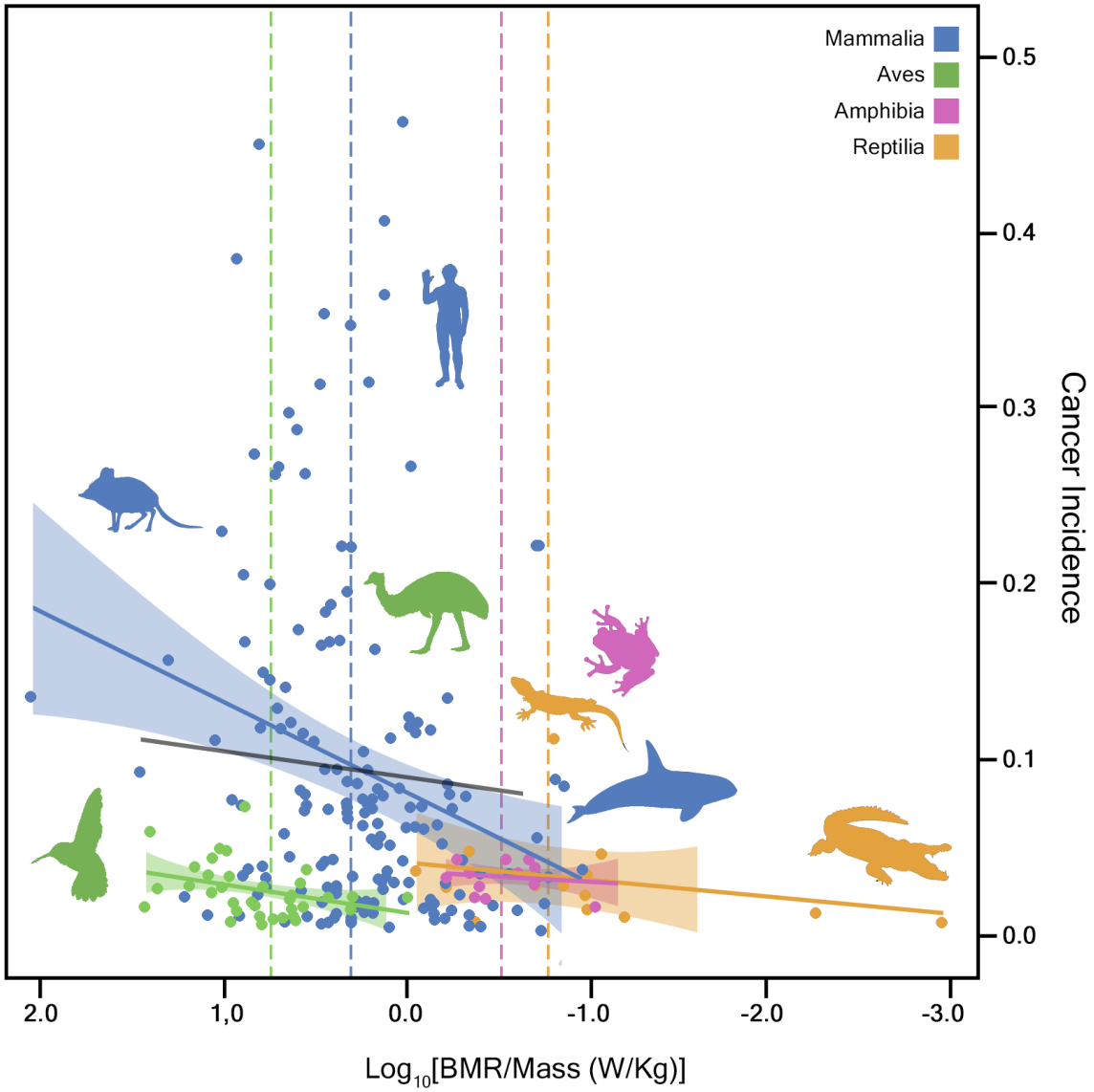


Fig. 2: Phylogenetically-weighted logistic regressions between cancer incidence and mass-specific basal metabolic rate (msBMR) across all 240 vertebrate species in four classes (where n > 15); mammals (177), birds (36), reptiles (14) and amphibians (13). Dashed vertical lines indicate phylo-weighted mean msBMR per taxon, shaded area modelled upper and lower confidence interval ranges. There is no significant difference in the slopes of the regression lines between birds, amphibians and reptiles as determined by pANCOVA ($p>0.1$). When mammal orders at the extreme ends of the body mass spectrum are removed (Cetacaea and Soricidae), the relationship between cancer and ms-BMR in mammals likewise follows a near identical trajectory to the other taxa (black line; $p>0.1$). Large ungulates and cetaceans have equivalent msBMR values to typically poikilothermic amphibians and reptiles, and share similar low lifetime cancer incidence rates. Crocodilians and giant tortoises likewise have particularly low observed msBMR, exhibiting degrees of negligible senescence and lifespans approx. 100-200 years, with little records of observed neoplasia (Stark *et al.*, 2018; Miller, 2001). Illustrations taken from PhyloPic (Keesey, 2020).

3.3. Mammals

A total of 9,888 tumours were found in 248 mammal species (where n>15), representing 19 taxonomic orders with mass spanning five orders of magnitude, from the African pygmy mouse (*Mus minutoides*; weighing 7.2g, with a maximum lifespan of 4.3 years) to the blue whale (*Balaenoptera musculus*; with an estimated mean weight of ~136,000kg, and living ~110 years).

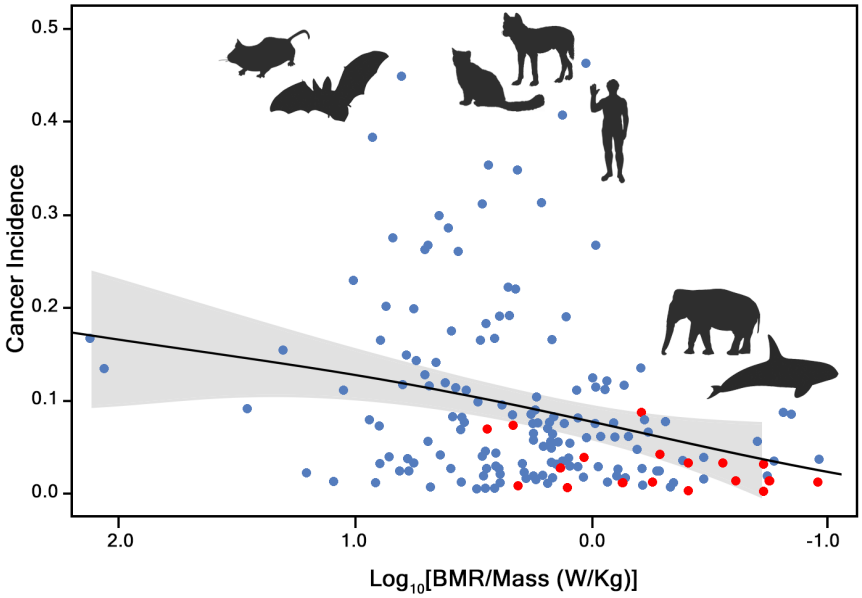
Cancer incidence was found to scale invariantly with size, longevity and basal metabolic rate across 270 species of mammal, supporting Peto's hypothesis (see Fig.3). No significant relationship was found between cancer incidence and any combination of body mass, life span or basal metabolic rate, across the class Mammalia as a whole, following a series of phylogenetically-and-count-weighted logistic regression analyses. Removing the effect of phylogenetic bias between species did invoke a significant relationship between cancer and lifespan, body mass and mass-specific metabolic rate ($P<0.05$), thereby suggesting not only does cancer incidence not increase with increasing size and longevity, but scales negatively. The interpretive relevance of said non-phylogenetically weighted models without accounting for bias is spurious at this taxonomic (Class) resolution however, given both independent and dependent variables, and residual errors, tested strongly for phylogenetic signal bias (see Methods 2.3. and SI.3.6; Revell, 2010).

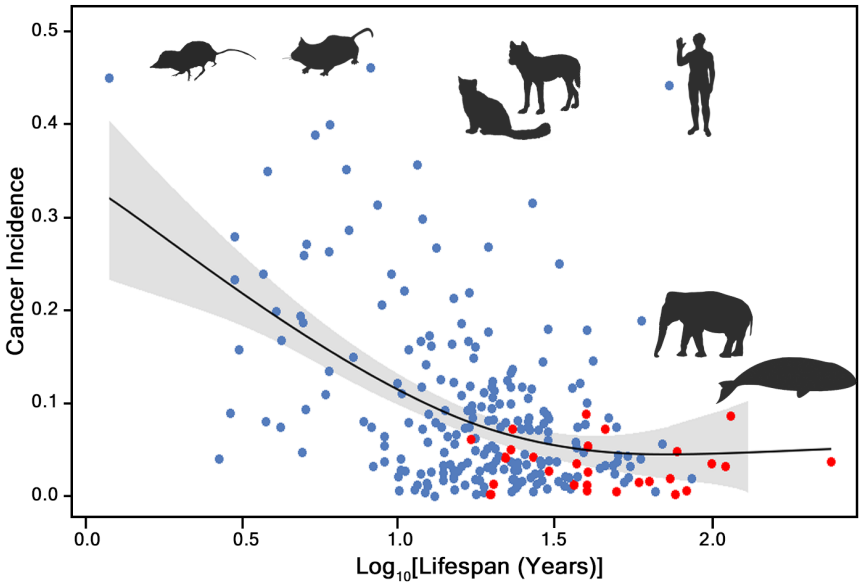
Models with and without inclusion of *a priori* estimated over-influential species - human, domesticated species, and an ecotoxicological-biased subset of data from two cetacean species (humpback whale, and St. Lawrence estuary population of beluga whale) - yielded no discernible difference in output for Mammalia as a whole. However, sensitivity analyses testing for influential clades on model parameter estimates revealed a number of taxa with a disproportionately large effect on predictive outcome ($P<0.01$, following randomisation test); of note, cetaceans had the largest disproportional weighted influence on mass and lifespan and when removed from the dataset and a phylogenetically-weighted logistic model reapplied, a significant relationship between cancer incidence and lifespan and mass-specific metabolic rate ($P<0.05$) was observed across 228 and 177 mammalian species respectively (BMR values being available for some, but not all, species).

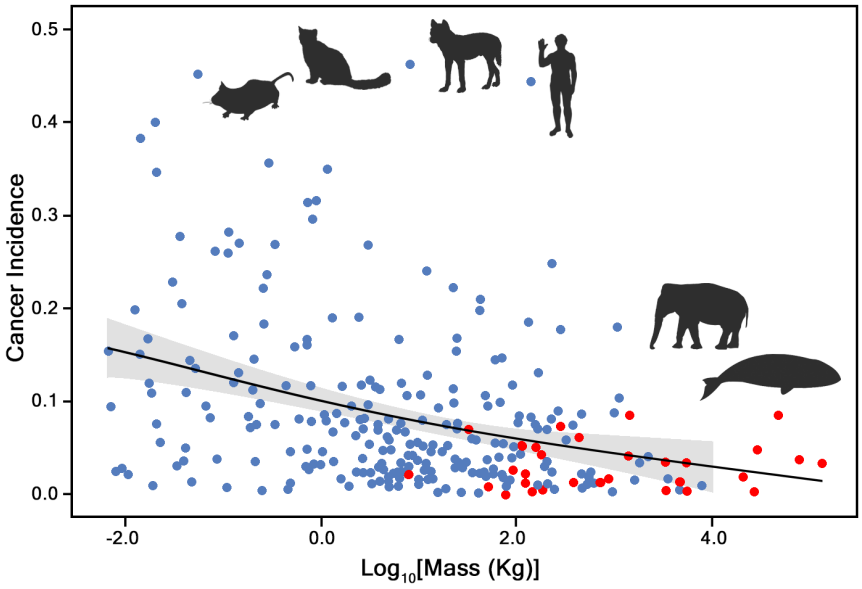
The high variance of cancer incidence, particularly in small mammals with short lifespans, makes precise predictive model fitting difficult at this taxonomic rank, and, as above, influential clade analysis suggests assorted taxa, between infraorder and family level, may have wildly varying relationships

between cancer and life history traits. The following sections investigate the relationship between cancer and both primary life history traits, and some taxon-specific secondary traits, at these levels.

Fig. 3 (overleaf): Quantitatively supporting Peto's paradox, increasing cancer incidence is not associated with increase in longevity (b) nor mass (c) across 270 mammal species [excluding cetacea (red)] as shown by phylogenetically weighted logistic regression (pseudo-R^2: 0.206 and 0.23 respectively); while cancer incidence scales invariantly with mass (p=0.29), lifespan is a significantly negative inter-species predictor of cancer incidence ($p<0.05$). Likewise, though (non-adjusted) cancer incidence significantly correlates negatively with decreasing mass-specific metabolic rate ($p<0.05$), it remains a poor predictor of cancer incidence as shown by a (a) phylogenetically weighted logistic model across 228 species (p=0.16, pseudo-R^2: 0.28), Regression undertaken on non-transformed data; x-axis log-transformed for visualisation purposes. Illustrations taken from Phylopic (Keesey, 2020).


Cancer Incidence
0.5
0.4
0.3
0.2
0.1
0.0
2.0
1.0
0.0
-1.0
Log10[BMR/Mass (W/Kg)]


Cancer Incidence
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.5
1.0
1.5
2.0
Log10[Lifespan (Years)]


Cancer Incidence
0.5
0.4
0.3
0.2
0.1
0.0
-2.0
0.0
2.0
4.0
Log10[Mass (Kg)]

Mammal Taxa

3.3.1. Monotremes & Marsupials

329 tumours were noted in 2,497 individuals from 36 marsupial and a single monotreme species - the short-beaked echidna (see Table 4). Reports of neoplasia in prototherian mammals (monotremes) are rare; as of 2019 only eight total cases have been recorded worldwide (Tong, 2019; Ladds, 2009); six in short-beaked echidna (*Tachyglossus aculeatus*; three of which were lymphoid proliferations) and two in platypus (*Ornithorhyncus anatinus*).

Here, only a single instance of neoplasia was recorded in the short-beaked echidna, providing an estimated cancer incidence rate of 2.35% (CI: 2.03-2.68) for the species. Though caution must be taken given the low sample size (n=83), if this estimate is consistent with true incidence rates, it is particularly low for such a long-lived mammal for its given size-class, compared to the placental mammals. Of note, the short-beaked echidna has a disproportionately low mass-specific basal metabolic rate (msBMR), at 0.517 W/kg, which is lower than almost all mammals, except the very largest Artiodactyla, and akin to that of poikilothermic reptiles and amphibians (despite being homeothermic itself). Echidnas are semi-fossorial, spending the majority of their time underground (~83.3-86.3% of total measured time budget; Clemente *et al*., 2016), and are likewise noted for their physiological adaptations to tolerate hypoxic and hypercapnic (high CO_2) environments (Waugh *et al*., 2006; Griffiths, 1989). Adaptations to hypoxia and hypercapnia independently evolved in unrelated (and long-lived, relatively metabolically inactive) blind mole-rats (Spalacidae) and naked mole-rats (Heterocephalidae) and is thought to be responsible for the respective high cancer resistance in both groups (see Chapter 2: 5.1). Low cancer incidence in the short-beaked echidna might also be the outcome of a similar evolutionary trajectory shared by mole-rats, given some similarities in life history (though msBMR in the echidna is still an order of magnitude lower than in both mole-rats). At time of writing (May 2020), the echidna genome is currently being assembled, and may elucidate more insight in this regard.

In contrast to monotremes, neoplasia is frequently reported in marsupials in the literature, with most marsupial taxa considered moderately or markedly susceptible to neoplasia. Dasyurids, including the Tasmanian devil (*Sarcophilus harrissii*), cat-like quolls and mouse-like antechinus, are known to be at particular risk, and it is not uncommon for individuals in this clade to present clinically with

multiple independent neoplasms (Tong, 2019; Michael & Sangster, 2010; Canfield, Hartley & Reddacliff; 1990). Despite this, it is often noted that fatal and/or aggressive epithelial cancers are observed more frequently in eutherian mammals, compared to metatherian, a leading hypothesis suggesting the development of an increasingly invasive placenta and feto-maternal immune tolerance favours the emergence of neoplasia and subsequent metastasis in eutherian mammals (see Chapter 2: 4.2). Assorted contemporary marsupial taxa represent a wide range of placental phenotypes, exhibiting degrees of development and invasiveness allowing for an examination into any relationship between cancer incidence and placentation.

Here, contrary to the literature, marsupials collectively presented higher cancer incidence rates relative to eutherians (a sub-set, selected from within the minimum and maximum bounds of marsupial mass, lifespan and metabolic rate entries) (see Fig. 4). This held even when removing entries from the dasyurid family, which exhibit an inverse relationship between cancer incidence and all three life history traits (see Fig. 3). Of note, this removal of Dasyuridae from analysis suggested a significant negative relationship between cancer incidence and body size ($p<0.01$), as opposed to an invariant relationship otherwise presented including all taxa ($p=0.906$).

Beyond relationship with the primary life history traits, phylogenetically-weighted analysis of covariance (pANCOVA) including additional life history variables, reveals the most significant influence on model prediction and cancer incidence are those associated with degree of placentation. For all given metrics of placental development, those more closely aligned with the eutherian condition (highly invasive placenta) exhibit significantly higher cancer incidence compared to those species presenting a more distant maternal-fetal interaction (see Fig. 5).

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Tachyglossus aculeatus	Short-beaked echidna	1	83	4.5	50	2.3270	2.35	2.03	2.68
Aepyprymnus rufescens	Rufous bettong	1	23	2.81	8	3.7041	8	6.94	9.06
Antechinus minimus	Swamp antechinus	1	5	0.053	2	-	66.7	25.2	31.9
Antechinus stuartii	Brown antechinus	2	12	0.029	3	0.189	28.6	19.3	23.6
Bettongia penicillata	Brush-tailed rat-kangaroo	0	3	1.184	6	3.132	21.4	16.5	23.5
Burramys parvus	Mountain pygmy possum	13	97	0.044	12	0.205	20	13.5	14.8
Dasycercus cristicauda	Mulgara	1	1	0.1	7	0.26	14.1	61.3	72
Dasykaluta rosamondae	Little red antechinus	1	1	0.039	3	-	66.7	61.3	72
Dasyuroides byrnei	Kowari	13	47	0.109	7	0.439	28.6	27.3	29.8
Dasyurus geoffroii	Northern quoll	14	39	1.107	5	1.501	36.6	35.1	38.1
Dasyurus maculatus	Spotted-tailed quoll	7	28	3.284	5	3.142	26.7	25.1	28.2
Dasyurus viverrinus	Eastern quoll	14	41	1.101	6.8	2.26	34.9	33.5	36.3
Didelphis virginiana	Virginia opossum	32	168	2.442	5	5.299	19.4	18.8	20
Dorcopsis muelleri	Dorcopsis wallaby	1	6	5.371	7.6	-	25	22	28
Isoodon macrourus	Northern brown bandicoot	1	21	1.506	3	3.202	8.7	7.54	9.85
Isoodon obesulus	Southern brown bandicoot	2	36	0.825	3.75	1.239	7.89	7.04	8.75
Macropus agilis	Agile wallaby	1	10	11.816	12	-	16.7	14.6	18.8
Macropus giganteus	Eastern grey kangaroo	2	144	33.41	24	-	2.05	1.82	2.28
Macropus parma	Parma wallaby	1	73	4.157	10	-	2.67	2.3	3.03
Macropus parryi	Whiptail wallaby	1	5	12.63	14	-	28.6	25.2	31.9
Macropus robustus	Common wallaroo	1	51	25.979	24	33.056	3.77	3.26	4.29
Macropus rufogriseus	Red-necked wallaby	3	80	16.85	19	-	4.88	4.41	5.34
Macropus rufus	Red kangaroo	10	178	38.968	30	31.353	6.11	5.76	6.46
Macrotis lagotis	Bilby	1	7	1.23	10	2.4	22.2	19.5	24.9
Monodelphis domestica	Gray short-tailed opossum	39	150	0.093	6	0.335	26.3	25.6	27
Parantechinus apicalis	Dibbler	2	2	0.071	3	-	75	70.8	79.2
Parantechinus bilarni	Sandstone antechinus	2	6	23.1	3	-	37.5	34.1	40.9
Petaurus breviceps	Sugar glider	24	205	0.121	14	0.517	12.1	11.6	12.5
Petrogale xanthopus	Yellow-footed rock wallaby	0	2	8.5	12	-	25	20.8	29.2
Phascogale tapoatafa	Brush-tailed phascogale	2	2	0.193	5	0.694	75	70.8	79.2
Phascolarctos cinereus	Koala	26	154	6.529	20	5.744	17.3	16.7	17.9
Planigale maculata	Common planigale	4	23	0.012	4	0.067	20	18.4	21.6
Potorous tridactylus	Long-nosed potoroo	0	1	1.055	12	2.556	33.3	28	38.7
Pseudantechinus macdonnellensis	Fat-tailed antechinus	7	27	0.037	3	0.252	27.6	26	29.2
Sarcophilus harrisii	Tasmanian devil	70	151	8.202	8.2	8.664	46.4	45.6	47.2
Sminthopsis crassicaudata	Fat-tailed dunnart	3	12	0.016	4.3	0.121	28.6	26.2	30.9
Vombatus ursinus	Common wombat	2	38	26	26.1	-	7.5	6.68	8.32

Table 4: Cancer incidence for 36 marsupial and 1 monotreme species (orange). 27 species met the inclusion criteria for analysis (> 15 necropsies). Of note, tumours were observed in 46.4% of examined Tasmanian devils (*Sarcophilus harrisii*; CI: 45.6-47.2), this without including any entries for individuals diagnosed with the transmissible cancer devil facial tumour disease (DFTD). More broadly, as frequently noted in the literature, other species in the dasyrid family (Dasyuridae) to which it belongs likewise exhibit high incidence of cancer for their size, such as the Eastern quoll (*Dasyurus viverrinus*; 34.9%) and Northern quoll (*Dasyurus geoffroii*; 35.1%).

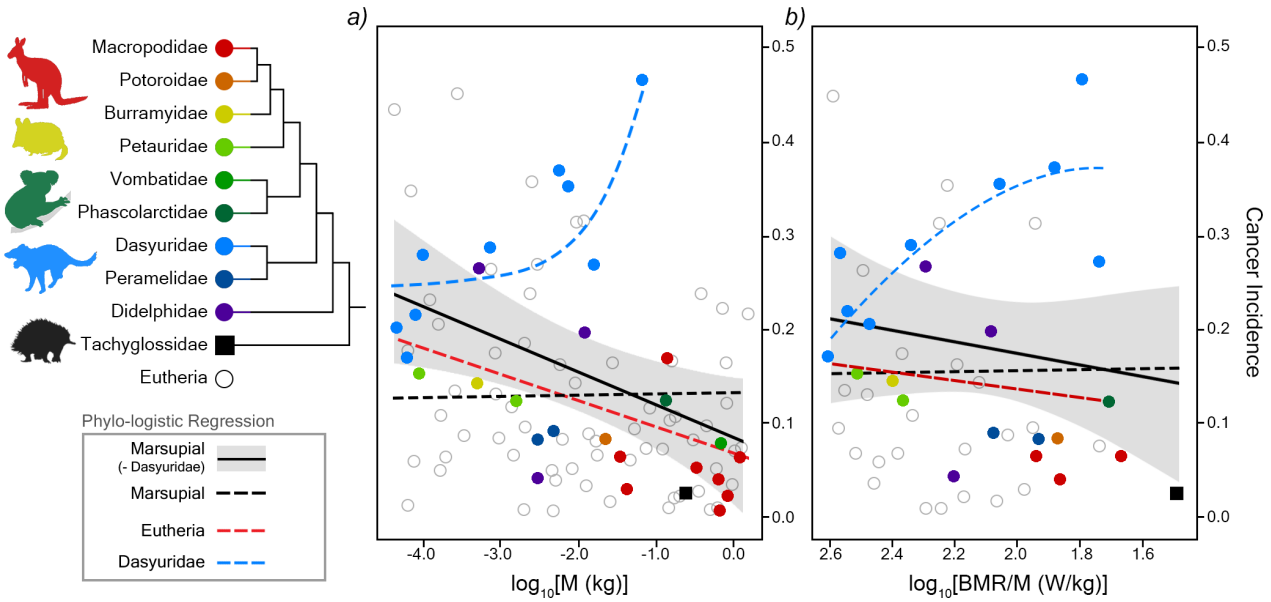


Fig. 4: Cancer prevalence in relation to (a) body mass and (b) mass-specific basal metabolic rate is significantly higher in marsupials (black line, with Dasyuridae removed) relative to incidence rates in eutherian mammals of comparative size (red dashed line), as compared by Chow test ($p<0.05$, F: 3.30), controlling for phylogenetic bias. Dasyurid marsupials have particularly high rates of cancer compared to other marsupial taxa, with increasing body mass (a), at least in this model, a significant predictor of intra-clade cancer incidence (blue dashed line; p¡0.01, eq. $R^2 = 0.003$), with strong phylogenetic bias ($\lambda \sim 1$) - this without any entry of transmissible cancers in Tasmanian devils (*Sarcophilus harrissii*).

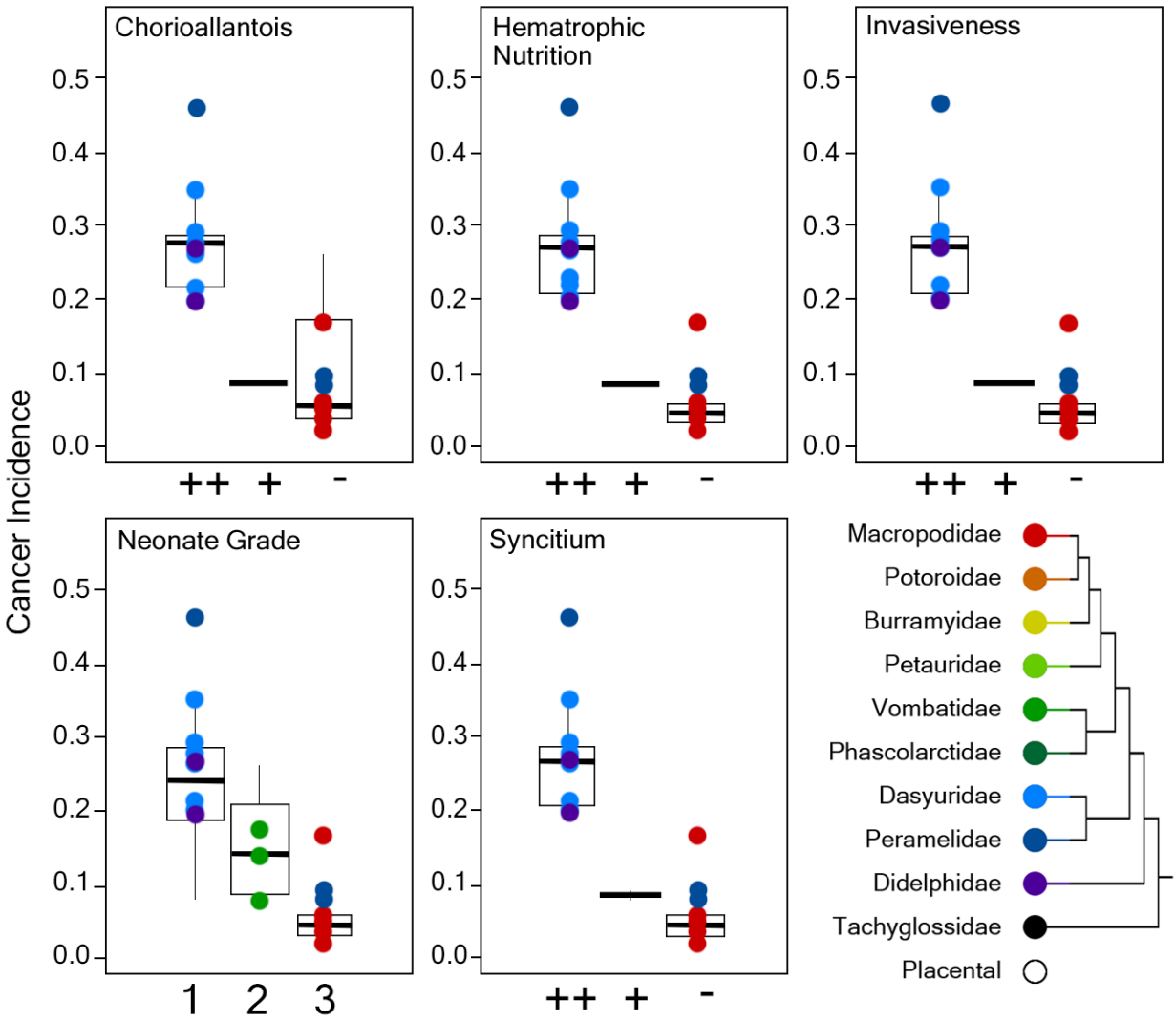


Fig. 5: Cancer prevalence observed in marsupial taxa by major characteristics of placental and neonate development. Boxplots represent presence of (a) chorioallantois (chorioallantoic membrane, if only transitory), (b) evidence of hematrophic nutrition, (c) degree of placental invasiveness (i.e. evidence of an endotheliochorial placenta), and (e) the presence of syncytium (see Chapter 2: 4.2 & Fig.1 for more information), in 19 species. Columns refer to relative similarity to basal eutherian condition, with + representing close to the condition, ++ representing much closer, and - representing a trait further from the condition present in eutherians (as given by Smith, 2015). Neonate grade (d) represents the altricial precocial spectrum for 22 species, as shown in Fig. 1a (Chapter 2); 1 representing the most altricial neonates, and 3 the most developed. For all metrics, families exhibiting condition most similar to eutherian mammals have the highest rates of cancer incidence (Dasyuridae and Didelphidae).

3.3.2. Afrotheria

Afrotheria is a diverse superorder, here represented by 13 species with body mass spanning 7 orders of magnitude, from the short-eared elephant shrew (*Macroscelides proboscideus*), weighing 52g, to the largest terrestrial animal, the African bush elephant (*Loxodonta africana*), weighing an average of 4,800kg and living up to 65 years. 71 tumours were observed in 882 examined individuals, from 13 afrotherian species.

With exception to tenrecs, for which pathology reports are common, published records on neoplasia in Afrotheria is rare. Neoplasia is not common in aardvark (Tubulidentata) nor hyraxes (Hyracoidea) and, with exception to benign viral papillomas recorded in captive manatees, is likewise rarely recorded in Sirenia, and not once in dugongs (Agnew *et al*., 2018; Woods *et al*., 2008; Bossart *et al*., 2001). Published reports of neoplasia in elephants (Proboscidea) are likewise few, limited to single case reports. This lack of published data in the past has led many to infer a lower-than-expected rate of neoplasia in elephants, an observation supporting subsequent study into resistance mechanisms (Landolfi & Terrell, 2018), including this one. Abegglen *et al*. (2015) calculated an age-adjusted estimate of cancer incidence at ~3.11% (CI: 1.74-4.47%), based on 644 annotated elephant deaths sourced from the Elephant Encyclopaedia database, although as highlighted in response by Pessier *et al*. (2016), said database was unreliable, and inclusion criteria were poorly defined. Pessier *et al*. estimated a lifetime prevalence of 33% based on 12 necropsies undertaken at San Diego Zoo Global (between 1987-2015), suggesting elephants had an age-adjusted incidence like that of humans.

Following this review of 360 elephant necropsies, lifetime cancer incidence rates of 1.86% and 0.67% were observed for Asiatic (*Elephas maximus*) and African bush elephants (*Loxodonta africana*) respectively, comparable to the estimate calculated by Abegglen *et al*. (2015), thereby supporting the proposition that large, long-lived species do exhibit low incidence rates. Cancer prevalence is particularly high within tenrec species; lesser hedgehog tenrecs (*Echinops telfairi*) at 26.7% (CI: 25.8-27.5), and greater hedgehog tenrecs (*Setifer setosus*) at 22.2% (CI: 20.7-23.8). Of note, these rates are similar to those observed in their ecological-niche (insectivore) analogues, Eulipotyphla (i.e. hedgehogs; *Hemiechinus auritus* at ~35%).

Across Afrotheria, a strong negative relationship was observed between cancer incidence and all three primary life history traits (see Fig.6). *LIF* gene family expansion at the base of Paenungulata (see Chapter 2: 2.2) had no influence on outcome, nor did influence of additional life history traits.

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Dugong dugon	Dugong	0	16	295	70	58.5	5.56	4.5	6.61
Echinops telfairi	Lesser hedgehog tenrec	27	103	0.15	13	0.75	26.7	25.8	27.5
Elephantulus rufescens	Rufous elephant shrew	12	94	0.05	6	5.98	13.5	12.9	14.2
Elephas maximus	Asian elephant	3	213	3,178	86.5	2,336.5	1.86	1.68	2.04
Hemicentetes semispinosus	Streaked tenrec	0	3	0.13	2.7	0.38	20	16.5	23.5
Loxodonta africana	African bush elephant	0	147	4,800	65	-	0.67	0.54	0.802
Macroscelides proboscideus	Short-eared elephant shrew	7	37	0.04	8.7	0.29	20.5	19.2	21.8
Orycteropus afer	Aardvark	0	34	56.18	24	34.28	2.78	2.24	3.31
Petrodromus tetradactylus	Four-toed elephant shrew	0	4	0.2	3	0.85	16.7	13.7	19.6
Procavia capensis	Rock hyrax	12	137	2.95	14	4.95	9.35	8.87	9.84
Rhynchocyon petersi	Black and rufous elephant shrew	0	2	0.42	3	-	25	20.8	29.2
Setifer setosus	Greater hedgehog tenrec	5	25	0.26	10.5	0.57	22.2	20.7	23.8
Trichechus manatus	West Indian manatee	5	67	467.32	30	70.01	8.7	8.03	9.36

Table 5: Cancer incidence for 13 afrotherian species, 10 of which met inclusion for analysis.

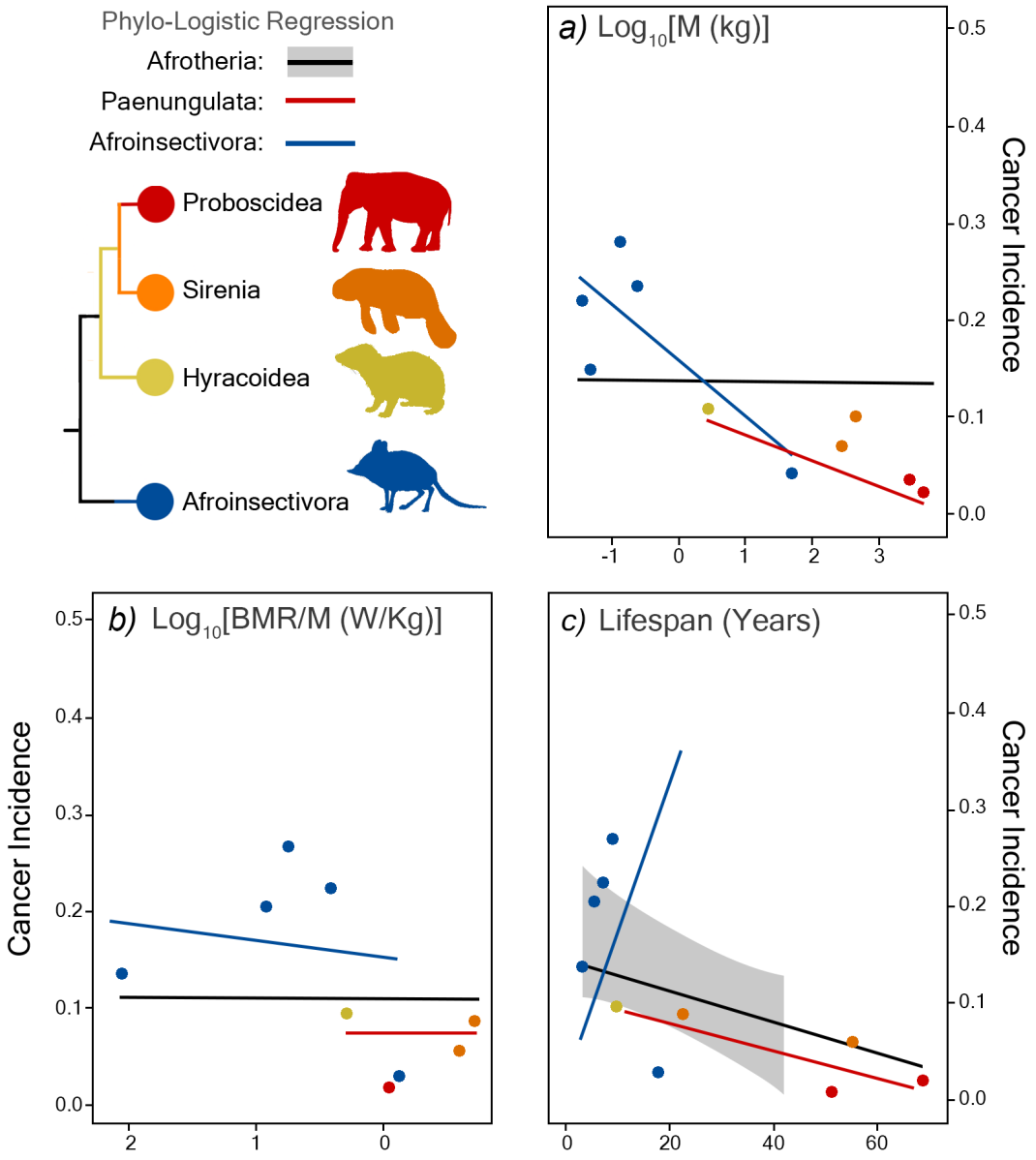


Fig. 6: Cancer prevalence is significantly negatively predicted by (a) body mass (p=0.009, prd. R^2 = 0.685), (b) mass-specific basal metabolic rate (msBMR; p=0.02, prd. R^2=0.685), and (c) lifespan (p = 0.012, prd. R^2 = 0.697) in 10 afrotherian species (9 for msBMR), as shown by phylogenetic logistic regression. There is no significant difference in incidence relationships between Paenungulata (having undergone *LIF* expansion) and Afroinsectivora (which hasn't; see Chapter 2: 2.2, and Chapter 5: 1.2), determined via Chow test (p>0.05). Analysis undertaken without log-transformation; plots (except (c) lifespan) log-transformed along x-axis for visualisation purposes only.

3.3.3. Rodentia

Rodentia is the most diverse group in terms of species (about 40% mammals; 2,277 species) and morphological disparity, particularly in body size. Here represented by 42 species across 18 families, body mass spans four orders of magnitude, from the African pygmy mouse (*Mus minutoides*), weighing only 6.4g, to the capybara (*Hydrochoerus hydrochaeris*), weighing in at ~48kg. 1,177 tumours were observed in 13,409 necropsies.

Neoplasms are a common occurrence in inbred domesticated mice and rat lines, prevalence higher in some strains than others, often by design for research purposes. This high prevalence, though artificially inflated in domesticated strains, is nevertheless likewise frequently observed in wild rodents, disproportionately more so than in other mammalian orders. Notable exceptions include several long-lived subterranean species, such as naked mole-rats, blind mole-rats and bamboo rats, whose adaptations to hypoxic environments are thought to confer cancer and other age-associated disease resistance, and are frequently used as animal models in carcinogenesis research. Rare cases of neoplasia have been reported in naked mole-rats, though there are no recorded tumours found in Spalacidae (unrelated mole-rats and bamboo rats), unless artificially induced. Similarly, only three instances of cancer have been reported to date in capybara (*Hydrochoerus hydrochaeris*), the world's largest rodent (Herrera-Álvarez *et al*., 2020).

Analysis of this dataset reveals no significant relationship between cancer incidence and life history traits, except in the instance of lifespan ($p<0.01$, see Fig. 7) with inclusion of capybara (without, $p = 0.14$), where cancer incidence scales negatively with increasing lifespan. High variance in recorded incidence across many rodent families prevented high-scoring model-fitting, further relationships thereby potentially obscured by non-independence bias and statistical noise.

Table 6 (overleaf): Cancer prevalence in 42 rodent species, 29 of which met inclusion criteria for analysis. Domesticated species are noted accordingly. In the case of brown rats (*Rattus norvegicus*), data was split to differentiate between domesticated lines ('fancy rats'; *Rattus norvegicus domestica*) and wild individuals. Living approximately ~15 months, California voles (*Microtus californicus*) have the highest recorded incidence of cancer in the entire dataset, at ~45% (CI: 44.4-45.7%), bar the Tasmanian devil (*Sarcophilus harrissii*.

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Acomys cilicicus	Turkish spiny mouse	0	14	0.04	4	-	6.25	5.06	7.44
Arvicola amphibius	European water vole	119	454	0.12	5	0.613	26.3	25.9	26.7
Castor fiber	Eurasian beaver	0	17	19	25	-	5.26	4.26	6.27
Cavia porcellus	(Domesticated) Guinea pig	216	1316	0.73	14.8	2.13	16.5	16.3	16.7
Chinchilla lanigera	Long-tailed chinchilla	4	698	0.48	11.3	1.31	0.71	0.652	0.777
Coendou prehensilis	Prehensile-tailed porcupine	1	37	4.12	17.3	5.123	5.13	4.44	5.82
Cricetomys gambianus	Gambian rat	2	10	1.27	8.4	6.024	25	22.6	27.4
Cricetus cricetus	Common hamster	0	3	0.43	4	1.251	20	16.5	23.5
Cuniculus paca	Paca	1	17	9	16.3	15.323	10.5	9.15	11.9
Cynomys ludovicianus	Black-tailed prarie dog	54	173	0.8	8.5	2.358	31.4	30.7	32.1
Dasyprocta azarae	Azara's agouti	0	100	2.98	11.8	10.521	0.98	0.789	1.17
Dinomys branickii	Pacarana	4	19	12.5	9.4	-	23.8	22	25.6
Dipodomys merriami	Merriam's kangaroo rat	1	8	0.04	2	0.246	20	17.5	22.5
Dolichotis patagonum	Patagonian mara	2	92	8	14.4	-	3.19	2.84	3.55
Ellobius talpinus	Northern mole vole	5	123	0.04	5	-	4.8	4.43	5.17
Glaucomys volans	Southern flying squirrel	0	2	0.07	12	0.414	25	20.8	29.2
Graphiurus murinus	Woodland dormouse	6	61	0.02	5.8	0.225	11.1	10.3	11.9
Heterocephalus glaber	Naked mole rat	6	62	0.04	10	0.128	10.9	10.2	11.7
Hydrochoerus hydrochaeris	Capybara	3	222	48.14	12	36.75	1.79	1.61	1.96
Hypogeomys antimena	Malagasay jumping rat	0	11	1.18	12.6	-	7.69	6.24	9.14
Hystrix cristata	Crested porcupine	1	77	13.41	28	-	2.53	2.19	2.88
Lagostomus maximus	Plains viscacha	0	6	4.66	9.4	10.623	12.5	10.2	14.8
Lagurus lagurus	Steppe lemming	60	174	0.02	3.8	-	34.7	34	35.4
Marmota monax	Woodchuck	1	15	3.88	10	3.696	11.8	10.2	13.3
Mesocricetus cricetus	Golden hamster	361	8,207	0.11	3.5	0.69	4.41	4.37	4.45
Micromys minutus	Old World harvest mouse	9	105	0.01	5	0.201	9.35	8.79	9.9
Microtus californicus	California vole	108	240	0.06	1.2	0.367	45	44.4	45.7
Mus minutoides	African pygmy mouse	15	101	0.01	3.1	0.129	15.5	14.8	16.2
Mus musculus	House mouse	9	23	0.02	6	-	40	38.1	41.9
Myocastor coypus	Coypu	6	40	6.36	12	16.88	16.7	15.5	17.8
Notomys alexis	Spinifex hopping mouse	1	1	0.03	5.2	0.252	66.7	61.3	72
Pedetes capensis	Springhare	1	27	2.55	14.5	4.427	6.9	5.97	7.82
Peromyscus maniculatus	Deer mouse	1	24	0.03	8.3	0.219	7.69	6.67	8.72
Peromyscus polionotus	Oldfield mouse	44	115	0.01	5.5	0.12	38.5	37.6	39.3
Phloeomys pallidus	Slender-tailed cloud rat	9	123	1.74	10.7	-	8	7.52	8.48
Phodopus sungorus	Siberian dwarf hamster	16	72	0.03	3.2	0.313	23	22	23.9
Rattus norvegicus	(Wild) Brown rat	17	74	0.28	3.8	-	23.7	22.7	24.6
Rattus norvegicus	(Domesticated) Brown rat	89	463	0.28	3.8	1.404	19.4	19	19.7
Rattus rattus	Black rat	0	10	0.14	4.2	0.77	8.33	6.77	9.9
Spermophilus citellus	European souslik	0	2	0.4	6.7	-	25	20.8	29.2
Spermophilus richardsonii	Richardson's ground squirrel	1	6	0.33	3	0.788	25	22	28
Tamiasciurus hudsonicus	Red squirrel	4	65	0.2	12	1.615	7.46	6.83	8.09

3.3.4. Ungulata (Perissodactyla and non-Cetacean Artiodactyla)

With few exception (i.e. oncogenic fibroma viruses prevalent in some deer species), neoplasia is uncommonly reported in ungulates, even in domesticated species with considerable veterinary and industrial attention. This dataset analysis supports neoplastic rarity in this group (mean incidence: 3.9-4.5%; median: 2-2.3%), with incidence scaling invariantly with all life history traits (see Fig. 8). Species with higher rates of cancer here typically represent deer species subject to high oncogenic delta-papillomavirus morbidity, or endangered species whose current captive populations are descended from an extremely small number of founder individuals (e.g. critically-endangered addra gazelle, and Père David's deer currently classified as extinct in the wild).

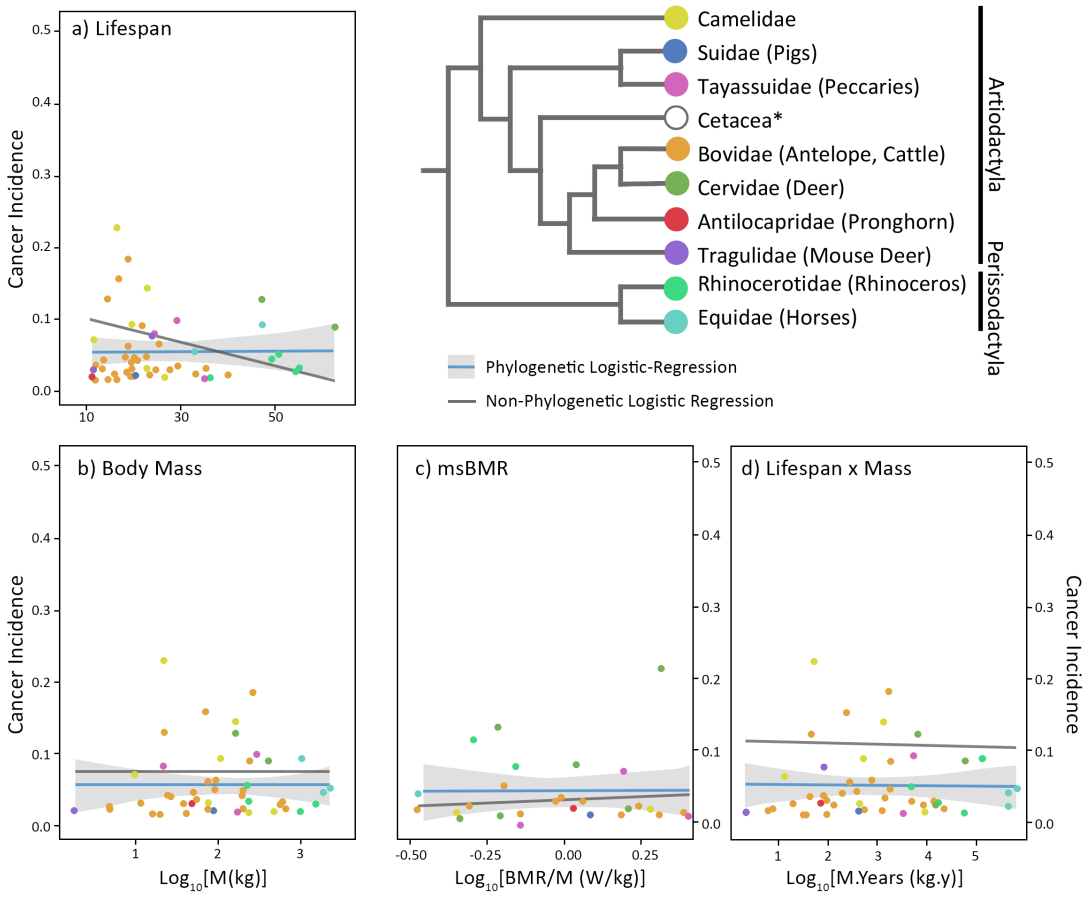


Fig. 8: Following phylogenetic regression, cancer incidence in 52 ungulate species (28 only with values for BMR) scales invariantly with (a) lifespan (p=0.018, prd. R^2 = 0.302), (b) body mass (p=0.286, prd. R^2 = 0.317), (c) mass-specific basal metabolic rate (p=0.561, prd. R^2 = 0.226), and (d) lifespan x mass (p=0.21, prd. R^2 = 0.291). Analysis undertaken without log transformation.

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Addax nasomaculatus	Addax	1	36	96.08	25.67	-	5.26	4.55	5.97
Alces alces	Moose	0	114	461.901	27	286.847	0.86	0.694	1.03
Antilocapra americana	Pronghorn	2	153	47.45	12	50.973	1.94	1.72	2.15
Bison bison	American bison	3	316	624.577	33	-	1.26	1.14	1.38
Bos taurus	Domesticated cow*	98	4,407	618.642	20	306.77	2.25	2.2	2.29
Camelus bactrianus	Bactrian camel*	4	245	475	35.4	-	2.02	1.85	2.2
Camelus dromedarius	Dromedary camel*	0	90	492.714	40	224.779	1.09	0.875	1.3
Capra hircus	Domesticated goat*	47	1,344	47.386	20.8	46.414	3.57	3.47	3.67
Capreolus capreolus	Roe deer	33	152	22.502	17	46.347	22.1	21.4	22.7
Cervus elaphus	Red deer	0	189	240.867	26.8	112.43	0.52	0.421	0.626
Choeropsis liberiensis	Pygmy hippopotamus	8	415	230	55	-	2.16	2.02	2.3
Connochaetes taurinus	Blue wildebeest	4	166	198.62	21.5	230.073	2.98	2.72	3.23
Elaphurus davidianus	Père David's deer	30	230	165.989	23.3	102.43	13.4	12.9	13.8
Eudorcas thomsonii	Thompson's gazelle	12	107	22.907	15.17	-	11.9	11.3	12.5
Gazella dorcas	Dorcas gazelle	0	283	15.636	17.42	-	0.35	0.282	0.42
Gazella leptoceros	Sand gazelle	0	31	24.646	14	-	3.03	2.45	3.62
Gazella spekei	Speke's gazelle	0	241	20	12.67	-	0.41	0.331	0.492
Gazella subgutturosa	Saudi goitered gazelle	0	32	26.981	20	25.65833	2.94	2.37	3.51
Giraffa camelopardalis	Reticulated giraffe	4	753	964.655	36.3	-	0.66	0.604	0.72
Hippopotamus amphibius	Hippopotamus	2	178	1,536.31	54.5	-	1.67	1.48	1.85
Hippotragus niger	Sable antelope	6	86	236.406	22.25	-	7.95	7.39	8.52
Kobus ellipsiprymnus	Waterbuck	2	303	204.393	19.92	148.949	0.98	0.873	1.09
Lama glama	Llama*	3	249	78.323	28.3	148.94	1.59	1.44	1.75
Madoqua kirkii	Kirk's dik-dik	1	165	4.833	16.5	11.966	1.2	1.03	1.36
Nanger dama	Addra gazelle	12	75	71.425	17.33	-	14.8	14	15.5
Nanger granti	Grant's gazelle	1	80	55.464	12.67	-	2.44	2.11	2.77
Nanger soemmerringii	Soemmerring's gazelle	0	267	41.583	15.5	-	0.37	0.299	0.444
Odocoileus virginianus	White-tailed deer	6	347	75.901	23	123.447	2.01	1.86	2.15
Okapia johnstoni	Okapi	0	22	230.001	33	-	25	20.8	29.2
Oreamnos americanus	North American mountain goat	2	56	72.105	19.17	46.414	5.17	4.6	5.74
Oryx dammah	Scimitar-horned oryx	3	423	199.799	20.42	-	0.94	0.849	1.03
Ovis aries	Domesticated sheep*	9	615	39.098	19.17	-	1.62	1.52	1.72
Ovis canadensis	Bighorn sheep	0	93	74.645	24	114.674	1.05	0.847	1.26
Pecari tajacu	Collared peccary	15	220	21.134	24.7	33.165	7.21	6.87	7.55
Philantomba monticola	Blue duiker	0	106	4.896	12	10.075	0.93	0.745	1.11
Pudu puda	Southern pudu	0	15	9.643	12.5	-	5.88	4.76	7
Rangifer tarandus	Reindeer*	0	210	109.089	20.2	119.66	8.33	6.77	9.9
Raphicerus campestris	Steenbok	0	51	11.662	14	20.619	1.89	1.52	2.25

Cont. overleaf

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Sus scrofa	Domesticated pig*	3	450	84.472	21	104.15	0.89	0.799	0.971
Syncerus caffer	Cape buffalo	2	133	592.666	29.5	-	2.22	1.97	2.47
Taurotragus oryx	Eland	6	430	562.593	25	190.209	1.62	1.5	1.74
Tragelaphus eurycerus	Bongo	20	116	270.998	19.42	-	17.8	17.1	18.5
Tragelaphus imberbis	Lesser kudu	5	150	94.32	18.92	-	3.95	3.64	4.26
Tragelaphus strepsiceros	Greater kudu	4	131	206.056	23	-	3.76	3.44	4.08
Tragulus javanicus	Lesser malay mouse deer	0	124	1.89	12	4.883	0.79	0.639	0.949
Ceratotherium simum	White rhinoceros	6	177	2,285.939	50	774.7228	3.91	3.63	4.19
Dicerorhinus sumatrensis	Sumatran rhinoceros	0	4	1,046.156	35	-	16.7	13.7	19.6
Diceros bicornis	Black rhinoceros	3	46	995.941	47	-	8.33	7.55	9.12
Equus asinus	Domesticated donkey*	171	1,444	164.998	47	85.6	11.9	11.7	12.1
Equus caballus	Domesticated horse*	65	830	403.599	62	283.5114	7.93	7.75	8.12
Rhinoceros unicornis	Greater one-horned rhinoceros	1	55	1,843.656	49	-	3.51	3.03	3.99
Tapirus bairdii	Baird's tapir	11	133	293.782	29.6	-	8.89	8.41	9.37
Tapirus terrestris	Brazilian tapir	1	306	169.497	35	-	0.65	0.56	0.739

* Indicates domesticated species

Table 7: Cancer prevalence in 53 ungulate species; 8 odd-toed (Perissodactyla; bottom/orange), 45 even-toed (Artiodactyla), all but one - the rare Sumatran rhinoceros (*Dicerorhinus sumatrensis*) - meeting criteria for analysis. Reliably measuring basal metabolic rate (BMR; calculated from inactive, thermoneutral, post-absorptive adults) is problematic in ungulate species, given near-persistent microbial fermentation in their guts delays or outright prohibits entrance into a post-absorptive state, a condition required for accurate basal measurement. As such, BMR values are unavailable for most species.

3.3.5. Cetacea

Recorded incidence of neoplasia in cetaceans, outside captivity and the St. Lawrence Estuary case-study, are rare. Excluding neoplasia found in bottlenose dolphin, orca and beluga whale, publicly published records of tumours in free-living cetacean species totals only 88 individual cases (see SI.3.1).

Included in this dataset are 108 confirmed tumours, from 8,096 individual necropsy reports, reaffirming the low incidence of observed neoplasia in cetaceans (mean incidence: ~3.1.3.7%; median: 2.5-3%). No relationship was observed between cancer incidence included in this dataset and any life history trait (see Fig. 9); cancer prevalence scales invariantly with both body size ($p = 0.379$) and longevity ($p = 0.19$). In addition, no relationship between cancer-incidence and mass-specific metabolic rate was found, neither across 5 species with accurately measured BMR ($p = 0.49$), nor 20 species with estimated BMR values ($p = 0.64$). Cetaceans as a whole did present a qualitative lower rate of incidence relative to other ungulates (both Artiodactyla and Perissodactyla), though this was not significant following pANCOVA analysis.

Table 8 (overleaf): Cancer prevalence in 41 cetacean species, 29 meeting criteria for analysis (excluding <15 necropsies, St. Lawrence Estuary beluga population, and the humpback whale, *Megaptera novaeangliae*, with biased pathological sampling). Reliable data on longevity in most cetacean species is unavailable (see Methods 2.2) and should be interpreted with caution. Basal metabolic rate (BMR) data in cetaceans is likewise scarce, and has only been reliably measured in captive animals, incidated by ‡. All other BMR values are based upon metabo-ecological calculated estimates provided in the literature (See SI.3.1). Beluga whale entries here have been split to differentiate between the St. Lawrence estuary population known to be subject to heavy industrial contamination, and beluga whales from other, globally distributed populations.

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Balaena mysticetus	Bowhead whale	1	52	79,691.18	240	8847.6	3.7	3.2	4.21
Balaenoptera acutorostrata	Common minke whale	0	250	5,587.094	50	2,197.21	0.4	0.319	0.474
Balaenoptera bonaerensis	Antarctic minke whale	0	5	5,600	60	-	14.3	11.7	16.9
Balaenoptera borealis	Sei whale	2	160	22,106.25	74	3,994.48	1.85	1.64	2.06
Balaenoptera musculus	Blue whale	3	119	136,000	110	25,675.99	3.31	2.99	3.62
Balaenoptera physalus	Fin whale	12	149	47,506.01	116	6,878.04	8.61	8.16	9.06
Delphinapterus leucas	Beluga (St. Lawrence Estuary)*	67	524	1,381.641	40	**845.48** ‡	12.9	12.6	13.2
Delphinapterus leucas	Beluga (Other) †	4	55	1,381.641	40	**845.48** ‡	8.77	8.04	9.51
Delphinus capensis	Common dolphin	0	80	125	40	-	1.22	0.982	1.46
Delphinus delphis	Short-beaked common dolphin	0	749	79.272	20	-	0.13	0.107	0.159
Eschrichtius robustus	Gray whale	1	864	27,324.02	77	5,122.31	0.23	0.199	0.263
Feresa attenuata	Pygmy killer whale	0	18	170	23	-	5	4.04	5.96
Globicephala macrorhynchus	Short-finned pilot whale	2	221	726	63	412.97	1.35	1.19	1.5
Globicephala melaena	Long-finned pilot whale	3	277	800	59	-	1.43	1.29	1.57
Grampus griseus	Risso's dolphin	0	81	387.5	20	287.21	1.2	0.97	1.44
Hyperoodon ampullatus	Northern bottlenose whale	0	27	3,393.361	37	970.33	3.45	2.78	4.11
Inia geoffrensis	Amazon river dolphin	2	107	121.431	30	166.54	2.7	2.4	3
Kogia breviceps	Pygmy sperm whale	1	31	431.5	17	294.19	6.06	5.25	6.87
Kogia sima	Dwarf sperm whale	0	2	183.067	22	-	25	20.8	29.2
Lagenodelphis hosei	Fraser's dolphin	0	1	164	18	-	33.3	28	38.7
Lagenorhynchus acutus	Atlantic white sided dolphin	8	216	186.518	27	205.49	4.13	3.86	4.39
Lagenorhynchus albirostris	White-beaked dolphin	0	170	186.635	40	-	0.58	0.468	0.695
Lagenorhynchus obliquidens	Pacific white sided dolphin	9	367	100	40	-	2.71	2.54	2.88
Lagenorhynchus obscurus	Dusky dolphin	2	2	127.5	25	-	75	70.8	79.2
Lipotes vexillifer	Baiji	0	6	112.138	24	-	12.5	10.2	14.8
Megaptera novaeangliae	Humpback whale	3	16	30,000	95	5,631.31	22.2	20.3	24.1
Mesoplodon bidens	Sowerby's beaked whale	0	215	3,400	84	-	0.46	0.371	0.551
Mesoplodon densirostris	Blainsvilles beaked whale	1	7	2,300	27	777.41	22.2	19.5	24.9
Mesoplodon europaeus	Gervais' beaked whale	0	7	5,600	48	-	11.1	9.06	13.2
Mesoplodon mirus	True's beaked whale	0	1	2,100	49	-	33.3	28	38.7
Monodon monoceros	Narwhal	1	14	938.126	40	573.81	12.5	10.9	14.1
Neophocaena phocaenoides	Finless porpoise	1	26	32.5	23	**90.24** ‡	7.14	6.19	8.1
Orcinus orca	Killer whale	2	85	5,628.759	100	**2,274.9**	3.45	3.06	3.83
Phocoena phocoena	Harbor porpoise	18	2,856	52.731	15	**107.67** ‡	0.67	0.635	0.695
Physeter macrocephalus	Sperm whale	7	165	28,500	77	-	4.79	4.47	5.11
Pseudorca crassidens	False killer whale	0	22	1,360	22	680.06	4.17	3.37	4.97
Stenella coeruleoalba	Striped dolphin	0	244	142.103	50	179.73	0.41	0.327	0.486
Stenella frontalis	Atlantic spotted dolphin	1	35	110	40	-	5.41	4.68	6.13
Steno bredanensis	Rough-toothed dolphin	1	6	130	32	172.13	25	22	28
Tursiops truncatus	Bottlenose dolphin	23	320	281.041	46	**604.15**‡	7.45	7.17	7.74
Ziphius cavirostris	Cuvier's beaked whale	0	68	4,775	36	1,182.96	1.43	1.15	1.71

* Beluga population under high oncogenic pollutant exposure; † Globally distributed beluga sample; ‡ Experimentally validated BMR values

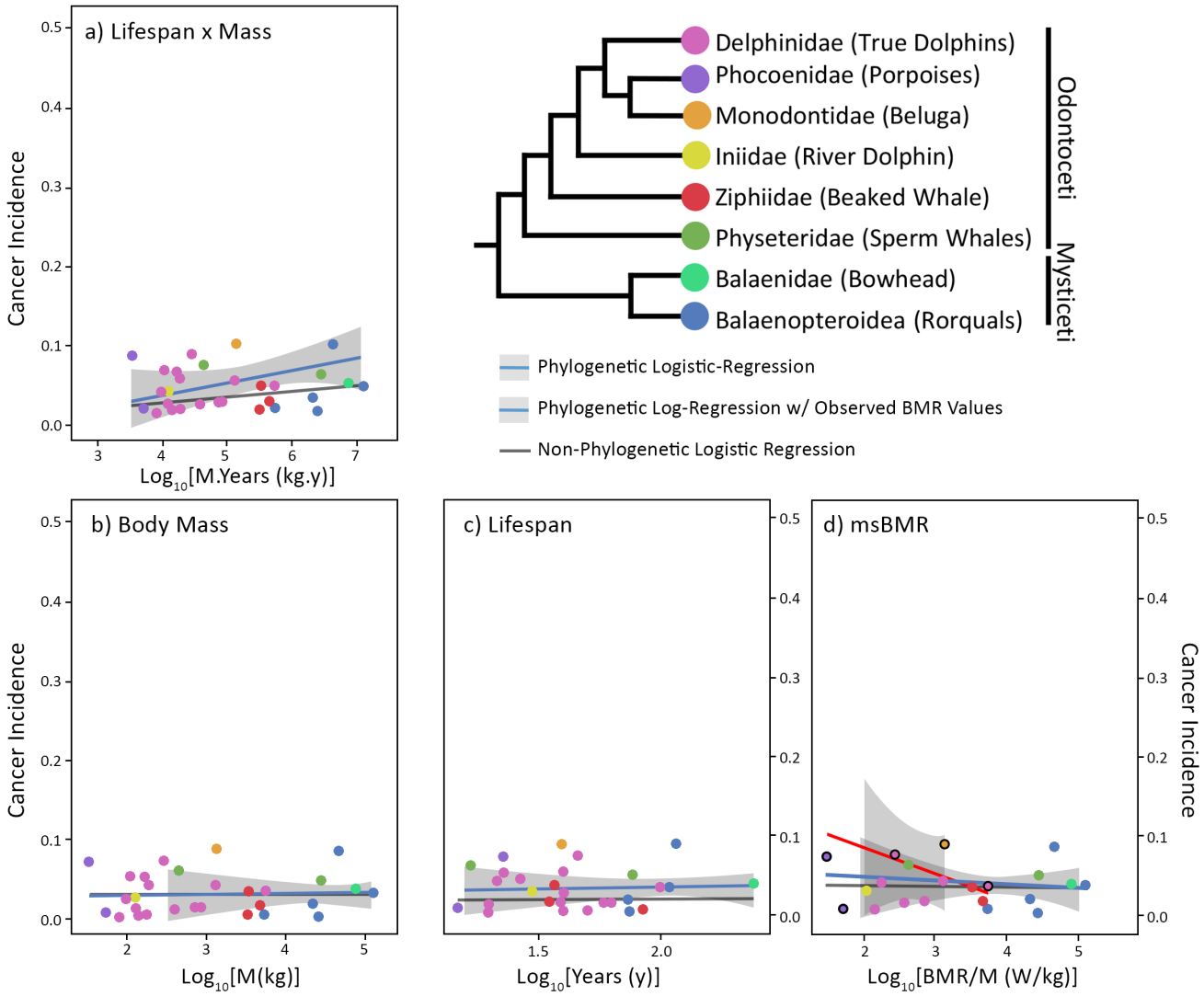


Fig. 9: Cancer scales invariantly with (a) lifespan times mass (p=0.2); (b) body size (p=0.379, prd. R^2 = 0.404); (c) lifespan (p=0.19, prd. R^2 = 0.361); and (d) mass-specific basal metabolic rate (p=0.647, prd. R^2 = 0.0.317) in 29 cetacean species. For msBMR (d), BMR values were extrapolated from available data and established scaling laws (See Methods 2.2.1); analysis using measured BMR values for 5 species (red line; highlighted points) similarly indicate no significant relationship (p=0.495, prd. R^2 = 0.44).

3.3.6. Chiroptera

Chiroptera represents the second most diverse group of mammals, the bats (at 1,200+ described species), notable for being the only mammal taxon capable of true flight. Given the high metabolic demands of flight, as with rodents, one would expect cancer prevalence to be high. In contrast, reports of neoplasia in bats are rare. As such, as with cetaceans, this seems paradoxical - not only due to their high metabolic requirements, but also due to their longevity. On average, the recorded lifespans of bats are approx. 3.5 times greater than non-flying mammals of similar size, and records of individuals living more than 30 years in the wild exist for at least five species (including Brandt's bat, at only 4.5g; in 2005, one individual was recorded at 41 years of age). Bats further present a suitable clade for study into Peto's paradox given the breadth of body size – ranging from the giant golden-crowned flying fox (*Acerodon jubatus*) at 1.4kg, to the smallest mammal in the world, the bumblebee bat (*Craseonycteris thonglongyai*) at ~1.9g. Any potential underlying tumour suppressive mechanisms remain poorly understood, though in 2019, Koh *et al.* determined bat cells are more resistant to DNA damage induced by genotoxic compounds compared to humans, owing to increased expression of *ABCB1*, a gene involved in genotoxic compound efflux.

Only 28 neoplasias were recorded in 1,209 bat entries (~2.31%) (see Table 9), providing continued support to the notion cancer in bats is rare. No significant relationship was modelled between cancer incidence and any life history trait (see Fig.10)

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Antrozous pallidus	Pallid bat	1	33	0.022	9.1	0.104	5.71	4.95	6.48
Artibeus jamaicensis	Jamaican fruit-eating bat	0	86	0.044	10	0.359	1.14	0.915	1.36
Carollia perspicillata	Seba's short-tailed bat	1	169	0.019	12.4	0.24	1.17	1.01	1.33
Cynopterus brachyotis	Lesser short-nosed fruit bat	0	28	0.034	8	0.262	3.33	2.69	3.98
Desmodus rotundus	Vampire bat	5	186	0.033	19.5	0.194	3.19	2.94	3.44
Eptesicus fuscus	Big brown bat	1	15	0.017	20	0.113	11.8	10.2	13.3
Glossophaga soricina	Pallas's long-tongued bat	1	86	0.01	10	0.164	2.25	1.94	2.56
Macrotus californicus	California leaf-nosed bat	0	5	0.012	10.4	0.082	14.3	11.7	16.9
Myotis lucifugus	Little brown bat	0	38	0.008	34	0.051	2.5	2.02	2.98
Phyllostomus discolor	Pale spear-nosed bat	0	25	0.037	9	0.267	3.7	2.99	4.42
Phyllostomus hastatus	Greater spear-nosed bat	0	25	0.091	10	0.559	3.7	2.99	4.42
Plecotus auritus	Long eared bat	1	6	0.008	30	-	25	22	28
Pteropus giganteus	Indian flying fox	2	85	0.825	31.4	1.622	3.45	3.06	3.83
Pteropus hypomelanus	Variable flying fox	4	42	0.436	20.3	1.618	11.4	10.4	12.3
Pteropus livingstonii	Livingstone's fruit bat	5	116	0.734	15	-	5.08	4.69	5.48
Pteropus poliocephalus	Gray-headed flying fox	0	34	0.703	23.6	1.768	2.78	2.24	3.31
Pteropus pumilus	Little golden-mantled flying fox	1	22	0.184	17.2	0.705	8.33	7.23	9.44
Pteropus rodricensis	Rodrigues fruit bat	0	135	0.256	30	0.753	0.73	0.587	0.872
Pteropus vampyrus	Large flying fox	0	21	1.028	28.9	4.486	4.35	3.51	5.18
Rousettus aegyptiacus	Egyptian fruit bat	6	52	0.134	22.9	0.684	13	12.1	13.9

Table 9: Cancer prevalence in 20 species of bat, 18 of which met criteria for inclusion in analysis. Bats display demonstrably estimated rates of cancer compared to similar non-flying mammals: for example, despite the large flying fox (*Pteropus vampyrus*) having a basal metabolic rate of 4.486W, three times higher than that of the sunda slow loris (*Nycticebus coucang*) at 1.504W, an animal of comparable mass (0.92kg to the bat's 1.028kg) and lifespan (26.5 years to 28.9 years), it presents an incidence estimate nearly a tenth of the slow loris' tumour incidence rate (flying fox: 3.51-5.18%; slow loris: 30.6-32.4%).

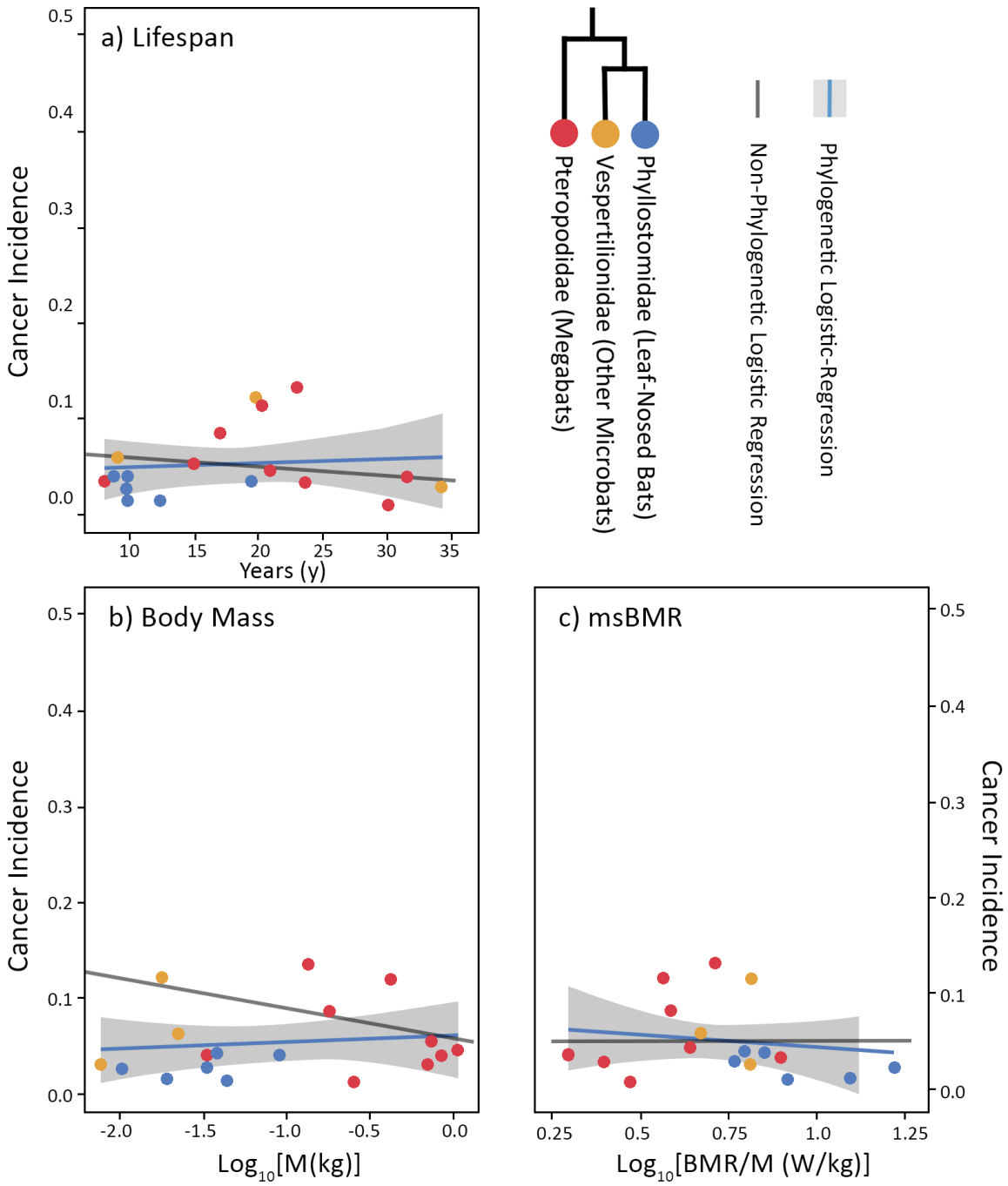


Fig. 10: Phylogenetic logistic regression demonstrates cancer scales invariantly with (a) lifespan (p=0.9623, prd. $R^2 = 0.168$); (b) body mass (p=); and (c) mass-specific basal metabolic rate (p=0.344, prd. $R^2 = 0.105$) in 18 bat species.

3.3.7. Carnivora

With notable exceptions, tumours in non-domestic canids, felids and ursids are represented by small numbers of case reports. Neoplasia is more prevalent in smaller carnivorans, especially mustelids, with wild populations experiencing similar incidence to domesticated varieties.

The highest cancer rates presented are indicative of species recently having undergone a genetic bottleneck (i.e. endangered black-footed ferrets and red wolves; their entire extant populations descended from 18 and 14 individuals respectively) and those subject to environmental carcinogenic pollution (i.e. Californian sealion and Pacific walrus). It is well documented wild pinnipeds suffer high rates of cancer; interestingly, in this study, captive-managed harbour seals displayed the lowest rates of cancer (1 in 130 individuals; ~1.5%); likewise, within the captive subset of Californian sealion entries (380 individuals), no incidence of neoplasia was reported. This despite their larger mass and longer lifespans relative to the rest of the clade. Many pinniped adaptations are associated with the physiological stresses of prolonged diving, including a high tolerance to tissue hypoxia and hypercapnia. As discussed in Chapter 2: 5.1, hypoxia tolerance is associated with known mechanisms of cancer resistance in rodents, and may perhaps result in similar resistance in pinnipeds. As of time of writing, several pinniped genomes are in the process of being assembled and may elucidate insight into potential hypoxia-associated resistance mechanisms.

Across the clade as a whole, no significant relationship was found between incidence and any life history trait (see Fig. 11).

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Acinonyx jubatus	Cheetah	56	506	50.58	19	61.77	11.2	10.9	11.5
Ailuropoda melanoleuca	Giant panda	0	7	117.5	36.8	-	11.1	9.06	13.2
Ailurus fulgens	Red panda	26	366	5.17	14	4.989	7.34	7.07	7.6
Aonyx cinereus	Asian small-clawed otter	8	398	3	23	-	2.25	2.1	2.4
Arctictis binturong	Binturong	9	78	13	22.7	12.747	12.5	11.8	13.2
Arctocephalus australis	South American fur seal	2	28	67.98	21	-	10	8.93	11.1
Canis familiaris	Domesticated dog*	5,300	14,770	40	24	-	35.9	35.8	36
Canis lupus	Mexican gray wolf	14	424	26.63	20.6	35.2	3.52	3.35	3.7
Canis mesomelas	Black-backed jackal	1	44	8.25	14	21.533	4.35	3.76	4.94
Canis rufus	Red wolf	37	175	36.29	15	-	21.5	20.9	22.1
Chrysocyon brachyurus	Maned wolf	17	106	21.5	16.8	-	16.7	16	17.4
Crocuta crocuta	Spotted hyena	9	67	63.37	41.1	-	14.5	13.7	15.3
Enhydra lutris	Sea otter	6	96	27.41	30	98.479	7.14	6.63	7.65
Felis catus	Domesticated cat*	256	22,993	2.88	34	-	1.12	1.1	1.13
Felis chaus	Jungle cat	5	255	7.16	12	-	2.33	2.15	2.52
Felis concolor	Mountain lion	13	260	64	20	-	5.34	5.07	5.62
Felis margarita	Sand cat	3	173	2.82	13.9	-	2.29	2.06	2.51
Felis rufus	Bobcat	1	71	8.6	32.3	23.446	2.74	2.37	3.11
Gulo gulo	Wolverine	0	35	12.79	18	31.765	2.7	2.18	3.23
Helogale parvula	Dwarf mongoose	0	1	0.28	10.9	-	33.3	28	38.7
Hyaena hyaena	Striped hyaena	5	94	35.07	24	31.954	6.25	5.77	6.73
Leopardus pardalis	Ocelot	5	112	11.88	20.3	17.368	5.26	4.85	5.67
Leopardus wiedii	Margay	1	35	3.27	20	5.227	5.41	4.68	6.13
Leptailurus serval	Serval	3	201	13.35	22.4	-	1.97	1.78	2.16
Lontra canadensis	North American river otter	1	1	8.09	25	-	66.7	61.3	72
Lutra lutra	European otter	5	140	8.87	22	25.104	4.23	3.89	4.56
Lycaon pictus	African wild dog	31	194	22	11	33.01	16.3	15.8	16.8
Lynx pardinus	Iberian lynx	4	60	11.05	13	-	8.06	7.39	8.74
Martes foina	Stone marten	3	262	1.68	18.1	-	1.51	1.36	1.66
Martes martes	Pine marten	0	13	1.3	17	4	6.67	5.4	7.93
Meles meles	European badger	7	250	11.88	16.2	16.647	3.16	2.95	3.38
Melursus ursinus	Sloth bear	15	202	100	40	47.064	7.84	7.47	8.21
Mephitis mephitis	Striped skunk	12	1	2.4	12.9	1.674	7.74	7.33	8.14
Mirounga angustirostris	Northern elephant seal	1	16	1,112.39	20.3	-	10.5	9.15	11.9
Mustela erminea	Stoat	0	8	0.28	7.1	-	10	8.14	11.9
Mustela nigripes	Black-footed ferret	137	461	0.91	12	4	29.8	29.4	30.2
Mustela putorius furo	Domesticated polecat*	58	420	0.98	14	-	14	13.7	14.3
Mustela putorius	European polecat	0	27	0.98	14	-	3.45	2.78	4.11

Cont. overleaf

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Nasua narica	White-nosed coati	3	61	4.58	17.6	6.733	6.35	5.75	6.95
Nasua nasua	South American coati	6	394	3.78	17.7	5.591	1.77	1.64	1.9
Neofelis nebulosa	Clouded leopard	25	271	14.95	17	-	9.52	9.18	9.87
Odobenus rosmarus	Pacific walrus	17	98	1,043	40	-	18	17.2	18.8
Panthera leo leo	Asiatic lion	10	55	158.62	30	94.58	19.3	18.3	20.3
Panthera leo	Lion	41	529	158.62	30	94.58	7.91	7.68	8.14
Panthera onca	Jaguar	41	356	83.94	23	62.419	11.7	11.4	12.1
Panthera pardus	Leopard	21	227	52.4	23	-	9.61	9.23	9.99
Panthera tigris	Tiger	60	828	161.91	26.3	133.859	7.35	7.17	7.53
Panthera uncia	Snow leopard	32	2	32	21.2	-	12.5	12.1	12.9
Phoca vitulina	Harbour seal	1	131	87.32	40	73.29	1.5	1.3	1.71
Potos flavus	Kinkajou	2	37	2.44	29	4.294	7.69	6.86	8.53
Procyon lotor	Common raccoon	17	239	6.37	21	10.428	7.47	7.14	7.8
Proteles cristata	Aardwolf	0	16	8.14	25	11.563	5.56	4.5	6.61
Puma yagouaroundi	Jaguarundi	1	76	6.88	10.7	9.69	2.56	2.21	2.91
Suricata suricatta	Slender-tailed meerkat	6	40	0.73	12.5	1.729	16.7	15.5	17.8
Taxidea taxus	American badger	0	45	7.84	26	15.062	2.13	1.72	2.54
Urocyon cinereoargenteus	Gray fox	0	35	3.83	15	-	2.7	2.18	3.23
Ursus arctos	Brown bear	12	180	196.29	50	-	7.14	6.77	7.52
Ursus maritimus	Polar bear	9	130	371.7	38.2	-	7.58	7.12	8.03
Vulpes lagopus	Arctic fox	4	70	3.58	15	7.665	6.94	6.36	7.53
Vulpes macrotis	Kit fox	0	6	4.5	20	4.928	12.5	10.2	14.8
Vulpes velox	Swift fox	2	29	2.09	20	4.948	9.68	8.64	10.7
Vulpes vulpes	Red fox	0	49	4.82	15	13.731	1.96	1.58	2.34
Vulpes zerda	Fennec fox	3	54	1.25	16.3	2.693	7.14	6.47	7.82
Zalophus californianus	Californian sealion	246	1,350	137.19	30	-	18.3	18.1	18.5

* Indicates domesticated species

Table 10: Cancer incidence in 64 carnivoran species, 52 of which met criteria for analysis (32 for BMR). 6,556 tumours were reported in 42,115 veterinary reports; excluding domesticated species, 1,000 tumours were observed in 9,908 non-domesticated records. The ~45% incidence rate in Californian sea lion (*Zalophus californianus*) is the second highest rate in the entire dataset, and predominantly based on wild individuals.

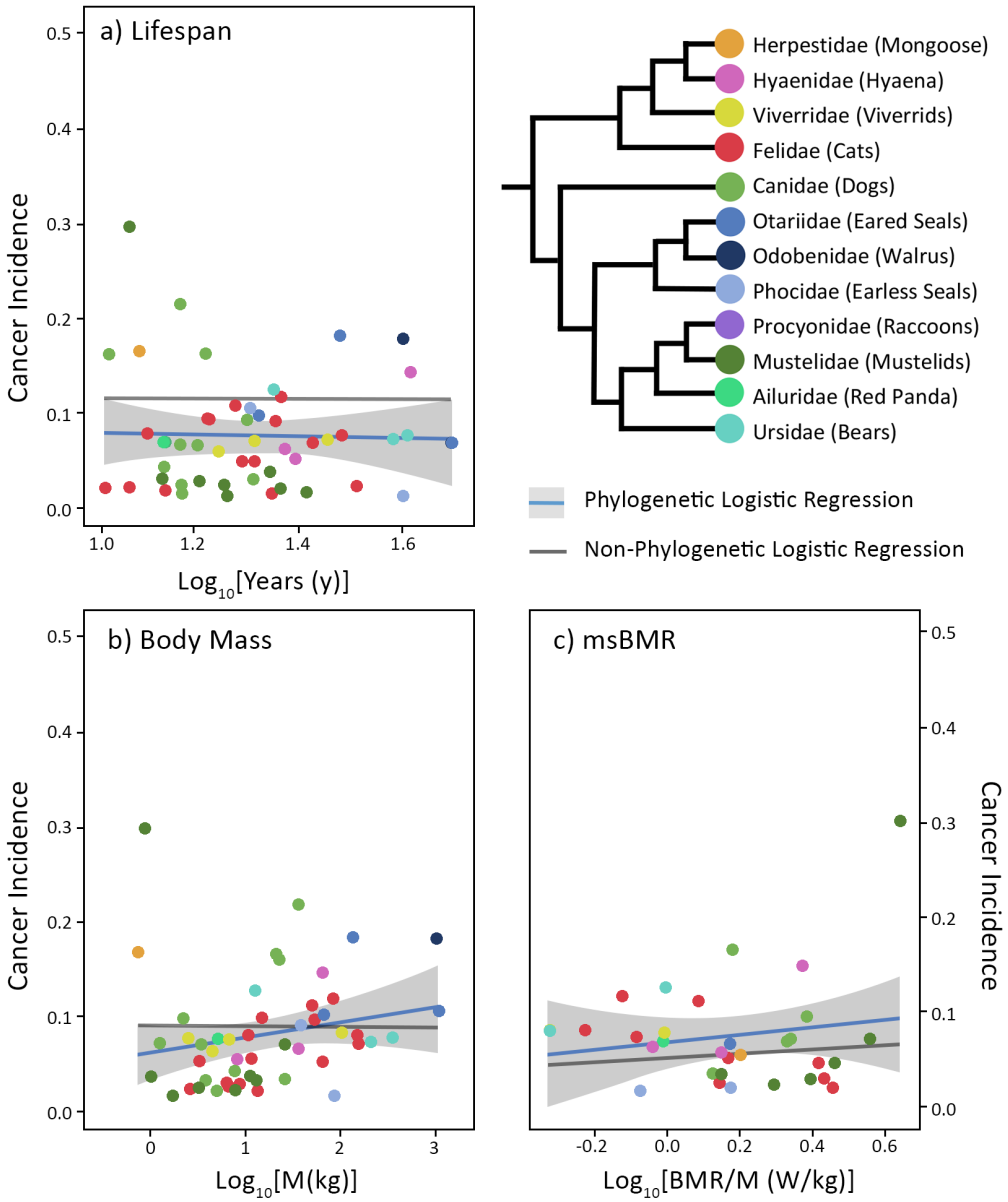


Fig. 11: Cancer scales invariantly with (a) lifespan (p=0.492, prd. R^2 = 0.0.364); and (b) body mass (p=0.291, prd. R^2 = 0.277) for 52 carnivoran species; and (c) mass-specific basal metabolic rate (p=0.09, prd. R^2 = 0.044) in 32 species with known BMR values.

3.3.8. Primates

64 primate species across 13 families were reviewed, from the gray mouse lemur (*Microcebus murinus*; weighing ~69g) to the Western lowland gorilla (*Gorilla gorilla*; mean ~112.5kg, up to 275kg in adult males), of which 50 had sufficient data required for analysis (20 for BMR). This group includes humans, with a maximum known lifespan of ~122.5 years, though here an upper mean expectancy bound of 81.16 years is used for analysis. 498 tumours were noted in 7,878 veterinary reports.

Malignancy in prosimians is widely reported in both wild and captive individuals, most commonly in the liver and digestive system, the causal basis remaining poorly understood (see Remick *et al*., 2009 for comprehensive review). The data here likewise demonstrates relatively high incidence, averaging 15% across the clade. Incidence in both New World and Old World monkeys was comparatively low, with exception to, often gastronintestinal, tumours in callitrichid (tamarin and marmoset) species (e.g. *Callimico goeldii* at ~16%). Similar incidence is reported in the literature; however by comparing pathology between captive and wild individuals, Wood *et al*. (1999) determined incidence was only a significant factor in captive colonies, suggesting environmental influence as a causal mechanism. Nutrition and food sensitivity has indeed historically, and remains, a major concern in captive callitrichid husbandry - species in the wild having diverse, difficult to fully replicate diets - and is associated with a wide range of captive-specific pathology (e.g. wasting marmoset syndrome; Yoshimoto *et al*., 2016), supporting the notion higher incidence recorded here is likewise associated with captivity, rather than life history.

Cancer is rarely reported in non-human apes, with exception to female reproductive malignancies in lowland gorilla (Lowenstine *et al*., 2016; Varki & Varki, 2015). Viral-induced leukemia and lymphoma were once widely reported in captive gibbon populations, due to widespread prevalence of Gibbon ape leukemia virus (GaLV), however following pathogen identification and management, is now rare (Brown & Tarlinton, 2016). Data collected here corroborates the literature; with exception to chimpanzee (~18.8%), no non-human species report incidence rates above ~5%.

The disparity between incidence in closely related chimpanzee, bonobo and human is, at a glance, intriguing. In the wild, chimpanzee lifespan is reported to average ~25 years, though occasionally up to 60; in captivity, the median lifespan is ~35, with the oldest confirmed individual at 59.4 years, but with some reports up to 80 years (Hakeem *et al*., 1996; Weigl, 2005). Review of the veterinary records collected here indicates most cases of neoplasia occurred in old individuals (>35-40 years),

suggesting the high incidence rate an artefact of artificially extended lifespan due to captive conditions, rather than any biological consideration (e.g. as with humans, a diet involving carnivory, in contrast to all other ape species) - and thereby outside the scope of typical selection - however the low number of bonobo data points (n=20) precludes robust analysis.

No significant relationship was modelled between incidence and any life history trait (see Fig. 12).

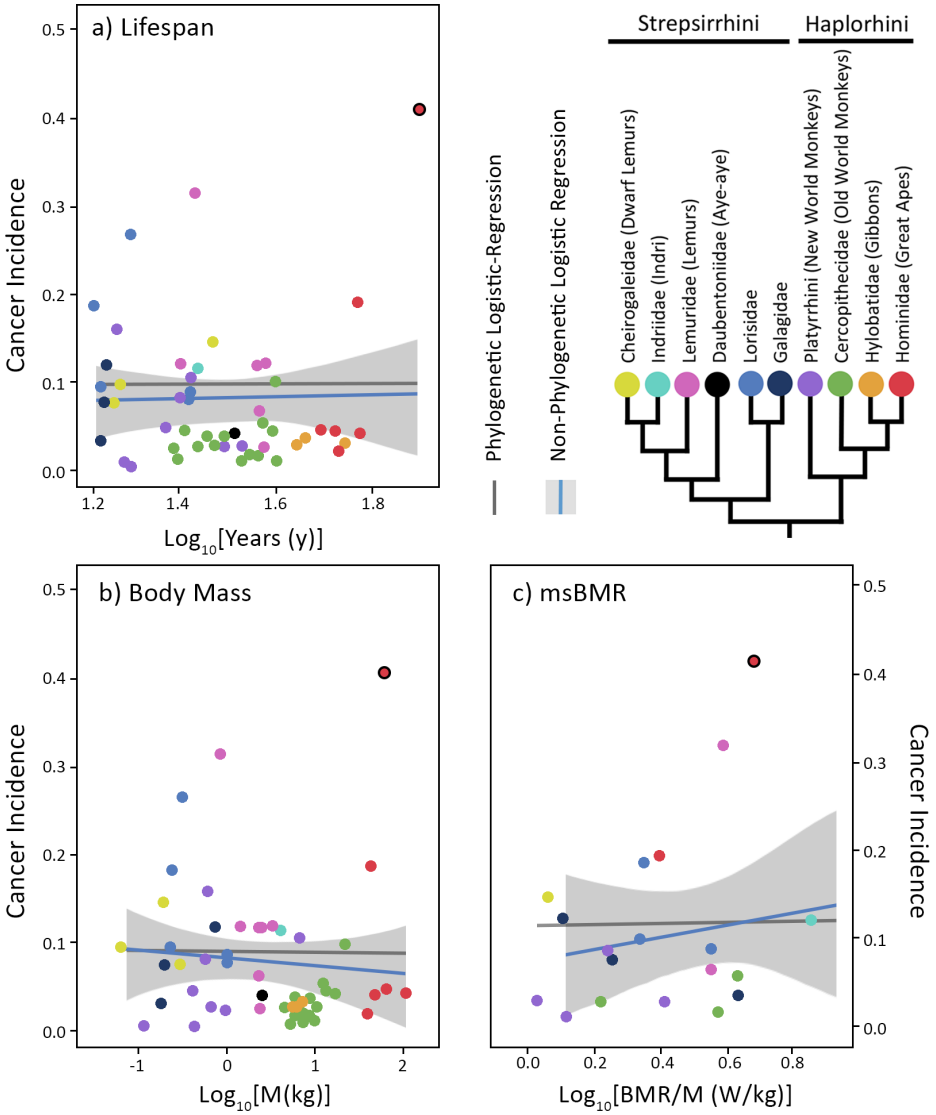


Fig. 12: Cancer scales invariantly with (a) lifespan times mass (p=0.2); (b) body size (p=0.379, prd. R^2 = 0.404); (c) lifespan (p=0.19, prd. R^2 = 0.361); and (d) mass-specific basal metabolic rate (p=0.647, prd. R^2 = 0.0.317) in 29 cetacean species. For msBMR (d), BMR values were extrapolated from available data and established scaling laws (See Methods 2.2.1); analysis using measured BMR values for 5 species (red line; highlighted points) similarly indicate no significant relationship (p=0.495, prd. R^2 = 0.44). Human (*Homo sapiens*) data point highlighted in bold.

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Alouatta caraya	Black howler	0	3	5.57	32.4	-	20	16.5	23.5
Aotus griseimembra	Grey-handed night monkey	0	1	1.009	33.8	2.49	33.3	28	38.7
Ateles belzebuth	White-fronted spider monkey	0	2	6.69	28	-	25	20.8	29.2
Ateles hybridus	Brown spider monkey	1	17	7.1	26	-	10.5	9.15	11.9
Callicebus cupreus	Coppery titi	1	2	1.12	26	-	50	45.1	54.9
Callimico goeldii	Goeldi's monkey	151	947	0.558	17.9	-	16	15.8	16.3
Callithrix geoffroyi	White-fronted marmoset	0	9	0.342	17.6	0.848	9.09	7.39	10.8
Cheirogaleus medius	Fat-tailed dwarf lemur	0	13	0.19	29	1.088	6.67	5.4	7.93
Daubentonia madagascariensis	Aye-aye	0	14	2.73	32.3	-	6.25	5.06	7.44
Eulemur coronatus	Crowned lemur	0	1	1.7	27	-	33.3	28	38.7
Eulemur flavifrons	Blue-eyed black lemur	0	1	1.8	20	-	33.3	28	38.7
Gorilla gorilla	Western lowland gorilla	8	114	112.59	54	-	7.26	6.8	7.71
Hapalemur alaotrensis	Alaotran gentle lemur	2	15	1.62	23	-	42.9	39.2	46.5
Homo sapiens	Human	**450***	**1000***	62.04	**122.5 (81.16)** †	82.78	40.9	34.1	40.9
Lemur catta	Ring-tailed lemur	2	6	2.63	37.3	-	37.5	13.7	19.6
Leontopithecus chrysomelas	Golden-headed lion tamarin	0	4	0.57	21.3	-	16.7	45.1	54.9
Leontopithecus chrysopygus	Black lion tamarin	1	2	0.66	17.9	-	50	7.23	8.99
Leontopithecus rosalia	Golden lion tamarin	2	35	0.59	24.8	2.125	8.11	17.1	20
Loris lydekkerianus	Grey slender loris	4	25	0.25	16	0.714	18.5	8.79	10.1
Loris tardigradus	Red slender loris	6	72	0.25	16.4	0.714	9.46	1.41	1.63
Macaca mulatta	Rhesus macaque	6	458	6.46	36	-	1.52	0.789	1.17
Macaca silenus	Lion-tailed macaque	0	100	5.99	40	-	0.98	35.7	44.3
Mandrillus leucophaeus	Drill	1	3	14.253	39	-	40	16.5	23.5
Microcebus murinus	Grey mouse lemur	0	3	0.069	18.2	-	20	30.7	32.4
Nycticebus coucang	Sunda slow loris	34	109	0.92	26.5	1.504	31.5	25.9	27.4
Nycticebus pygmaeus	Pygmy slow loris	35	133	0.34	19.3	-	26.7	35.7	44.3
Pan troglodytes	Chimpanzee	123	487	45	60	114.52	25.4	25	25.7
Perodicticus potto	Potto	2	33	1.081	26	1.948	8.57	7.64	9.5
Pithecia pithecia	White-faced saki	0	2	1.67	36	-	25	20.8	29.2
Pongo pygmaeus	Bornean orangutan	5	70	53.41	60	-	7.89	7.29	8.5
Propithecus coquereli	Coquerel's sifaka	2	24	4.19	27	3.738	11.5	10.3	12.8
Saguinus bicolor	Pied tamarin	0	22	0.465	19	-	4.17	3.37	4.97
Saguinus imperator	Emperor tamarin	0	6	0.40	23.7	-	12.5	10.2	14.8
Saguinus oedipus	Cotton-top tamarin	12	276	0.462	23.1	-	4.68	4.43	4.92
Saimiri sciureus	Common squirrel monkey	1	9	0.75	30.32	4.429	18.2	15.9	20.5
Trachypithecus auratus	Javan langur	0	1	9.72	31.1	-	33.3	28	38.7
Trachypithecus cristatus	Silvered leaf monkey	0	24	6.185	28.3	-	3.85	3.11	4.59
Varecia variegata	White-belted ruffed lemur	0	3	3.85	32	-	20	16.5	23.5

* For analysis purposes, values represent incidence as determined by (Smittenaar *et al.*, 2016; Cancer Research UK, 2020) † Maximum recorded lifespan and (in brackets) maximum life expectancy based on current human life characteristic value (δ); the latter was used in analysis.

Table 11: Cancer prevalence in 64 species of primate, 50 of which met criteria for inclusion in analysis. Due to taxonomic error using the methodology described in R, data for the Alaotran gentle lemur (*Hapalemur alaotrensis*) could not be included in analysis, though the high incidence remains of note. The species is critically endangered; all individuals included are captive (13 at the Duke Lemur Centre, NC, USA), descended from a very small founder population, preventing ruling out of higher incidence due to heritable factors via the founder effect.

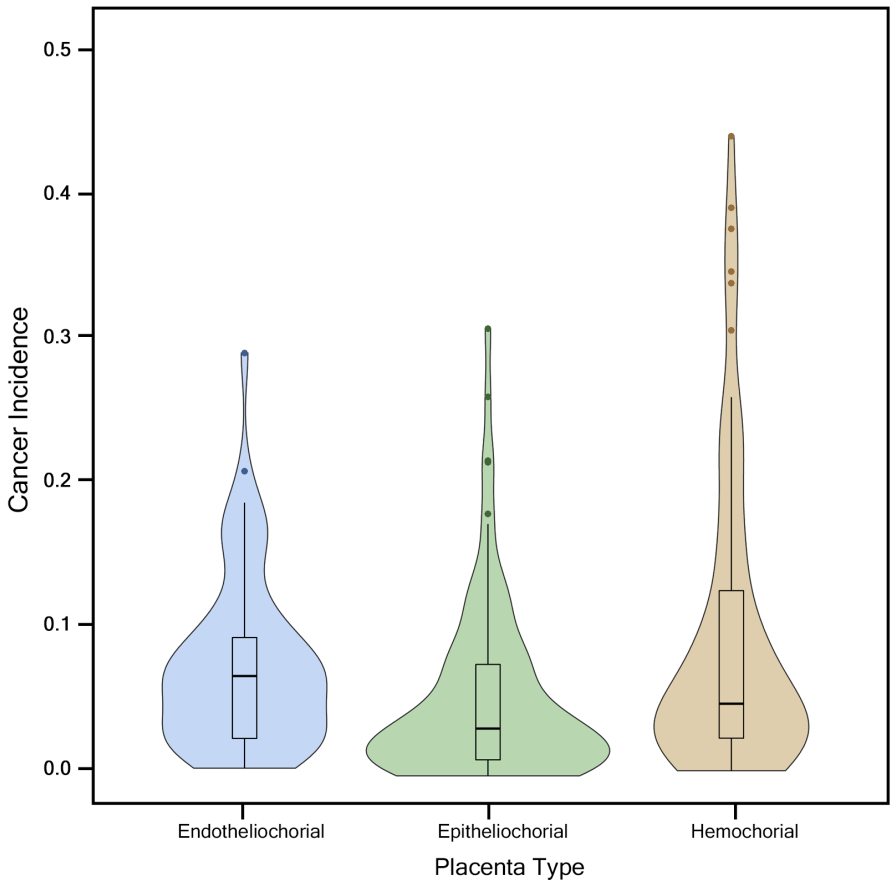
3.3.9. Further Life History Traits

3.3.9.1. Placentation

Across eutherian (non-marsupial) mammals, degree of placentation as classified based upon number of tissues separating maternal and fetal blood (in decreasing order, and increasing invasiveness: endothelio-, epithelia- and hemochorial) conferred no significant influence on modelled outcomes on the relationship between cancer prevalence and body mass, lifespan and msBMR, as determined by pANCOVA following weighted regression (p=1). Despite cancer cells recapitulating many of the same aspects of embryogenesis - rapid proliferation, replication stress and host-immune evasion - and indeed often re-expressing trophoblast-specific genes in order to advance tumorigenesis, at the broad scale selection for increased invasiveness in mammals doesn't seem to indicate increased malignancy risk.

With reference to marsupial placentation analysed and discussed above, it is however clear that stages of increasing evolution towards a eutherian placenta from the monotreme-like quasi-placental condition does confer greater neoplastic risk - indeed, placental condition was the most predictive modelled indicator of cancer prevalence in marsupials. Likewise of note, though not statistically significant, across categories of increasing placental invasiveness in eutherians, higher maximum observed neoplasia prevalence was observed (see Fig. 13). Given higher degrees of fetal-maternal interaction evolved independently in several taxa, alongside further lineage-specific placental adaptations, emphasis must be made on the generalised and simplified nature of this eutherian analysis. With more germane phenotypic classification and additional placental trait consideration, further investigation may reveal a relationship between selection for invasiveness, and associated host-immune-suppressive mechanisms, and neoplastic risk.

Fig. 13 (overleaf): Cancer incidence presented with respect to placental classification. From left to right with increasing invasiveness: endotheliochorial, epitheliochorial and hemochorial. Width of violin plots indicate species sample size. The most malignancy-prone species qualitatively seem to exhibit increasing placental invasiveness; the high hemochorial outliers were small rodents and Eulipotyphla (shrews and hedgehogs). As entries for said taxa were also the least relatively well-sampled of all mammalian taxa, sampling bias may have influence over the comparative analysis undertaken here.



3.3.9.2. Gestation, Litter Size & Diet

No relationship was observed between cancer incidence and litter size, following analysis on a subset of 247 mammal species with available trait data ($p = 0.67$). Similarly, no relationship was found between diet and cancer incidence, following pANCOVA analysis.

3.4. Birds

Outside domesticated species, despite often presenting relatively high longevity, cancer is rarely reported in birds - both captive and wild-examined. Here 269 tumours were noted in 28,498 records (<1%), across 173 species - 88 of which were suitable for analysis (see Table 12). No significant relationship was modelled between cancer incidence and any life history trait at class level, however within three orders with sufficient species data points (>10-15 species), a negatively scaling relationship was observed with increasing lifespan (see Fig.14).

Hornbills, unusually, are one of the few groups in which cancer is commonly observed - typically squamous carcinoma in the decorative bill or casque. Indeed, approximately ~40% of hornbill records collated here involved carcinoma, however no species data met the requisite >15 minimum records required for robust analysis. Incidence was otherwise highest in domesticated (or semi-domesticated) species, notably those intensively bred for the pet and ornamental trade (e.g. *Melopsittacus undulatus*, the budgerigar, at ~14.4%). The literature similarly describes the highest rates of avian neoplasia in these species, which is thought to be related to in-breeding and potential viral tumour induction, as occurs, for example, in psittacid herpesvirus papillomas (Reavill & Dorrestein, 2018).

Despite their longevity and relatively higher metabolic rates, birds have significantly lower cancer incidence rates compared to mammals of either equivalent size, longevity or basal metabolic rate. Given the independent origin of endothermy in birds, perhaps this disparity can be explained by difference in intrinsic metabolic processes and pathways. Indeed, Jimenez *et al.* (2019) experimentally demonstrated, despite birds having higher mass-specific metabolic rates compared to equivalently sized mammals (and, under a metabolic hypothesis, therefore be at higher risk of cancer), at a cellular level, basal metabolism, total antioxidant capacity, circulating lipid damage, and catalase activity were significantly lower. Further, oxidative stress parameters in birds, unlike in mammals, did not correlate with increasing longevity. It's speculated differences at a cellular level make small contributions to higher level metabolic scaling, and that higher avian whole-organism mass-specific metabolic rate manifests itself at a higher organisational level - for example, via organ size and activity (birds have proportionally larger active organs, such as the heart or liver, compared to similarly sized mammals). If changes in tissue architecture, organ activity, and similar, can result in high mass-specific metabolism without proportionally increasing cellular metabolism, and ROS production, it

may present a trajectory towards increased longevity and cancer resistance more broadly, including within mammals. For example, given many of the tissue- and organ-level metabolic adaptations in birds may have evolved to facilitate flight, it may be the disproportionally low rates of cancer and increased longevity in bats, despite similar mass-specific metabolic rate to other mammals, may be due to similar adaptation.

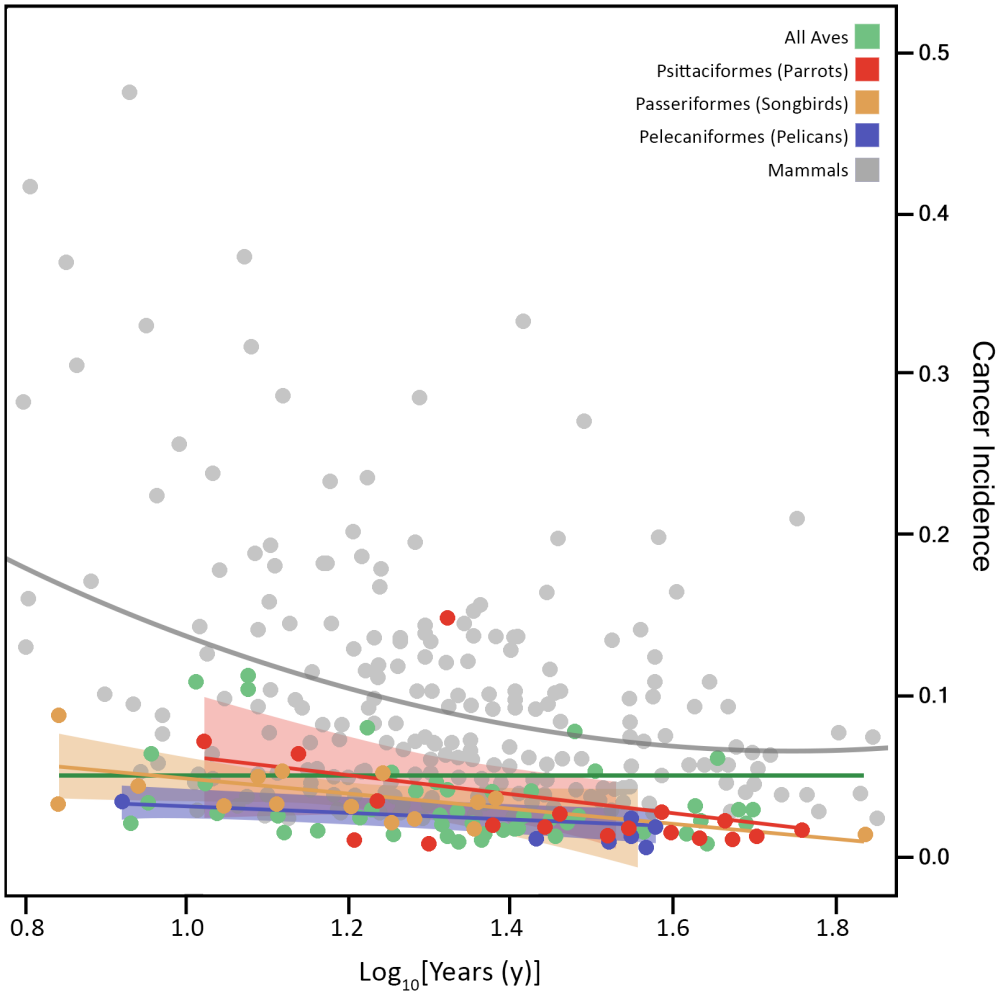


Fig. 14: Cancer incidence modelled against lifespan indicates no scaling relationship across birds as a class; lower taxon-level phylogenetic analysis similarly indicates no relationship, with three exceptions: across the Psittaciformes ($p=0.049$, prd. $R^2 = 0.24$), Passeriniformes ($p=0.006$, prd. $R^2 = 0.0403$), and Pelecaniformes ($p=0.0079$, prd. $R^2 = 0.819$). Of note, the family-level datasets for most other orders are too small to undertake equivalent analysis; they may present similar relationships, however the data provided is not sufficient to test them. Compared to mammals (grey), birds have significantly lower incidence rates via pANCOVA ($p<0.01$).

Species	Common Name	# Tumours	# Necropsies	Mass (Kg)	Lifespan (years)	BMR (W)	Tumour (%)	Lower 95% CI	Upper 95% CI
Spheniscus humboldti	Humboldt penguin	4	1757	4.379	20	-	0.28	0.259	0.309
Gallus gallus	Red junglefowl*	119	1639	0.7798	30	6.005	7.31	7.19	7.44
Eudocimus ruber	Scarlet ibis	6	1627	0.615	33.2	-	0.43	0.398	0.461
Phoenicopterus chilensis	Chilean flamingo	2	1521	2.277	36.7	-	0.2	0.175	0.219
Pavo cristatus	Indian peafowl	6	1509	4.2	23.2	-	0.46	0.429	0.498
Phoenicopterus roseus	Greater flamingo	3	989	2.5	44	-	0.4	0.364	0.443
Anas platyrhynchos	Mallard	11	797	1.0481	29.1	4.068	1.5	1.42	1.59
Aix sponsa	Wood duck	6	763	0.4528	22.5	2.247	0.92	0.848	0.983
Nymphicus hollandicus	Cockatiel	6	685	0.09	35	-	1.02	0.944	1.09
Threskiornis aethiopicus	Sacred ibis	0	680	1.53	37	-	0.15	0.118	0.175
Trichoglossus haematodus	Rainbow lorikeet	2	676	0.1151	16.1	-	0.44	0.393	0.492
Ara ararauna	Blue-and-yellow macaw	2	541	1.03	43	-	0.55	0.49	0.615
Phoeniconaias minor	Lesser flamingo	2	441	1.5	41	-	0.68	0.601	0.754
Ara macao	Scarlet macaw	1	401	1.04	33	-	0.5	0.428	0.565
Myiopsitta monachus	Monk parakeet	0	345	0.0817	22.1	0.5189	0.29	0.232	0.345
Eclectus roratus	Eclectus parrot	2	332	0.44	28.5	-	0.9	0.797	0.999
Bubo scandiacus	Snowy owl	3	290	2.0425	28	4.2244	1.37	1.24	1.5
Nesoenas mayeri	Pink pigeon	1	285	0.35	18	-	0.7	0.601	0.793
Aptenodytes patagonicus	King penguin	5	283	13.0368	26	25.889	2.11	1.94	2.27
Ara chloropterus	Red-and-green macaw	1	254	1.22	50.1	-	0.78	0.673	0.889
Psittacula krameri	Rose-ringed parakeet	1	242	0.128	34	-	0.82	0.707	0.933
Melopsittacus undulatus	Budgerigar	33	234	0.03	21	-	14.4	14	14.9
Eolophus roseicapilla	Galah	1	223	0.3	40	1.248	0.89	0.766	1.01
Struthio camelus	Ostrich	4	214	111	50	-	2.31	2.11	2.52
...	...	...	...	...	...	...	...	...	...

* Domesticated species.

Table 12: Cancer incidence in an example selection of 20 bird species (with highest sample size) from a dataset of 173 species (see Supplementary Materials X.X for full list). 88 species met criteria for analysis.

3.5. Other Taxa

3.5.1. Reptiles & Amphibians

Only 99 observations of neoplasia were reported in 5,112 necropsy records representing 75 reptile species, 20 of which were in the semi-domesticated corn snake (*Pantherophis guttatus*), with an incidence of ~26.9%. Very few, if any, tumours were recorded in any long-lived chelonian (turtles and tortoises) or crocodilian species, highest in the Eastern box turtle (*Terrapene carolina*) at 0.052%, and American alligator (*Alligator mississippiensis*), at 0.068%,respectively. Reports in the broader literature are similarly rare; as far as can be reviewed, only 24 tumours have been reported in crocodilian species, and, with exception to herpesvirus-induced neoplasia, and fibrous-tissue carcinomas in marine turtles, chelonians likewise present neoplasia extremely infrequently. Cancer is likewise rare in saurian lizards; of 1,297 submissions to a private diagnostic clinic, only ~6% were noted as neoplastic, usually melanophore and other skin-associated growth in pet trade species.

Conversely neoplasia in snakes is commonly reported in the veterinary literature, typically associated with snakes approaching (or often exceeding previously) reported maximum lifespan - suggestive of an artefact of captivity. Results collated here present similar rates between ~3-5% as that reported in the literature.

Phylogenetic regression analysis modelled no significant relationship between cancer incidence and any life history trait, including whether corn snake data was included/excluded.

11 cases of neoplasia were recorded for amphibian species, from 1,208 necropsy reports representing 17 species. Incidence in the axolotl (*Ambystoma mexicanum*) was high, at ~18% (n=3/20). No data was presented regarding whether the individuals were wild-type specimens, collected recently for conservation purposes, or of typical lab / pet-trade stock, where the entire global population remaining descended from only five individuals shipped to Paris in 1863 (Voss *et al*., 2015). If the latter, low genetic diversity may explain such an incidence. Otherwise, no amphibian species presented significant incidence; the next highest count was two tumours (of 412 necropsies) reported in the Dyeing poison frog (*Dendrobates tinctorius*).

3.5.2. Fish

Despite the common misconception that sharks do not get cancer, neoplasia does occur in chondrichthyans (cartilaginous fish), albeit rarely. A review of elasmobranch pathology by Garner (2013) observed 0.4% of submissions were neoplastic in origin. As for teleost fish, cancer is commonly reported, predominantly through the lens of aquacultural and ecotoxicological interest, given fish serve as good models for environmental carcinogenic contamination (not unlike marine mammals). Several types of tumour are unique to fish, though in general tumours of the same type and degree of differentiation generally behave similarly in both fish and other vertebrates, although the spread of malignant tumours via metastasis is far less common in fish, the reason for which remains unclear (Frasca Jr. *et al*., 2018).

Following review of institutional databases as part of this investigation, though neoplasia was observed in a number of species, no species presenting neoplasia met the requisite criteria for any further life history trait analysis. In-depth necroptic investigation following mortality was rarely reported. Screening the ZIMS global database at higher taxonomic ranking however revealed 723 cases of neoplasia reported in 9,648 necropsies (see Table. 13.), predominantly in Perciformes ('perch-like' fish). No further analysis in the interest of this investigation was pursued.

5.3. Invertebrates

Cancer has been documented in all but four invertebrate taxa to date: Ctenophora (comb jellies), Placozoa, Porifera (sponges) and Hemichordates (acorn worms) (Aktipis *et al*., 2015). There remain many challenges documenting neoplasia reliably given the characteristics used to define neoplasia in vertebrates often cannot be fulfilled with respect to markedly different invertebrate physiologies. Here, data from 23 species was collated and two neoplasms were recorded: an exophytic papilloma in a freshwater mussel (*Margaritifera margaritifera*) and a hamartoma on the carapace of a white-clawed crayfish (*Austropotamobius pallipes*).

	Taxon	Description	# Tumours	# Necropsies	Tumour (%)	Lower 95% CI	Upper 95% CI
	Selachimorpha	Sharks	1	382	0.521	0.449	0.593
a)	Batoidea	Rays	1	303	0.656	0.565	0.746
	Rajiformes	Skates	0	25	3.7	2.99	4.42
b)	Lepidosireniformes	Lungfish	0	21	4.35	3.51	5.18
	Acipenseriformes	Sturgeon	0	54	1.79	1.44	2.13
	Anguilliformes	Eels	1	24	7.69	6.67	8.72
	Atheriniformes	Silversides	3	574	0.694	0.627	0.762
	Characiformes	Tetra, Piranha, Tigerfish	8	1671	0.538	0.503	0.573
	Clupeiformes	Herring, Anchovy	0	92	1.06	0.856	1.27
	Cypriniformes	Carp, Minnow, Loaches	22	1,925	1.19	1.15	1.24
	Cyprinodontiformes	Killifish, Pupfish, Livebearers	8	371	2.36	2.21	2.51
	Gadiformes	Cod	2	22	12.5	11.2	13.8
c)	Gasterosteoidei	Sticklebacks	2	28	10	8.93	11.1
	Syngnathiformes	Seahorses, Pipefish	4	173	2.86	2.61	3.1
	Osteoglossiformes	Butterflyfish, Arowana, African Knifefish	1	32	5.88	5.09	6.67
	Gymnotiformes	South American Knifefish, Electric Eel	3	11	30.8	28.3	33.3
	Perciformes	‘Perch-like’ fish	645	3,353	19.3	19.1	19.4
	Pleuronectiformes	Flatfish	3	45	8.51	7.71	9.31
	Salmoniformes	Salmon, Trout, Chars	9	159	6.21	5.84	6.58
	Scorpaeniformes	Scorpionfish, Lionfish, Sculpins	2	199	1.49	1.32	1.66
	Siluriformes	Catfish	4	104	4.72	4.31	5.12
	Tetraodontiformes	Pufferfish, Sunfish, Boxfish	4	80	6.1	5.58	6.62

Table 13: Cancer prevalence in Chondrichthians (a; cartilaginous fish), Dipnomorpha (b; lungfish) and Teleosts (c; bony-fish) as logged on the ZIMS global necropsy database. Mean and median cancer incidence (where n¿15) are 4.07% (CI: 3.62-4.52) and 2.86% (CI: 2.64-3.1) respectively.

Conclusion

This investigation presents the first large-scale quantitative assessment of cancer prevalence across the animal kingdom and the results the first confirmation that Peto's hypothesis holds across it, regardless of taxon: increase in size and longevity does not correlate with increase in cancer risk, and therefore larger, longer-lived organisms must have developed mechanisms to prevent cancer.

Collecting reliable data on cancer prevalence is an incredible challenge for most animal species in the wild. Here, records predominantly from captive individuals were collated in order to calculate estimated (non-age-adjusted) lifetime incidence risk of cancer per species. Though data from captive specimens may present differently to animals living in their natural environment (see 4.2: Limitations), this dataset nonetheless presents a broadscale approximation of prevalence rates across animal taxa.

In order to explore the relationship between cancer and life history theory, I modelled the three primary life history trade-offs, or traits, - body mass, lifespan and mass-specific basal metabolic rate (msBMR) - against cancer prevalence in 715 vertebrate species, of which 522 were mammals. Consistent with the hypothesis presented by Peto's paradox, cancer incidence scales invariantly with lifespan and body mass across all taxa, with exception to a few lower taxonomically ranking clades whereby it scaled negatively (as discussed above). In those taxa where cancer scaled negatively, in most cases the best modelled predictor of the negatively scaling trajectory was msBMR, particularly in several groups representing large ranges in body size; notably rodents and afrotheria, clades representing three and four orders of magnitude in size difference respectively. Cetaceans observe a similar change in size across the clade, however in this analysis, msBMR data for larger whales was based purely on estimates, preventing thorough analysis and a more robust inference - one suspects if reliable values for scaling msBMR were possible, it may likewise be shown that cancer prevalence may follow much the same trajectory as in afrotheria and rodents.

4.1. Life History Traits

4.1.1. Placentation

To date, there has been little investigation into malignancy risk and placentation depth in mammals. One limited study on four domesticated species suggested an association between high invasiveness (epitheliochoriality in horses and cows) and lower cancer risk, though this did not take into account additional covariables (Souza & Wagner, 2014). Here, a broad analysis of the relationship between cancer and invasiveness was undertaken. Whilst a categorised trajectory towards increased placentation in marsupials did indicate invasiveness a significant covariable in predicted cancer risk, more broadly across eutherian mammals no significant relationship was discovered. A tentative qualitative assessment hints at a potential association however, supporting the need for further investigation.

Here, clade-specific categorisation of placentation depth was undertaken only at the most generalised level, given a focus on placental invasiveness was beyond the main scope of this project. Given evolution towards placentae of varying depth has been undertaken independently multiple times across multiple taxa (and further, secondarily reduced independently across several taxa), use of generalised categorisation in analysis is likely insufficient to explore any relationship between cancer and placentation. Understanding the physiological and molecular routes arrived by assorted mammals towards their current degree of placentation, and forming analyses based on these more nuanced adaptive trajectories may yet reveal a relationship between placental invasiveness and broader cancer risk, as potentially demonstrated by the simple analysis of marsupial placentation undertaken here.

4.2. Limitations

All entries in this dataset came from necropsy and veterinary records, undertaken by trained veterinary professionals. However, it must be recognised neither methodology associated with examining specimens and recording observed pathology, nor follow-up histo- and immunopathological investigation to determine diagnosis, was standardised, as entries originated from a wide breadth of sources. This measurement error was controlled for where possible (see Methods 2.1), though, even within a single given institution, procedure may be inconsistent. In addition, selective bias may be present, as depth of investigative pathology may be undertaken on some species of interest moreso than others.

Both sources of bias here may have be mitigated partially by the large number of sources used.

Cancer is not a singular disease, but a cellular behaviour presenting a wide variety of different phenotypes. For the purposes of this study, this diversity was modelled as a simple determinative score, and those phenotypes presenting as neoplastic were considered equivalent and included in analysis. Other phenotypes (e.g. hyperplasia, fibroma etc.) were excluded from analysis. However, with appropriate standardisation in classifying cancer phenotypes, this dataset does present future opportunity as a foundation for further in-depth study on the proliferation of different cancer types in animal species throughout lifespan, and across taxonomic lineages.

Life history trait data was taken as presented by the literature sources cited. Though data on age and weight at time of death was recorded for many entries, as these individual entries were pooled with entries lacking said information, they were not used when calculating lifespan and body mass estimates for each species).

The majority of the entries in this dataset originate from captive or semi-captive individuals. Individual animals living in captivity may lead markedly different lives to their wild counterparts, both exposed to or otherwise protected from factors influencing disease risk, resulting in bias in the dataset. Indeed, as per Tidiere *et al.* (2016), the majority of captive mammals live longer lives than their wild counterparts, particularly those species with shorter lifespans, higher reproductive rates and higher mortality in the wild ('r-selected' species). Given cancer is an age-associated disease, this dataset likely skews towards a higher cancer prevalence rate in such smaller, faster-living species (e.g. rodents) compared to their wild counterparts. Age standardisation was not undertaken on this dataset, owing to lack of reliable data for the majority of entries, however if undertaken on appropriate data (in this instance, more recorded entries from managed populations), age-adjusted estimates of cancer incidence would provide a more accurate representation of cancer pathology in wild animal species. At the other end of the scale, comprehensive necroptic investigation is likely to be skewed towards underestimating cancer incidence in larger mammals, especially in non-managed species including most cetaceans given the challenge of performing comprehensive necropsy on such large animals. Despite this recognised bias, the degree of change in under- and over-representing cancer incidence respectively in larger and smaller animals required to alter or contradict the results of this investigation are significant - tripling observed number of tumours in baleen whales, for example, has no effect on the statistical outcome. We can therefore expect the results of this investigation to remain valid.

4.3. Summary & Next Steps

Cancer prevalence varies across the animal kingdom. In this chapter, incidence was established for 793 species, providing the first large and comprehensive dataset of cancer prevalence across all vertebrate taxa. Life history trait analysis provided a quantitative establishment of Peto's paradox; cancer incidence scales largely invariantly with body size and longevity, confirming the implication that the biology of large, long-lived mammals must account for the fundamental constraint cancer presents on morphological evolution.

Peto's original hypothesis was based upon, and paradoxical, from the perspective of on multistage models of human cancer progression. Solutions have been proposed based upon adaptive evolution (e.g. Nunney, 1999), however analysis undertaken here suggests a correction to the problem may also be arrived, at least partly, by considering intrinsic life history factors, notably metabolic rate. In the following chapters, I will explore both avenues.

First, if it can be established that intrinsic traits, such as basal metabolic rate, can reduce cancer incidence in a biologically relevant way, much of the variance, and therefore supposed paradox, in cancer incidence across lifespan and body mass may be explained, and a solution to Peto's paradox found. In the next chapter a series of mechanistic models of cancer progression incorporating the influence of metabolic rate on DNA mutation and cell division rate are therefore explored.

5. References

- Abegglen, L.M., Caulin, A.F., Chan, A., Lee, K., Robinson, R., Campbell, M.S., Kiso, W.K., Schmitt, D.L., Waddell, P.J., Bhaskara, S., Jensen, S.T., Maley, C.C. & Schiffman, J.D. (2015) Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. *JAMA*. 314 (**17**), 1,850-1,860

- Agnew, D., Nofs, S., Delaney, M.A. & Rothenburger, J.L. (2018) Xenartha, Erinacoemorpha, some Afrotheria, and Phloidota. In: Terio, K.A., McAloose, D. & Leger, J.S. (eds.) *Pathology of Wildlife and Zoo Animals*. Academic Press. pp 517-532

- Aktipis, C.A., Boddy, A.M., Jansen, G., Hibner, U., Hochberg, M.E., Maley, C.C. & Wilkinson, G.S. (2015) Cancer across the tree of life: cooperation and cheating in multicellularity. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 370 (**1673**), e20140219

- Andervont, H.B. & Dunn, T.B. (1962) Occurrence of Tumours in Wild House Mice. *Journal of the National Cancer Institute*. 28 (**5**), 1,153-1,163

- Bäcklin, B.M. & Olovsson, E.M. (2003) Histology of Uterine Leiomyoma and Occurrence in Relation to Reproductive Activity in the Baltic Gray Seal (*Halichoerus grypus*). *Veterinary Pathology*. 40 (2), 175-180

- Barrowclough, G.F., Cracraft, J., Klicka, J. & Zink, R.M. (2016) How many species of birds are there and why does it matter? *PLoS One*. 11 (**11**)

- Baumann, P.C. (1984) Cancer in Wild Freshwater Fish Populations with Emphasis on the Great Lakes. *Journal of Great Lakes Research*. 10 (**3**), 251-253

- Baumann, P.C. (1998) Epizootics of cancer in fish associated with genotoxins in sediment and water. *Mutation Research / Reviews in Mutation Research*. 411 (**3**), 227-233

- BirdLife International (2020) Accessed online at: https://www.birdlife.org

- Bossart, G.D., Reidarson, T.H., Dierauf, L.A. & Duffield, D.A. (2001) Clinical Pathology. In: Gulland, D.A. (ed.) *CRC Handbook of Marine Mammal Medicine*. CRC Press, Boca Raton, Florida, US. pp. 383-430

- Bowser, P.R., Wolfe, M.J., Forney, J.L. & Wooster, G.A. (1988) Seasonal Prevalence of Skin Tumors from Walleye (*Stizostedion vitreum*) from Onedia Lake, New York. *Journal of Wildlife Diseases*. 24 (**2**), 292-298

- Brown, K. & Tarlinton, R.E. (2016) Is gibbon ape leukemia virus still a threat?. *Mammal Review*. 47 (**1**), 53-61

- Burgin, C.J., Colella, J.P., Kahn, P.L. & Upham, N.S. (2018) How many species of mammals are there? *Journal of Mammology*. 99 (**1**), 1-14

- Cancer Research UK (2020) Cancer Statistics. [online] Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk [Accessed: 10/06/20]

- Canfield, P.J., Hartley, W.J. & Reddacliff, G.L. (1990) Spontaneous proliferations in Australian marsupials - a survey and review; 2. Dasyurids and Bandicoots. *Journal of Comparative Pathology*. 103 (**2**), 147-158

– Caulin, A.F. & Maley, C.C. (2011) Peto's Paradox: evolution's prescription for cancer prevention. *Trends in Ecology & Evolution*. 26 (**4**), 175-182

– Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J., Tzovaras, B.G., Marchand, P., Tran, V., Salmon, M., Li, G. & Grenié, M. (2020) taxize: Taxonomic information from around the web. R package version 0.9.98

– Chang, P.H., Liu, C.H. & Jeng, C.R. (2012) Spontaneous neoplasms in zoo mammals, birds, and reptiles in Taiwan – A 10-year survey. *Animal Biology*. 62 (**1**), 95-110

– Clemente, C.J., Cooper, C.E., Withers, P.C., Freakley, C., Singh, S. & Terrill, P. (2016) The private life of echidnas: using accelerometry and GPS to examine field biomechanics and assess the ecological impact of a widespread, semi-fossorial monotreme. *Journal of Experimental Biology*. 219 (**20**), 3,271-3,282

– Coffee, L.L., Casey, J.W. & Bowser, P.R. (2013) Pathology of Tumors in Fish Associated With Retroviruses: A Review. *Veterinary Pathology*. 50 (**3**), 390-403

– CSIP: Cetacean Stranding Investigation Programme (2020) Available online at: http://ukstrandings.org/csip-reports/

– Dang, C.V. (2015) A metabolic perspective of Peto's paradox and cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 370 (**1673**), e20140223

– Delaney, M.A., Ward, J.M., Walsh, T.F., Chinnadurai, S.K., Kerns, K., Kinsel, M.J. & Treuting, P.M. (2016) Initial Case Reports of Cancer in Naked Mole-rats (*Heterocephalus glaber*). *Veterinary Pathology*. 53 (**3**), 691-696

– Delaney, M.A., Nagy, L., Kinsel, M.J. & Treuting, P.M. (2013) Spontaneous histologic lesions of the adult naked mole rat (*Heterocephalus glaber*): a retrospective survey of lesions in a zoo population. *Veterinary Pathology*. 50 (**4**), 607-621

– Dobson, J.M., Samuel, S., Milstein, H., Rogers, K. & Wood, J.L.N. (2006) Canine neoplasia in the UK: estimates of incidence rates from a population of insured dogs. *Journal of Small Animal Practice*. 43 (**6**)

– Duignan, P.J., Gibbs, N.J. & Jones, G.W. (2004) Autopsy of cetaceans incidentally caught in commercial fisheries, and all beachcast specimens of Hectorís dolphins, 2001/02. *DOC Science Internal Series*. 176

– Ernest, S.K.M. (2003) Life History Characteristics of Placental Nonvolant Mammals. *Ecological Archives*. 84 (12), 3,402

– Frasca Jr, S., Wolf, J.C., Kinsel, M.J., Camus, A.C. & Lombardini, E.D. (2018) Osteichthyes. In: Terio, K.A., McAloose, D. & Leger, J.S. (eds.) *Pathology of Wildlife and Zoo Animals*. Academic Press. pp. 953-1,001

– Froese, R. & Pauly, D. (2019) Fishbase. Available online at: https://www.fishbase.se

– Garland Jr, T., Bennet, A.F. & Rezende, E.L. (2005) Phylogenetic approaches in comparative physiology. *Journal of Experimental Biology*. 16, 3,015-3,035

– Garner, M.M. (2013) A Retrospective Study of Disease in Elasmobranchs. *Veterinary Pathology*. 50 (**3**), 377-389

– Griffiths, M. (1989) Tachyglossidae. In: Walton, D.W. & Richardson, B.J. (eds.) *Fauna of Australia*. Mammalia, Canberra, Australia. pp. 407–435

– Hakeem, A., Sandoval, G.R., Jones, M. & Allman, J. (1996) Brain and life span in primates. In: Birren, J. (ed.) *Handbook of the Psychology of Aging*. Academic Press, San Diego, USA. pp. 78-104

– Herrera-Álvarez, S., Karlsson, E., Ryder, O.A., Lindblad-Toh, K. & Crawford, A.J. (2020) How to Make a Rodent Giant: Genomic Basis and Tradeoffs of Gigantism in the Capybara, the World's Largest Rodent. *Molecular Biology and Evolution*. msaa285

– Ho, L.T. & Ané, C. (2014) A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systems Biology*. 63 (**3**), 397-408

– Hubbard, G.B., Schmidt, R.E. & Fletcher, K.C. (1983) Neoplasia in Zoo Animals. *The Journal of Zoo Animal Medicine*. 14 (**1**), 33-40

– ITIS: Integrated Taxonomic Information System: Available online at: https://www.itis.gov

– Ives, A.R. & Garland Jr, T. (2010) Phylogenetic logistic regression for binary dependent variables. *Systematic Biology*. (**1**), 9-26

– Jimenez, A.G., O'Connor, E.S., Tobin, K.J., Anderson, K.N., Winward, J.D., Fleming, A., Winner, C., Chinchilli, E., Maya, A., Carlson, K. & Downs, C.J. (2019) Does Cellular Metabolism from Primary Fibroblasts and Oxidative Stress in Blood Differ between Mammals and Birds? The (Lack-thereof) Scaling of Oxidative Stress. *Integrative & Comparative Biology*. 59 (**4**), 953-969

– Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L., Safi, K., Sechrest, W., Boakes, E.H., Carbone, C., Connolly, C., Cutts, M.J., Foster, J.K., Grenyer, R., Habib, M., Plaster, C.A., Price, S.A., Rigby, E.A., Rist, J., Teacher, A., Bininda-Emonds, O.R.P., Gittleman, J.L., Mace, G.M. & Purvis, A. (2009) PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecological Archives*. 90 (**9**), 2,648

– Keesey, M.T. (2020) Phylopic. Available online at: https://www.phylopic.org [Accessed: 01/09/2020]

– Kleiber, M. (1932) Body size and metabolism. *Hilgardia*. 6 (**11**), 315-353

– Knottenbelt, D.C., Snalune, K. & Patterson-Kane, J. (2016) *Clinical Equine Oncology*. Elsevier.

– Koehl, D. (2020). Elephant Encyclopedia. Available online at: https://www.elephant.se

– Koh, J., Itahana, Y., Mendenhall, I.H., Low, D., Soh, E.X.Y., Guo, A.K., Chiong, Y.T., Wang, L.F. & Itahana, K. (2019) *ABCB1* protects bat cells from DNA damage induced by genotoxic compounds. *Nature Communications*. 10, 2820

– Kolokotrones, T., Savage, V., Deeds, E.J. & Fontana, W. (2010) Curvature in metabolic scaling. *Nature*. 464 (**7289**), 753-756

– Ladds, P. (2009) *Pathology of Australian Native Wildlife*. CSIRO Publishing, Melbourne, Australia.

– Landolfi, J.A. & Terrell, S.P. (2018) Proboscidae. In: Terio, K.A., McAloose, D. & Leger, J.S. (eds.) *Pathology of Wildlife and Zoo Animals*. Academic Press. pp. 413-431

– Lowenstine, L.J., McManamon, R. & Terio, K.A. (2016) Comparative Pathology of Aging Great Apes: Bonobos, Chimpanzees, Gorillas, and Orangutans. *Veterinary Pathology*. 53 (**2**), 250-276

– Madsen, T., Arnal, A., Vittecoq, M., Bernex, F., Abadie, J., Labrut, S., Garcia, D., Faugere, D., Lemberger, K., Beckmann, C., Roche, B., Thomas, F. & Ujvari, B. (2017) Chapter 2: Cancer Prevalence and Etiology in Wild and Captive Animals. In: Ujvari, B., Roche, B. & Thomas, F. (eds) *Ecology and Evolution of Cancer*. 1st Edition. Elsevier Academic Press. p.p. 11-46

– Manov, I., Hirsh, M., Iancu, T.C., Malik, A., Sotnichenko, N., Band, M., Avivi, A. & Shams, I. (2013) Pronounced cancer resistance in a subterranean rodent, the blind mole-rat, Spalax: in vivo and in vitroevidence. *BMC Biology*. 91

– Martineau, D., Lemberger, K., Dallaire, A., Labelle, P., Lipscomb, T.P., Michel, P. & Mikaelian, I. (2002) Cancer in wildlife, a case study: beluga from the St. Lawrence estuary, Québec, Canada. *Environmental Health Perspectives*. 110 (**3**), 285-292

– McAloose, D. & Newton, A.L. (2009) Wildlife cancer: a conservation perspective. *Nature Reviews Cancer*. 9, 517-526

– McNab, B.K. (2008) An analysis of the factors that influence the level and scaling of mammalian BMR. *Comparative biochemistry and physiology, part A: Molecular & integrative physiology*. 151 (**1**), 5-28

– Michael, S.A. & Sangster, C.R. (2010) Pulmonary mycobacteriosis caused by Mycobacterium intracellulare in a Tasmanian devil (*Sarcophilus harrisii*). *Australian Veterinary Journal*. 88 (**7**), 280-282

– Miller, J.K. (2001) Escaping senescence: demographic data from the three-toed box turtle (Terrapene carolina triunguis). *Experimental Gerontology*. 36 (**4-6**), 829-832

– Myers. M.S., Landahl, J.T., Krahn, M.M., Johnson, L.L. & McCain, B.B. (1990) Overview of studies on liver carcinogenesis in English sole from Puget Sound; evidence for a xenobiotic chemical etiology I: Pathology and epizootiology. *Science of The Total Environment*. 94 (**1-2**), 33-50

– NCBI: National Center for Biotechnology Information (2020). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. Available online at: https://www.ncbi.nlm.nih.gov

– Nunney, L. (1999) Lineage selection and the evolution of multistage carcinogenesis. *Proceedings of the Royal Society B: Biological Sciences*. 266 (**1418**), 493-498

– Parr, C.S., Wilson, N., Leary, P., Schulz, K.S., Lans, K., Walley, L., Hammock, J.A., Goddard, A., Rice, J., Studer, M., Holmes, J.T.G. & Corrigan, R.J. (2014) The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth. *Biodiversity Data Journal 2*. e1079

– Paterno, G.B., Penone, C. & Werner, G.D.A. (2018) sensiPhy: An r-package for sensitivty analysis in phylogenetic comparative methods. *Methods in Ecology and Evolution*. 9, 1,461-1,467

– Pessier, A.P., Stern, J.K. & Witte, C.L. (2016) TP53 Gene and Cancer Resistance in Elephants. *JAMA*. 315 (**16**), 1,789

– Peto, R., Roe, F.J., Lee, P.N. & Clack, J. (1975) Cancer and ageing in mice and men. *British Journal of Cancer*. 32, 411-426

– Pompei, F., Polkanov, M. & Wilson, R. (2001) Age distribution of cancer in mice: the incidence turnover at old age. *Toxicology & Industrial Health*. (**17**), 7-16

– Reavill, D.R. & Dorrestein, G. (2018) Psittacines, Coliiformes, Musophagiformes, Cuculiformes. In: Terio, K.A., McAloose, D. & Leger, J.S. (eds.) *Pathology of Wildlife and Zoo Animals*. Academic Press.

– Rees, J. & Cranston, K. (2017) Automated assembly of a reference taxonomy for phylogenetic data synthesis. *Biodiversity Data Journal 5*. e12581.

– Remick, A.K., Van Wettere, A.J. & Williams, C.V. (2009) Neoplasia in prosimians: case series from a captive prosimian population and literature review. *Veterinary Pathology*. 46 (**4**), 746-772

– Revell, L.J. (2010) Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution*. 1 (**4**), 319-329

– Revell, L.J. (2012) phytools: An R package for phylogenetic comparative biology)and other things). *Methods in Ecology and Evolution*. 3, 2,173-2,223

– Roskov, Y., Ower, G., Orrell, T., Nicolson, D., Bailly, N., Kirk, P.M., Bourgoin, T., DeWalt, R.E., Decock, W., Nieukerken, E. van, Zarucchi, J. & Penev L. eds. (2019). *Species 2000 & ITIS Catalogue of Life, 2019 Annual Checklist*. Available online at: https://www.catalogueoflife.org/annual-checklist/2019

– Smith, K.K. (2015) Evolution of the placenta in therian mammals. In: Dial, K.P., Shubin, N. & Brainerd, E.L. (eds.) *Great transformations in Vertebrate Evolution*. University of Chicago Press, Michigan, USA.

– Smittenaar, C.R., Peterson, K.A., Stewart, K. & Moitt, N. (2016) Cancer incidence and mortality projections in the UK until 2035. *British Journal of Cancer*. 115 (**9**), 1,147-1,155

– Souza, A.W.D. & Wagner, G.P. (2014) Malignant cancer and invasive placentation: A case for positive pleiotropy between endometrial and malignancy phenotypes. *Evolution, Medicine, and Public Health*. 1, 136-145

– Stark, G., Tamar, K., Itescu, Y., Feldman, A. & Meiri, S. (2018) Cold and isolated ectotherms: drivers of reptilian longevity. *Biological Journal of the Linnaean Society*. 125 (**4**), 730-740

– Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V. E. & de Magalhaes, J. P. (2018) Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research*. 46

– Tidière, M., Gaillard, J.M., Berger, V., Müller, D.W.H., Lackey, L.B., Gimenez, O., Clauss, M. & Lemaître, J.F. (2016) Comparative analysis of longevity and senescence reveal variable survival benefits of living in zoos across mammals. *Scientific Reports*. 6, 36361

– Tong, L. (2019) Neoplasia. In: *Current Therapy in Medicine of Australian Mammals*. Vogelnest, L. & Portas, T. (eds.). CSIRO Publishing, Australia.

– Uetz, P., Freed, P. & Hošek, J. (2020) The Reptile Database. Available online at: https://www.reptile-database.org

– Varki, N.M. & Varki, A. (2015) On the apparent rarity of epithelial cancers in captive chimpanzees. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 370 (**1673**), 20140225

– Voss, S.R., Palumbo, A., Nagarajan, R., Gardiner, D.M., Muneoka, K., Stromberg, A.J. & Athippozhy, A.T. (2015) Gene expression during the first 28 days of axolotl limb regeneration I: Experimental design and global analysis of gene expression. *Regeneration*. 2 (**3**), 120-136

– Waugh, C.A., Grigg, G.C., Booth D.T. & Beard, L.A. (2006) Respiration by buried echidnas Tachyglossus aculeatus. *Journal of Experimental Biology*. 209 (**5**), 938-944

– Weigl, R. (2005) *Longevity of Mammals in Captivity; from the Living Collections of the World*. Kleine Senckenberg-Reihe, Stuttgart, Germany.

– Wilson, D.E. & Reeder, D.M. (2005) *Mammals of the World: A Taxonomic and Geographic Reference* (3rd ed.). Johns Hopkins University Press, USA.

– Wood, J.D., Peck, O.C., Tefend, K.S., Rodriguez, M.A., Rodriguez, J.V., Hernández, J.I., Stonerook, M.J. & Sharma, H.M. (1998) Colitis and colon cancer in cotton-top tamarins (*Saguinus oedipus oedipus*) living wild in their natural habitat. *Digestive Disease and Sciences*. 43 (**7**), 1,443-1,353

– Woods, R., Ladds, P.W. & Blyde, D. (2008) Dugongs. In: Vogelnest, L. & Woods, R. (eds.) *Medicine of Australian Mammals*. CSIRO, Victoria Australia. pp. 615-628

– Yanai, T., Sakai, H., Goto, S., Murata, K. & Masegi, T. (2002) Tumors in Zoo Animals. *Japanese Journal of Zoo and Wildlife Medicine*. 7 (**1**), 45-51

– Yoshimoto, T., Niimi, K. & Takahashi, E. (2016) Tranexamic Acid and Supportive Measures to Treat Wasting Marmoset Syndrome. *Comparative Medicine*. 66 (**6**), 468-473

6. Bibliography

– Greig, D.J., Gulland, F., & Johnson, C.K. (2005) A Decade of Live California Sea Lion (*Zalophus californianus*) Strandings Along the Central California Coast: Causes and Trends, 1991-2000. *Aquatic Mammals*. 31 (**1**), 11-22

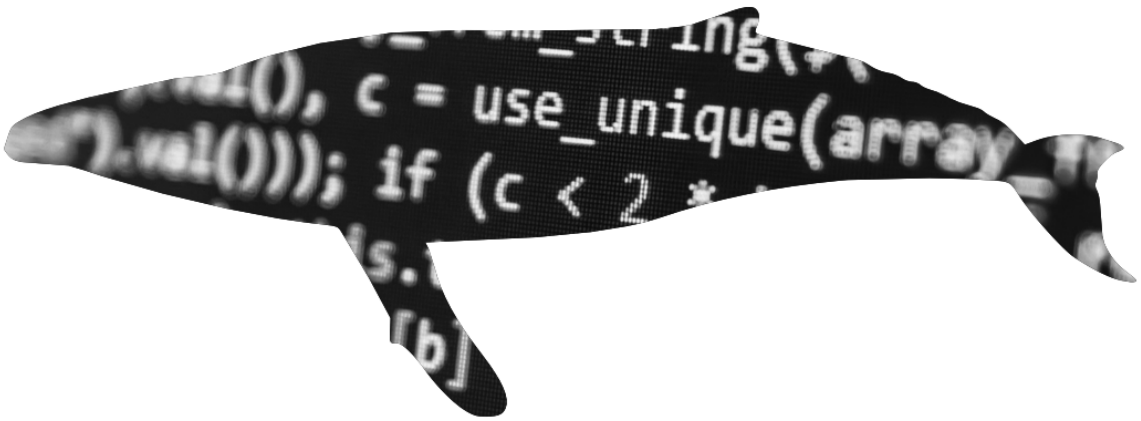
– Roberts, R.M., Green, J.A. & Schulz, L.C. (2016) The Evolution of the Placenta. *Reproduction*. 152 (**5**), R179-R189

– Terio, K.A., McAloose, D. & Leger, J.S. (eds.) (2018) *Pathology of Wildlife and Zoo Animals*. Academic Press.

– Wildman, D.E., Chen, C., Erez, O., Grossman, L.I., Goodman, M. & Romero, R. (2006) Evolution of the mammalian placenta revealed by phylogenetic analysis. *PNAS*. 103 (**8n**), 3203-3208

Chapter 4

Modelling Cancer Resistance in Mammals



Abstract

Metabolic theory suggests cellular metabolic rate and, subsequently, oxidative stress, decreases with increasing body size and longevity, potentially providing an intrinsic physiological protective mechanism against carcinogenesis. Here, three models were developed to assess this hypothesis; an i) algebraic (Model I), and ii) stochastic model (Model II) of multistage cancer progression; and iii) a mechanistic differential model of cellular oxidative stress (Model III).

Models I and II demonstrate that necessary scaling changes in metabolic rate are sufficient to explain the variance in incidence rates across most mammal taxa - supported by predicted outcomes of Model III which substantiate a mechanistic relationship between metabolic cellular factors and DNA damage - with exception to those species at the extreme ends of metabo-physiological scaling. Notably in Cetacea, the incorporation of allometrically scaling metabolic parameters was insufficient to either present nor explain the relatively low cancer incidence rates observed in wild populations, highlighting how additional protective factors must be necessary and present in cetacean species.

In contrast, increase in the number of rate-limiting steps - consistent with natural selection towards greater tumour-suppressor gene or network redundancy - did demonstrably suppress modelled cancer incidence risk in large, long-lived mammals, supporting the hypothesis that increase in genetic control can sufficiently explain observed incidence rates in cancer resistant species. This result supports the possibility that informed study of cetacean genetic and epigenetic control may reveal novel anticancer mechanisms.

Introduction

1.1. Multistage Model of Carcinogenesis

Originally developed in the 1950s, notably by Nordling (1953) and Armitage & Doll (1954), and modified numerous times since (see Hornsby, Page & Tomlinson, 2007, for review), the multistage model of carcinogenesis remains the foundation upon which our understanding of cancer initiation still stands. Though details may differ, all multistage models are based on the principle that cancer arises following the accumulation of several requisite somatic mutations within a single cell, usually gained serially over several generations. When used to quantify cancer incidence as a function of mutation risk resulting from increase in longevity and/or body size, the multistage model can be defined by Eq.1 (per Nunney, 1999):

$$P_t = 1 - \left(1 - \prod_{i=1}^{m_d} \left\{1 - exp(-(1 + D_i)^{u_i tk)}\right\}\right)^C \tag{1}$$

Eq. 1: In tissue consisting of C cells, which exhibits minimal post-developmental growth, and self-replenishing via cell division at rate k, the probability of cancer initiation (Pt) at a given time t is approximated, where m_d is the required number of driver mutations for cancer initiation, and u_i is the per allele rate of mutation (where $D_i = 1$ if the mutant is dominant, as in proto-oncogenes, and $D_i = 0$ when recessive, as with tumour-suppressor genes).

Peto's paradox arises from a false prediction presented by the multistage model: that, all else being equal, increase in somatic mutation (via larger bodies or longer lifespans) results in an increase in cancer incidence. In order to therefore resolve Peto's paradox, either; i) the basic assumptions of the multistage model are incorrect - i.e. that neither increased longevity nor body size drive increased somatic mutation risk; or ii) a solution lies in the alteration of key parameters resulting from evolutionary processes. Epidemiological evidence demonstrates increasing cancer risk with longevity. Further, within a single species, evidence from studies on both domesticated dogs (Fleming, Creevy & Promislow, 2011) and humans (Nunney, 2018; Green *et al.*, 2011) indicates a positive size-incidence relationship. Consequently, the assumptions of the multistage model are valid, and a resolution to Peto's paradox lies elsewhere - in appropriate parameterisation.

In the previous chapters, both intrinsic basal metabolic rate (BMR), tied to life history strategy, and molecular modification to increase rate-limiting steps (i.e. increased tumour-suppressor gene redundancy) were considered as potential resolutions to Peto's paradox. Computational models of cancer risk can be used to estimate the influence that changing parameter variables based on the above hypotheses has on lifetime cancer risk, and, by comparing against known variables and incidence data, evaluate the plausibility of said hypotheses.

1.2. Comparative Oncology *in silico*

Many mathematical models of cancer progression and estimated risk have been published, evolving the original multistage models to those more elaborate, incorporating multiple hits, rate-limiting steps and genomic instability (Beerenwinkel *et al*., 2007; Calabrese, Tavaré & Shibata, 2004; Frank, 2007), though only a couple have, to date, been applied in a broad evolutionary or comparative oncological context.

Repurposing a simple multistage algebraic model (Fig.1a) described by Calabrese & Shibata (2010) to model colorectal cancer risk, Caulin *et al.* (2015) explored altering parameter values for mutation rates (u), number of rate-limiting mutations (m_d), and size and division variables for colonic crypts (~C and k), to demonstrate change required for estimated lifetime colorectal cancer risk in blue whales to present similarly to humans - all else being equal, either 8 rate-limiting steps m_d, a lowered mutation rate u of 3.13×10^{-10}, or a replication cycle of ~13 days is enough to suppress colorectal cancer risk in an organism with a x1000 body size that of a human. Further, beyond a simple algebraic model, Beerenwinkel *et al.* (2007) developed a more complex model of colorectal cancer progression incorporating multiple cancer-associated genes, mutant selective advantage, and cellular growth dynamics and death, providing more realistic time-associated estimates of cancer risk via a Wright-Fisher process. As with the algebraic model, Caulin *et al.* (2015) likewise applied a simplified version of the Wright-Fisher model (Fig.1b) to simulate cancer progression and estimate lifetime cancer risk in a blue whale - predicting ~30% of surviving colorectal stem cells having initiated cancer by age 90.

Beyond defining broad parameter value space ('working backwards') and demonstrating possible values lie within biologically plausible ranges, no published literature applies cancer progression models using known, or robustly derived, initial values ('working forwards') widely across taxa in order to

evaluate hypotheses. Further, no causal mechanistic models have similarly been developed and applied across taxa with respect to non-linear scaling of metabolic processes and the associated influence on DNA damage and mutation rates.

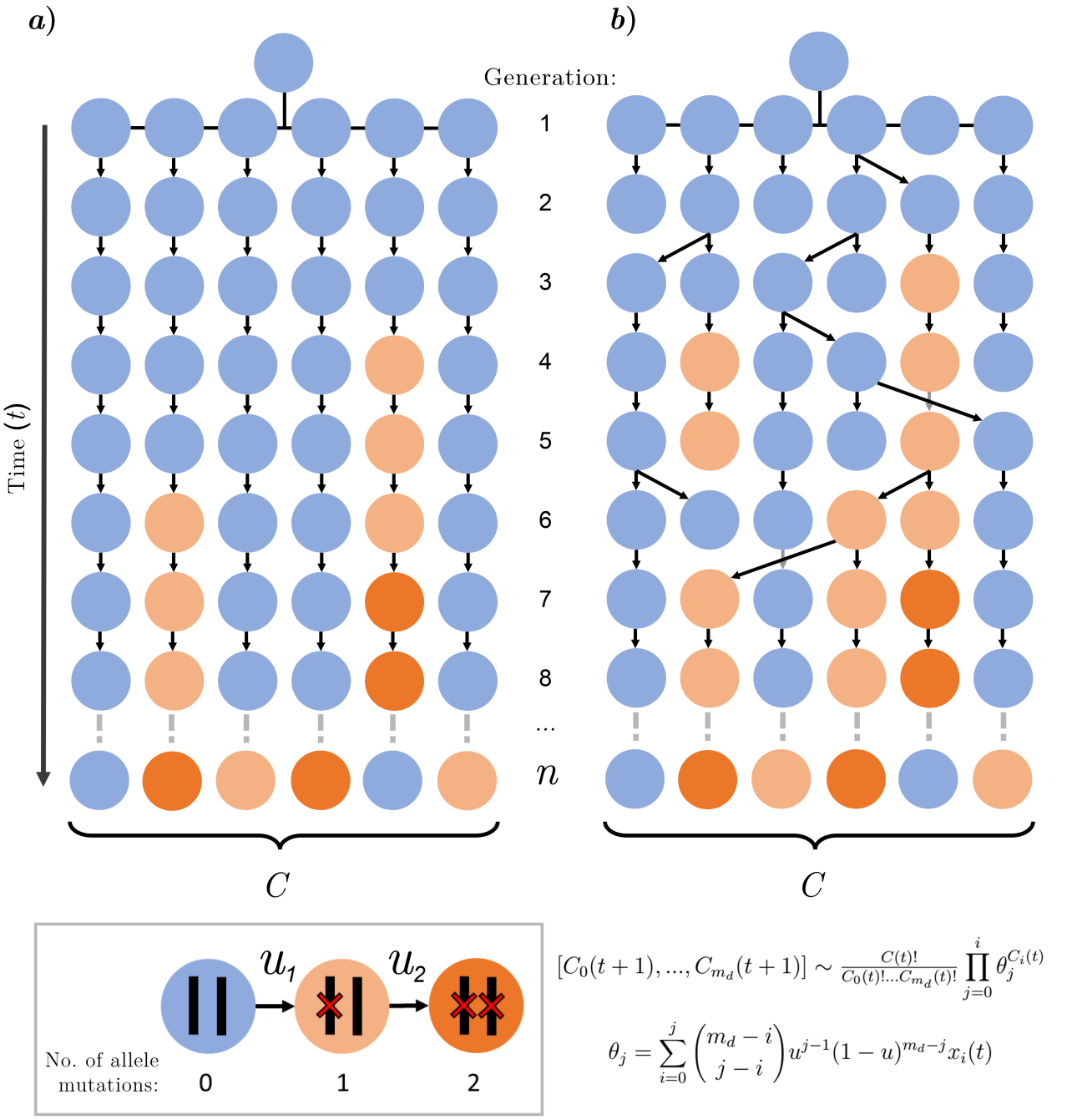


Fig. 1: Two alternative multistage models of cancer progression, each representing a fixed population of cells C, with discrete non-overlapping generations, where cells self-renew at rate k, producing a single surviving daughter cell, where mutations accumulate in a lineage at rate u_n until generation n at time t. **(a)** A simple algebraic model (e.g. per Calabrese & Shibata, 2010, and Model I utilised in this chapter; see Methods 2). **(b)** A Wright-Fisher model of neutral genetic drift. In each generation, parents are chosen via random sampling with replacement. At a given time $(t+1)$, the proportional probability of cells having between no (C_0) and maximum (i.e. cancer initiating) mutations (C_{m_d}) is described by a multinomial distribution, where θ_j is the probability that a cell in the population will have j mutations (and x_i is the proportion of cells at time t with i mutations). As implemented in this chapter, tumour-suppressor gene mutation rate u differs in value depending on whether it is the first u_1 or second u_2 hit.

1.3. Metabolic Processes, Mitochondria & ROS

As overviewed in Chapter 2, hypothesised causal mechanisms relating metabolic processes to cancer incidence, though diverse, remain only partially understood (for full review, see Dang, 2015).

A primary driver of genetic and epigenetic mutation driving tumorigenesis is exposure to reactive oxygen species (ROS) (see comprehensive reviews by Ježek *et al*., 2018; Rodic & Vincent, 2017; and Idelchik *et al*., 2017), a group of short-lived, highly reactive oxygen-containing molecules (e.g. superoxides $^{-}O_2$, peroxides H_2O_2, hydroxyl radicals •OH) that directly induce DNA damage (Dizdaroglu *et al*., 2002). ROS are produced as normal, albeit hazardous, by-products of respirative metabolic processes (and indeed, under balanced control, provide important functional roles, e.g. regulating signalling pathways; Zhang *et al*., 2016; Sena & Chandel, 2012), however an imbalance results in damaging oxidative stress. ROS quantity and concentration in a cell, determined via production *vs* removal rate, therefore offers a simple modellable causal mechanism linking cellular metabolic rate (and subsequently whole-organism metabolic rate), which directly governs ROS production, to rates of DNA damage and cancer incidence.

ROS are neutralised by anti-oxidant molecules, ranging from glutathione and peroxiredoxins (Mari *et al*., 2009; Cao *et al*., 2007), through to anti-oxidant transcriptional responses that upregulate production of enzymes such as superoxide dimutates, which converts superoxide to hydrogen peroxide, and catalase, which then converts peroxide into harmless oxygen and water (Sun *et al*., 2015; Bai & Cedarbaum, 2001). Many of these transcriptional pathways are regulated by several tumour-suppressor genes (Leinonen *et al*., 2014; Kansanen *et al*., 2013), providing the basis for a positive feedback loop whereby damaged tumour-suppressor genes no longer subdue radical production as effectively, promoting further damage.

Mitochondria, endosymbiotic organelles responsible for energy production in the form of ATP, are the primary source of cellular ROS. Being endosymbiotic, mitochondria have retained their own genetic material (mtDNA), with each mitochondrion containing an estimated 1-15 mtDNA copies, or 4.6 on average (Wiesner *et al*., 1992; Satoh & Kuroiwa, 1991). This genetic material specifies proteins of the oxidative phosphorylation pathway, with all other mitochondrial proteins, including those required for mitochondrial replication, encoded in the nuclear genome (Taanman, 1999). Being in close proximity to the epicentre of oxidative metabolism, mtDNA is particularly vulnerable to oxidative damage,

resulting in mitochondrial dysfunction, higher ROS production, and membrane fragmentation. Given the nuclear origin of replicative machinery, a mitochondrion containing damaged mtDNA will still propagate and, over time, there is a risk the proportion of healthy mitochondria will be overtaken by damaged organelles. As damage to the mitogenome results in increased ROS generation (with highest rates at 10-30 fold normal respiratory levels; Turrens, 2003; Bandy & Davison, 1990), it follows the greater the collective damage to cellular mitochondria, the higher the risk of nuclear damage from ROS production, as reactive species diffuse out of the organelle into the rest of the cytosol - therefore, inhibitive processes aside (e.g. apoptosis in response to high ROS levels; Zou *et al*., 2017), the higher the risk of cancer.

Theoretical models of mitochondrial oxidative stress are well developed and often sophisticated, particularly within the context of free-radical and other theories of ageing (though not without criticism; see: Liochev, 2013; Sanz & Stefanatos, 2008; Wickens, 2001), however little to no work has applied nor integrated similar acute cellular and subcellular models in a broader life-history and evolutionary context.

In this chapter, quantitative approaches are utilised to explore these hypotheses utilising three separate models:

i) Model I: A simple algebraic multistage model is used to set a benchmark control for cancer incidence probability distributions across size and longevity morphospace (i.e. "default Peto's paradox") against which subsequent models can be compared. Further, a modified algebraic model is used examine the influence of allometrically scaling BMR and rate-limiting genetic alteration on cancer progression.

ii) Model II: A novel stochastic model of cancer progression utilising Bayesian and coalescence methodology is developed to, as in Model I, examine parameter morphospace required to suppress incidence to observed levels, with and without metabolic scaling. Using observed values from ~178+ mammal species as input parameters, both Models are then used to predict cross-taxa incidence rates and compared to known values.

iii) Model III: A mechanistic differential model of subcellular mitochondrial dynamics and ROS production is used to estimate DNA damage risk as a function of changing cellular metabolism.

Methods

All mathematical modelling and computational simulation was, unless otherwise stated, written and run in R-3.3.0 and later. See SI.4 for annotated R scripts, set-up and running instructions, and required package details and citation. Code sequences referenced in the main text are cited by script ID and line number.

Undertaking meaningful cross-species comparison by estimating whole organism cancer progression is unlikely to be informative, for one given body tissue composition (and therefore metabolically active, routinely dividing cell populations more likely to undergo carcinogenesis) can differ considerably between taxa - e.g. comparing a human with mean body fat ~13% to a baleen whale, where up to ~30% of total body mass alone can comprise of metabolically inert adipose tissue, or blubber (Nishiwaki 1950; Lockyer 1976). Instead, crude estimates aside, focus was made on modelling defined cancers in specific tissue.

In both models I and II, I focused on simulating two cancer types: colorectal and lung adenocarcinomas - given the relative abundance and availability of data, differing between each type, with which to calibrate input parameters; and these cancers ubiquitous pathological record across taxa.

Epithelial tissue, from which the above cancers originate, is the most replicatively active tissue in the body, comprised of actively cycling cells with high turnover throughout adult life (Pastula & Marcinkiewicz, 2019). Organs are not homogenous masses of cells - rather, they are rapidly partitioned off during embryogenesis into compartments at the tissue level, with limited intercompartmental exchange between them (Pierson, 2012). Epithelial tissue is compartmentalised into functional units derived from a single origin (e.g. crypts of Lieberkühn in the gastrointestinal tract, bronchioalveolar ducts in the lungs). Terminally differentiated epithelial cells, with short lifespans, comprise the bulk of cells within a compartment, derived from a small number of proliferative transit amplifying (TA) cells (Blanpain, Horsley & Fuchs, 2008). These, in turn, are derived from a smaller population of multipotent epithelial stem cells (ESCs), defined by their ability to self-perpetuate long-term, and given their longevity are considered the progenitors of the majority of double-hit TSG mutants and therefore cancer (Carulli *et al.*, 2014; see Barker, 2014 for full review), in contrast to, for example, differentiated epithelia which don't persist for long enough for mutations to typically accumulate

(Komarova & Wang, 2004; Cairns, 2002). Estimates of cancer incidence therefore need only focus on modelling carcinogenic progression in stem cell niches, as all other (derived) cells do not contribute significantly towards cancer risk. Unless otherwise specified (i.e. incorporating developmental growth), when simulating organ systems, compartment number per species is assumed fixed, and no budding nor exchange of cells occurs between compartments.

2.1. Parameters & Scaling Relationships

All models required life history trait values for mean adult body mass, maximum longevity and basal metabolic rate (BMR) for a range of mammals. When analyses prompted use of data from animal taxa, these values were obtained from the dataset established in the previous chapter (see Chapter 3: 2.2, and S.I.3.4 and 3.8 for details and data), else estimates were derived from established scaling relationships. Further, values were required for cellular metabolic rate, division rate, tissue compartment size and number, and proto-oncogene and tumour-suppressor gene number and allele-specific mutation rates.

The relationship between body size and BMR has historically been defined by a linear log-log relationship (Eq.2), the precise value of the slope subject to intense debate (typically between 0.65 and 0.75; White & Seymour, 2004). More recently, a curvilinear relationship has been adopted (see Eq.3) (Bueno & López-Urrutia, 2014; Kolokotrones *et al.*, 2010) and will be used here, given the better fit across a wide range of taxa and, consequently, change in metabolic rate with body size is more pronounced for larger mammals compared to smaller ones.

$$\log B = \alpha + B_0 \log M \tag{2}$$

$$\log B = \beta_0 + \beta_1 \log M + \beta_2 (\log M)^2 + \epsilon \tag{3}$$

$$x = \beta_1 + 2\beta_2 (\ln M) \tag{4}$$

$$V_c \propto m_c \propto M^0 \ , \ B_c \propto M^{-(1-x)} \ , \ N_c \propto M \ , \ t_c \propto M^{1-x} \tag{5}$$

Eq. 2: Kleiber (1932) suggested basal metabolic rate B scaled with body mass M across animal species, concluding the log-log relationship was described by a linear power (B_0) approximately equal to ¾, since known as Kleiber's Law: $B \propto M^{3/4}$.

Eq. 3 & 4: : Kolokotrones *et al.* (2010) conclude the relationship between basal metabolic rate B and body mass M is described by convex curvature and a quadratic polynomial model (3), where β_0 is the logarithm of B_0 in Eq.2, and ϵ is the error term. This equation implies the log-log slope for B vs. M is a non-constant value that increases with body size (4): the local scaling exponent, defined as the derivative of the scaling relationship (x), increases significantly when applied to a mammalian dataset - from 0.57 to 0.87 - and accounts for a significantly higher amount of variance than linear models (AIC(Eq.3)=-415.13 vs. AIC(Eq.2)=-331.28; see SI.3.4 for details).

Eq. 5: Per Savage *et al.* (2007), average cell size and average cellular metabolic rate cannot both remain constant with body size due to the well-established body-size dependence of whole-organism metabolic rate. Given that cellular metabolic rate B_c and cell mass m_c cannot both remain constant as body mass M varies, cells adopt one of two alternative strategies ('fast' and 'slow'). Epithelial cell lineages adopt the 'fast' strategy whereby average cell mass remains fixed (along with cell volume V_c), whereas average *in vivo* cellular metabolic rate and generation time t_c varies, so cell number scales linearly with body mass, whilst cellular metabolism scales inversely.

Under the assumption that stem cell size and density is invariant with body size; cellular metabolic and division rate, and stem cell compartment size and number were estimated using the scaling relationships defined by Savage *et al.* (2007) (Eq.5), where the exponent x, otherwise described by a metabolically-associated constant slope value (i.e. ¾ or ⅔), is substituted by the non-constant exponent $(\beta_1 + 2\beta_2(\ln M))$ from Eq.4. It is assumed organ size, and therefore number of compartments N containing c cells, scales proportionally with body size. Human data was used as a baseline from which to derive estimates - scaling coefficients calibrated where available with known values from mouse model data - and to define a standard for comparison.

Per Gilloly *et al.* (2005, 2007), as cellular division rate has the same size and temperature dependence as cellular metabolic rate, the mutation rate per generation is independent of metabolic rate. For all simulations the rate of mutation of a single gene of ~1,000nt base pairs is, unless otherwise stated, given as 10^{-6} per generation for proto-oncogenes and the first TSG allele, and 10^{-5} per generation for the second TSG allele - the second TSG allele is more likely to mutate since point mutations, small insertions and deletions, structural changes of the chromosome or entire chromosomal loss can constitute the first hit, whereas all of these events plus mitotic recombination can occur as the second hit. Large deletions resulting in the loss of both alleles in a single event are not considered, as large homozygous deletions are often lethal and no viable daughter cells will be produced (Pleasance *et al.*, 2010).

For the colon, Pickhardt *et al*. (2005) estimate the number of colonic crypts in a human as ~$1.5x10^7$, whereas, in a mouse, they estimate around $3x10^4$ colonic crypts, with crypt density taken at approximately 5,000 crypts per cm^2. In the same manner, using stereological approaches, Ochs *et al*. (2004) and Knust *et al*. (2009), estimate ~$4.8x10^7$ alveoli in the human lung, and ~$2.3x10^6$ in the mouse, with stem cells occurring in the cleft between alveoli and bronchiole (Pine *et al*., 2008; Lindstedt & Schaffer, 2002).

2.2. Algebraic Multistage Model (Model I)

2.2.1 Basic Model

The algebraic multistage model used here, henceforth denoted Model I, considers and assumes the following biological properties (illustrated in Fig.1a):

- Epithelial tissue is comprised of independent compartments.
- Compartment number and size remains constant throughout lifespan (no tissue growth).
- Cells turnover at a constant rate per unit time
- Cell lineages are immortal, do not die, and are not replaced. A single cell only produces one surviving daughter cell.
- Mutations arise and accumulate in a cell at a constant probability rate per unit time.
- Mutations present no selective advantage nor other change in cellular behaviour.
- Only once a defined number of mutations have accumulated is cancer considered to have initiated in a cell. Once initiated, cancer persists in a cell lineage thereafter.

$$P_t = 1 - \left(1 - \prod_{i=1}^{m_d} \left\{1 - exp(-(1 + D_i)^{u_i t k})\right\}^c\right)^N \tag{6}$$

$$N \propto M \quad , \quad k' \propto M^{1-x} \quad , \quad u' \propto B_c \propto M^{-(1-x)} \tag{7}$$

Eq. 6 & 7: The basic multistage model (Eq.1) is first modified to include tissue compartmentalisation of given number N, each compartment comprising an independent population of stem cells c (6), where u and k act as metabolic-associated variables in order to test hypotheses. From (Eq.5), given cellular metabolic rate scales as $B_c \propto M^{1-(1-x)}$, time scales determined by cellular metabolism, including cell lifespan and cell-cycle time, must scale inversely as $t_c \propto M^{1-x}$ (e.g. cell turnover is slower in larger organisms with lower mass-specific cellular metabolic rates). In order to test the influence of intrinsic metabolism, we can thus substitute k' and u' into Eq.6, where $k \equiv k' = t_c$ and, assuming mutation rate is proportional to cellular metabolic rate, $u \equiv u' = B_c$ (7), and both are estimated using a local metabolic scaling exponent x from (Eq.4).

2.2.2 Implementation & Hypothesis Testing

In order to account for variation in proto-oncogene and TSG number (influencing D_i), and the different associated value for u_i per TSG allele, this model (Eq.6) was computationally expressed via a looping function (SI.4.2, lines 5-21). Parameter values for k',u' and N were similarly estimated (see SI.4.1, lines 12-21) for use in simulations (see Table SI.4.8 for details).

Baseline parameters were defined for mouse, human and whale models (and further speculatively estimated for an enlarged dataset of 178 mammal species), and simulations were run: i) without any metabolic modification (control); ii) only scaling division rate; iii) only scaling mutation rate; and iv) scaling both. To test the significance of scaled parameter influence on the output cancer incidence distribution, either two-sample or n-sample Kolmogorov–Smirnov paired tests were undertaken, depending on variable number comparison. For each species model, further parameter exploration was undertaken to evaluate plausible value morphospace, including alteration to m_d (the required number of driver mutations), which was undertaken by rearranging the computational implementation of Eq.6 to solve for the closest minimum integer value of m_d following an input P_t equivalent to both: i) human data (i.e. assumed incidence is invariant with body size); and ii) observed cancer incidence rates per species (collated in Chapter 3).

2.3. Stochastic Multistage Bayesian Model (Model II)

2.3.1. Basic Model (IIa)

For comparison against Model I, a more sophisticated stochastic model is developed, henceforth denoted as Model II, considering the following biological properties (see also Fig. 2):

- Epithelial tissue is comprised of independent compartments, each a finite population of cells.
- Unlike in Model I, cell generations can overlap, with the processes of cell removal and replacement modelled via an implementation of the Gillespie algorithm (Gillespie, 1977), which allows for computationally inexpensive reproduction of exact stochastic trajectories of a system.
- Mutations similarly accumulate in a cell line at a constant rate per unit time, per generation within a lineage, likewise modelled using a Gillespie algorithm.
- Mutations present no selective advantage nor other change in cellular behaviour.
- Only once a defined number of mutations have accumulated is cancer considered to have been initiated in a cell. Once initiated, cancer persists in a cell lineage thereafter, unless removed.

In variations of the model:

- Compartment number and size either remains i) constant from onset; or ii) chronologically, compartment size begins at size 1 and grows at a given growth rate G_r until a compartment size C is reached at the end of a growth phase.
- Cells may or may not differentiate and be removed from the compartment population, and the resultant lineage modelled (see 3.3 / Model IIb).

2.3.2. Standard Stochastic Model (IIa) Implementation

Model IIa is simulated in two main steps: As we are concerned with modelling rate of cancer incidence through until maximum lifespan (and to explore onset age-incidence outcomes), and in order to minimise computational power required, first, a complete compartmental genealogy is modelled starting from the final population of cells C, working backwards to the most recent common ancestor (MRCA) at the original division event. This coalescence process is neutral, derived from the Unified Neutral Theory (Hubbell, 2001). This is scripted as an iterative looping function (see SI.4.3, lines

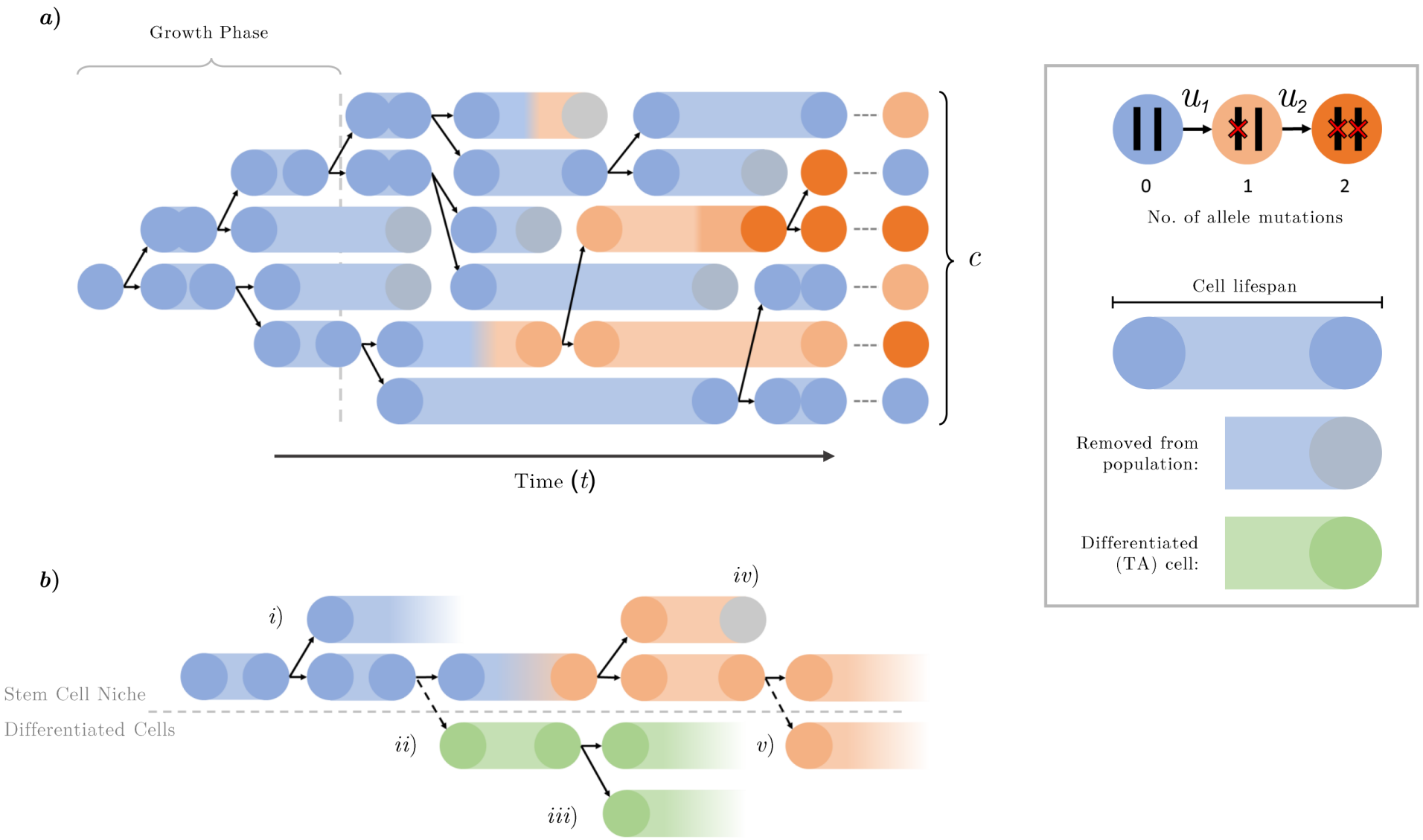
x-x), where the probability of a given cell lineage coalescing is decided by drawing a random number from a uniform distribution (0-1) and compared with the probability of coalescence, λ, (see Eq. 8):

$$\lambda = \frac{n-1}{j-1} \cdot \frac{n}{j} \tag{8}$$

Eq. 8: Probability of a coalescence event occurring (λ), where n is the ratio of sampled linages, and j the compartment size. Each time a lineage coalesces, two daughter cells are chosen at random from the population, and assigned a parent. The sample size is reduced by one, and lambda is re-calculated. If no coalescence event occurs, branch length increases by 1, and the process is repeated (see SI.4.3).

Once a complete lineage is established, working forwards in time on the pruned tree, a mutation rate u is applied. At division events, cells chosen to replicate have a probability of acquiring mutations to n haploid genes, drawn as a Poisson random variable with mean u_1 (proto-oncogene, and first hit TSG) and u_2 (second hit TSG) respectively. Daughter cells inherit all mutations. For the purposes of this implementation, mutations are considered selectively neutral. At the end of a single run, the data for each point along the lineage, and when and where mutations arise, is stored for later analysis.

Fig. 2 (overleaf): (a) Illustration of a stochastic non-spatial simulation of mutation accumulation in a self-renewing finite cell population C. In this model, each cell has an individual generational lifespan, independent of other cells within the compartment. As before, removal and replacement is undertaken by neutral sampling, and mutations accumulate in each lineage at rate u until generation n at time t (end of lifespan). The tumour-suppressor gene mutation rate for the first hit u_1 differs in value from the rate of the second hit u_2. In some implementations of the model, a growth phase is included, beginning at $t = 0$ until population size C is reached, governed by growth rate G_r. Else, simulations begin as in Model I (Fig. 1), with a full population at outset. In the basic model (Model IIa), compartments reflect biological stem cell niches, with the condition of cells removed from the population having no significance (whether mutated or not). In models including differentiated cells (Model IIb) (b), cells may (i) as before, duplicate symmetrically following a removal event, but also; (ii) asymmetrically divide without prior removal of an adjacent cell into one stem and one differentiated daughter, and (iii) differentiated daughters continue to divide within their niche. In the basic model, cells removed following mutation (iv) are not considered; however in differentiated models, a proportion of removal events model division and differentiation, where the daughter differentiated cell retains mutations into the subsequent lineage (v).


a)
Growth Phase
c
Time (t)
b)
i)
ii)
iii)
iv)
v)
Stem Cell Niche
Differentiated Cells
u_1
u_2
0
1
2
No. of allele mutations
Cell lifespan
Removed from population:
Differentiated (TA) cell:

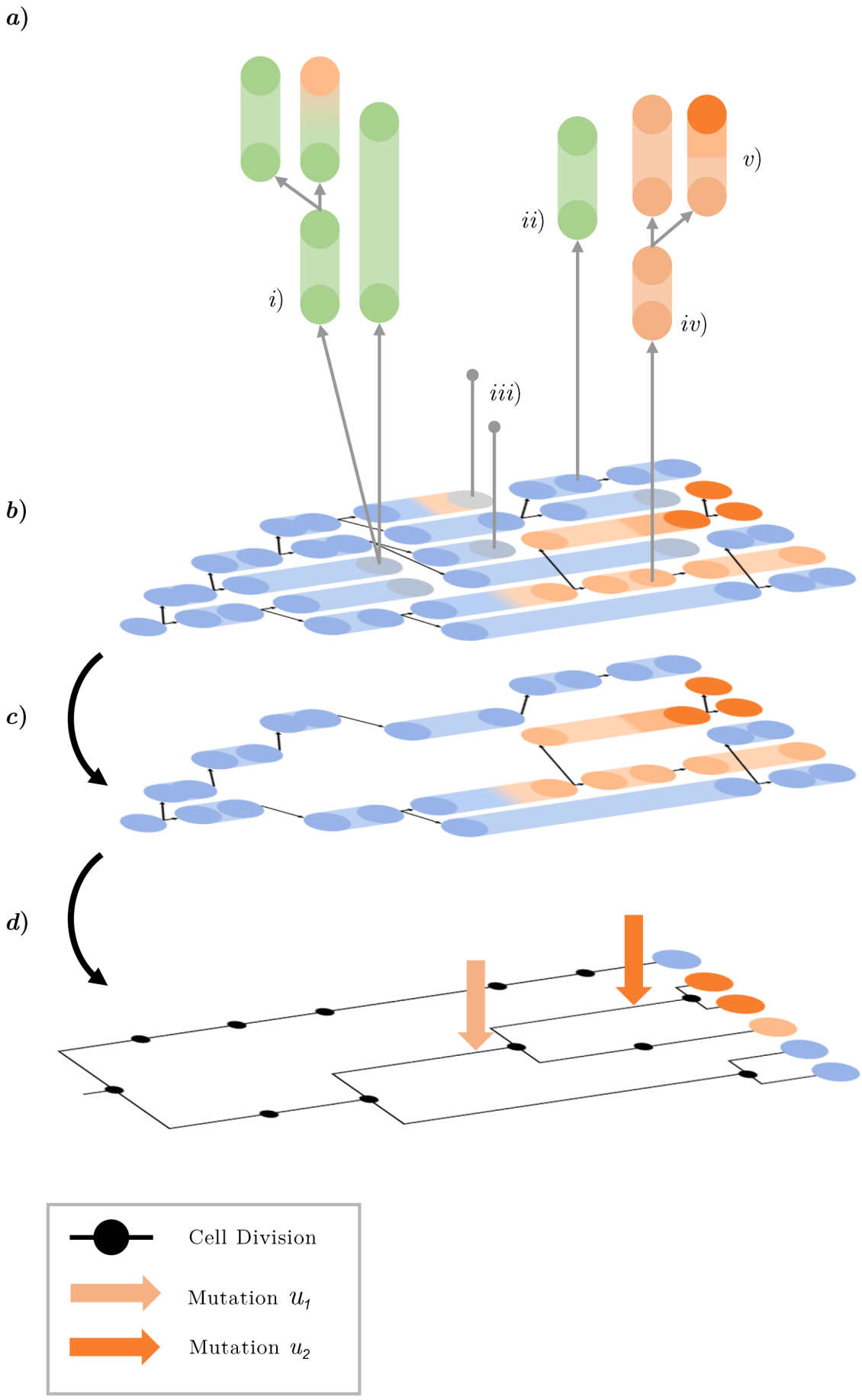

a)
i)
ii)
iii)
iv)
v)
b)
c)
d)
Cell Division
Mutation u_1
Mutation u_2

Fig. 3 (overleaf): A standard compartmental model (b) can be supplemented with additional elements, such as including symmetric and asymmetric division into populations of differentiated cells (as in IIb) (a), albeit with the cost of additional computational expense. Under the standard stochastic model (IIa), given lineages destined to be removed from the population are not considered, simulations are instead modelled in inverse via coalescent theory (c), as a generated cladogram (d) where only cell division and mutation events pertinent to the final compartment population, drawn from probabilistic distributions, need to be computed. (a) By default, stem cells divide asymmetrically into two daughter stem cells, or symmetrically following a removal event, but may also; i) divide symmetrically into differentiated cells; or ii) divide asymmetrically into one stem and one differentiated daughter; if not iii) removed altogether. Differentiated cells can iv) inherit mutations accumulated in the stem cell niche, and v) further accumulate mutations with subsequent division events. Given the need to model removal / differentiation events from the stem cell niche population into individual differentiated cell lineages, a stripped down coalescent approach is less suitable, and simulations are undertaken via a Moran process.

2.3.3. Stem Cell Dynamics and Differentiation (Model IIb)

In order to appropriately scale computation, the standard stochastic model (IIa) assumes a fixed stem cell population size, and cell lineages lacking progeny that persist until end of lifespan have no consequence - that neutral selection and removal of said lineages, even if they would otherwise host accumulated mutations, are biologically non-significant. As such, in effect, cells within the population divide asymmetrically to maintain a fixed population size, punctuated by moments of stochastic lineage removal where a single cell divides symmetrically to replace the removed lineage. These dynamics however are not representative of the understood dynamics of crypt stem cells. Recent evidence from genetic labelling has demonstrated symmetric division contributes significantly to maintaining a fluctuating stem cell population, and numerical analyses incorporating stochastic symmetric division for proliferation and differentiation suggest intestinal crypt two-hit mutation rates are significantly reduced compared to strictly (or near-strictly, as with a metabolic model) asymmetric division (Shahriyari & Komarova, 2013; Lopez-Garcia *et al*., 2010).

Furthermore, despite being proliferative, it is often argued that transit-amplifying (TA) cells are unimportant for neoplastic initiation, as they differentiate, get sloughed away, and are replenished too rapidly for any two-hit mutants to arise and be of clinical significance (Komarova & Wang, 2004;

Cairns, 2002). It has been shown, however, that the small probability of any individual TA cell producing a second hit is outweighed by their relative abundance with respect to stem cells, and can contribute cancer-progressing mutations under sets of reasonable biological conditions (Shahriyari & Komarova, 2013), perhaps having significant effect on estimated cancer incidence.

In order to explore what effect, if any, inclusion of additional stem cell dynamics and differentiation has on estimates for cancer incidence under a stochastic model, the basic model (Model II) will be supplemented to include appropriate dynamics for comparison (Model IIb).

2.3.4. Differentiation Model (IIb) Implementation

Unlike in Model IIa, where lineages are modelled via neutral coalescent theory, given the need to consider lineage branching from stem to TA cell populations, a forward-moving Moran-like process is utilised (Moran, 1958). Crypts are compartmentalised into populations of stem and TA cells, which, as before, can each undergo mutations with a given probability u_1 and u_2 depending on prior allele set condition. Each compartment is therefore composed of six different cell types: stem cells (wild-type i_0, one-hit mutants, i_1, and two-hit mutants, i_2) and TA cells (wild-type j_0, one-hit mutants, j_1, and two-hit mutants, j_2) (see Eq. 9):

$$N = (i_0 + j_0) + r_1(i_1 + j_1) + r_2(i_2 + j_2) \tag{9}$$

Eq. 9: Total compartment population N remains fixed as each removed TA cell j_n is replaced by a daughter cell of one of six cell types, with a likelihood proportional to their relative abundance. It is assumed stem and TA cells are considered well-mixed, such that the choice of dividing cell (mutant or not) to replace the outgoing TA cell is random (or selectively neutral) per division event. The probability of a wild-type TA cell dividing is j_0/N; of a one-hit mutant dividing is $r_1 j_1/N$; of a two-hit mutant dividing is $r_2 j_2/N$. If a wild-type TA cell divides, it produces another wild-type cell with probability $1 - u_1$, mutating to a one-hit TA cell with probability u_1. If a one-hit mutant divides, it produces another one-hit mutant with probability $1 - u_2$ and a two-hit mutant with probability u_2.

Stem cells are capable of symmetric and asymmetric division (see Fig. 2b). The probability of a stem cell dividing symmetrically is given as parameter σ, whereby $\sigma = 0$ results in all divisions occurring asymmetrically, and $\sigma = 1$ results in all divisions occurring symmetrically. When stem cells divide symmetrically, two outcomes are possible: proliferation or differentiation. The likelihood of either is governed by a regulatory mechanism to ensure stable long-term populations of stem cells (see Eq.

10).

$$p = \frac{(i_0+j_0)^{10}}{S^{10}+(i_0+j_0)^{10}} \tag{10}$$

Eq. 10: The probability of stem cell proliferation is taken to be p, where S is a constant parameter which measures the expected number of stem cells in the system, acting as a regulatory mechanism to ensure a self-renewing stem cell population (Shahriyari & Komarova, 2013).

These updates were performed sequentially for the given expected number of stem cell generations for the maximum lifespan of the species.

2.3.5. Computation and Parallelisation

To increase simulation efficiency, a Bayesian approach is utilised where after a defined threshold value, simulations switch from performing neutral coalescence to sampling using the Gillespie algorithm (Gillespie, 1977). A Poisson point process is used to determine branch lengths between coalescence events, approximated by iteratively drawing random numbers and obtaining corresponding values from a cumulative exponential distribution, allowing for accelerated simulation. In order to further decrease simulation time, the final function was wrapped in a parallelisation process (see sIX), allowing multiple simulations to be run simultaneously utilising multiple CPU cores, using the foreach (Wallig, 2020a) and doMC (Wallig, 2020b) packages.

Simulations were run in blocks on high performance computing (HPC) suites, between Imperial College London's cx1 cluster and the University of Oxford's arcus-b cluster. Owing to their lower tractability, a limited number of IIb runs were initiated, and cross-compared with preliminary results from equivalent IIa runs, in order to determine IIa assumption validity. Thereafter IIa runs were undertaken in two series of experiments; i) as with Model I, using baseline parameter values defined for mouse, human and whale models, to explore parameter morphospace with and without metabolic scaling; and ii) on the available parameter dataset of 178 mammal species in order to estimate cross-taxa cancer incidence rates.

2.3.6. Hypothesis Testing and Analysis

By combining the four incidence datasets for each species - Model I default and scaled, and Model II default and scaled - in different ways, setting model type as an indicator dummy variable in each

case, a series of logistic regression fitting exercises could be undertaken and analysed. Unless otherwise stated, as we are testing to see whether model results differ significantly from a default state of 'Peto's paradox' (here defined as the default non-scaled results from Model I for each species), and not testing against a standard null hypothesis (slope / β_0 = 0), each analysis was modified in R using the 'offset' function, taking the default null logistic model ($\ln \frac{\pi}{1-\pi} = \beta_0$) and substituting in the predictor and β_n coefficients associated with the "Peto's null" Model I. The influence of alteration to division rate and cellular metabolic rate on each model was tested independently using a Wald test (where if $p>0.05$, removing the variable does not harm the fit of the model), else both together using likelihood ratio tests on combined data and a dummy variable signifying model type. Given the algebraic nature of Model I, the error variance differs appreciably between results from Model I and II, though comparison between the degree of influence of metabolic scaling between each model was not undertaken.

2.4. Algebraic Differential Mitochondrial Model (Model III)

A mechanistic model, defined as a series of differential equations, exploring the relationship between cellular metabolism, ROS production, and cancer incidence risk via DNA damage, hereafter known as Model III, takes into account and addresses the following biological properties (see also Fig.4):

- Mitochondria are polyploid.
- Mitochondria produce ATP.
- Mitochondria produce ROS.
- Mitochondria suffer mtDNA damage depending on level of exposure to ROS.
- ROS production increases, whilst ATP production decreases, following mtDNA damage.
- Mitochondria multiply via division and daughter organelles receive a random distribution of mitochondrial genomes.
- Mitochondrial growth rate, and therefore ultimately division rate, is governed by cellular ATP concentration.
- There is a turn-over of mitochondria via degradation and replacement.
- ROS can diffuse out of the mitochondria and attack nuclear DNA, at a given diffusion constant.
- Anti-oxidant neutralisation mechanisms remove ROS from the system at a given rate.
- Growth rate and ATP consumption rate are proportional to mass-specific basal metabolic rate.

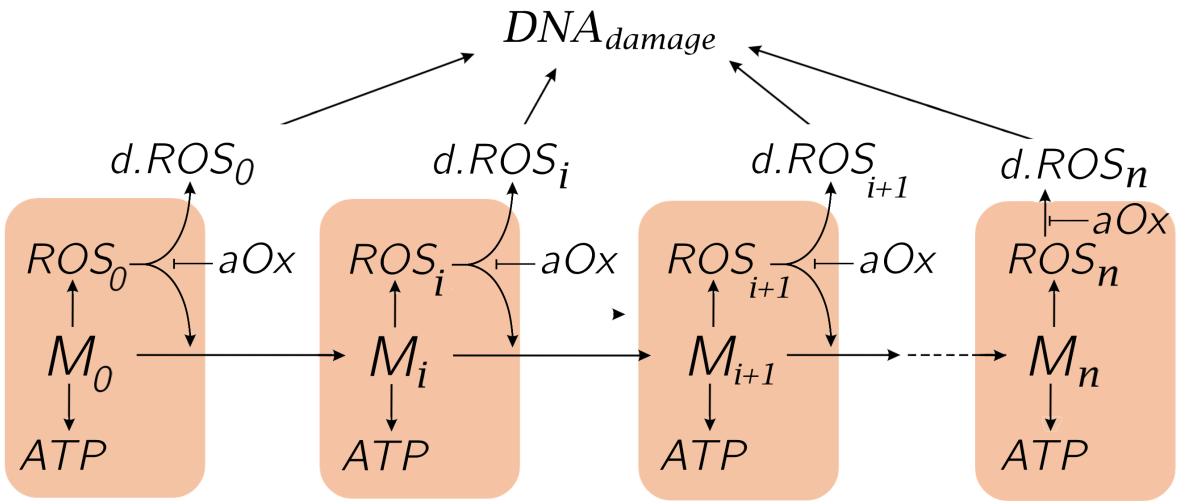


Fig. 4: Mitochondria produce ATP and ROS, with ROS contributing to degradation of both mtDNA and nuclear DNA. As individual mtDNA molecules become damaged, mitochondira are described as moving to a higher damage class ($M_{0\to n}$), with differing production rates of ATP and ROS ($ROS_{0\to n}$). ROS diffuse out of the mitochondria ($d.ROS_i$) and promote nuclear DNA damage. Anti-oxidant mechanisms remove ROS from the system (aOx). Cellular metabolic rate determines rate of ATP production and depletion.

2.4.1 Preliminary Equations

The polyploid mitochondrial genome consists of *n* copies of mtDNA, each mtDNA being regarded as a target for damage, occurring at a certain probability, giving rise to *n*+1 classes of mitochondria: from M_0 denoting no mtDNA damage, through M_1... M_n denoting $1-n$ copies damaged respectively.

2.4.1.1. Mitochondrial Damage and Transcription Rate

Synthesis of active proteins used in respiration, replication and other biological processes decreases as a consequence of damage to individual mtDNA molecules (Cline, 2012; Cui, Kong & Zhang, 2012). It is assumed mitochondria attempt to compensate for this decrease by feedback mechanisms that enhance transcription rates (Pich *et al.*, 2004), described by Eq.11:

$$IT = \frac{i}{n}(IT_{max} - 1) + 1 \tag{11}$$

Eq. 11: Alteration to transcription rate (*IT*) following oxidative stress, if $\frac{i}{n}$ is the fraction of damaged mtDNA molecules (0-1), and assuming transcription rate increases linearly with $\frac{i}{n}$. $IT_{\max}$ is the maximum amount by which the transcription rate can be increased due to biological limits, for $\frac{i}{n}$ = 1. As a consequence of altered transcription rates, the amount of ATP synthesised by a given mitochondria is a fraction of the ATP synthesised by a healthy mitochondrion (M_0), given by $(1 - \frac{i}{n}).IT$

2.4.1.2. Mutation and Radical Production Rates

Undamaged mitochondria ($\frac{i}{n} = 0$) produce oxygen radicals at a rate ROS_0 per second. For those mitochondria in which all mtDNA is damaged ($\frac{i}{n} = 0$), the increased radical production is proportional to a constant radical enhancement factor (*REF*) (describing the factor of increase) and alteration to transcription and synthesis rates (*IT*), giving ($ROS_0.IT.REF$). Substituting *IT* from Eq.11 gives Eq.12:

$$ROS_{mito} = \frac{i}{n}.ROS_0(IT.REF - 1) + ROS_0 \tag{12}$$

Eq. 12: ROS production increases non-linearly with degrees of $\frac{i}{n}$. A completely damaged mitochondrion ($\frac{i}{n} = 1$) produces *REF* times more ROS than a healthy, undamaged mitochondrion ($\frac{i}{n} = 0$).

ROS have high diffusion rates, due to their small size (see Eq. 20), however a major species, the

superoxide radical, cannot easily cross mitochondrial membranes due to its negative charge (Mumbengegwi *et al.*, 2008). Therefore, different ROS levels will be present in different damage class types of mitochondria - here ROS_0 - ROS_n are associated with classes M_0 - M_n respectively.

Given that ROS instigates mtDNA damage, an increase in radical production is linearly proportional to the probability p that one copy of mtDNA is damaged per unit time, with a ROS level of ROS_0. While p is the probability of one damaged mtDNA per unit time, it is not the same as the probability that a given mitochondrion becomes a mitochondrion of a higher damage class. Due to the polyploid nature of the mitogenome, the probability of a mitochondrion of class i entering class i+1 is:

$$P_i = (n-i).p.(1-p)^{n-1-i} \tag{13}$$

Eq.13: Probability of a mitochondrion of class i entering class i+1. With regards to the dynamics of a given mitochondrion class, this mutation rate is added to the population (with $p.ROS_{i-1}$) when describing a mitochondrion mutating into said class, whilst being subtracted (with $p.ROS_i$) when mtDNA is mutated to a higher class (see Eq.20-23 & Fig. X for details).

2.4.1.3. Growth and Replication

It can be reasonably assumed that the replication of mitochondria in the cell is governed by the availability of energy, determined by the cellular ATP level. Furthermore, damaged mitochondria have a growth and replicative disadvantage proportional to their degree of damage, due to the requirement of mitochondria to transport nuclear-coded mitochondrial proteins through the mitochondrial membranes (Dudek, Rehling & van der Laan, 2013). As damaged mitochondria produce less energy, with further impairments to the membrane potential, said proteins are less able to enter the mitochondrion in order to facilitate growth and replication. This difference in replicative ability between different classes of damaged mitochondrion is defined as Growth Difference (*GD*). Assuming a linear relationship between damage and growth ability, those mitochondria with a given damage of $\frac{i}{n}$ grow $\frac{i}{n}.(GD-1)+1$ times slower than healthy, undamaged mitochondria. During mitochondrial replication, there is a binomial segregation of mtDNA copies to the daughter mitochondria (Taanman, 1999). Although the copy number of mtDNA remains constant within the daughters, the damage level for each respective daughter may be different, depending on their inherited mtDNA. It therefore holds that the probability for a M_i mitochondrion to produce M_j daughters is:

$$p_{i\to j} = \frac{\binom{2n-2i}{n-j}.\binom{2i}{j}}{\binom{2n}{n}} \tag{14}$$

Eq. 14: Where $\binom{a}{b}$ is the standard binomial coefficient notation term for the possible number of selections when choosing *b* from set *a*. For convenience, in the following set of differential equations, ${}^{a}C_{b}$ has been used to describe the combinatorial term $\binom{a}{b}$.

2.4.2 Model Differential Equations: ATP

$$\frac{dATP}{dt} = ATP_0.\sum_{i=0}^{n}(1-\frac{i}{n})IT_i.M_i$$

$$-\sum_{i=0}^{n}\frac{n}{i.(GD-1)+n}.\frac{GR}{1+(\frac{ATP}{ATP_c})^f}rM_i \tag{15}$$

$$-ATP_d\frac{ATP}{ATP+ATP_t}$$

Eq. 15: Concentration of ATP over time can be summarised as: $\frac{dATP}{dt}$ = ATP synthesis by different classes of mitochondria - mitochondrial ATP consumption via growth and replication - cellular ATP consumption. The latter term is required as the overall levels of ATP within the cell as a whole governs the growth and replication of mitochondria via feedback mechanisms (cells low on ATP will invest less in growth), and therefore the behaviour of the model.

This differential equation can be broken down as follows:

2.4.2.1. ATP Synthesis

$$\frac{dATP}{dt}_{synth} = ATP_0.\sum_{i=0}^{n}(1-\frac{i}{n})IT_i.M_i \tag{16}$$

Eq. 16: Here ATP is measured in multiples of ATP_0. As mitochondrial damage increases ($M_0 \to M_n$), despite biological feedback mechanisms which increase ATP production via increased transcription ($IT \to IT_{max}$) (see Eq. 8), ATP synthesis decreases ($ATP = 0$ for M_n)

2.4.2.2. Mitochondrial ATP Consumption

$$\frac{dATP}{dt}_{mit} = -\sum_{i=0}^{n} \frac{n}{i.(GD-1)+n} \cdot \frac{GR}{1+(\frac{ATP}{ATP_c})^f} rM_i \tag{17}$$

Eq. 17: *GD* denotes Growth Difference, *GR* denotes Growth Rate (upper limit reachable) of the entire population per day (approximately $0.06d^{-1}$ or 6% daily), *r* is the level of ATP required for mitochondrial replication and *f* is the negative feedback associated with mitochondrial division. ATP_c refers to the minimum cellular ATP level that the cell tries to maintain, with this steady state level varying with species and tissue. As damaged mitochondria grow more slowly, they use less ATP, reflected in the first term with *GD*. The second term allows for the fact that the higher the level of ATP over the minimum maintained level, the more that is available for growth and replication, and is used accordingly. For both terms, as $M_0 \rightarrow M_n$ and $\frac{i}{n}$ approaches 1, the value of ATP decreases.

2.4.2.3. Cellular ATP Consumption

$$\frac{dATP}{dt}_{cell} = ATP_d \frac{ATP}{ATP + ATP_t} \tag{18}$$

Eq. 18: It is assumed that cellular ATP consumption depends on cellular ATP level, reaching a maximum of ATP_d if there is a surplus of ATP. This value is calculated under the assumption that five molecules of ATP are generated per molecule of oxygen, with specific metabolic rate and cell volume per species governing the calculated value (e.g. for humans, 0.25ml of O_2 per g/h and a cell size of $4000\mu m^3$ give a value of $3.8 \ . \ 10^{12}$. See: Müller *et al.*, 2013). ATP_t acts as a feedback threshold and controls for the decline in cellular activity due to a low level of cellular energy. If the ATP level decreases to ATP_t, then cellular ATP consumption is decreased to 50% its maximum denoted by ATP_d.

2.4.3. Model Differential Equations: Radical Production and DNA Damage

$$\frac{dROS}{dt} = M_0 + \sum_{i=0}^{n} ROS_i \cdot M_i - aOx.ROS \tag{19}$$

$$\frac{dROS_{cyto}}{dt} = d_0.M_0 + \sum_{i=0}^{n} ROS_i.M_i.d_i - aOx.ROS \tag{20}$$

$$ROS_i = \frac{i}{n}(IT_i.REF - 1) + 1 \tag{21}$$

Eq. 19-21: Given in multiples of ROS_0 for comparison with ROS output of an intact mitochondrion. aOx denotes the proportion of radical species removed by antioxidant mechanisms. Due to the accumulation of superoxides, the diffusion potential of ROS for each class of mitochondrion is different, and needs to be taken into account when calculating ROS available for nuclear damage in the cytosol (Eq.11). The proportion of escaped ROS is given by d. ROS is given in multiples of ROS_0 for convenience (86.4×10^6).

2.4.4. Model Differential Equations: Mitochondrial Damage and Population Dynamics

Derived from the preliminary equations, Eq.22-25 describe a general set of differential equations governing the population dynamics of ($M_0 \rightarrow M_n$) mitochondria, with respect to mtDNA mutation, growth and replication, and degradation.

Model Differential Equations: Mitochondrial Damage and Population Dynamics

$$\frac{dM_0}{dt} = \frac{GR}{1+(\frac{ATP}{ATP_c})^f}.(M_0 + \frac{n}{i+1.(GD-1)+n}.\frac{^{2n-2i+1}C_{n-j}\,^{2i+1}C_j}{^{2n}C_n}.M_{i+1} + \frac{n}{i+2.(GD-1)+n}.\frac{^{2n-2i+2}C_{n-j}\,^{2i+2}C_j}{^{2n}C_n}.M_{i+2} - (n-i).p.(1-p)^{n-1-i} \tag{22}$$

$$\frac{dM_i}{dt} = \frac{GR}{1+(\frac{ATP}{ATP_c})^f}. + (\frac{n}{i.(GD-1)+n}.\frac{^{2n-2i}C_{n-j}\,^{2i}C_j}{^{2n}C_n}.M_i + \frac{n}{i+1.(GD-1)+n}.\frac{^{2n-2i+1}C_{n-j}\,^{2i+1}C_j}{^{2n}C_n}.M_{i+1} + (\frac{n}{i+2.(GD-1)+n}.\frac{^{2n-2i+2}C_{n-j}\,^{2i+2}C_j}{^{2n}C_n}.M_{i+2} - ((n-i).ROS_i.p.(1-ROS_i.p)^{n-1-i} + mtd)Mi + (n-i).p.(1-p)^{n-1-i} \tag{23}$$

$$\frac{dM_i}{dt} = \frac{GR}{1+(\frac{ATP}{ATP_c})^f}. + (\frac{n}{i.(GD-1)+n}.\frac{^{2n-2i}C_{n-j}\,^{2i}C_j}{^{2n}C_n}.M_i + \frac{n}{i+1.(GD-1)+n}.\frac{^{2n-2i+1}C_{n-j}\,^{2i+1}C_j}{^{2n}C_n}.M_{i+1} + (\frac{n}{i+2.(GD-1)+n}.\frac{^{2n-2i+2}C_{n-j}\,^{2i+2}C_j}{^{2n}C_n}.M_{i+2} - ((n-i).ROS_i.p.(1-ROS_i.p)^{n-1-i} + mtd)M_i + (n-i).ROS_{i-1}.p.(1-ROS_{i-1}.p)^{n-1-i}.M_{i-1} \tag{24}$$

+...+

$$\frac{dM_n}{dt} = \frac{GR}{1+(\frac{ATP}{ATP_c})^f}. + (\frac{n}{n.(GD-1)+n}.\frac{^{2n-2n}C_{n-j}\,^{2n}C_j}{^{2n}C_n}.M_n + \frac{n}{n-2.(GD-1)+n}.\frac{^{2n-2n-2}C_{n-j}\,^{2n-2}C_j}{^{2n}C_n}.M_{n-2} + (\frac{n}{n-1.(GD-1)+n}.\frac{^{2n-2n-1}C_{n-j}\,^{2n-1}C_j}{^{2n}C_n}.M_{n-1} + (n-i).ROS_{n-1}.p.(1-ROS_{n-1}.p)^{n-1-i}.M_{n-1} - mtd.M_n \tag{25}$$

Eq. 22-25: Series of differentials governing the population dynamics of mitochondria of given damage classes (M_0- M_n), and therefore directly impacting ATP and radical change, depends entirely on the size of *n* - each set requires inclusion of the full possible scope of parent mitochondria from which the given mitochondrion class (*i*) can inherit mtDNA. With exception to the first and last equations (Where i = 0 or n for M_i), each can be described as follows: $\frac{dM_i}{dt}$ = Growth and replication + likelihood of biogenesis of mitochondrial class from one of several parents - likelihood of additional mtDNA damage removing mitochondria to a higher class ($M_i \rightarrow M_{i+1}$) + likelihood of mtDNA damage adding mitochondria from a lower class ($M_{i-1} \rightarrow M_i$). For M_0, being healthy and undamaged, those terms describing promotion to this class via damage are removed, alongside the inclusion of direct replication from equivalently healthy mitochondria. Vice-versa for M_n. As mitochondria degrade, and must be replaced, a first order decay term is used ($mtd.M$) to ensure a continuous degradation and thereby turn-over the mitochondrial population, based upon observed mitochondrial half-lives.

2.4.5 Model III Implementation

A script was written in R-3.2.2 to run a basic simulation, using a range of input parameters sampled from distributions of reasonable biological limits (see Table 1), in order to discover and analyse stable steady-states (see SI.4.7).

A second series of simulations was run in order to directly model ROS production scaling with metabolic rate in three mammal models - mouse, human, and whale. Using their respective values for specific metabolic rate, cell size and estimated daily mitochondrial growth rate (*GR*), ATP_d (ATP consumption per day per cell) was calculated, assuming a synthesis rate of 5 ATP per O_2 molecule (Hinkle, 2005). See Table 2 for parameter values used. ATP_c (basal ATP level requiring maintaining) remained constant, however this ATP steady state is known to vary between tissue and species. Using these parameter values, the model was used to simulate radical production. Direct rate of DNA damage were not modelled, however proportion of escaped ROS was used as an approximate proxy for potential DNA damage. It must be emphasised that this method does not take into account spatial considerations of ROS diffusion and cellular structure (see Discussion). Number of initial M_0 was calculated using the scaling strategies developed by Savage *et al.* (2007) when cell size was altered (see SI.4.7). For all other parameters, values were sampled from stable distributions produced during the basic simulation.

Parameter	Symbol	Values	Reference
Maximum Growth Rate	GR	0.06	Species-specific, otherwise 6% per day
Degradation Rate	mtd	0.05	Milwa *et al.*, 2008; Gottlieb *et al.*, 2011
mtDNA Mutation Rate	p	10^{-4}	Haag-Liautard *et al.*, 2008
ATP required for replication	r	$4x10^{9}$	See SI.4.9
Biogenesis Negative Feedback	f	3	-
Minimum ATP Level	ATP_t	10^{9}	-
Growth Difference	GD	1.1	-
Radical Enhancement Factor	REF	20	Turrens, 2003; Bandy & Davison, 1990
Maximum Transcription Increase	IT_{max}	2	-
Maintained Cellular ATP Level	ATP_c	$10x10^{9}$	-
ATP Consumption Rate	ATP_d	$3.8x10^{12}$	Metabolism-specific / Müller *et al.*, 2013
ROS Production Rate	ROS_0	$86.4x10^{6}$	Baria, 2007

Table 1: Parameter values used during Model III simulation described in Fig.10a, representing reasonable biological values expected for a human being. A $\pm$10% alteration in any parameter offers no significant differences in simulated outcome, with exception to *GR*, *p* and *mtd* (see Fig. 10b).

Species	Lifespan (Years)	Mass (Kg)	GR	ATP_d	mtd	M_0
Mus musculus	4	0.019	0.24	$18.1x10^{12}$	0.23	950
Homo sapiens	80	60.5	0.06	$3.8x10^{12}$	0.05	1000
Balaenoptera musculus	110	154321.3	0.002	$0.0x10^{12}$	0.0015	1100

Table 2: Values calculated using scaling equations developed by Savage *et al.* (2007) for approximate cell size and mitochondrial surface area (therefore population size, M_0) and quadratic model developed by Kolokotrones *et al.* (2010) to scale metabolic rate and ATP consumption. Known cell replication cycle times used as a linearly proportional proxy for mitochondrial growth rate, with *GR* scaled accordingly. Recorded lifespan used as taken in Chapter 3. Emphasis placed on the approximate nature of values, used purely for proof of concept only.

Results

3.1. Model I

A simple algebraic model was used to i) set a baseline for comparison with Model II in order to ii) test the influence of metabolically-scaled division and cellular mutation rate (separate and together) on estimated cancer incidence rates; and iii) explore potential parameter values that would support observed incidence data (and acting as a benchmark to evaluate any model). Model I successfully modelled lifetime incidence rates for three model organisms (see Fig. 5a), and was further used to predict cancer incidence rates in 174 species of mammal with known data (see Fig. 7a).

3.1.1. Metabolic Scaling

Following simulation, logistic regression fitting and analysis was undertaken on the complete dataset for each of the three model species (each scaling condition set as indicator dummy variables), followed by likelihood ratio tests for each curve. In the whale model, neither altered division nor metabolic rate alone were considered significant perturbations from the null non-scaled curve, confirmed via pairwise analysis (p=0.95), however both combined significantly reduced incidence rate over lifespan (reduced from ~100%, to 75.1%, at 100 years) ($p<0.05$, via ANOVA) (see Fig.5a.iv). In contrast, the influence of either metabolically-scaled parameter presented a marked change in incidence rates in the mouse model, change in mutation rate alone (u=5.5 $10^{-6} \rightarrow$ 3x10^{-5}) resulting in near ~100% incidence by ~5 years (see Fig.5.a.v). Despite incorporating metabolically-scaled parameters in Model I, incidence rate in whales remains high, and under the assumptions of the algebraic model, the influence of lower mass-specific metabolic rate is insufficient to explain Peto's paradox.

3.1.2. Driver Mutations

Increasing the number of rate-limiting steps - i.e. increasing the number of driver mutations required to initiate cancer - results in marked change in estimated incidence rate. Under biologically plausible parameters, a two-fold increase in driver mutation number was sufficient to reduce lifetime incidence rate from ~100% to 0% in a whale model (see Fig.5.a.vi). The influence of any combination of metabolically-scaling parameters was found inconsequential to required change in the value of m_d in order to reduce incidence risk, despite incidence reducing success following an exponential decay distribution, with each additional rate-limiting step resulting in diminishing impact (see Fig.5.f).

Independent of other parameters, multivariate Kolmogorov-Smirnov analysis, for a whale model, indicates a significant reduction in incidence rate for increasing values of m_d from 2 through 5.

3.1.3. Parameter Space

Multivariate parameter space models were produced to explore potential plausible ranges of division rate and metabolic rate sufficient to maintain low risk of neoplasia through lifespan (see Fig.5d). Table SI.4.9 reviews minimum increase in parameters to confer independent cancer resistance with respect to mouse baseline. As observed in Fig.7, parameter estimates, required to maintain lifetime incidence risk below 25% and 5% respectively in the three model organisms, produced by Model I (with m_d=4), were incongruous with observed (and in the case of whale, estimated) data for the respective values in division and mutation rate. For example, a mouse model, with stem cells dividing approximately once every 12 hours, requires a mutation rate suppressed to ~10^{-12} in order to avoid neoplasia for four years - else a division rate >72 hours for a mutation rate along observed rates of ~3x10^{-5}. No evidence to date corroborates such required estimates; thus metabolic scaling under Model I, without altering m_d, is insufficient to predict known cancer incidence rates.

3.2. Model IIa

Model IIa was developed using Bayesian and coalescence methodology in order to tractably simulate stem cell and mutation dynamics across species. Model IIb represents an investigative side-branch undertaken during development, in order to test hypotheses and ensure basic assumptions made by IIa were valid (i.e. that mutations in stem cell populations represent the overwhelming majority of neoplastic progression risk). Given the results of IIb proved negative (see 3.3.1-3.3.2), the standard version of Model II (IIa) was used to, as in Model I, investigate the influence of metabolic scaling on incidence, explore potential parameter value space, and compare against known incidence data.

3.2.1. Metabolic Scaling

Though qualitatively, metabolically-scaled incidence estimates appear to more accurately reflect expected rates compared to known values (higher than non-scaled in mice; vice-versa in whales; see Fig 6i/v), following likelihood ratio analysis on a logistic regression of the combined datasets, using model type as categorical indicator variables, no quantitative significant difference was found (P>0.05 in all cases). Assessing the degree of divergence from null (default Model I) Peto's paradox,

Fig. 5 (overleaf): **(a)** Estimated colorectal cancer incidence from Model I using parameters approximating mouse, human and whale models (10^6, 10^8, and 10^{11} dividing stem cells respectively; where u=3x10^{-5} and k=96) (i) without metabolic scaling; (ii) metabolically-scaled cell division rate (k=17, 96 and 256 hours); (iii) metabolically-scaled mutation rate (1.25x10^{-5} and 7.1x10^{-7} in mouse and whale respectively); and (iv) both. Inclusion of metabolic scaling increases estimated incidence in the mouse model considerably; as seen in a shorter time frame in (v). The number of rate-limiting steps (required TSG mutations), m_d, is 5 by default; except in (vi) where m_d=3, 4 and 5 in a whale model including both metabolically-skewed parameters (k=260,u=7.1x10^{-7}). In these examples, parameterisation is undertaken via a single proportionality coefficient using human data as baseline and calibrated using mouse data from the literature (see SI.4.1, lines x-x for details).

(b) The influence of both scaling parameters on time-series incidence estimates can be visualised in three-dimensional morphospace, using colour gradient to denote incidence; here continuous parameters demarcated in units of 10 in order to see depth. **(c)** The threshold between $P_t = 0$ and $P_t = 1$ for continuous parameter variables can be isolated and used to construct threshold matrices (d).

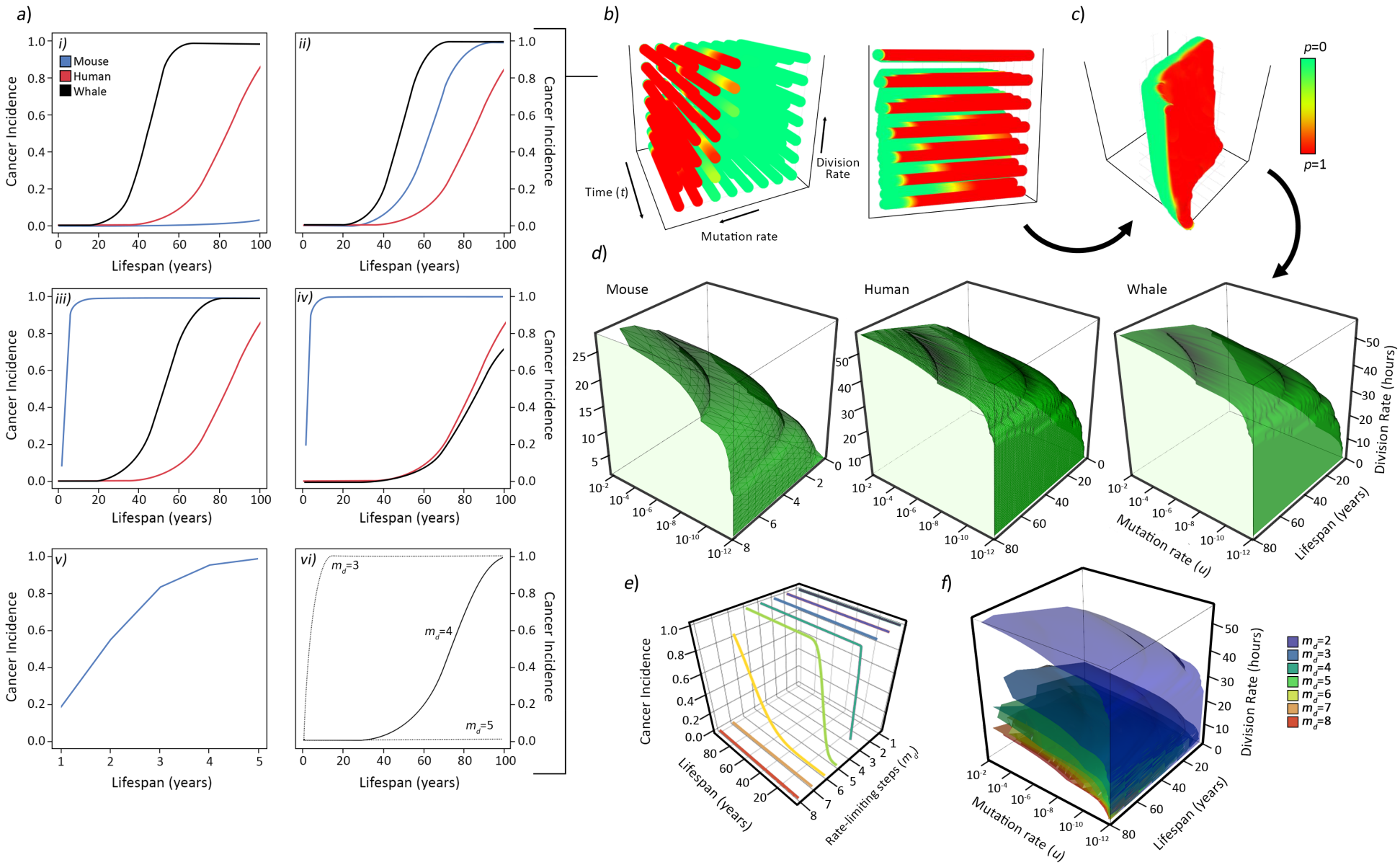
(d) Visualisation of possible metabolically-scaled parameters sufficient for lifetime estimated incidence rate, under Model I, to remain invariant at $P_t = 0.25$ across approximations for mouse, human and whale models. Shaded region indicates parameter combinations for lifetime incidence to engender $P_t > 0.25$ - e.g. under Model I, a human-sized organism at ~60 years of age requires a division rate > 50 hours and $u < 10^{-7}$ for $P_t < 0.25$. Here m_d = 5.

(e) Parallel time-series for different values of rate-limiting m_d can likewise be visualised in three-dimensional morphospace as in (f), illustrating gradient shift in two-dimensional S-curves, as in (a(vi)), in a whale model for values of m_d between 2-8. in a whale model where $P_t < 0.25$.

---

via ROC curve analysis (varImp in caret R package; loess smoothing fit between outcome and predictor. Kuhn, 2008; Gelman & Hill, 2006), for model type (i.e. II) and inclusion of either altered division rate, metabolic rate, or both together, indicated model type was the predictor with the highest relative effect on incidence estimates (t-equiv.=100). Neither independent inclusion of division rate nor mutation rate resulted in significant improvement to model fitting (F=0.09 and 0.94 respectively, P>0.05, via Wald test), and when combined resulted in P=0.056, after controlling for model type. m_d = 4 for all analyses; no variations were tested against Model I.

Though the implementation of the stochastic model itself resulted in significant divergence from a "Peto's paradox" baseline, inclusion of scaling parameters based on intrinsic mass-specific metabolic rate didn't confer significant quantifiable difference in either mouse or whale experiments.


a)
i)
Mouse
Human
Whale
ii)
iii)
iv)
v)
vi)
m_d=3
m_d=4
m_d=5
Cancer Incidence
Lifespan (years)
b)
Time (t)
Mutation rate
Division Rate
c)
p=0
p=1
d)
Mouse
Human
Whale
Mutation rate (u)
Lifespan (years)
Division Rate (hours)
e)
Cancer Incidence
Lifespan (years)
Rate-limiting steps (m_d)
f)
Mutation rate (u)
Lifespan (years)
Division Rate (hours)
m_d=2
m_d=3
m_d=4
m_d=5
m_d=6
m_d=7
m_d=8

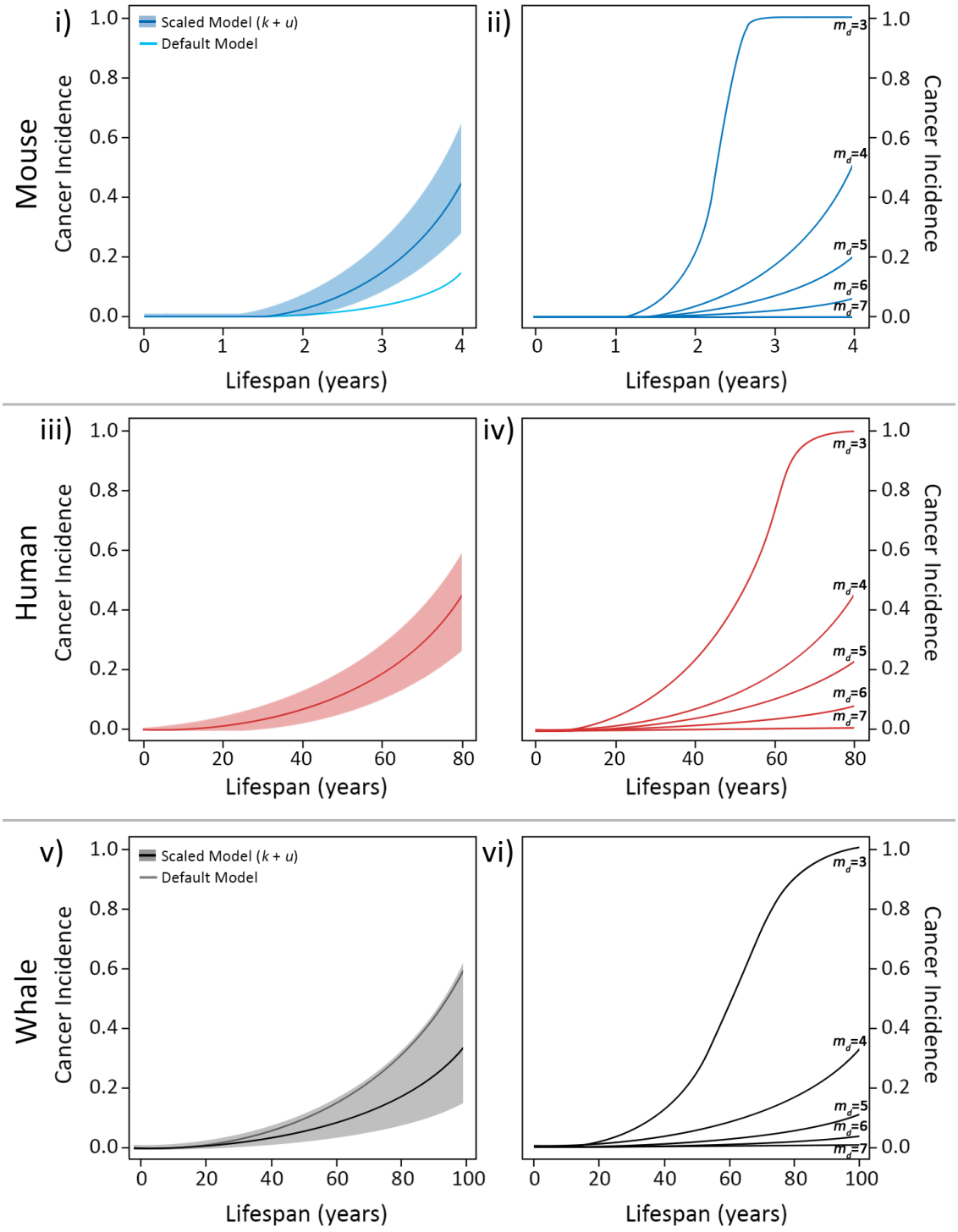


Fig. 6: (a) Cancer incidence estimates simulated using Model II for mouse, human and whale approximations. (i), (iii) and (v) compare the inclusion of metabolically-scaled values of division rate (k) and mutation rate (u) - mouse: k=12, u=3x10^{-5}; human: k=96, u=5.5x10^{-6} (same as default, human taken as baseline); whale: k=260, u=7.1x10^{-7} (m_d = 4 in all cases). (ii), (iv) and (vi) compare alteration in rate-limiting steps (m_d 3→7) for each species.

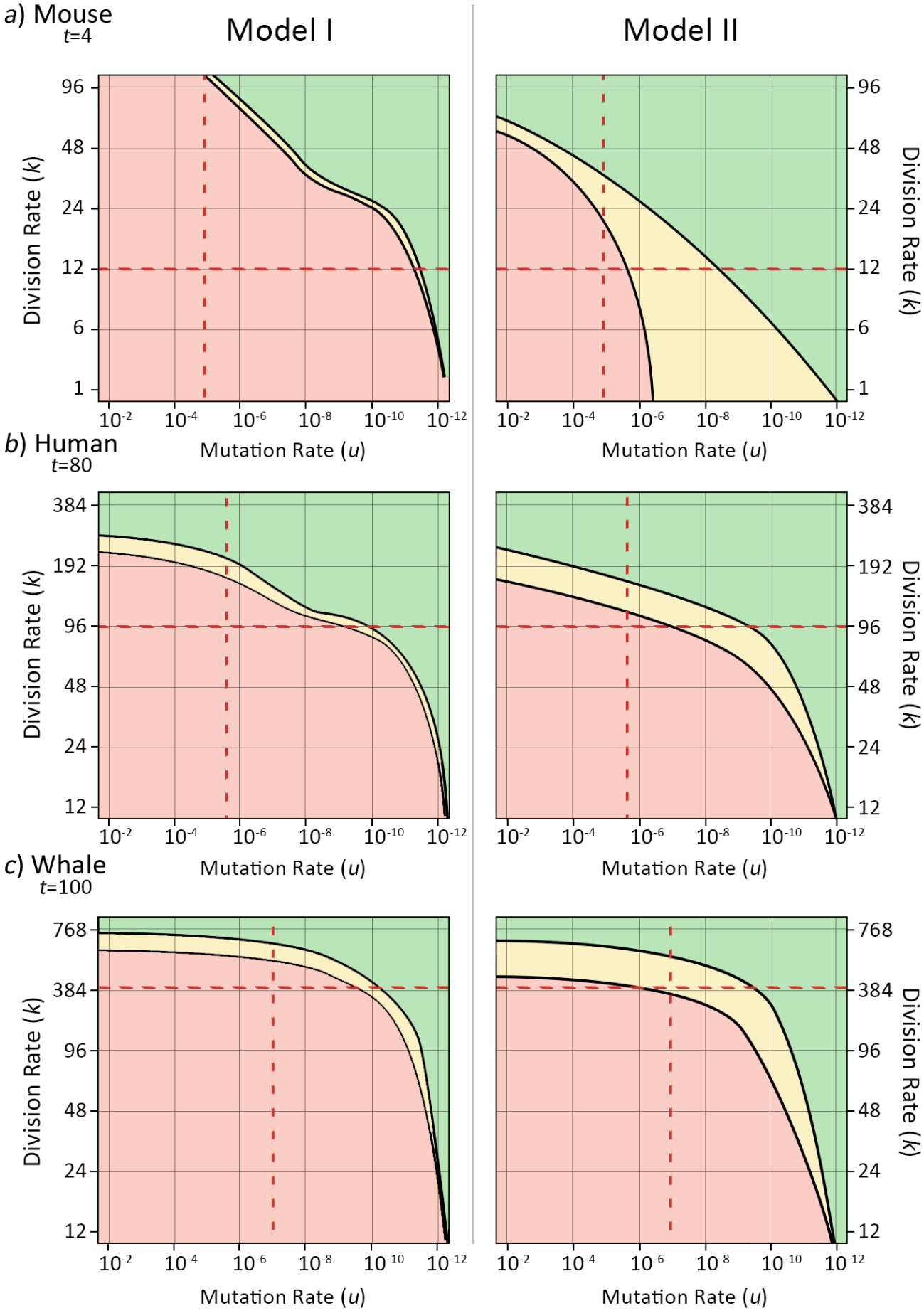
3.2.2. Driver Mutations

As in Model I, increasing the number of rate-limiting driver mutations ($m_d \uparrow$) presents marked decrease in estimated lifetime incidence risk. In a whale model, an increase in m_d of 3 (4→7), with metabolic scaling, is sufficient to reduce incidence estimates to <5% (see Fig. 6.vi) Without metabolic scaling, an increase in m_d of 4 is required (see Fig. 6b.iii) - as such, despite alteration to k and u presenting no measured effect where P<0.05, when m_d is constant, Model II demonstrates the metabolic effect can nonetheless allow for a lesser degree in m_d change required for resistance to manifest. No multivariate experiment testing changing m_d with respect to continuous change in k and u was undertaken.

3.2.3. Parameter Space

As seen in Fig.7, Model II presents a wider range of potential parameter values sufficient to maintain low risk of neoplasia compared to Model I, approaching values compatible with known biological data (though nonetheless still relatively high; >5% in whale, and >25% in human and mouse models). Demonstrative of the lower slope gradients produced for lifetime incidence risk, the space between threshold risk values is larger, allowing greater scope for plausible parameter values - e.g. under the assumptions of Model II, a humpback whale (*Megaptera novaeangliae*; dashed line estimates in Fig.6c) could exhibit a value for u between ~1x10^{-7} and ~6x10^{-9} to present a lifetime incidence risk invariant with size at 0.05> P_t <0.25.

Fig. 7 (overleaf): Parameter value space estimated to support a lifetime (*t*) incidence rate of *P*>0.25 (red) and *P*>0.05 (yellow) under Model I and Model II for three species approximates: (a) mouse (C =); (b) human (C=2x10^8); and (c) whale (C=). m_d = 4 in all cases. Red dashed lines indicate observed values for mouse and human, and estimated based on scaling relationships in whale (using the humpback whale, *Megaptera novaeangliae*, as reference). Width between incidence P value thresholds is indicative of slope gradient; the steeper logistic 'S-curve' for Model I presenting a narrow range between high and low incidence rates.


a) Mouse
t=4
Model I
Model II
b) Human
t=80
c) Whale
t=100
Division Rate (k)
Mutation Rate (u)
96
48
24
12
6
1
384
192
768
10-2
10-4
10-6
10-8
10-10
10-12

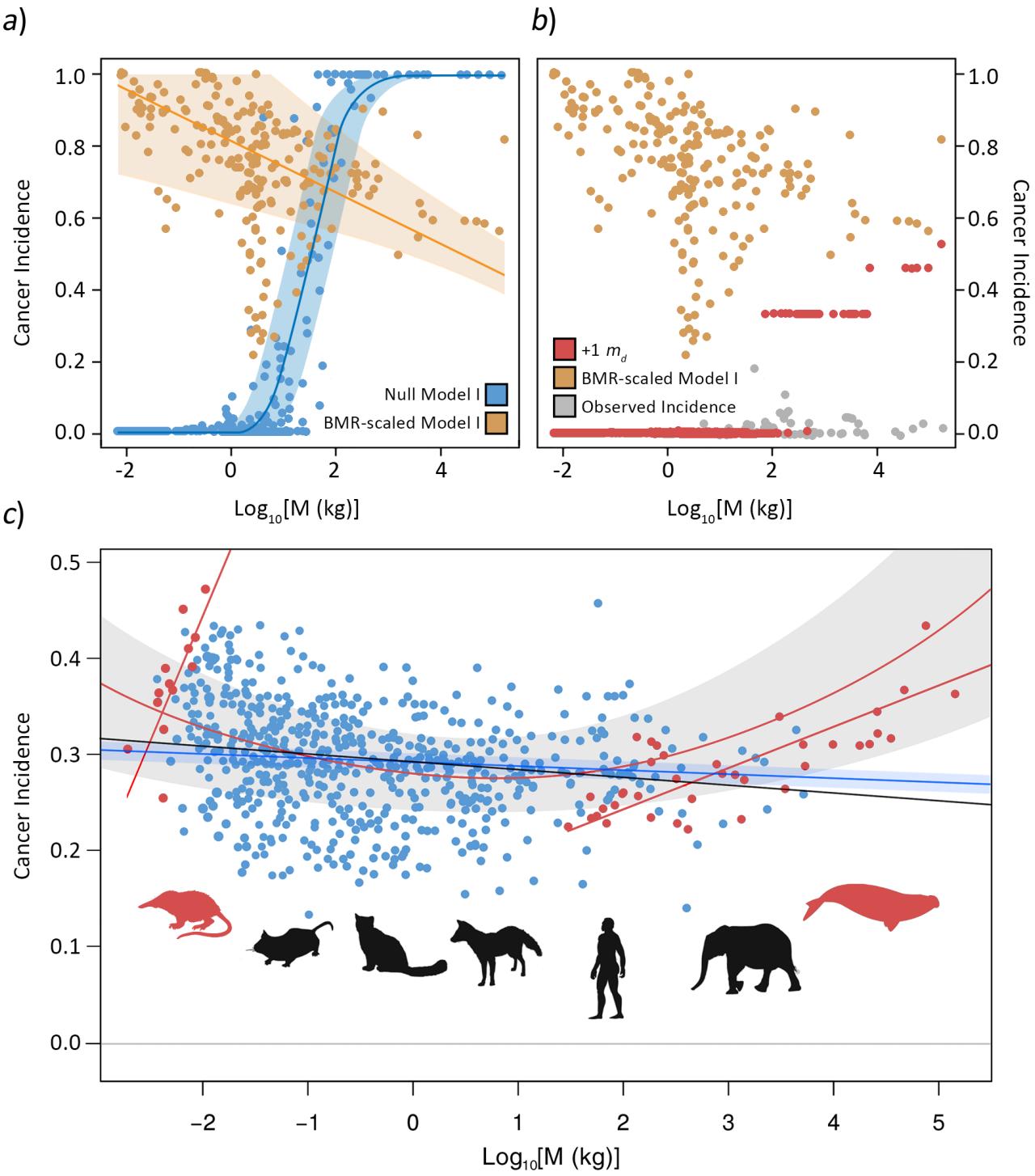


Fig. 8: **(a)** Model I applied to observed data for 174 mammal species, using estimated values for division and mutation rate ($\pm$ 10%) based on known BMR values, calibrated using human and mouse data (see SI.3.4 and SI.4.1 for scaling) - blue without scaling (using default values; see SI.4.0 for data) (R^2=0.96), orange with scaling (R^2=0.32). **(b)** An increase in driver mutation (M_d) of only one has a considerably higher impact on reducing estimated incidence - towards observed incidence data - under Model I than the incorporation of metabolic scaling. **(c)** Model II applied to observed data for 174 mammal species, and an additional 128 species for which life history traits (e.g. BMR)

was estimated (e.g. BMR values for large baleen whales; see Methods 2.1). Non-phylogenetically weighted quadratic regression (red curved line), phylogenetic linear regression (blue line), null model (black; invariant with msBMR) applied. Adj. R^2 (0.071) for the quadratic model applied to all mammals is extremely low; taxa at the tails ends of the mass scale, notably Soricidae (shrews) and Cetacea (whales) (taxon-specific linear regressions in red), skews model-fitting. In comparison to Model I, Model II presents estimated incidence rates qualitatively more similar to observed incidence for most taxa (see Chp.3 Fig.3).

2.3. Cross-Taxa Comparison

Both Model I and Model IIa were applied to a dataset of parameter variables for 174 and 302 species respectively, division and mutation rate values estimated based on either known (or in the latter case, estimated via scaling) life-history trait values (see SI.3.4). Though the aim was not to recapitulate known approximations of cancer incidence rates, modelling was undertaken to assess whether the qualitative trajectory of cancer incidence scaling would manifest based upon the model assumptions. Under Model I, application of metabolically-scaled values did markedly alter projected cancer incidence rates across taxa (Fig 8a), presenting a negative (albeit poorly fitted; R^2=0.32) downward trajectory with increasing body size. However, in comparison to a neutrally-scaled model with a stepwise increase in rate-limiting steps (i.e. increasing number of required TSG inactivations), the influence of metabolic scaling on reducing lifetime incidence risk seems relatively poor (Fig 8b).

Application of cross-taxa metabolically-scaled parameters on Model IIa demonstrates a clear qualitative improvement towards estimating more realistic size-incidence scaling compared to known incidence values than Model I (Fig 8c). Phylogenetically-weighted regression model comparison in this instance presents no significant deviation in the predictive ability of a metabolically-scaled model from a msBMR neutral model, suggesting a negligible influence of metabolism on incidence estimates - at the broadest cross-taxon scale. However, within lineages, the influence of metabolic scaling becomes increasingly pronounced as one moves further towards the extreme ends of body mass values. The high metabolic rates of shrews (with estimated k as low as ~9.12 hours on the lower bound) predicts high mass/lifespan-specific sensitivity on incidence rate outcomes compared to null. Similarly, low msBMR in large cetaceans isn't sufficient to align size-incidence scaling trajectories with an observed null (or negative) relationship.

3.3. Model IIb

Here we investigated whether inclusion of populations of differentiated, mutation accumulating transit-amplifying (TA) cells, alongside the requisite altered stem cell replication dynamics, contributed significantly towards lifetime incidence risk. The developed model was capable of successfully simulating the dynamics of cell populations within compartments (see Fig. 9a), albeit in a relatively intractable manner, owing to the high computational load.

3.1. Stem Cell Dynamics: Symmetric and Asymmetric Division

Due to the computational expense, a limited set of simulations were run for models approximating mouse, human and whale at ~10% compartment number, repeated in blocks until two-hit mutants arose at least 1000 times for each species (see Table 3 for input parameters). There was no observed temporal effect on the onset of two-hit mutant development, compared to equivalent control (IIa) simulation runs (see Fig.9b / 9c): for all species, the likelihood of a two-hit mutant arising was comparable for all given fractions of symmetric division (σ).

Though no temporal effect was observed, two-hit incidence rates were significantly reduced with increasing propensity for symmetrical division (see Fig 9d), reaching their lowest point at $\sigma = 1$, when all divisions represent stem cells dividing symmetrically - either into two daughter stem cells following a removal event, or two daughter TA cells. However, the influence this had on lifetime incidence risk was small ($p>0.05$ against control IIa data, via linear regression analysis) for biologically reasonable mean values for σ (~0.3-0.7); given the additional simulation complexity, further development of Model II did not include a σ coefficient.

3.2. Transit-Amplifying Cell Mutation and Cancer Incidence

Though hypotheses have suggested mutations accumulating in TA cell populations may significantly contribute towards accumulation of two-hit mutants and onset of neoplasia, the modelled approach utilised here failed to support such a hypothesis. The proportion of two-hit mutants arising within TA cell populations was negligible in all species models and for all tested values of σ (mean proportion for mouse: 0.0015; human: 0.0042; whale: 0.0001) - and not a single double-hit mutant arose in any TA population in whale models when $\sigma = 1$. Under the assumptions of Model IIb, no evidence was found to support the inclusion of TA populations, or a coefficient thereof, in subsequent mainline (IIa) simulations for Model II.

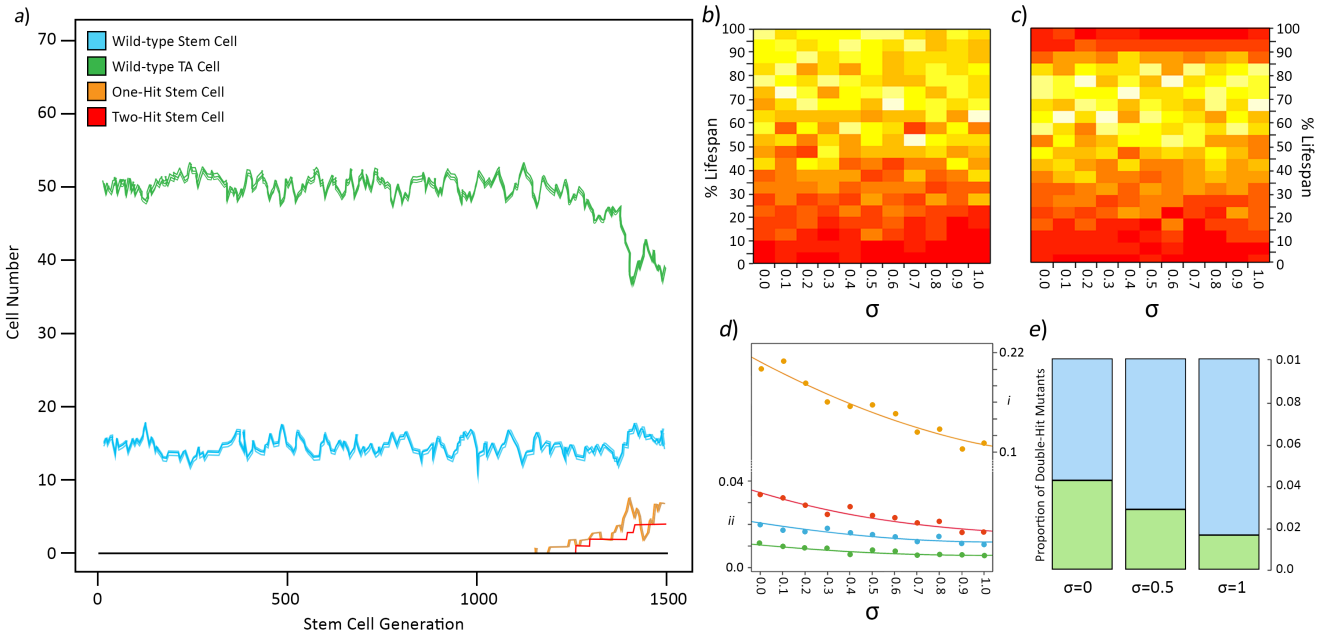


Fig. 9: **(a)** Qualitative example of the cell dynamics within a human tissue compartment through until end of lifespan simulated using Model IIb. Wild-type stem and TA cell populations persist in a fluctuating steady-state (~16-19 stem cells, ~49-51 TA cells). Single-hit stem cell mutants (i_1) may come and go without progressing to a double-hit mutant (i_2); given the high removal rate in TA populations, the likelihood of mutant arising *de novo* is low (none observed in this run). **(b)** A change in σ does not result in a temporal shift, nor delay, in first incidence of a two-hit mutation in any species model, here (b) mouse, and (c) whale. The likelihood of two-hit mutation increases through life, though in longer-lived species (c) a threshold age is eventually reached after which hit rate rapidly decreases. **(d)** As the propensity for a stem cell to divide symmetrically, as opposed to asymmetrically, increases ($\sigma = 0 \rightarrow 1$), the probability of a two-hit mutation arising within a species' lifespan decreases. The effect becomes less pronounced as lifespan increases and basal metabolic rate per unit mass decreases - orange (mouse), red (chimpanzee), blue (human), green (whale). **(e)** Two-hit mutations arise from stem cell populations (blue) in the overwhelming majority of instances, whilst TA cell derived two-hit mutants (green) for up to 1,000 double-hits were in single digits, if any. In this human approximated model, which displayed the highest TA double-hit mutant proportions across all three values of σ, the proportion remains below <5%. In the above runs, $u_1 = 10^{-6}$, $u_2 = 10^{-5}$ (Haag-Liautard *et al.*, 2008; Helleday & Nik-Zainal, 2014), $r = 1.1$ and s being set to the initial value for i_0.

Species	Lifespan (Years)	BMR (W)	Stem Cell Generation Time (Hours)	Crypt Number
Mus musculus	4	0.271	6	18.1×10^{12}
Pan troglodytes	59.4	60.5	66.47	1.2×10^{7}
Homo sapiens	80	82.78	72	1.5×10^{7}
Balaenoptera musculus	110	44370	169.4	0.0×10^{12}

Table 3: Parameter values used during simulations detailed in Fig. 9. Generation time was estimated using scaling equations developed by Savage *et al.* (2007) (see Eq. 5), with BMR for *Balaenoptera musculus* from the quadratic model developed by Kolokotrones *et al.* (2010) (see Eq. 3). Crypt number estimated by assuming comparative crypt density across mammals, and calculating internal surface area from colon lumen diameter and colon length (see SI.4.1).

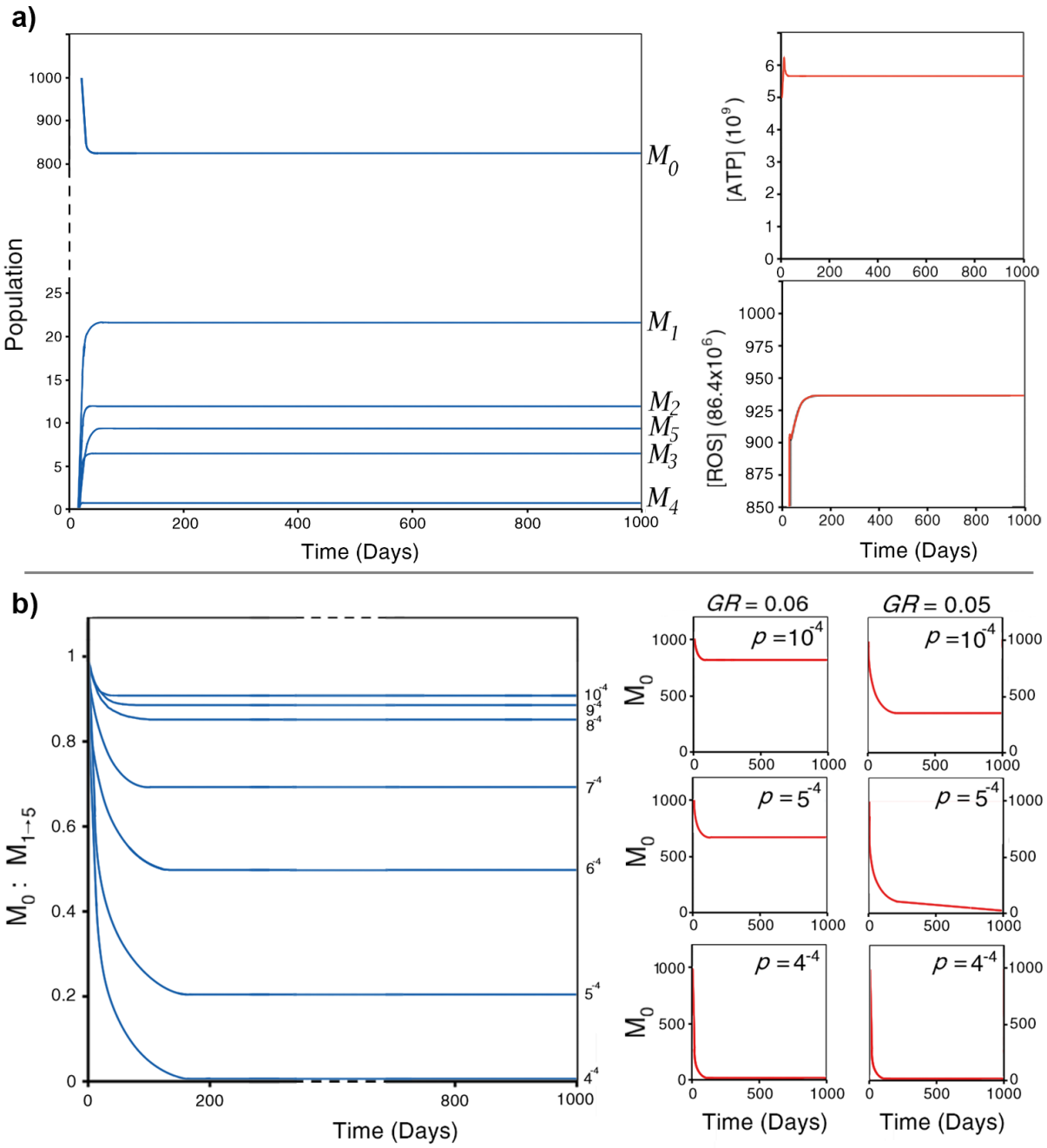
3.4. Model III

4.1. Baseline Simulation

Simulations were run to establish a steady-state solution for the differential model in order to analyse which parameters, if perturbed, influence the model to a significant degree. Given the average polyploid number of mitochondria is 4.6 (Wiesner *et al.*, 1992; Satoh & Kuroiwa,1991), when n=5, the system quickly converges to a steady state provided a standard set of initial concentrations and parameter values (see Fig.10a). In the simulation described in Fig.10, undamaged mitochondria reach stability at 826, with less than 6% of total mitochondria containing damaged mtDNA ($M_{1\rightarrow5}$). M_4 mitochondria remain at the lowest levels, with a higher accumulation of M_5, due to there being fewer potential mechanisms by which M_5 can be removed from the system (cannot mutate into a higher damage class). Total ROS produced falls within the expected range (938, given that the majority of mitochondrion are class M_0 and ROS is measured in multiples of ROS_0).

Following sensitivity analyses for all parameters ($\pm$10% within distribution ranges, see SI.10), changes to parameter *p*, mtDNA mutation rate, and *GR*, maximum growth rate, have the most substantial effect on the stability of the system. Using a numerical example; over an increase in the likelihood of mutation from 10^{-4} through to 3^{-4}, the ratio of damaged to undamaged mitochondria reduces to zero and the system collapses (see Fig.10b). Similarly, a threshold is reached when reducing *GR* to $0.05d^{-1}$, at which point the population of mitochondria collapses. The effects of *p* are modulated by *GR* and a fatal system collapse can be mitigated by an increase in the latter - for example, increasing *GR* from 0.05 to 0.06 renders an otherwise non-stable mutation rate of 5^{-4}, stable.

Fig. 10 (overleaf): (a) Standard simulation run over 1000 days using the parameter values described in Table 1. ATP and ROS concentrations are given in multiples of 10^9 and 86.4x10^6 respectively. The initial concentrations were: ATP=5, M_0=1000, $M_{1\rightarrow5}$=0 and ROS=0. The system quickly converges to a stable, stead state by day 50. Here M_0=826, M_1=22, M_2=12, M_3=6, M_4=1, M_5=9, ATP=5.79 and ROS=938. **(b)** Increasing mtDNA rate between 10^{-4} to 4^{-4} results in a decreasing ratio of undamaged to damaged mitochondria ($M_0 : M_{1\rightarrow5}$). As $p = 10 \rightarrow 4^{-4}$, values for $M_0 : M_{1\rightarrow5}$ are 0.928, 0.901, 0.873, 0.747, 0.526, 0.212 and 1.5x10^{-5} respectively, to 3s.f. Between p=5^{-4} and p=4^{-4} the system collapses completely, with the population M_0 reducing rapidly to zero (red), when *GR*=0.06. Reducing *GR* to 0.05 destabilises the mitochondrial population, with a collapse when p=5^{-5}. Increasing *GR* therefore has a stabilising effect on mutation rate.



4.1. Metabolic Scaling

Model III was run using parameters described in Table 2 for three mammal species, covering 8 orders of magnitude in size (0.019 to 154,321.3Kg). Fig.11 describes ROS production at a stable steady-state for each species, controlling for variation in mitochondrial number in order to display ROS production per size of mammal. Mice produce ~25% more ROS per cell per unit time than humans, and ~28% more than blue whales. Assuming rate of ROS production has equivalency to rate of DNA damage, Model III demonstrates how feedback processes associated with mitochondrial ROS production and mitochondrial damage can result in cross-taxon divergence in base mutation rate.

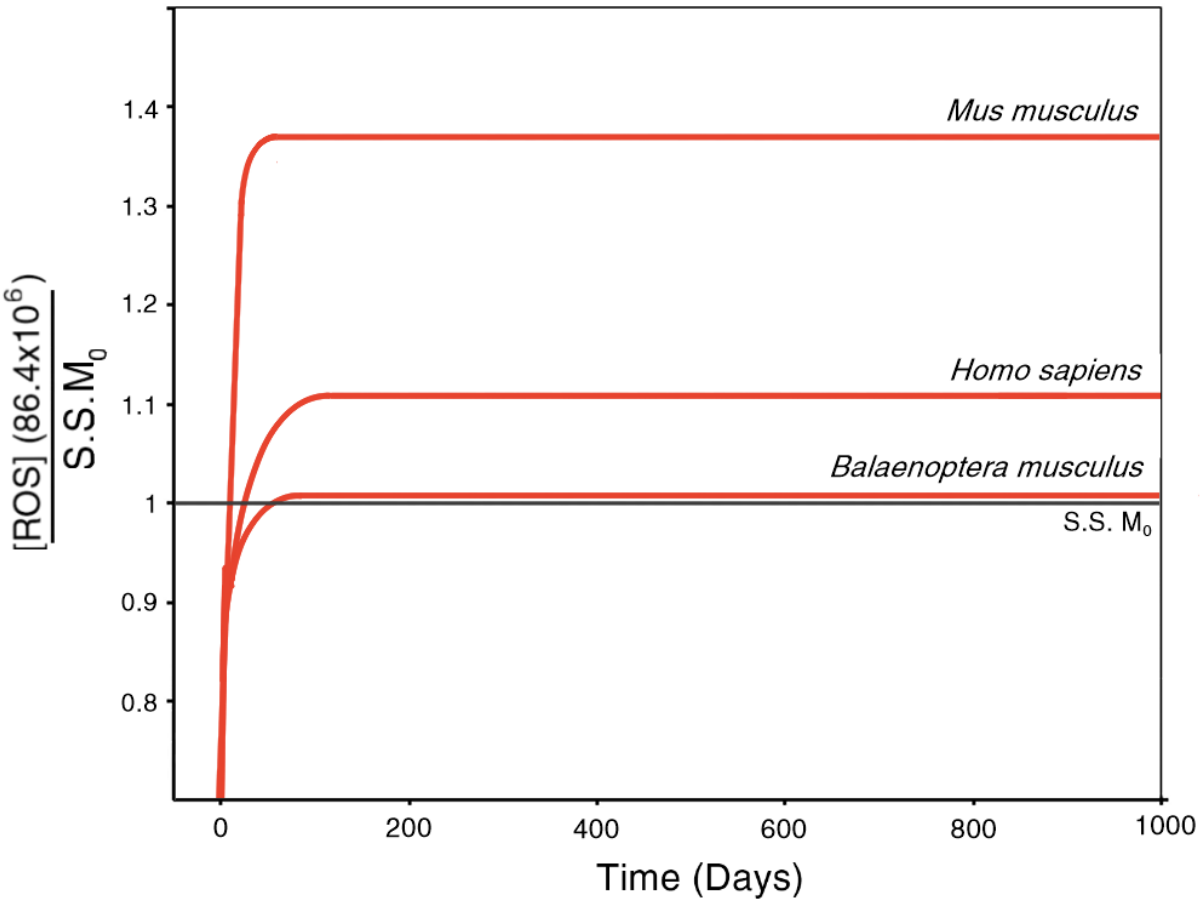


Fig. 11: ROS production is described with respect to steady-state levels of *S.S.* M_0 for each modelled species. In a steady state, mice (*Mus musculus*) produce approximately 25% more ROS, proportional to mitochondrial level, per unit time than humans (*Homo sapiens*) (1.375 and 1.1196 respectively), and 28% more than blue whales (*Balaenoptera musculus*) (1.007). Here M_0 for mice = 764; humans = 826; and blue whale = 1086); with ROS levels: mice = 1050.5, humans = 938 and blue whale = 1093.6.

Discussion

"All models are wrong...
... but some are useful"
- George Box, 1976

Cancer is a complex disease; each tumour its own individual ecosystem and microcosm of evolution, and no model, however sophisticated, can accurately represent the complexity of cancer progression across all its diversity. Nonetheless, a modelling approach, even when simplified considerably, can present useful results, and, in broad strokes, inform further avenues for investigation. Here, two models (I and II) were used to: i) determine whether intrinsic, non-adaptive changes in mass-specific metabolic rate can explain divergent rates of neoplasia across mammals; ii) determine the degree to which adaptive (whether directly or indirectly) alteration to the number of rate-limiting steps required for carcinogenesis could suppress cancer; and iii) determine the parameter space within which plausible biological values could result in the approximately invariant trajectory incidence rates seem to present as from observation. Further, a third, mechanistic model (III) was developed to explore the relationship between cellular metabolic rate and mitochondrial damage and dynamics, ROS production, and base rates of nuclear DNA damage - in order to quantitatively establish the plausibility of the metabolic hypothesis from first principles; that alteration to cellular metabolism does have a significant influence on mutation rate, a major premise of Model I and II.

At the outset, Model I was successfully able to reproduce Peto's hypothesis - that, all else being equal, larger, longer-lived mammals are disproportionately more susceptible to cancer risk, to such a degree such species should not exist (in this instance, 100% of whales succumbing to neoplasia by age 60; see Fig.5a). The results mirror those of the model produced by Caulin *et al.* (2015), and here used as a baseline 'null' trajectory against which subsequent models are compared.

4.1. Metabolic Scaling

A plausible explanation to Peto's paradox, as described by Dang (2015), could be that larger, longer-lived organisms are better protected from risk of neoplasia simply due to their intrinsically low mass-specific metabolic rate. Requiring less energy to maintain homeothermy via ATP production and utilisation, larger species produce less ROS as a byproduct, and are thereby exposed to less risk of DNA damage. This hypothesis was examined using two multistage models of cancer progression (Model I and Model II) by modifying values for cellular turnover rate, and intrinsic mutation rate, with respect to null values (expected under Peto's paradox). In both instances, though a qualitative shift in estimated incidence trajectories did more germanely reflect known trajectories, for example those seen in Chapter 3, the incorporation of metabolic rate alone was insufficient to quantitatively explain Peto's paradox in large mammals.

Metabolic scaling applied to whale parameters in Model I essentially shifted time required to 100% incidence rate by approximately 20-25 years, but otherwise still produced scenarios under which no whale could survive until end of lifespan without further strategies to reduce risk. Model II was more successful in recapitulating an age-incidence trajectory comparable to known incidence curves, though when compared to Model I, the intrinsic design of Model II - incorporating more accurate cellular growth and replication behaviour - proved to be the more influential factor, rather than any benefit provided by the inclusion of scaled parameters (no cross-taxon difference between msBMR null and metabolically-scaled models).

When applying Model II using scaled values for 302 mammal species, a linear cross-class analysis did indicate an invariant relationship between incidence and body size, tentatively rejecting Peto's hypothesis. However, the degree of variance was high, and the fit relatively poor - taxon-specific analyses indicated those groups further towards the extremes of the body mass scale (and similarly lifespan, not shown) presented an age-incidence relationship expected under Peto's rule, further indicative, in both cases, that, if the assumptions of Model II are reasonable, then intrinsic metabolic rate isn't sufficient to confer adequate resistance in large mammals, and indeed considerably elevates risk in the smallest mammals. For most mammal groups however, application of Model II with metabolic scaling presented an invariant relationship, as observed in nature.

Whether this was due to the model successfully behaving in a similar manner to real cancer progression, and thereby producing general applicable results, or an artefact of parameterisation, is difficult to

determine without further experimentation, or far more numerous and more accurate data for species values across a range of size, longevity and metabolic scales. Calibration for scaled values was done using established human and mouse data from the literature in order to extrapolate a proportionality coefficient, and values for the human model were used as baseline for comparison - as such, the human data point may act as a pivot around which other species' results were disproportionately rotated / transformed (i.e. in Model I, Fig8a, the degree to which scaled results are different to non-scaled increases with distance away from human).

4.2. Cellular Metabolism and ROS Production

The basis for metabolic scaling in Model II was the assumption that a change in cellular metabolic rate produced a proportional change in mutation rate. Model III attempted to test this in a quantitative, mechanistic fashion, by establishing a working model of subcellular metabolism – incorporating mitochondria dynamics, ROS production, and DNA damage - to test whether the predicted metabolism-specific ROS-damage steady state distributions followed similar scaling trajectories to those used in Model II between species.

Model III was successfully able to simulate ATP and ROS production, and population levels of associated damaged and undamaged mitochondria. By undertaking robustness analyses, it was determined net growth rate (*GR-mtd*) and mtDNA mutation rate (p) were critical in determining the stability of the system, with only a slight perturbation to either (without associated change in the other) resulting in a system collapse. This dependence on the fine-tuning of parameters is not surprising; previous work on mathematical models governing mammal cell proliferation in lab culture likewise concluded population density was governed by opposing proliferation rate and accumulation of genetic damage (Tan, 1994; Zhang, 1991). Once stable simulation parameters were established to provide a baseline state, the degree to which metabolic alteration to the system could be tested.

With increasing mass-specific metabolic rate (whale $\rightarrow$ mouse), Model III predicted a disproportionate increase in ROS production with respect to number and degree of damaged mitochondria, suggesting positive feedback loops play a more important role in determining mutation rate in smaller, shorter-lived mammals with fast metabolisms – i.e. the slight increase in cellular ATP and ROS production rates in fast metabolism species, if not quenched by antioxidant and other suppressive mechanisms, triggers earlier and more accelerated degradation of mitochondria (and associated accelerating ROS production and damage) compared to slow metabolism species. Though only three

species models were used, limiting the degree to which scaling rules can be extrapolated, the exponent approximated -0.3, not dissimilar to established -0.25/-0.33 metabolic scaling exponents for Kleiber's law (Brody & Procter, 1932; Savage *et al*., 2004).

Though only an early stage developmental model, using only known measured input parameters (ml O2 p/t), directly proportional to ATP consumption rate, the dynamics of the model were able to recapitulate allometric scaling in ROS production and nuclear DNA damage rate, supporting the notion accumulation of damaged mitochondria within a cell lineage represents a hallmark of cancer, often associated with the Warburg hypothesis

4.3. Driver Mutations

Both models I and II present evidence suggesting the most compelling (and likely easiest) target in the multistage framework to reduce cancer risk was the required number of genetic controls, or rate-limiting driver mutations (m_d). In both models, only a relatively small number of additional rate-limiting steps (between 2 and 4) were required for an organism with a body mass x1000 a human to present an incidence phenotype $<5\%$ (with or without metabolic scaling). As discussed in Chapter 2, a substantial body of evidence supports this result, the most obvious means via an increase in TSG redundancy - as seen by *TP53* and *LIF* gene expansion in elephants (Fig 12i). However, if under the assumptions of Model IIa only two additional TSG duplications are sufficient to reduce cancer risk to $<5\%$ in elephants (Fig 12ii-iv), why then does the elephant tentatively have upwards of ~20 or more redundant TSG copies?

One explanation is that they don't - at least not functionally. Another is that they're expressed only within narrow domains outside the broader remits of their canonical parent genes. Indeed, as discussed in the following chapter (5.1.2), there remains little evidence to suggest the majority of the duplicate sequences are functional, and, further, those which are translated seem to present some degree of neo- or subfunctionalisation, limited temporo-spatially to specific niches. Given the complexity of the neoplastic phenotype, there are numerous pathways to cancer, and perhaps each supernumerary retrogene in the Probiscidean clade has remained having specialised in reducing risk along one or more subpathways, as opposed to a broader level of protection typical of the function of most TSGs. Nonetheless, as much as the degree of expansion as seen in elephants seems superficially unnecessary, at least until it is further understood, the recruitment of additional driver mutations remains a critical means of reducing incidence risk.

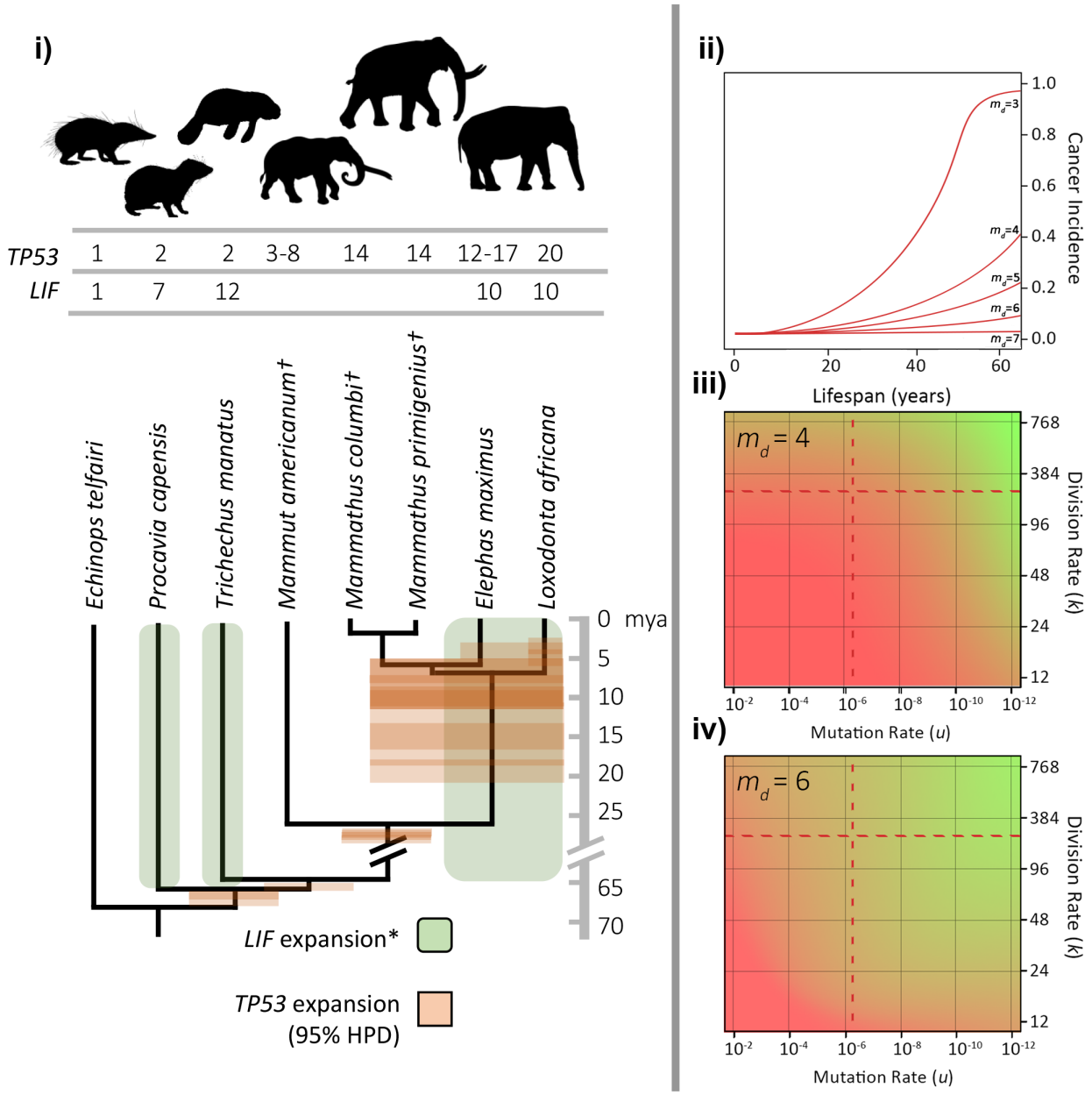


Fig. 12: i) *TP53* and *LIF* expansion occurs within Paenungulata; a single canonical duplication basally, with subsequent repeated retrotransposition into the Probiscidean clade († denotes extinct species), associated with body size increase within the last ~25mya. 95% highest posterior density (HPD) of estimated divergence dates are shown for *TP53* (from Sulak *et al.*, 2016); *LIF* expansion dates (and presence in extinct Probiscideans) uncertain (denoted *). **ii)** $m_d < 6$ is sufficient to reduce lifetime incidence risk to within expected ranges based upon Model IIa without metabolic scaling in the African elephant. An additional two driver mutations (m_d) from baseline of 4 **(iii)** to 6 **(iv)** increases potential available parameter space substantially (simplified elephant estimates given by red dashed line).

4.4. Cell Dynamics and Tissue Architecture

Despite previous hypotheses, simulations incorporating the cellular dynamics of symmetrically and asymmetrically dividing stem cells, and subsequent pools of transit-amplifying cells, did not produce results supporting any quantitative bearing on model outcomes, and were thus discarded. Similarly (though not shown) there was no behavioural difference in outcomes for either cancer subtype modelled (colorectal and lung).

Input parameters differ not only between species, but between intra-organismal tissue and cell types, Here, fast-dividing tissue types (colonic crypts) were selected given that such 'fast' tissues are most at risk of mutation. A speculative model (not shown) based on organismal surface area in order to simulate melanoma progression was begun, but not pursued; though such a model may be more appropriate when comparing cross-taxa and against observed data, as reliably scaling organ systems without adequate data is problematic (see also Chapter 6: 2.2).

4.5. Conclusion

As much as the intrinsic size-dependent scaling of metabolic rate can demonstrably reduce cancer incidence, it alone is not sufficient to explain the variance in observed cancer incidence across the animal kingdom, nor the higher and lower observed incidence of cancer respectively in the smallest and largest mammals compared to levels expected under Peto's paradox. In contrast, enhanced genetic control, either through modification of the number of rate-limiting steps or by other means, is potentially sufficient alone to remove cancer as a constraining factor on evolutionary trajectories that would otherwise considerably increase cancer risk.

It's likely that both factors contribute towards cancer resistance; any suppressive benefit gained by intrinsic scaling offsetting the degree to which adaptive genetic control would have to be modified in order to present a phenotype, either directly or indirectly, within cancer-associated constraints.

In the next chapter, taking inspiration from analyses of the elephant genome, comparative genomic analyses are undertaken on the numerous available cetacean genomes to see whether any evidence of similar genetic changes that, for example, increase the number of rate-limiting steps in elephants, can help resolve Peto's paradox in cetaceans.

5. References

– Armitage, P. & Doll, R. (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer*. 8 (**1**), 1-12

– Bai, J. & Cederbaum, A.I. (2001) Mitochondrial catalase and oxidative inquiry. *Biological Signals & Receptors*. 10 (**3-4**), 189-199

– Bandy, B. & Davison, A.J. (1990) Mitochondrial mutations may increase oxidative stress: implications for carcinogensis and aging. *Free Radical Biology & Medicine*. 8 (**6**), 523-539

– Baria, G. (2007) Mitochondrial oxygen consumption and reactive oxygen species production are independently modulated: implications for aging studies. *Rejuvination Research*. 10 (**2**), 215-24

– Barker, N. (2013) Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nature Reviews Molecular Cell Biology*. 15, 19-33

– Beerenwinkel, N., Antal, T., Dingli, D., Trauslen, A., Kinzler, K.W., Velculescu, V.E., Vogelstein, B. & Nowak, M.A. (2007) Genetic Progression and the Waiting Time to Cancer. *PLoS Computational Biology*. 3 (**11**), e225

– Blanpain, C., Horsley, V. & Fuchs, E. (2008) Epithelial Stem Cells: Turning over New Leaves. *Cell*. 128 (**3**), 445-458

– Box, G.E.P. (1976) Science and Statistics. *Journal of the American Statistical Association*. 71 (*356*), 791-799

– Brody, S. & Procter, R.C. (1932) Relation between basal metabolism and mature body weight in different species of mammals and birds. *Research bulletin (University of Missouri. Agricultural Experiment Station)*.166, 89–101

– Bueno, J. & López-Urrutia, A. (2014) Scaling up the curvature of mammalian metabolism. *Frontiers in Ecology and Evolution*. 2

– Cairns, J. (2002) Somatic stem cells and the kinetics of mutagenesis and carcinogenesis. *PNAS*. 99 (**16**), 10,567-10,570

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

– Calabrese, P., Tavaré, S. & Shibata, D. (2004) Pretumor Progression: Clonal Evolution of Human Stem Cell Populations. *The American Journal of Pathology*. 164 (**4**), 1,337-1,346

– Cao, Z., Lindsay, J.G. & Isaacs, N.W. (2007) Mitochondrial peroxiredoxins. *Subcellular Biochemistry*. 44, 295-315

– Carulli, A.J., Samuelson, L.C. & Schnell, S. (2014) Unraveling intestinal stem cell behavior with models of crypt dynamics. *Integrative Biology*. 6 (**3**), 243-257

– Caulin, A.F., Graham, T.A., Wang, L.S. & Maley, C.C. (2015) Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. *Proceedings of the Royal Society B: Biological Sciences*. 370 (**1673**), e20140222

– Cline, S.D¿ (2012) Mitochondrial DNA damage and its consequences for mitochondrial gene expression. *Biochimica et Biphysica acta*. 1,819 (**9-10**), 979-991

– Cui, H., Kong, Y. & Zhang, H. (2012) Oxidative stress, mitochondrial dysfunction, and aging. *Journal of Signal Transduction*. e646354

– Dang, C.V. (2015) A metabolic perspective of Peto's paradox and cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 370 (**1673**), e20140223

– Dizdaroglu, M., Jaruga, P., Birincloglu, M. & Rodriguez, H. (2002) Free radical-induced damage to DNA: mechanisms and measurement. *Free Radical Biology and Medicine*. 32 (**11**), 1,102-1,115

– Dudek, J., Rehling, P. & van der Laan, M. (2013) Mitochondrial protein import: Common principles and physiological networks. *BBA: Molecular Cell Research*. 1833 (**2**), 274-285

– Fleming, J.M., Creevy, K.E. & Promislow, D.E.L. (2011) Mortality in North American Dogs from 1984 to 2004: An Investigation into Age-, Size-, and Breed-Related Causes of Death. *Journal of Veterinary Internal Medicine*. 25 (**2**), 187-198

– Frank, S.A. (2007) Chp.4: History of Theories. In: *Dynamics of Cancer: Incidence, Inheritance, and Evolution*. Princeton University Press, Princeton (NJ), USA.

– Gelman, A. & Hill, J. (2006) *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

– Gillespie, D.T. (1977) Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry*. 81 (**25**), 2340-2361

– Gilloly, J.F., Allen, A.P., West, G.B. & Brown, J.H. (2005) The rate of DNA evolution: Effects of body size and temperature on the molecular clock. *PNAS*. 102 (**1**), 140-145

– Gilloly, J.F., McCoy, M.W. & Allen, A.P. (2007) Effects of metabolic rate on protein evolution. *Biology Letters*. 3 (**6**)

– Gottlieb, R.A. & Gustafsson, A.B. (2011) Mitochondrial turnover in the heart. *BBA: Molecular Cell Research*. 1813 (**7**), 1295-1301

– Green, J., Cairns, B.J., Casabonne, D., Wright, F.L., Reeves, G. & Beral, V. (2011) Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *The Lancet Oncology*. 12 (**8**), 785-794

– Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M. Charlesworth, B. & Knightley, P.D. (2008) Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biology*. 6 (**8**)

– Helleday, T., Eshtad, S. & Nik-Zainal, S. (2014) Mechanisms underlying mutational signatures in human cancers. *Nature Reviews Genetics*. 15, 585-598

– Hinkle, P.C. (2005) P/O ratios of mitochondrial oxidative phosphorylation. *BBA: Bioenergetics*. 1706 (**1-2**), 1-11

– Hornsby, C., Page, K.M. & Tomlinson, I.P.M. (2007) What can we learn from the population incidence of cancer? Armitage and Doll revisited. *The Lancet Oncology*. 8 (**11**), 1,030-1,038

– Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press.

– Idelchik, M.D.P.S., Begley, U., Begley, T.J. & Melendez, J.A. (2017) Mitochondrial ROS control of cancer. *Seminar in Cancer Biology*. 47, 57-66

– Ježek, J., Cooper, K.F. & Strich, R. (2018) Reactive Oxygen Species and Mitochondrial Dynamics: The Yin and Yang of Mitochondrial Dysfunction and Cancer Progression. *Antioxidants (Basel)*. 7 (**1**), 13

– Kansanen, E., Kuosamen, S.M., Leinonen, H. & Lenoven, A.L. (2013) The Keap1-Nrf2 pathway: Mechanisms of activation and dysregulation in cancer. *Redox Biology*. 1 (**1**), 45-49

– Kleiber, M. (1932) Body size and metabolism. *Hilgardia*. 6 (**11**), 315-353

– Knust, J., Ochs, M., Gundersen, H.J. & Nyengaard, J.R. (2009) Stereological estimates of alveolar number and size and capillary length and surface area in mice lungs. *The Anatomical Record*. 292 (**1**), 113-22

– Kolokotrones, T. Savage, V., Deeds, E.J. & Fontana, W. (2010) Curvature in metabolic scaling. *Nature*. 464, 753-756

– Komarova, N.L. & Wang, L. (2004) Initiation of colorectal cancer: where do the two hits hit?. *Cell Cycle*. 3 (**12**), 1,158-1,565

– Kuhn, M. (2008) Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*. 28 (5)

– Leinonen, H.M., Kansanen, E., Pölönen, P., Heinäniemi, M. & Levonen AL. (2014) Role of the Keap1-Nrf2 pathway in cancer. *Advances in Cancer Research*. 122, 281-230

– Lindstedt, S.L. & Schaeffer, P.J. (2002) Use of allometry in predicting anatomical and physiological parameters of mammals. *Laboratory Animals*. 36 (**1**), 1-19

– Liochev, S.I. (2013) Reactive oxygen species and the free radical theory of aging. *Free Radical Biology & Medicine*. 60, 1-4

– Lockyer, C. (1976) Body weights of some species of large whales. *Journal du Conseil / Conseil Permanent International pour l'Exploration de la Mer*. 36 (**3**), 259-273

– Lopez-Garcia, C., Klein, A.M., Simons, B.D. & Winton, D.J. (2010) Intestinal stem cell replacement follows a pattern of neutral drift. *Science*. 330 (**6005**), 822-825

– Mari, M., Morales, A., Colell, A., Garcia-Rulz, C. & Fernandez-Checa, J.C. (2009) Mito- ´ chondrial Glutathione, a Key Survival Antioxidant. *Antioxidants & Redox Signaling*. 11 (**11**), 2,685-2,700

– Milwa, S., Lawless, C. & von Zglinicki, T. (2008) Mitochondrial turnover in liver is fast in vivo and is accelerated by dietary restriction: application of a simple dynamic model. *Aging Cell*. 7 (**6**), 920-923

– Moran, P. A. P. (1958). Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*. 54 (**1**), 60–71

– Müller, M.J., Wang, Z., Heymsfield, S.B., Schautz, B. & Bosy-Westphal, A. (2013) Advances in the understanding of specific metabolic rates of major organs and tissues in humans. *Current Opinion in Clinical Nutrition & Metabolic Care*. 16 (**5**), 501-508

– Mumbengegwi, D.R., Li, Q., Li, C., Bear, C.E. & Engelhardt, J.F. (2008) Evidence for a Superoxide Permeability Pathway in Endosomal Membranes. *Molecular and Cellular Biology*. 28 (**11**), 3700-3712

– Nishiwaki, M. (1950) On the Body Weight of Whales [online]. [online] Available at: tknk.io/ZdyG

– Nordling, C.O. (1953) A New Theory on the Cancer-inducing Mechanism. *British Journal of Cancer*. 7 (**1**), 68-72

– Nunney, L. (2018) Size matters: height, cell number and a person's risk of cancer. *Proceedings of the Royal Society B: Biological Sciences*. 285 (**1889**), e20181743

– Nunney, L. (1999) Lineage selection and the evolution of multistage carcinogenesis. *Proceedings of the Royal Society B: Biological Sciences*. 266 (**1418**), 493-498

– Ochs, M., Nvengaard, J.R., Jung, A., Knudsen, L., Voigt, M., Wahlers, T., Richter, J. & Gundersen, H.J. (2004) The number of alveoli in the human lung. *American Journal of Respiratory and Critical Care Medicine*. 169 (**1**), 120-124

– Pastula, A. & Marcinkiewicz, J. (2019) Cellular Interactions in the Intestinal Stem Cell Niche. *Achivum Immunologiae et Therapiae Experimentalis*. 67, 19-26

– Pich, M.M., Raule, N., Catani, L., Fagioli, M.E., Faenza, I., Cocco, L & Lenaz, G. (2004) Increased transcription of mitochondrial genes for Complex I in human platelets during ageing. *FEBS Letters*. 558 (**103**), 19-22

– Pickhardt, P.J., Halberg, R.B., Taylor, A.J., Durkee, B.Y., Fine, J., Lee, F.T. & Weichert, J.P. (2005) Microcomputed tomography colongraphy for polyp detection in an in vivo mouse tumor model. *PNAS*. 102, 3419-3422

– Pierson, R.N. (2012) *Quality of the Body Cell Mass: Body Composition in the Third Millenium*. 2nd Edition. Serono Symposia USA.

– Pine, S.R., Marshall, B. & Varticovski, L. (2008) Lung cancer stem cells. *Disease Markers*. 24 (**4-5**), 257-266

– Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M.L., Ordonzes, G.R., Bignell, G.R., Ye, K., Alipaz, J., Bauer, M.J., Beare, D., Butler, A., Carter, R.J., Chen, L., Cox, A.J., Edkins, S., Kokko-Gonzales, P.I., Gormley, N.A., Grocock, R.J., Haudenschild, C.D., Hims, M.M., James, T., Jia, M., Kingsbury, Z.,Leroy, C., Marshall, J., Menzies, A., Mudie, L.J., Ning, Z., Royce, T., Schulz-Trieflaff, O.B., Spiridou, A., Stebbings, L.A., Szajkowski, L., Teague, J., Williamson, D., Chin, L., Ross, M.T., Campbell, P.J., Bentley, D.R., Futreal, P.A. & Stratton, M.R. (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*. 463, 191-196

– Rodic, S. & Vincent, M. (2017) Reactive oxygen species (ROS) are a key determinant of cancer's metabolic phenotype. *International Journal of Cancer*. 142 (**3**), 440-448

– Sanz, A. & Stefanatos, R.K.A. (2008) The Mitochondrial Free Radical Theory of Aging: A Critical View. *Current Aging Science*. 1 (**1**), 10-21

– Satoh, M. & Kuroiwa, T. (1991) Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Experimental Cell Research*. 196 (**1**), 137-140

– Savage, V.M., Allen, A.P., Brown, J.H., Gillooly, J.F., Herman, A.B., Woodruff, W.H. & West, G.B. (2007) Scaling of number, size, and metabolic rate of cells with body size in mammals. *PNAS*. 104 (**11**), 4,718-4,723

– Savage, V.M., Gillooly, J.F., Woodruff, W.H., West, G.B., Allen, A.P., Enquist, B.J. & Brown, J.H. (2004) The predominance of quarter-power scaling in biology. *Functional Ecology*. 18 (**2**), 257-282

– Sena, L.A. & Chandel, N.S. (2012) Physiological roles of mitochondrial reactive oxygen species. *Molecular Cell*. 48 (**2**), 158-167

– Shahriyari, L. & Komarova, N.L. (2013) Symmetric vs. asymmetric stem cell divisions: an adaptation against cancer?. *PLoS One*. 8 (10), e76195

– Sun, J., Ren, X. & Simpkins, J.W. (2015) Sequential Upregulation of Superoxide Dismutase 2 and Heme Oxygenase 1 by tert-Butylhydroquinone Protects Mitochondria during Oxidative Stress. *Molecular Pharmacology*. 88 (**3**), 437-449

– Taanman, J.W. (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et biophysica acta*. 1410 (2), 102-123

– Tan, Z. (1994) DNA damage and the proliferation and aging of cells in culture: a mathematical model with time lag. *Mathematical Biosciences*. 122 (**1**), 67-88

– Turrens, J.F. (2003) Mitochondrial formation of reactive oxygen species. *The Journal of Physiology*. 552 (**2**), 335-344

– Wallig, M. (2020a) *foreach: Provides Foreach Looping Construct*. [online] Available at: https://CRAN.R-project.org/package=foreach

– Wallig, M. (2020b) *doMC: Foreach Parallel Adaptor for 'parallel'* [online] Available at: https://CRAN.R-project.org/package=doMC

– White, C.R. & Seymour, R.S. (2004a) Mammalian basal metabolic rate is proportional to body mass $^{2/3}$. *PNAS*. 100 (**7**), 4,046-4,049

– White, C.R. & Seymour, R.S. (2004b) Does basal metabolic rate contain a useful signal? Mammalian BMR allometry and correlations with a selection of physiological, ecological, and life-history variables. *Physiological and Biochemical Zoology*. 77 (**6**), 929-941

– Wickens, A.P. (2001) Ageing and the free radical theory. *Respiration Physiology*. 128 (**3**), 379-391

– Wiesner, R.J., Rüegg, J.C. & Morano, I. (1992) Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues. *Biochemical and Biophysical Research Communications*. 183 (**2**), 553-559

– Zheng, T. (1991) A mathematical model of proliferation and aging of cells in culture. *Journal of Theoretical Biology*. 149 (**3**), 287-315

– Zhang, J., Wang, X., Vikash, V., Ye, Q., Wu, D., Liu, Y. & Dong, W. (2016) ROS and ROS-Mediated Cellular Signaling. *Oxidative Medicine and Cellular Longevity*. e4350965

– Zou, Z., Chang, H., Li, H. & Wang S. (2017) Induction of reactive oxygen species: An emerging approach for cancer therapy. *Apoptosis*. 22, 1,321–1,335

Bibliography

– Armitage, P. (1985) Multistage Models of Carcinogenesis. *Environmental Health Perspectives*. 63, 195-201

– Chen, H. & Chan, D. (2017) Mitochondrial Dynamics in Regulating the Unique Phenotypes of Cancer and Stem Cells. *Cell Metabolism*. 26 (**1**), 39-48

– Nunney, L. (2015) Commentary: The multistage model of carcinogenesis, Peto's paradox and evolution. *International Journal of Epidemiology*. 45 (**3**), 649-653

– Nunney, L. (2013) The real war on cancer: the evolutionary dynamics of cancer suppression. *Evolutionary Applications*. 6 (1), 11-19

– Vyas, S., Zaganjor, E. & Haigis, M. (2016) Mitochondria and Cancer. *Cell*. 166 (**3**), 555-566

– Wallace, D.C. (2012) Mitochondria and cancer. *Nature Reviews Cancer*. 12 (**10**), 685-698

– Zong, W.X., Rabinowitz, J.D. & White, E. (2016) Mitochondria and Cancer. *Molecular Cell*. 61 (**5**), 667-676

Chapter 5

Comparative Genomics of Cetaceans

Abstract

Under the multistage model of cancer progression, the number of driver mutations required for carcinogenesis presents a target for adaptive selection; increase in the genetic redundancy of tumour-suppressor genes (TSGs) resulting in lower risk of neoplasia. Here, a comparative genomic approach is undertaken, presenting evidence of signatures of genetic change which may engender low cancer incidence rates in large, long-lived mammals, and thereby explain Peto's paradox.

Utilising 1,737 query cancer-associated genes and the 27 publicly available cetacean genomes, signatures of tumour-suppressor duplication and both size- and lineage-associated accelerated selection in cetacean lineages were explored. 62 TSG duplicates were uncovered, 17 previously unreported (*CASP3*, *FAT4*, *HINT1*, *DUSP22* etc.), though unlike in elephants - where repetitive expansion of a basal *TP53* may explain low incidence rates across the entire elephantine family - most are lineage-specific single duplicates, and few are basal in cetacean clades having evolved gigantism. Modelled dN/dS rates further presented evidence of 32 cancer-associated genes having undergone positive selection, 25 of which identified as associated with substantial increase in body mass; presenting potential candidate genes for further targeted anti-cancer investigation.

In general, results provide evidence that alteration to genetic control mechanisms are present in these large, long-lived mammals, which could explain low incidence rates, thereby providing insight into adaptive trajectories towards cancer resistance.

Introduction

Published in 2003, the Human Genome Project presented the first reference sequence of the human genome, having taken 13 years and over $450 million to complete (Mardis, 2011). Today, the cost to sequence, assemble and annotate a full reference genome is comparatively trivial, in recent years causing a veritable explosion in high-quality genomic data available for comparative analysis.

In previous chapters, intrinsic life-history factors, such as metabolic rate, were discovered to be likely insufficient to correct for Peto's paradox alone, thus the selective recruitment of additional means of genetic control may instead present a solution. As reviewed in Chapter 2, this can be achieved via suppressing somatic mutation and variation - either through duplication of tumour suppressor genes (TSGs) or elimination of proto-oncogenes - or by suppressing somatic advantage and proliferation - via increase in DNA damage response sensitivity, shorter telomeres and other mechanisms (see Chapter 2; 2.2-2.3 for full review). These means of additional control can present themselves by investigating the sequence-level data of a species' genome, utilising several methods.

1.1. Comparative Genomics

Comparative genomic methods rely on the assumption that homologous and functional genomic regions ought to remain evolutionarily conserved across species. This is a result of positive selection; the process by which evolutionary forces favour the retention in a population of those mutations in a gene that present a fitness benefit - in contrast to neutral processes acting on genetic changes which confer neither advantage nor disadvantage to the organism. DNA sequences corresponding to functional homologous regions are therefore expected to be more similar to each other between species than non-functional sequences, and that the similarity between non-functional sequences correlates with the evolutionary distance between two species (upon which the molecular clock hypothesis is based).

Most genes originate via gene duplication, a common mechanism in molecular evolution resulting from unequal crossing over events, retrotransposition or chromosomal (or genome) duplication (Magadum *et al*., 2013). When a gene duplicates, it creates a paralogous copy, which may then acquire a new function (neofunctionalization) or maintain the original function of the parent gene (subfunc-

tionalisation). Gene copies across species derived from the same ancestral gene are orthologs, and usually maintain similar functions across species.

The most common method to ascertain sequence similarity in comparative genomics is to use search algorithms such as BLAST. Retrieved sequences are usually then aligned to orthologs from one or more species, and the alignment searched for mutations of functional (i.e. non-synonymous, or amino-acid changing) importance. Positive selection can be detected by signals of accelerated evolution, often indicative of lineage-specific adaptation (Yang & Nielsen, 2002). The rate of protein evolution is measured using the ratio of synonymous (*dS*) and non-synonymous nucleotide changes per respective site (dN/dS, usually referred to as ω). If ω =1, selection is neutral, with ω >1 and ω <1 representing positive and purifying (negative) selection respectively; the greater the divergence from 1, the greater the selection (Miyata & Yasunaga, 1980).

Positive selection seldom occurs universally in all species within a phylogeny, or across all regions of a gene - the likelihood is positive selection only occurs in some branches, some sites, or, most realistically, only in specific sites in specific branches. These possibilities are formalised into models - 'branch models', 'site models', and 'branch-site' models respectively.

1.2. Cancer Suppression in Elephants

Utilising comparative methods, several studies have investigated duplication events and positive selection in cancer-associated genes in elephants. These investigations have revealed that African elephants (*Loxodonta africana*) encode a single *TP53*, and 19 duplicated *TP53* retrogenes (TP53RTGs), 14 of which retain sequence potential to encode truncated proteins (Sulak *et al*., 2016; Abegglen *et al*., 2015). Similarly, elephants encode a single canonical *LIF* gene (involved in pro-apoptotic response and *TP53* regulation), and 10 additional pseudogenes (Vazquez *et al*., 2018) (see Fig.1).

Comparative genomic analysis of the Paenungulata (one of two sister branches in Afrotheria, containing modern elephants, sea cows and hyraxes) indicates an initial duplication of TP53 via retrotransposition occurred at the base of the lineage approximately 64mya. No further retrotranspositions nor segmental duplications were fixed in the genome until ~40mya, correlating strongly with body size expansion in the Proboscidean lineage, resulting in estimated TP53RTG numbers of 12-17 in the extant Asian elephant (*Elephas maximus*), 14 in the extinct woolly (*Mammathus primigenius*) and

Columbian (*Mammathus columbi*) mammoths, and 3-8 in the extinct American mastodon (*Mammut Americanum*; diverging from the rest ~25mya) respectively. As for *LIF*, based on reconstructed gene trees and reconciliation models it appears duplicates have arisen and become fixed frequently within Paenungulata (at least 17 duplication, and 14 loss events) via serially repeating segmental duplication.

Do these loci encode functional genes and, if so, are they causally associated with body size expansion? To date, it isn't entirely clear. Comparative bioinformatic methodology is limited in its ability to assess functional phenotypic change, and requires experimental validation.

Experimental evidence via RNAseq on extant elephant cell lines indicates at least five, but not all, TP53RTGs are transcribed - *TP53RTG5*, "*12*, "*13*, and "*18-19* - with evidence of translation in both *TP53RTG12* and *TP53RTG19* in elephant fibroblasts. *LIF6* is similarly transcribed and translated in fibroblasts, and the only duplicate associated with a *TP53* response element. Experimental evidence confirmed enhanced apoptotic response to induced DNA damage was present in elephant cell lines, at lower damage thresholds compared to control, and this increased apoptotic activity, mediated by a hyperactive *TP53* signalling pathway which was dependent on *TP53RTG12* activity - when silenced via siRNA, a comparable response was not observed. Similar siRNA experiments confirmed *LIF6* activity was crucial in presenting increased apoptotic sensitivity. Unlike the canonical *LIF*, or spliced variants (i.e. *LIF-T*), which has diffuse cytoplasmic and nuclear localisation, *LIF6* presents unique (compared to control lines) localisation and activity limited to mitochondrial dysfunction.

Given many duplicate genes are only preserved as they evolve with either spatial or temporal-specific expression patterns (subfunctionalisation), or new functions altogether (neofunctionalization), it is unclear whether any of the other TP53RTGs are functional, or merely present as nonfunctional artefacts of selectively neutral duplication processes. *LIF6* is unique in retaining a *TP53* transcription binding factor, and given it is deeply nested within the *LIF* duplicate clade, suggests it has re-evolved into a functional gene from a pseudogenic ancestor. This is corroborated via trans-Probiscidean positive selection analysis - it is the only duplicate to present evidence of increasing positive selection, consistent with a model of refunctionalisation.

Analysing the genomes of elephants has made it clear that functional, selectively retained duplicate TSGs are likely involved in preventing cancer incidence in elephant species. Might evidence of similar changes be present in the genomes of cetaceans?

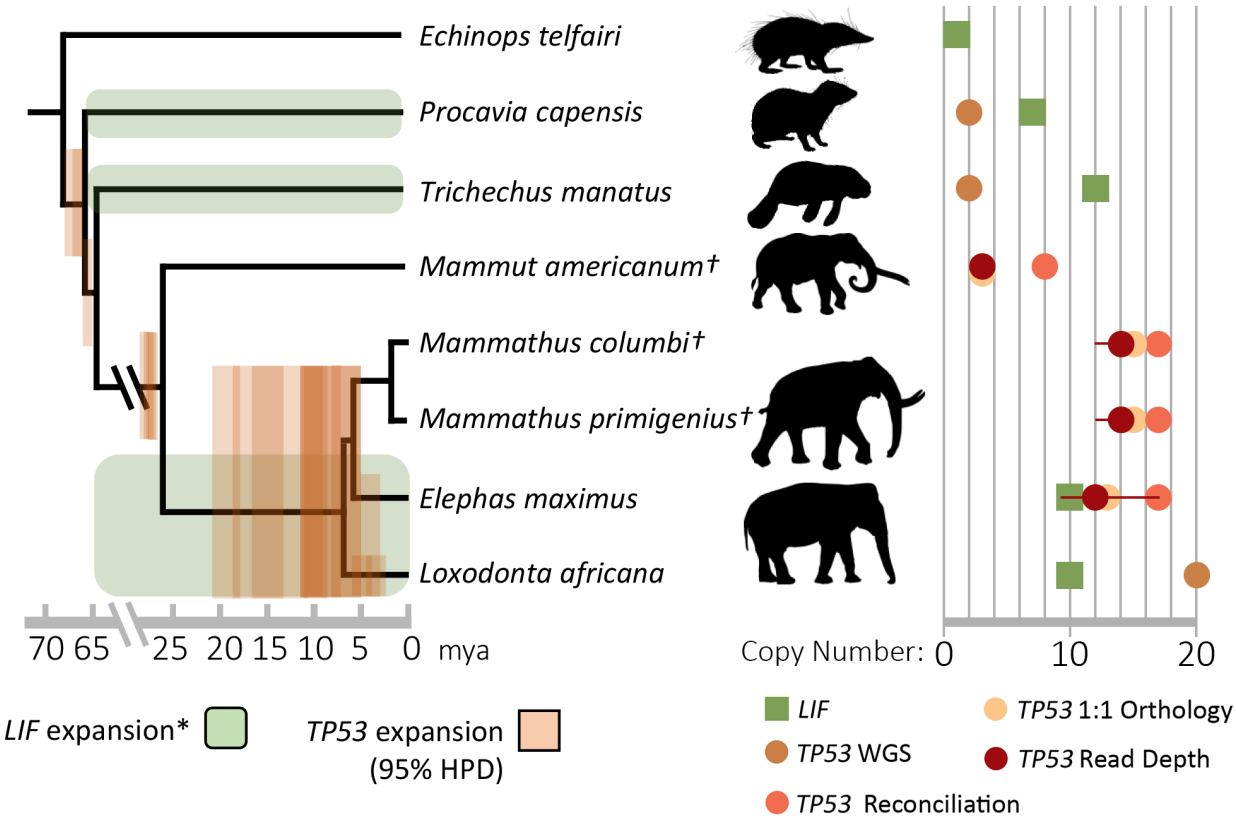


Fig. 1: Estimated *TP53/TP53RTG* and *LIF* copy number inferred from complete genome sequencing data (WGS), 1:1 orthology, gene tree reconciliation, and normalized read depth from genome sequencing data. Whiskers on normalized read depth copy number estimates show the 95% confidence interval of the estimate (adapted from Sulak *et al.*, 2016). Initial *TP53* expansion occurred basally in Paenungulata, with subsequent iterative retrotransposition in the Probiscidean clade coinciding with body size increase in the last ~25mya (95% highest posterior density (HPD) shown for TP53RTGs).

With recent advances in genome sequencing, and the availability of cetacean genomes representing multiple species across the lineage with diverse body sizes, only recently has the opportunity become available to investigate molecular mechanisms underlying cancer suppression in cetaceans. At project onset, no study had yet leveraged this opportunity. Given body size has increased several times independently not only within cetaceans, but across mammals in general, duplication of a single specific gene is unlikely to explain Peto's paradox – indeed, following the discovery of *TP53* duplication in elephants, attempts have failed to find a similar *TP53* duplication in cetaceans. It is likely multiple independent genetic changes have arisen across multiple lineages; here, 27 whole genome sequences and associated shotgun and transcriptomic libraries will therefore be screened for homologs of all reasonably established human (as query) TSGs and proto-oncogenes. The resulting cetacean sequences will then be used to test the hypotheses that, in lineages which have present an increase in longevity

and/or body size since the ancestral condition:

- TSG duplication events have occurred, and there may be cross-taxa positive relationships between duplicate copy number and body size and/or longevity.
- Gene deletion or loss-of-function in proto-oncogene homologs may be present
- Evidence of positive / accelerating selection will be present
- Similar non-synonymous sequence changes may have arisen independently in the same homologs in two or more lineages (parallel / convergent evolution).

Methods

2.1. Query Sequences

A database of human cancer-associated genes was collated from entries in the TSGene 2.0 (Zhao *et al*., 2016) and NCG 6.0 (Repana *et al*., 2019) published databases. This collated dataset included information on gene name and all aliases, type (proto-oncogene or TSG), associated public database IDs and a binary categorisation based on experimental support; those genes whose involvement in cancer has robust experimental support were defined as 'confirmed', whilst those whose somatic alterations have predicted cancer driver roles but lacked further experimental support were noted as 'unconfirmed'. In total, 1,737 human cancer-associated gene entries were collected, including 711 genes with experimentally validated cancer association.

Entrez IDs for each gene were converted to RefSeq IDs using DAVID (Huang, Sherman & Lempicki, 2009), presenting 27,719 available sequence entries. An initial filtering process removed sequences tagged as non-coding, along with those retaining incomplete codon sequences and/or unsuitable open-reading frames (i.e. those non-divisible by three) and with premature (internal) stop codons (though the latter were binned for later analysis into gene deletion). Transcripts of a considerable number of isoforms, produced via alternative splicing, remained; canonical DNA sequences were therefore determined by cross-comparison with 3,069 protein sequences defined as canonical by UniProt (see: `uniprot.org/help/canonical_and_isoforms`). Of the remaining 219 genes lacking successful UniProt mapping, the sequence chosen for any given gene was determined via maximum length. Once canonical coding sequences for all 1,737 query genes were established, the associated mRNA .fasta sequences were downloaded via the UCSC Table Browser (Karolchik *et al*., 2004).

2.2. Genomes

Whole-genome assemblies for 7 cetacean species were used as reference genomes (see Table 1), alongside unassembled shotgun libraries for a further 2 species. In addition, representative genomes of the hippopotamus (*Hippopotamus amphibius*), domesticated cow (*Bos taurus*), and human (*Homo sapiens*) were used as outgroups and/or query genomes for a reciprocal BLAST. The genome sequences for each species were downloaded in full from the NCBI and the Bowhead Whale Genome Project websites (Keane *et al*., 2015), for local (as opposed to web query) analysis.

2.3. Reciprocal BLAST Analysis and Copy Number Estimation

In order to analyse genomic changes, homologs of the query human cancer-associated genes were identified - and, crucially, results differentiated between orthologs (original sequences that diverged from one another at the most recent common ancestor of both species) and paralogs (sequences that diverged at another time - i.e. via a subsequent duplication event). Not only are we interested in confirming potential duplication events, but also in order to determine rates of protein evolution, comparisons can only be made between sequences which are strictly orthologous (as they are compared proportionally against relative evolutionary rates since time of species divergence) (see Fig.2a). Here, in order to identify and differentiate both orthologs and paralogs in cetacean genomes, an automated pipeline was developed to run a reciprocal best Basic Local Alignment Search Tool (BLAST) (RBB), utilising Wall *et al*.'s (2003) reciprocal smallest distance algorithm (RSD) to test orthology. Unlike typical RBB methodology, RSD uses global sequence alignment (using clustalW; Thompson, Higgins & Gibson, 1994) and maximum likelihood estimation of evolutionary distances (using PAML; Yang, 2007) to reduce the exclusion of authentic orthologous pairs and avoid being misled by the presence of close paralogs (see Fig. 2b).

FASTA mRNA sequences were first translated into amino acid sequences for BLASTp searches (as nucleotide BLAST searches - BLASTn - fare relatively poorly for finding protein-coding sequences), and a housekeeping script then appropriately pre-formats both query and genome sequence files for use with RSD and PAML. For a BLASTp hit to count as an independent homologous instance of a given gene query, it had to first meet threshold criteria for coverage and significance - by testing the e-values, % identity, and bit-scores. Here the union of all hits to that sequence in the subject's genome must exceed $>30\%$ across the entire length of the human query gene, and at least one of the BLAST hits in the region must have an e-value $<10^{-5}$, with all other hits counting towards the $>30\%$ coverage having e-values $<10^{-3}$.

Returned hits that met inclusion thresholds were binned, and subsequently separately aligned against the query sequence using clustalW (default settings). If the alignable region(s) exceed a threshold fraction of the total alignment length (>0.8), codeml via PAML was used to obtain maximum likelihood estimates of the number of substitutions separating the sequences. Of all returned sequences for which an evolutionary distance is estimated, the sequence yielding the shortest distance is then used for a reciprocal BLAST against the original query genome (human), and if any of the returned

Latin Name	Common Name	Assembly Name	GenBank ID	RefSeq Assembly ID	Coverage	Assembly Level	Sequencing Technology
Balaena mysticetus	Bowhead Whale	Bowhead Genome	PRJNA194091	-	x150	Scaffold	Illumina HiSeq
Eubalaena glacialis	North Atlantic Right Whale	Illumina sequencing of north atlantic right whale: cell culture	SRX2901265:	-	-	Shotgun	Illumina HiSeq 2500
Eubalaena japonica	North Pacific Right Whale	EubJap_v1_BIUU	GCA_004363455.1	-	x40	Scaffold	Illumina HiSeq
Balaenoptera acutorostrata	Common Minke Whale	BalAcu1.0	GCA_000493695.1	GCF_000493695.1	x60	Scaffold	Illumina HiSeq 2000
Balaenoptera bonaerensis	Antarctic Minke Whale	ASM97880v1	GCA_000978805.1	-	x92	Scaffold	Illumina HiSeq 2000
Eschrichtius robustus	Gray Whale	ASM218922v1	GCA_002189225.1		x11	Scaffold	Illumina HiSeq
Megaptera novaeangliae	Humpback Whale	megNov1	GCA_004329385.1	-	x102	Scaffold	Illumina HiSeq
Balaenoptera physalus	Fin Whale	Baphy	GCA_008795845.1	-	x40	Scaffold	Illumina
Balaenoptera musculus	Blue Whale	mBalMus1.alt.v2	GCA_008658375.2	-	x51.16	Scaffold	Illumina NovaSeq
Balaenoptera borealis	Sei Whale	Illumina sequencing of sei whale: cell culture	SRX2901260	-	N/A	Shotgun	Illumina HiSeq 2500
Physeter catodon	Sperm Whale	ASM283717v2	GCA_002837175.2	GCF_002837175.2	x248	Chromosome	BGISEQ-500
Kogia breviceps	Pygmy Sperm Whale	KogBre_v1_BIUU	GCA_004363705.1	-	x38.8	Scaffold	Illumina HiSeq
Platanista minor	Indus River Dolphin	PlaMin_v1_BIUU	GCA_004363435.1	-	x28	Scaffold	Illumina HiSeq
Ziphius cavirostris	Cuvier's Beaked Whale	ZipCav_v1_BIUU	GCA_004364475.1	-	x30.6	Scaffold	Illumina HiSeq
Mesoplodon bidens	Sowerby's Beaked Whale	MesBid_v1_BIUU	GCA_004027085.1	-	x32.4	Scaffold	Illumina HiSeq
Lipotes vexillifer	Baiji	Lipotes_vexillifer_v1	GCA_000442215.1	GCF_000442215.1	x115	Scaffold	Illumina HiSeq
Pontoporia blainvillei	La Plata Dolphin	PonBla_v1_BIUU	GCA_004363935.1	-	x55.3	Scaffold	Illumina HiSeq
Inia geoffrensis	Amazon River Dolphin	IniGeo_v1_BIUU	GCA_004363515.1	-	x42.5	Scaffold	Illumina HiSeq
Monodon monoceros	Narwhal	MonMon_v1_BIUU	GCA004027045.1	-	x34.7	Scaffold	Illumina HiSeq
Delphinapterus leucas	Beluga	ASM228892v3	GCA_002288925.2	GCF_002288925.2	x117	Scaffold	Illumina HiSeq X
Neophocoena asiaorientalis	Finless Porpoise	Neophocaena_asiaeorientalis_V1 - representative genome	GCA_003031525.1	GCF_003031525.1	x106	Scaffold	Illumina HiSeq 2000
Phocoena sinus	Vaquita	mPhoSin1.pri	GCA_008692045.1	GCF_008692025.1	x67.26	Chromosome	Illumina NovaSeq
Phocoena phocoena	Harbour Porpoise	ASM307100v2	GCA_003071005.1	-	x87.2	Scaffold	Illumina HiSeq
Orcinus orca	Orca	Oorc_1.1	GCA_000331955.2	GCF_000331955.2	x200	Scaffold	Illumina HiSeq
Lagenorhynchus obliquidens	Pacific White-Sided Dolphin	ASM367639v1	GCA_003676395.1	GCF_003676395.1	x35.68	Scaffold	Illumina HiSeq
Globicephala melas	Long-Finned Pilot Whale	ASM654740v1	GCA_006547405.1	GCF_006547405.1	x36	Scaffold	Illumina HiSeq
Tursiops truncatus	Bottlenose Dolphin	mTurTru1.mat.Y	GCA_011762595.1	GCF_011762595.1	x63.7	Chromosome	Illumina NovaSeq
Tursiops aduncus	Indo-Pacific Bottlenose Dolphin	ASM322739v1	GCA_003227395.1	-	x180	Scaffold	Illumina HiSeq
Sousa chinensis	Indo-Pacific Humpback Dolphin	S_chinensis_fine_genome_map	GCA_003521335.2	-	x107.6	Scaffold	Illumina HiSeq
Hippopotamus amphibius	Hippopotamus	HipAmp_v2_BIUU_UCD	GCA_004027065.2	-	x77	Scaffold	Illumina HiSeq
Bos taurus	Domesticated Cow	ARS-UCD1.2	GCA_002263795.2	GCF_002263795.1	x80	Chromosome	Illumina HiSeq
Homo sapiens	Human	GRCh38.p13	GCA_000001405.27	GCF_000001405.39	N/A	Chromosome	N/A

Table 1: Details of genomic and transcriptomic data publicly-available for 29 cetacean species, alongside human (*Homo sapiens*), domesticated cow (*Bos taurus*) and hippopotamus (*Hippopotamus amphibius*). The assembly level "scaffold" refers to both unordered contigs and ordered scaffolds.

sequences is the original query, the sequence is considered the putative ortholog. All other sequences are binned separately, and those returned 'one-to-many' tagged as duplicates for downstream inspection.

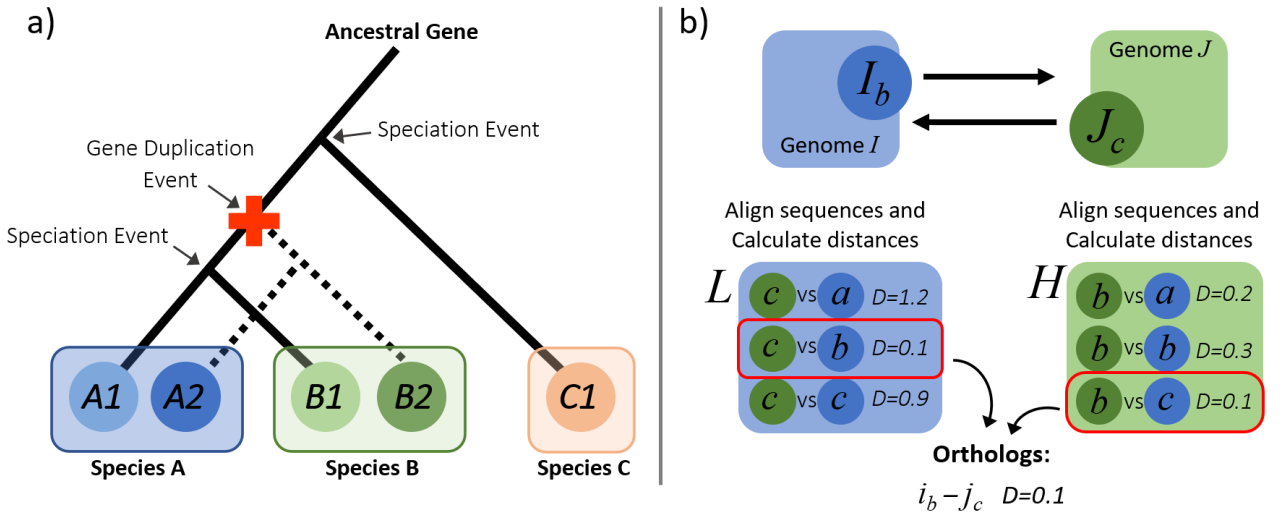


Fig. 2: **(a)** An ortholog is defined as a gene in two separate species that has evolved directly from a single gene in their most recent common ancestor – arising due to speciation, and (typically) retaining the original function in both species. Paralogs, by contrast, are sequences that diverged at some point subsequent to species divergence, and may retain similar or evolve new function (undergo positive selection), or lose function entirely via neutral genetic drift. Here, as example, an original duplication event results in two paralogous gene lineages. Though *A1* and *B1* are orthologs to *C1*, they're paralogs to *A2* and *B2*, and any measure of evolutionary divergence would be misrepresentative – only comparing *A2* and *B2*, or *C1* with *A1* or *B1*, for example, would yield germane estimates of rates of protein evolution.
(b) Reciprocal smallest distance algorithm (RSD) illustrated. Arrows denote bidirectional BLAST runs. After each run, hits are paired with the query to calculate evolutionary distances. If the same pair produces the smallest distance in both search directions, it is assumed to be orthologous. In summary: BLAST is undertaken with subject genome *J* and a query sequence *i* from genome *I*. A set of returned hits (*H*), exceeding inclusion criteria thresholds, is collated. Then, using clusalW, each sequence in *H* is aligned separately with the original query *i*. If the alignable region of both exceeds a pre-defined threshold, PAML is used to obtain maximum likelihood estimates for evolutionary distance between the two sequences. Of all sequences returned in *J*, only *j* (the sequence with the shortest distance) is retained, and then used as a query for a reciprocal BLAST against genome *I*, retrieving a set of high-scoring hits *L*. If any hit from *L* is the original query sequence (*i*), the distance between *i* and *j* are retrieved and the remaining hits from *L* are separately aligned with *j* and ML distance estimates calculated. If the protein sequence from *L*, producing the shortest distance to *j*, is the original query (*i*), it is assumed to be the true orthologous pairing. Adapted from Wall & DeLuca (2007).

2.4. Alignment and Gene Tree Generation

Codon-based nucleotide alignments were undertaken using PRANK (in codon mode) (Löytynoja, 2014), executed and monitored in parallel using multiple CPU cores to increase computational speed via doMC (Wallig, 2020) in R. The confidence of these was then assessed using two independent approaches. GUIDANCE2 (Sela *et al.*, 2015) was used to assess the sensitivity of alignments to perturbations in the associated guide tree - any low confidence scores resulted in manual review and either locale-specific masking (i.e. annotating simple repeats: ATATATAT; or regions overly-enriched with single letters: AAACAACAAA; to prevent spurious phylogenetic inference) or full alignment removal. T-Coffee (Notredame, Higgins & Heringa, 2000) was then used on translated PRANK alignments (with GUIDANCE2 masking) to assess alignment stability - this by independently re-aligning all possible pairs of sequences, scoring, and reviewing and/or removing alignments with low confidence scores. See SI.5 for alignment files.

Once alignments were constructed and quality controlled, phylogenetic trees were constructed in order to test evolutionary models. First, in order to generate a reference species tree, a housekeeping .pl script concatenates all masked alignments into a converted single PHYLIP alignment file, sorted by species, which is compatible with CodeML (i.e. stop codons removed, undetermined/masked codons - lowercase n - converted to CodeML ambiguous '?' etc.). Gene tree generation was then undertaken in RAxML (Stamatakis, 2014), using a GTR gamma model.

2.5. Modelling Natural Selection

To test for positive selection, indicated by evidence of accelerated evolution, ω (dN/dS) is estimated along different lineages root-to-tip from last common ancestor to see whether values of ω along some lineages are relatively high compared to other lineages, with a null model assuming all branches have equivalent values for ω (any genetic change confers no selective advantage nor disadvantage). All models used to estimate ω were undertaken using codeml (PAML) (see SI.5). In general, one of several models are used to calculate ω, and likelihood ratio tests (LRTs), which approximate chi-square distributions, are used to evaluate nested likelihood models (against null; M0). For each model, robustness of results was verified by a false discovery rate (FDR) process (Benjamini & Hochberg, 1995) using qvalue in R. In each instance, two steps were undertaken: the regression is repeated by i) excluding the point with largest residual error and compared to the original to check whether the

relationship is maintained (i.e. not skewed by outliers); and ii) iteratively excluding each species in order to calculate maximum possible p-value to ensure generalisability of the relationship (no single given species required to maintain slope and fit). Results which exceeded thresholds of $p<0.01$ and $p<0.05$ for each step respectively were collated.

An MCMC time-calibrated phylogenetic tree of the Cetacea clade (and additional outgroups) produced by McGowen *et al*. (2020) was used, adapted and trimmed (e.g discarding branches and internal nodes associated with non-applicable species) using phytools in R (Revell, 2012). Body mass and lifespan values for each species were taken from datasets established in Chapter 3.

2.5.1. Lineage-Specific Selection

To test whether positively selected genes were lineage-specific, both free-ratio (M1) and branch-site (M2) models were applied root-to-tip along each lineage and nodes of interest, and compared against each other and null using LRT and FDR. Given each model yields different ω estimates for each sequence, sequences were defined as being under robust lineage-specific positive selection if they exceeded significance thresholds (following FDR) under at least two maximum likelihood models.

2.5.2. Size-Dependent Selection

To test whether size-dependent selective pressures acted on cancer-associated genes in cetacean lineages with contrasting body sizes, the cetacean species were divided into three groups: those with $\log_{10}$mass>4 (Mysticetes and the sperm whale); <2.5 (dolphins and porpoises); and those in-between (beaked whales, orca, beluga and narwhal, and both minke whales) (see Fig.4a). Given the orca and minke whales have increased and decreased in size by an order of magnitude respectively, they were also partitioned into separate sub-groups for a more focused contrast between their lineage and parent clade.

A two-ratio model (M2) allowing for different estimates of ω within and between foreground and background branches was used to evaluate selective pressure in the size-grouped cetaceans by comparison with a one-ratio model (M0) used as null. Four experiments were conducted; using i) large whales and ii) small whales as foreground branches, and the remaining mid-sized whales as background; and using iii) the orca and large whales, and iv) minke and small whales as foreground, and the remaining as background. Genes were defined as being size-accelerated if their foreground ω exceeded that of background branches ($p<0.05$). Likelihood ratio tests (LRTs) were undertaken to

compare nested likelihood models, and a FDR process used to determine adjusted P values.

2.5.3. Genotype-Phenotype Association

Phylogenetically-weighted generalised least squares (PGLS) regressions were undertaken to assess the relationship between phenotypic traits and i) number of duplicate genes detected; and ii) several candidate genes flagged as being under strong positive selection. As in Chapter 3, given all species analysed are non-independent (i.e. evolutionarily associated), phylogenetic signal (λ) must be estimated (here, under a Brownian model) and tested via maximum likelihood ($\lambda = 0$ indicates no phylogenetic signal, with increasing influence as $\lambda \rightarrow 1$). PGLS was applied using the caper package in R (Orme, 2018) on $\log_{10}$-transformed ω estimated via branch models (root-to-tip ω generated by averaging independent ω estimates for each branch and node between each species and ancestral cetacean node) and $\log_{10}$-transformed values for species body mass. P-value of the PGLS regression slope was tested via FDR.

All annotated R, Python, Perl and bash scripts used in the methodology, alongside associated input and output data, are available in the Supplementary Information.

Results & Discussion

Since the onset of this research, a number of papers have been published independently by several groups including Tollis et al., (2019), Tejada-Martinez, de Magalhães & Opazo (2021), and Yuan *et al*. (2021) based on similar designs to my own. My findings in this chapter therefore largely independently corroborate previously published results. Results akin to those publicly available will be quickly summarised, with the bulk of the following discussion focusing instead on contrasting findings and, more broadly, the implications of what has been collectively uncovered analysing cetacean genomes.

3.1. Ortholog Identification and Genome Quality

A reciprocal smallest distance algorithm using 1,737 cancer-associated query human genes was able to successfully identify only 605 putative orthologs in all 27 cetacean species, which increased to 1,512 total if present in at least ten species. Further, when applied across the cetacean lineage, returned homologs (notably duplicated genes; see 3.2) are inconsistently placed, with the methodology applied unable to locate a considerable number of homologs which must (via parsimony) be present.

It has long been understood cetacean genomes comprise heavily of numerous repeat sequences (Árnason *et al*., 1988; 1984), and many are evolutionarily-recent large segmental duplications retaining a high number of duplicate genes under selection, as illustrated in Tollis *et al*. (2019) (though most are non-coding long terminal repeats or long-interspersed nuclear elements etc.). It is typical for up to a third or more of the entire length of a cetacean genome to comprise of these repetitive sequences, which can be extremely challenging to resolve without fragmentation or misassembly. Many of the early cetacean whole genome projects were assembled from short-read shotgun sequences, and even some of the best assemblies utilising the short-read methodology fare comparatively poorly against more recent long-read assemblies - for example, the sperm whale chromosome-level genome assembly has an estimated 1,513-9,978 gaps per scaffold (Fan *et al*., 2019), compared to the vaquita reference assembly (which has, by comparison, x140-475 longer contig N50s) with an average of 0-35 gaps per scaffold (Morin *et al*., 2020a; 2020b).

Given many available short-read assemblies utilised the original minke genome as reference (it with

67-73.7% reliable coverage), it's not surprising to suspect many of the assemblies used in this search have long sequences which are too heavily fragmented, unnecessarily collapsed, or are otherwise too misassembled for any genuine 'hits' to be identified considering inclusion thresholds. Indeed, manual searching for some duplicate TSGs reported in the literature (e.g. *FAT4* in sperm whales), but not flagged during my automated pipeline, revealed this to be the case, highlighting the difficulty in reliably detecting and differentiating between putative orthologs and paralogs using a large-scale automated process. A more germane approach, going forward, would be to focus analysis instead on the much smaller sample of high-quality reference genomes (or reassemble a broader range using lineage-specific repeat libraries to facilitate more accurate assembly) (see also Chapter 6 2.3).

3.2. Tumour-Suppressor Gene Redundancy

The recessive nature of TSG mutants means the number of mutants required for a deleterious phenotype to manifest is a rate-limiting step in the process of carcinogenesis - as demonstrated via modelling approaches in the previous chapter, a simple increase in allele number can have a markedly dramatic effect on incidence outcomes. Comparative analysis undertaken here identified 62 TSGs as having duplicated copies (based on inclusion criterion and confirmed by manual inspection) in one or more cetacean species, 17 of which previously unreported (see Fig3a for summary; SI.5.6 for full table).

Unlike in elephants, and more broadly Paenungulata, where large expansion in only a small number of TSGs is considered to result in increased cancer suppression, the number and functional diversity in cetacean-specific TSG duplications is large, often occurring along specific lineages. Many TSGs identified here have similarly been identified elsewhere - examples including *SALL4* in the sei whale, *UVRAG* in sperm and rorquals, *TGM3* in orca, several *H2AX* duplications across the clade, and *PDCD5* basally at cetaceans - and will not be elaborated on here (see Lagunas-Rangel, 2021; Tollis *et al*., 2019; and Tejada-Martinez, de Magalhães & Opazo, 2021, for full review). Previously unreported TSG duplications of note however include: an additional *HINT1* duplication - involved in the Wnt/β-catenin pathway (Genovese *et al*., 2012) - basally at mysticetes, with an additional duplication at least within minke whales; *DUSP22* - a suppressor of many target proteins involved in signalling pathways crucial to tumorigenesis and inflammation (i.e. EGFR and AR) (Lin *et al*., 2019) - duplicated in rorquals, narwhal, beluga and orca (all having undergone independent size increase); *FAT4* – which regulates the Hippo signalling pathway controlling cell proliferation and apoptosis (Ka-

toh, 2012) - duplicated in sperm and several times in baleen whales; three copies of *MAL* in rorquals - which induces apoptosis via the Fas signalling pathway (Mimori *et al*., 2003); and the *LAMTOR1* subunit gene - a component of the ragulator complex required for the activation of mTORC1 during autophagy (Martina *et al*., 2012) - separately in rorquals and the blue whale (see Fig.3c). Tollis *et al*. (2019) suspected a potential *TP53* duplication in the humpback whale, however manual examination here failed to reveal any additional *TP53* copies in any genome.

A qualitative overview of duplication events and their lineage relevance suggests a concentration of duplicate number in those lineages associated with rapid body mass expansion. PGLS analysis of collated TSG copy number per species against body size indicates a positive size-incidence relationship (see Fig.3b), supporting this interpretation, however the undue influence of sampling error due to low-quality assemblies or overly-sensitive threshold criterion must be considered; though a FDR robustness process was undertaken by serially removing the high-quality assemblies of the blue, bowhead and humpback whales from the analysis respectively, and the relationship in each instance maintained, the p-value (P_r) only just remained within acceptable boundaries. Using the data they collected, Tejada-Martinez, de Magalhães & Opazo (2021) undertook a TSG turnover analysis, comparing duplication and loss event rates between cetacean and terrestrial mammal clades, and found turnover rates of TSGs in cetaceans was more than twice as rapid (λ = 0.00074 $>$ 0.00031) than background in mammals, and is further accelerated in Mysticetes compared to Odontocetes (λ = 0.0019 $>$ 0.00022), providing support to the idea that higher rates of duplication are indeed present in large-bodied lineages and may contribute to cancer suppression. Evidence of positive selection (see 3.3) in duplicated genes further supports their adaptive benefit, as opposed to copy number being elevated through neutral birth-and-death processes rapidly generating large repeat sequences due to unrelated mechanisms. A major caveat however is the assumption here that gene duplicates retain ancestral function and enhance suppression via elevated expression or other dosage effects, or via simple rate-limiting redundancy. Many processes can result in divergence in gene function (e.g. sub- and neo-functionalisation), either directly through alteration to the protein sequence itself (detectable in sequence), or indirectly via alteration in transcription pathways.

Why are there so many different duplicate TSGs across lineages, and why not in numbers as high as *TP53* in elephants? Superficially it may seem like introducing as much single gene redundancy as possible may be disproportionately beneficial - why not duplicate all TSGs ten-fold - however, apart

from the lack of selective pressure to retain redundant genes if they aren't positively contributing to fitness in a statistically meaningful way, there are disadvantages to supernumerary TSGs. Evidence from *TP53* overexpression experiments in mouse lines suggest a trade-off between increased cancer resistance and numerous life-history traits including slower pre- and post-natal growth rates and reduced size (Maier *et al*., 2004), accelerated aging (Tyner *et al*., 2002), reduced fertility (Allemand *et al*., 1999; Maier *et al*., 2004), and developmental trade-offs including reduced proliferation, cellularity and atrophy across multiple organ and tissue systems (Dumble *et al*., 2007; Maier *et al*., 2004), defective uteric bud differentiation, and small kidneys (Godley *et al*., 1996).

Further, from a systems network perspective, as TSGs typically occupy roles regulating input and output at nodes in critical biomolecular pathways - malignant phenotypes arising if control of a node, or number of critical nodes, are lost - there are two potential ways of retaining network-wide stability: either prioritising control of only a small number of key nodes, minimising risk of loss by asserting maximum control (going 'tall'); or, increasing by other means the redundancy in the network by altering the number, or flexibility, of pathways, such that if one junction is 'lost', traffic can be rerouted along other paths to ensure system-wide function remains (going 'wide'); or some variation in-between. Given then the costs associated with duplicate expression, perhaps elephants, as mentioned previously, have arrived at a solution wherein duplicate TSG copies are rapidly partitioned to specific specialised roles with minimal expression profiles outside their target function, explaining why and how elephants may have gone 'tall' on specific TSG duplications, rather than possibly 'wide' as is seemingly the case in cetaceans (on the latter, perhaps metabolic adaptation to their aquatic lifestyle set the foundation for a comparatively radically altered metabo-biomolecular network allowing for this).

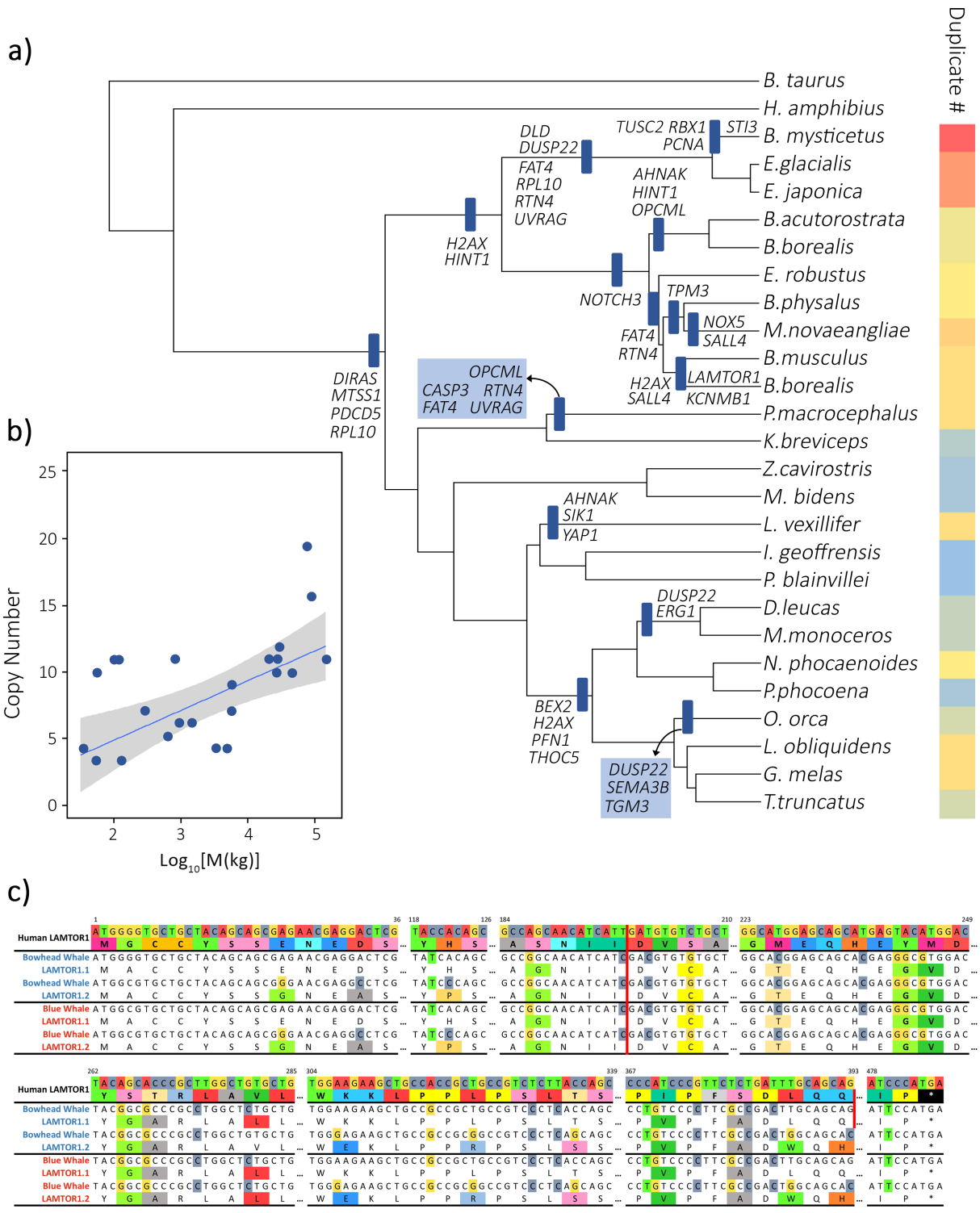
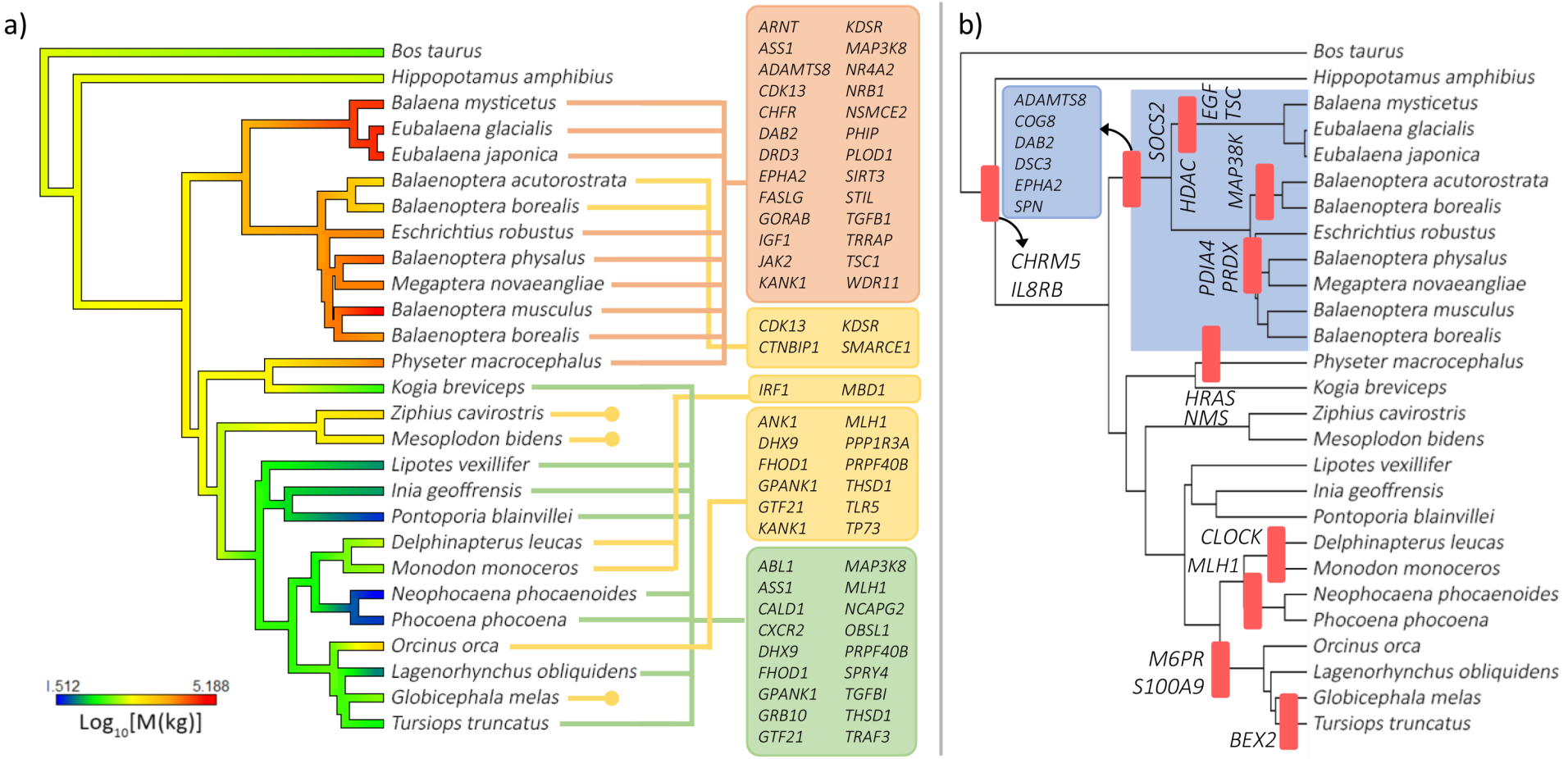


Fig.3: **(a)** Summary of notable TSG duplication events in Cetacea. Placement based on maximum parsimony following manual review of multiple alignment; heat map (right) indicating collated confirmed duplication events per species not including parsimony or controlling for genome quality. **(b)** PGLS between copy number expansion and $\log_{10}$-transformed body mass across cetaceans (adj.R^2 = 0.3255, P_r<0.05). **(c)** Segmental duplicaton (inc. introns) of *LAMTOR1* in bowhead (*Balaena mysticetus*) and blue (*Balaenoptera musculus*) whales; non-synonymous amino acid and nucleotide substitutions highlighted; splice regions in red. Amino acid substitutions at sites 81, 89, 94 occur within the functional domain of the Ragulator complex.


a)
Bos taurus
Hippopotamus amphibius
Balaena mysticetus
Eubalaena glacialis
Eubalaena japonica
Balaenoptera acutorostrata
Balaenoptera borealis
Eschrichtius robustus
Balaenoptera physalus
Megaptera novaeangliae
Balaenoptera musculus
Balaenoptera borealis
Physeter macrocephalus
Kogia breviceps
Ziphius cavirostris
Mesoplodon bidens
Lipotes vexillifer
Inia geoffrensis
Pontoporia blainvillei
Delphinapterus leucas
Monodon monoceros
Neophocaena phocaenoides
Phocoena phocoena
Orcinus orca
Lagenorhynchus obliquidens
Globicephala melas
Tursiops truncatus
1.512
5.188
Log10[M(kg)]
ARNT KDSR
ASS1 MAP3K8
ADAMTS8 NR4A2
CDK13 NRB1
CHFR NSMCE2
DAB2 PHIP
DRD3 PLOD1
EPHA2 SIRT3
FASLG STIL
GORAB TGFB1
IGF1 TRRAP
JAK2 TSC1
KANK1 WDR11
CDK13 KDSR
CTNBIP1 SMARCE1
IRF1 MBD1
ANK1 MLH1
DHX9 PPP1R3A
FHOD1 PRPF40B
GPANK1 THSD1
GTF21 TLR5
KANK1 TP73
ABL1 MAP3K8
ASS1 MLH1
CALD1 NCAPG2
CXCR2 OBSL1
DHX9 PRPF40B
FHOD1 SPRY4
GPANK1 TGFBI
GRB10 THSD1
GTF21 TRAF3
b)
Bos taurus
Hippopotamus amphibius
Balaena mysticetus
Eubalaena glacialis
Eubalaena japonica
Balaenoptera acutorostrata
Balaenoptera borealis
Eschrichtius robustus
Balaenoptera physalus
Megaptera novaeangliae
Balaenoptera musculus
Balaenoptera borealis
Physeter macrocephalus
Kogia breviceps
Ziphius cavirostris
Mesoplodon bidens
Lipotes vexillifer
Inia geoffrensis
Pontoporia blainvillei
Delphinapterus leucas
Monodon monoceros
Neophocaena phocaenoides
Phocoena phocoena
Orcinus orca
Lagenorhynchus obliquidens
Globicephala melas
Tursiops truncatus
ADAMTS8
COG8
DAB2
DSC3
EPHA2
SPN
CHRM5
IL8RB
SOCS2
HDAC
EGF
TSC
MAP38K
PDIA4
PRDX
HRAS
NMS
CLOCK
MLH1
M6PR
S100A9
BEX2

3.3. Selection in Tumour-Suppressor Genes

3.3.1. Lineage-Specific Positive Selection

Both branch-site and free-ratio models were used to elucidate signatures of positive selection in specific lineages. At least 32 TSGs and proto-oncogenes were flagged as positive ($\omega > 1$) under both models (see Fig.4b), including *CHRM5* and *IL8RB* (Cetacea vs outgroup); several (as also by Tejada-Martinez, de Magalhães & Opazo, 2021) within Mysteceti (vs other) including *CLCA1*, *COG8*, *SPN* and *TSC1*; in the independently large sperm whale (*Physter macrocephalus*) such as *HRAS*, *NMS* and *TSC1*; and several which were species-specific including *MAP3K8* in the Antarctic minke whale (*Balaenoptera bonaerensis*), *PD1AB* in the sei whale (*Balaenoptera borealis*), *MTUS1* and *PRUNE* in the bowhead whale (*Balaena mysticetus*), and *BEX2* in the bottlenose dolphin (*Tursiops truncatus*).

Fig.4 (overleaf): (a) Phylogeny of cetaceans; species with $\log_{10}$M > 4 in orange; < 2.5 in green; mid-sized (yellow) species used as background, subsetted into groups based on independent intra-clade size difference. Genes with noted size-dependent acceleration (LRT P_r <0.05 following FDR) listed - size-associated accelerated divergence occurring between sequences in large species (orange) and small species (green). Trait gradient mapping calculated via Bayesian MCMC using PhyTools in R (Revell, 2012), utilising MCMC time-calibrated tree from McGowen *et al.*, 2020. **(b)** Map of several notable lineage-specific positive selection identified here and in Tejada-Martinez, de Magalhães & Opazo, (2021)

3.3.2. Size-Associated Accelerated Selection

Following a two-ratio model root-to-tip analysis applied to orthologs categorised by species into groups defined by body mass, it was possible to identify TSGs having undergone lineage-controlled mass-specific accelerated evolution (see Fig.4a for summary). In total, 25 TSGs scored significantly (P_r <0.05 following FRD) in species with $\log_{10}$mass > 4 compared to mid-sized (background) and small-sized Cetacea; and 18 TSGs for species with $\log_{10}$mass < 2.5 (see Fig.4a for summary)

Follow-up analysis on three mid-sized groups having undergone independent evolutionary increase in body mass from within a smaller-bodied lineage revealed 4 genes in minke whales (*CDK13*, *CTNBIP1*, *KDSR*, *SMARCE1*), 2 genes in arctic whales (*IRF1* and *MBD1*), and 12 in the orca (*ANK1*, *DHX9*, *FHOD1*, *GPANK1*, *GTF21*, *KANK1*, *MLH1*, *PPP1R3A*, *PRPF40B*, *THSD1*, *TLR5* and *TP73*) have undergone divergent accelerated selection compared to relevant outgroups. No direct interac-

tions nor shared specific pathways were found following pathway review.

Tollis *et al.* (2019) undertaking a similar analysis comparing rates of evolution in genes between large- and small-bodied cetaceans reported evidence of positive selection in *CD274*, *ETNK1*, *IL21R*, *MYOD1* and *PHF6* in the humpback whale (*Megaptera novaeangliae*) compared to minke whales, and several including *BTG1*, *FANCD2*, *FAS*, *GPC3*, *MYB*, *RAD21*, *STIL* and *TAL1* in orca (*Orcinus orca*) compared to dolphins, amongst others.

By definition, all the TSGs used as query sequences are known to be involved in neoplastic phenotypes. However, interpreting signals of size- or lineage-associated accelerated or positive selection in their cetacean-equivalent orthologs as being directly associated with cancer-associated selection is potentially specious. Though many flagged TSGs, such as *CDK13* (ω = 4.08; P_r = 0.002 in stem minke whale; rate x4.6 against background) are known for their roles as 'master switches' in cell cycle control - their downstream targets affecting multiple cancer-associated pathways - others are likely to have been selected for other primary functions. *KDSR* (ω = 2.16; x3.16 against background) plays a crucial role in the synthesis pathway of sphingolipids, associated with cetacean metabo-ecological changes associated with diet and blubber formation. One of the most significant size-associated signals was detected in a *DRD* ortholog in baleen whales (ω = 4.96, x3.73 against background) encoding a D2-like dopamine receptor which responds to the neurotransmitter dopamine. Though (over-) activation of D2-like receptors are known to inhibit proliferation and induce apoptosis in highly-specific tumour types (Kline *et al.*, 2018; Wu *et al.*, 2018), their primary roles involve neurological processes - adaptive changes therefore perhaps driven more by the complex cognition and social behaviours of the clade (Marino *et al.*, 2007).

As much as TSG duplication and signatures of positive selection are present - and in high number - as hypothesised, sequence level data alone is insufficient to draw robust conclusions. Do these genes under adaptive selection confer enhanced cancer suppression? For that, at minimum, transcriptomics are required to verify duplicated TSGs are functionally transcribed (e.g. not pseudogenised), and ideally experimental validation *in vitro* must be pursued.

3.4. Genetic Change Validation & Experimental Design

3.4.1. Transcriptomics

Transcriptomes covering kidney and liver cells in the bowhead (*Balaena mystiectus*) and gray whale (*Eschrichtius robustus*) present an opportunity to investigate whether duplicate TSGs manifest themselves functionally. Screening the bowhead transcriptome for novel duplicate TSGs detected here did not reveal significant results (and only some of the orthologs returned hits); either said duplicates are invalid, or are differentially expressed elsewhere.

Analysis undertaken by Toren *et al.* (2020) however demonstrated elevated expression of cancer-associated genes involved in DNA maintenance and repair, and apoptosis in gray whale cells; while Seim *et al.* (2014) reported higher expression of the TSGs *E4F1*, *EGR1*, *GLTSCR2*, *PRKCDBP* and *ST7* in bowhead whale cells. Confusingly, the same analysis also uncovered bowhead whale liver cells present reduced expression of TSGs *GRB14*, *PERP* and *PGLYRP2*, whilst elevated differential expression of proto-oncogenic *CITED2*, *PDGFRB*, *RASIP1* and *WSB1*. Given the pleiotropic natures of these genes, and the basic assumption homologs retain function across taxa, it may simply be the cancer-promoting (in humans) differential expression of these genes have little to no influence on carcinogenesis in whales, but it is curious nonetheless.

3.4.2. Experimental Validation *in vitro*

It will only be possible to determine whether sequence duplication or selection results in enhanced cancer suppressive ability if it is tested *in vitro*. Those genes, for example, involved intimately with regulation of the apoptotic pathway lend themselves well to DNA damage response experiments; either through siRNA knock-out of target genes in whale cells, or CRISPR/Cas9 knock-in experiments using well-established model (human or mouse) cell lines, exposing either to DNA damage inducing exogenous agents. Apoptotic response can then be verified using flow cytometry on stained cells, with γ-H2AX foci counts being used to observe initial amount of DNA damage, with RNAseq establishing expression levels. A major limitation in this regard however is the availability of established cetacean cell lines, particularly non-immortalised lineages, and in those of species-specific interest. Marine mammal tissue banks operating in the UK, Belgium, Italy and the USA can intermittently provide samples which may be of some use; further, Burkard *et al.* (2015; 2019) have successfully developed fibroblast lines of the Humpback whale (*Megaptera novaeangliae*) - non-immortalised HuWa1 and

HuWa2 - for use in ecotoxicological assays, and available commercially.

But even if the logistical challenges are met, the question remains: which candidate genes? Unlike in elephants, where *TP53* and *LIF* are clear targets for experiment, the piecemeal expansion of TSGs in cetaceans, most with high divergence in pleiotropic function, and often along lineage-specific lines makes identifying clear candidates more of a challenge. Gene families expanding basally in large-bodied clades - such as *H2AX* and *LAMTOR1* - may be one option; a better one would be genes which have undergone independent duplication or convergent selection in multiple large-bodied lineages, such as *RTN4* and *UVRAG*.

UVRAG, duplicated independently in sperm and some baleen whales, may be particularly promising. Acting as both a proto-oncogene and tumour-suppressor, *UVRAG* encodes 'UV radiation resistance-associated gene protein' which, amongst other functions, patrols genetic integrity in response to types of exogenous DNA damage by directly interacting with and activating DNA-dependent protein kinases (Quach *et al*., 2019). Overexpression of *UVRAG* activates autophagy and suppresses tumour growth, whilst *UVRAG* knock-out cell lines fail to induce autophagy following DNA damage, and proliferate uncontrollably (Liang *et al*., 2006; Liang *et al*., 2008; Takahashi *et al*., 2007). Unfortunately, to date, cell lines for neither sperm nor rorqual whales have been established; perhaps knock-in replacement experiments in model cultures may however prove enlightening as a proxy to assess the consequences of the whale-specific amino acid changes, assuming functional equivalency between the genes in both organisms.

3.5. Re-Analysing Cetacean Genomes

Until experiments begin to validate hypotheses, the best clues we can get remain by analysing genome sequences. The research undertaken in this Chapter, though useful (and in many ways validated by published work), was limited in scope owing to numerous factors including methodology and available data, and failed to more incisively test the relationship between copy number expansion, body size and lifespan. Beyond simple PGLS regression on branch tip data, what is the relationship between copy number expansion and body size and lifespan evolution?

An appropriate method proposed here would be to assemble a tree of Cetacea, incorporating established time-calibrated fossil phylogenies into McGowen *et al*.'s (2020) molecular phylogeny, and

then jointly estimate rates of body mass evolution and reconstruct ancestral states using a modified Brownian motion model (selecting from a heavy-tailed distribution allowing for large jumps in traits; previously shown to outperform other models; Elliot & Mooers, 2014). PGLS methodology can similarly apply estimated ancestral lifespans (which lack calibration via fossil data) to the established tree.

With an appropriate evolutionary model of lifespan and body size evolution across the cetacean lineage developed, a similar process of TSG homolog identification as undertaken here can be carried out. Given the fragmentation of genes across multiple scaffolds in many cetacean genomes, unless better assembled genomes become available, a more robust method of estimating duplicate copy number can instead be undertaken using an 'estimated copy number by coverage' (ECNC) algorithm in conjunction with reciprocal BLAST/BLAT methodology (Poell *et al.*, 2019). Using the copy number data, maximum likelihood models can then be applied to infer gene duplication and loss events across the lineage using Bayesian ancestral state reconstruction method. Finally, to estimate evolutionary change in cancer risk, an appropriate multistage model can estimate incidence at all nodes (the change between a node and ancestor calculated by a $\log_2$-transformed ratio of incidence rates).

Such a method would present a means to more acutely test the hypothesis TSG expansion facilitated the evolution of increases in size and/or lifespan in cetaceans, and therefore enhanced cancer suppression in cetaceans.

Conclusion

Under the framework of the multistage model of cancer, increasing the number of driver mutations required for carcinogenesis to progress is a clear route to enhanced cancer suppression. If the risk of neoplasia has driven adaptive trajectories in large, long-lived whales by increasing the number of rate-limiting steps, signatures of genetic change associated with genes involved in neoplastic pathways should be detectable in the genome - such as TSG redundancy in relevant molecular pathways. The use of comparative genomic methodology applied here, and elsewhere, has revealed a considerable number of TSG duplication events within cetacenas - particular within the large baleen whales - alongside signatures of size-associated accelerated and/or positive selection in a number of cancer-associated genes, thereby providing a library of potential gene targets for further, more focused investigation.

Bioinformatic analyses such as that presented here have only been possible within the last few years; it goes without saying these are still the earliest of days. As the quality and availability of whole genome sequences improves to cover increasing numbers of taxa across the animal kingdom, one can only imagine what we will uncover in the coming years. Already similar investigations into Peto's paradox in carnivoran (Huang *et al*., 2020) and rodent (Vedelek *et al*., 2020) lineages have begun to provide insight, and with large-scale sequencing projects such as the Vertebrate Genome Project and Darwin Tree of Life gearing up to sequence hundreds of mammal genomes to the highest of standards, we will soon be entering a golden age of comparative genomics. Analysis of the soon available breadth of bat genomes - a lineage with paradoxical disproportionately long lifespans yet extremely high metabolic rates, and almost no recorded incidence of cancer - looks particularly promising.

Continuing with an outlook to the future, the next chapter will briefly summarise the research conducted as part of this project, and consider what we have learnt and where we should go next.

5. References

– Abegglen, L.M., Caulin, A.F., Chan, A., Lee, K., Robinson, R., Campbell, M.S., Kiso, W.K., Schmitt, D.L., Waddell, P.J., Bhaskara, S., Jensen, S.T., Maley, C.C. & Schiffman, J.D. (2015) Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. *JAMA*. 314 (**17**), 1,850-1,860

– Allemand, I., Anglo, A., Jeantet, A.Y., Cerutti, I. & May, E. (1999) Testicular wild-type p53 expression in transgenic mice induces spermiogenesis alterations ranging from differentiation defects to apoptosis. *Oncogene*. 18 (**47**), 6,521-6,530

– Árnason, Ú., Allderdice, P.W., Lien, J. & Widegren, B. (1988) Highly repetitive DNA in the Baleen whale generaBalaenoptera andMegaptera. *Journal of Molecular Evolution*. 27, 217-221

– Árnason, Ú. & Widegren, B. (1984) Different rates of divergence in highly repetitive DNA of cetaceans. *Hereditas*. 101 (**2**), 171-177

– Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*. 57 (**1**), 289–300

– Burkard, M., Nash, S.B., Gambaro, G., Whitworth, D. & Schirmer, K. (2019) Lifetime extension of humpback whale skin fibroblasts and their response to lipopolysaccharide (LPS) and a mixture of polychlorinated biphenyls (Aroclor). *Cell Biology and Toxicology*. 35, 387-398

– Burkard, M., Whitworth, D., Schirmer, K. & Nash, S.B. (2015) Establishment of the first humpback whale fibroblast cell lines and their application in chemical risk assessment. *Aquatic Toxicology*. 167, 240-247

– Duan, D.D., Xie, H., Shi, H.F., Huang, W.W., Ding, F., Hong, J.K., Fan, J.S., Hu, S.Y., Wang, Q.w. & Zhou, M.Q. (2019) Hint1 Overexpression Inhibits the Cell Cycle and Induces Cell Apoptosis in Human Osteosarcoma Cells. *OncoTargets and Therapy*. 13

– Dumble, M., Moore, L., Chambers, S.M., Geiger, H., van Zant, G., Goodell, M.A. & Donehower, L.A. (2007) The impact of altered p53 dosage on hematopoietic stem cell dynamics during aging. *Blood*. 109(**4**). 1,736-1,742

– Elliot, M.G. & Mooers, A. (2014) Inferring ancestral states without assuming neutrality or gradualism using a stable model of continuous character evolution. *BMC Ecology and Evolution*. 226

– Fan, G., Zhang, Y., Liu, X., Wang, J., Sun, Z., Sun, S., Zhang, H., CHen, J., Lv, M., Han, K., Tan, X., Hu, J., Guan, R., Fu, Y., Liu, S., Chen, X., Xu, Q., QUin, Y., Liu, L., Wang, Ou., Tang, J., Lu, H., Shang, Z., Wang, B., Hu, G., Zhao, X., Zou, Y., Chen, A., Gong, M., Zhang, W., Lee, S.M.Y., Li, S., Liu, J., Li, Z., Lu, Y., Sabir, J.S.M., Sabir, M.J., Khan, M., Hajrah, N.H., Yin, Y., Kristiansen, K., Yang, H., Wang, J., Xu, X. & Liu, X. (2019) The first chromosome-level genome for a marine mammal as a resource to study ecology and evolution. *Molecular Ecology Resources*. 19 (**4**), 944-956

– Genovese, G., Ghosh, P., Li, H., Rettino, A., Sioletic, S., Cittadini, A. & Sgambato, A. (2012) The tumor suppressor HINT1 regulates MITF and β-catenin transcriptional activity in melanoma cells. *Cell Cycle*. 11 (**11**), 2,206-2,215

– Godley, L.A., Kopp, J.B., Eckhaus, M., Paglino, J.J., Owens, J. & Varmus, J.E. (1996) Wild-type *p53* transgenic mice exhibit altered differentiation of the ureteric bud and possess small kidneys. *Genes & Development*. 10 (**7**), 836-850

– Huang, D.W., Sherman, D.T. & Lempicki, R.A. (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. 4, 44-57

– Huang, X., Sun, D., Wu, T., Liu, X., Xu, S. & Yang, G. (2020) New insights into the genetic mechanisms of body size evolution in Carnivora (pre-print). [Online] Available at: https://www.researchsqua 112815/v1

– Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. & Kent, W.J. (2004) The USCSC Table Browser data retrieval tool. *Nucleic Acids Research*. 32, 494-496

– Katoh, M. (2012) Function and cancer genomics of FAT family genes. *International Journal of Oncology*. 41 (**6**), 1,913-1,918

– Keane, M., Semeiks, J., Webb, A.E., Li, Y.I., Quesada, V., Craig, T., Madsen, L.B., van Dam, S., Brawand, D., Marques, P.I., Michalak, P., Kang, L., Bhak, J., Yim, H.S., Grishin, N.V., Nielsen, N.H., Heide-Jorgensen, M.P., Oziolor, E.M., Matson, C.W., Church, G.M., Stuart, G.W., Patton, J.C., George, J.C., Suydam, R., Larsen, K., Lopez-Otin, C., O'Connell, M.J., Bickham, J.W., Thomsen, B. & de Magalhaes, J.P. (2015) Insights into the evolution of longevity from the bowhead whale genome. *Cell Reports*. 10, 112-120.

– Kline, C.L.B., Ralff, M.D., Lulla, A.R., Wagner, J.M., Abbosh, P.H., Dicker, D.T., Allen, J.E. & El-Deiry, W.S. (2018) Role of Dopamine Receptors in the Anticancer Activity of ONC201. *Neoplasia*. 20 (**1**), 80-91

– Lagunas-Rangel, F.A. (2021) Deciphering the whale's secrets to a long life. *Experimental Gerontology*. 151, 111425

– Liang, C., Feng, P., Ku, B., Dotan, I., Canaani, D., Oh, B.H. & Jung, J.U. (2006) Autophagic and tumour suppressor activity of a novel Beclin1-binding protein UVRAG. *Nature Cell Biology*. 8, 688-698

– Liang, C., Lee, J.S., Inn, K.S., Gack, M.U., Li, Q., Roberts, E.A., Vergne, I., Deretic, V., Feng, P., Akazawa, C. & Jung, J. U. (2008) Beclin1-binding UVRAG targets the class C Vps complex to coordinate autophagosome maturation and endocytic trafficking. *Nature Cell Biology*. 10, 776-787

– Lin, H.P., Ho, H.M., Chang, C.W., Yeh, S.D., Su, Y.W., Tan, T.H. & Lin, W.J. (2019) DUSP22 suppresses prostate cancer proliferation by targeting the EGFR-AR axis. *FASEB*. 33 (**12**), 14,653-14,667

– Löytynoja, A. (2014) Phylogeny-aware alignment with PRANK. *Methods in Molecular Biology*. 1079, 155-170

– Magadum, S., Banerjee, U., Murugan, P., Gangapur, D. & Ravikesavan, R. (2013) Gene duplication as a major force in evolution. *Journal of Genetics*. 92 (**1**), 155-161

– Maier, B., Gluba, W., Bernier, B., Turner, T., Mohammad, K., Guise, T., Sutherland, A., Thorner, M. & Scrable, H. (2004) Modulation of mammalian life span by the short isoform of *p53*. *Genes & Development*. 18 (**3**), 306-319

– Mardis, E.R. (2011) A decade's perspective on DNA sequencing technology. *Nature*. 470 (**7,333**), 198-203

– Marino, L., Connor, R.C., Fordyce, R.E., Herman, L.M., Hof, P.R., Lefebyre, L., Lusseau, D., McCowan, B., Nimchinsky, E.A., Pack, A.A., Rendell, L., Reidenberg, J.S., Reiss, D., Uhen, M.D., der Gucht, E.V. & Whitehead, H. (2007) Cetaceans Have Complex Brains for Complex Cognition. *PLoS Biology*. 5 (**5**), 139

– Martina, J.A., Chen, Y., Gucek, M. & Puertollano, R. (2012) MTORC1 functions as a transcriptional regulator of autophagy by preventing nuclear transport of TFEB. *Autophagy*. 8 (**6**), 903-914

– McGowen, M.R., Tsagkogeorga, G., Álvarez-Carretero, S., dos Reis, M., Struebig, M., Deaville, R., Jepson, P.D., Jarman, S., Polanowski, A., Morin, P.A. & Rossiter, S.J. (2020) Phylogenomic Resolution of the Cetacean Tree of Life Using Target Sequence Capture. *Systematic Biology*. 69 (**3**), 479-501

– Mimori, K., Shiraishi, T., Mashino, K., Sonoda, J., Yamashita, K., Yoshinaga, K., Masuda, T., Utsunomiya, T., Alonso, M.A., Inoue, H. & Mori, M. (2003) MAL gene expression in esophageal cancer suppresses motility, invasion and tumorigenicity and enhances apoptosis through the Fas pathway. *Oncogene*. 22, 3,453-3,471

– Miyata, T. & Yasunaga, T. (1980) Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of Molecular Evolution*. 16, 23-36

– Morin, P.A., Archer, F.I., Avila, C.D., Balacco, J.R., Bukhman, Y.V., Chow, W., Fedrigo, O., Formenti, G., Fronczek, J.A., Fungtammasan, A., Gulland, F.M.D., Haase, B., Heide-Jorgensen, M.P., Houck, M.L., Howe, K., Misuraca, A.C., Mountcastle, J., Musser, W., Paez, S., Pelan, S., Philippy, A., Rhie, A., Robinson, J., Rojas-Bracho, L., Rowles, T.K., Ryder, O.A., Smith, C.R., Stevenson, S., Taylor, B.L., Teilmann, J., Torrance, J., Wells, R.S., Westgate, A.J. & Jarvis, E.D. (2020a) Reference genome and demographic history of the most endangered marine mammal, the vaquita. *Molecular Ecology Resources*. 21 (**4**), 1,008-1,020

– Morin, P.A., Alexander, A., Blaxter, M., Caballero, S., Fedrigo, O., Fontaine, M.C., Foote, A.D., Kuraku, S., Maloney, B, McCarthy, M., McGowen, M.R., Mountcastle, J., Nery, M.F., Olsen, M.T., Rosel, P.E. & Jarvis, E.D. (2020b) Building genomic infrastructure: Sequencing platinum-standard reference-quality genomes of all cetacean species. *Marine Mammal Science*. 36 (**4**), 1,356-1,366

– Notredame, C., Higgins, D.G. & Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*. 302 (**1**), 205-217

– Orme, D. (2018) *The caper package: comparative analysis of phylogenetics and evolution in R* [Online] Available at: https://CRAN.R-project.org/package=caper

– Poell, J.B., Mendeville, M., Sie, D., Brink, A., Brakenhoff, R.H. & Ylstra, B. (2019) ACE: absolute copy number estimation from low-coverage whole-genome sequencing data. *Bioinformatics*. 35 (**16**), 2,847-2,849

– Quach, C., Song, Y., Guo, H., Li, S., Maazi, H., Fung, M., Sands, N., O'Connell, D., Restrepo-Vassalli, S., Chai, B., Nemecio, D., Punj, V., Akbari, O., Idos, G.E., Mumenthaler, S.M., Wu, N., Martin, S.E., Hagiya, A., Hicks, J., Cui, H. & Liang, C. (2019) A truncating mutation in the autophagy gene *UVRAG* drives inflammation and tumorigenesis in mice. *Nature Communications*. 10, 5681

– Repana, D., Nulsen, J., Dressler, L., Bortolomeazzi, M., Venkata, S.K., Tourna, A., Yakovlva, A., Palmieri, T. & Ciccarelli, F.D. (2019) The Network of Cancer Genes (NCG): a comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome Biology*. 20 (**1**)

– Revell, L.J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*. 3, 217-223.

– Seim, I., Ma, S., Zhou, X., Gerashchenki, M.V., Lee, S.G., Suydam, R., George, J.C., Bickham, J.W. & Gladyshev, V.N. (2014) The transcriptome of the bowhead whale *Balaena mysticetus* reveals adaptations of the longest-lived mammal. *Aging*. 6 (10), 879-899

– Sela, I., Ashkenazy, J., Katoh, K. & Pupko, T. (2015) GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Researc*h. 43, 7-14

– Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30 (**9**), 1,312-1,313

– Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N.P., Emes, R.D. & Lynch, V.J. (2016) *TP53* copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *eLife*. e11994

– Takahashi, Y., Coppola, D., Masushita, N., Cualing, H.D., Sun, M., Sato, Y., Liang, C., Jung, J.U., Cheng, J.Q., Mul, J.J., Pledger, W.J. & Wang, H.G. (2007) Bif-1 interacts with Beclin 1 through UVRAG and regulates autophagy and tumorigenesis. *Nature Cell Biology*. 9, 1,142-1,151

– Tejada-Martinez, D., de Magalhães, J.P. & Opazo, J.C. (2021) Positive selection and gene duplications in tumour suppressor genes reveal clues about how cetaceans resist cancer. *Proceedings of the Royal Society B: Biological Sciences*. 288 (**1,945**), 20202592

– Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Researc*h. 22 (**22**), 4,673-4,680

– Tollis, M., Robbins, J., Webb, A.E., Kuderna, L.F.K., Caulin, A.F., Garcia, J.D., Bèrubè, M., Pourmand, N., Marques-Bonet, T., O'Connell, M.J., Palsbøll, P.J. & Maley, C.C. (2019) Return to the Sea, Get Huge, Beat Cancer: An Analysis of Cetacean Genomes Including an Assembly for the Humpback Whale (*Megaptera novaeangliae*). *Molecular Biology and Evolution*. 36 (**8**), 1,746-1,763

– Tolstun, D.A., Knyazer, A., Tushynska, T.V., Dubiley, T.A., Bezrukov, V.V., Fraifeld, V.E. & Muradian, K.K. (2020) Metabolic remodelling of mice by hypoxic-hypercapnic environment: imitating the naked mole-rat. *Biogerontology*. 21 (**2**), 143-153

– Toren, D., Kulaga, A., Jethva, M., Rubin, E., Snezhkina, A.V., Kudryavsteva, A.V., Nowicki, D., Tacutu, R., Moskalev, A.A. & Fraifeld, V.E. (2020) Gray whale transcriptome reveals longevity adaptations associated with DNA repair and ubiquitination. *Aging Cell*. 19 (**7**), e13158

– Tyner, S.D., Venkatachalam, S., Choi, J., Jones, S., Ghebranious, N., Igelmann, H., Lu, X., Soron, G., Cooper, B., Brayton, C., Park, S.H., Thompson, T., Karsenty, G., Bradley, A. & Donehower, L.A. (2002) *p53* mutant mice that display early ageing-associated phenotypes. *Nature*. 415, 45-53

– Vazquez, J.M., Sulak, M., Chigurupati, S. & Lynch, V.J. (2018) A Zombie LIF Gene in Elephants Is Upregulated by TP53 to Induce Apoptosis in Response to DNA Damage. *Cell Reports*. 24 (**7**), 1,765-1,776

– Wall, D. P. & DeLuca (2007) Ortholog Detection Using the Reciprocal Smallest Distance Algorithm. In: Bergman N.H. (eds) *Comparative Genomics: Methods in Molecular Biology*. Humana Press.

– Wall, D.P., Fraser, H.B. & Hirsh, A.E. (2003) Detecting putative orthologs. *Bioinformatics*. 19, 1710-1711

– Wallig, M. (2020b) *doMC: Foreach Parallel Adaptor for 'parallel'* [online] Available at: https://CRAN.R-project.org/package=doMC

– Wu, X.Y., Zhang, C.X., Deng, L.C., Xiao, J., Yuan, X., Zhang, B., Hou, Z.B., Sheng, Z.H., Sun, L., Juang, Q.C. & Zhao, W. (2018) Overexpressed D2 Dopamine Receptor Inhibits Non-Small Cell Lung Cancer Progression through Inhibiting NF-κB Signaling Pathway. *Cellular Physiology and Biochemistry*. 48, 2,258-2,272

– Yang, Z. (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*. 24 (**8**), 1,586-1,591

– Yang, Z. & Nielsen, R. (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*. 19 (**6**), 908-917

– Yuan, Y., Zhang, Y., Zhang, P., Liu, C., Wang, J., Gao, H., Hoelzel, A.R., Seim, I., Lv, M., Lin, M., Dong, L., Gao, J., Yang, Z., Caruso, F., Lin, W., Fonseca, R.R., Wang, D., Wang, X., Rasmussen, M.H., Liu, M., Zheng, J., Zhao, L., Campos, P.F., Kang, H., Iversen, M., Song, Y., Guo, X., Guo, J., Qin, Y., Pan, S., Xu, Q., Meng, L., Yunga, A., Liu, S., Lee, S.M.Y., Liu, X., Xu, X., Yang, H., Fan, G. Wang, K. & LI, S. (2021) Comparative genomics provides insights into the aquatic adaptations of mammals. *PNAS*. 118 (**37**), e2106080118

– Zhao, M., Kim, P., Mitra, R., Zhao, J. & Zhao, Z. (2016) TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Research*. 44 (**D1**), 1,023-1,031

Chapter 6

Conclusion and Future Directions

Research Summary

Why do whales exist?

Since first being conceived of by Peto in 1975, the paradox between the hypothesis that larger, longer-lived organisms should be disproportionately more prone to cancer, and the observation that they're seemingly not, has remained largely unresolved and, more generally, as a problem worth solving, overlooked. It is only in the last ten years or so that interest and methodologies have developed enough for interest to be piqued and studies to be undertaken, and only in the last five where availability of data and promise of germane experimental methodology has driven the potential of comparative oncology. Many questions are starting to be resolved - still, these are early days.

Since the conception of this project in 2014, the problem of Peto's paradox has experienced an explosion in interest and published research; to which I hope the work presented as part of this project may contribute in some small part. To date, no published study has, at large scale, quantitatively verified the existence of Peto's paradox across taxa. In Chapter 3, this was undertaken utilising the previously untapped potential of data that has been quietly accumulating in zoological institutions for decades. Collecting and analysing said data reveals an invariant, if not negative, incidence relationship with size and lifespan across taxa is indeed present, and established support for a potential causal association with mass-specific metabolic rate. This was tested in Chapter 4 using several developed computational models. Under the assumptions presented, though analysis *in silico* was able to demonstrate some influence of metabolic rate on predicted cancer incidence outcomes, it was insufficient alone to explain Peto's paradox. Instead, the parameter that influenced outcome potential the most (and was similarly highly biologically plausible) was the number of rate-limiting steps, or driver mutations, required for carcinogenesis under the multistage framework. In Chapter 5, publicly available cetacean genomes were analysed in order to identify and examine signatures of genetic change that could act as rate-limiting steps, or otherwise suppress a neoplastic phenotype. At the onset of this project, comparative genomics of cetacean genomes had not been undertaken, until several parallel investigations (including this one) reported a considerable number of duplicated tumour-suppressor genes, and evidence for strong positive selection in several pathways associated with neoplastic phenotypes within lineages having increased in body size. Though no focused follow-up analysis was undertaken

here or (yet) elsewhere, nor the functional validity of any genetic change established with certainty, the collective body of evidence accumulated from analysing cetacean genomes highlights how Peto's paradox is a multivariate phenomenon and there are many evolutionary trajectories towards cancer suppression: *TP53* isn't the only solution.

Though much has been discovered, the research presented here, and more broadly in the field at large over the last several years has, if anything, only uncovered how little we still collectively understand about this phenomenon, and why we need to continue to push for answers.

Cancer isn't going anywhere. It's the inevitable outcome of a bargain our ancestral unicellular organisms made with one another aeons ago, when evolutionary forces drew them together to suppress their own individual fitness and collectively organise as a multicellular unit to pursue increased fitness, at least for the short-term. Cells will always attempt to cheat the system given the opportunity, and opportunities to do so only increase with time. For us multicellular organisms, it is a losing battle; though still, not a battle that ought to remain unfought. Nor one we wage without allies. The mere existence of whales and other large, long-lived mammals demonstrates there are solutions that can help us win our collective endeavour against cancer, and it is the hope that my research, and research like it in the field of comparative oncology, will continue to help provide the ammunition that may just turn the tide; one day at a time.

So where do we go next?

Considerations and Potential Directions

If one thing is for certain, there must be something fundamental about the biology of larger, longer-lived organisms that enhances their ability to suppress cancer. Solutions presented thus far are numerous and highly dependent on a range of factors. This project aimed to elucidate some of them, though the designs, results and interpretation relied on simplifications and assumptions that will be addressed here. Further research pursued in this field may benefit from the following considerations.

2.1. Lifespan and Cancer Selection

For all analyses, lifetime probability of cancer incidence has been used as a metric for inter-specific incidence comparison, alongside the metrics of total body mass and (humans aside) maximum estimates for lifespan, usually taken from captive individuals. Are these appropriate?

The multistage model of cancer progression first posited by Armitage & Doll (1954) assumes that cancer incidence is driven by time-associated accumulation of mutations (Chapter 1: Eq.1). Though a considerable body of evidence supports the multistage model, a key assumption implies a linear accumulation of mutations with age. However, evidence from human and murine models suggests up to 50% of all somatic mutations accumulate in cell lines before even sexual maturity is reached, and slows down significantly thereafter (Vijg *et al*., 2005; Horvath, 2013) - further evidence suggesting that this coincides with stem cell behaviour shifting from a growth phase to a long-term maintenance phase (see Fig.1a) (Bowie *et al*., 2006). Under the theoretical framework of the multistage model, this seems paradoxical - evidence supports incidence increasing exponentially with age, yet mutation accumulation occurs primarily during early development.

What the basic multistage framework doesn't take into account that may address this is temporally- and spatially-dynamic selection pressures. In other words, selection pressures acting on tissue-specific cell lineages undergoing carcinogenesis should differ significantly during different periods of life: for example, during pre-reproductive development, when cells are rapidly dividing, suppressive mechanisms ought to focus on preventing mutant lineages expanding into larger pools of dividing cells, thus reducing the likelihood of secondary or tertiary mutations furthering the progress towards neoplasia (i.e. keeping lineages narrow) – while in late life, when lineages must be wide to maintain tissue

complexes, suppressive mechanisms must put pressure on division rate, keeping it slow. Given then that selection for mutant cells is not linear with time but depends on several factors including many that are age-associated, physiological aging and simple chronological time are not interchangeable when considering models of cancer incidence. This is important when considering size- and lifespan-incidence relationships, especially across taxa.

Evolution acts to select for somatic maintenance strategies that maximise reproductive success (at the cost of long-term maintenance) in a fitness landscape shaped by external factors - investment in fitness is high early in life, and gradually relaxes as an organism ages (see Fig.1b). When an organism is removed from its natural environment however, where those external factors no longer drive selection, physiology nonetheless still (largely) behaves in ways as if shaped by those forces even when they are removed. Why is this?

The natural lifespan of animals is governed largely by external hazards such as predation, food availability, and disease. Maximum lifespan (the metric used in this study), observable only in captivity for most species, is instead limited by physiological aging. However, any cancer-associated selection pressure occurs only within the context of natural lifespans, and therefore using the relationship between cancer incidence and maximum lifespan may yield results that inadequately answer evolutionary questions. Here, paucity of available data meant we had to rely on lifetime incidence rates and values for maximum lifespan, however a more germane approach would be to collate a database of cross-taxa pre- and intra-reproductive timespans and apply time-sensitive data on onset of neoplasia to establish comparative rates of cancer across species. Veterinary data collected here did include information on age at which neoplasia arose (or became terminal), but it is unlikely that there is enough such data to robustly assess comparative neoplasia rates, especially given the lack of available (or at least collated) data on length of physiological timespans between species.

Future work establishing rates of pre- or intra-reproductive cancer across species would be useful in addressing cross-taxa incidence questions. Such a study would not only have to address and control for variance in pre- and intra-reproductive timespans that are species-specific, but also sex-specific. In humans it is well established that cancer susceptibility differs markedly between biological sexes (Li *et al.*, 2020; Dorak, 2012). More broadly across the animal kingdom, it is not uncommon for effective female reproductive timespans to be longer than males: though both sexes may arrive at

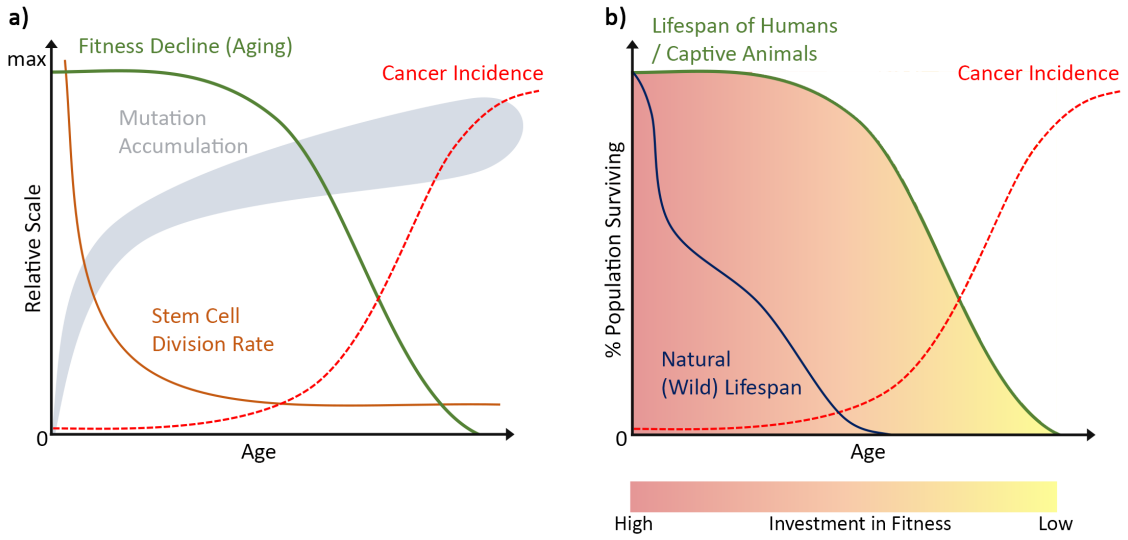


Fig. 1: Relationships between age-associated events: **(a)** approximately half of all somatic mutations (grey) occur before reproductive maturation (Horvath, 2013), at which point rate of stem cell division has declined (orange), switching from a growth phase to maintenance phase. Both rates of fitness decay (green) and cancer incidence (red) are delayed with respect to the mutation accumulation curve, but incidence does correlate with fitness decay. **(b)** Natural lifespan (blue) is governed by external hazards, and typically considerably shorter than captive lifespan (green) where those hazards are removed. Germline selection pressure, therefore investment in fitness, decreases with age; carcinogenesis (red) is suppressed during natural lifespan, only manifesting once body fitness selection pressure is relaxed during a period of maximum (post-wild) lifespan defined by rates of physiological aging (Adapted from Rozhok & DeGregori, 2017).

---

sexual maturity at approximately the same age, intra-sexual competition for mates between males favours older individuals (e.g. being larger, they succeed more; reproductively-capable males may wait years until they are successful), as is the case amongst elephants (Poole, 1989) and many baleen whales (Pack *et al.*, 2012), whereas females can successfully pass on genes immediately. As such, anti-cancer selection may favour males - there is more pressure to survive for longer - though perhaps this is balanced by sexual disparity in fecundity (a male can father several offspring simultaneously with multiple females; a female is limited to their immediate offspring). Might this influence cancer selection? This is potentially a good basis for a modelling approach - under a given framework of cancer progression, to what degree do perturbations in reproductive timespans between sexes and between species have on estimated incidence outcome?

In any case, age-associated variation in cancer selection pressures across different biologically-relevant timespans within an organism, and diversity in (perhaps sex-specific) timespans across taxa are relevant and need to be considered going forward.

2.2. Body Size and Tissue Heterogeneity

As with maximum lifespan, estimates for body size were used as a metric for size-incidence relationships out of simplicity. Animals however are not homogenous masses of identical cells, and tissue architecture can vary enormously between different species. For example, up to 40% or more of a baleen whale's body can comprise of relatively metabolically-inactive blubber (Nishiwaki 1950; Lockyer 1976), compared to 15-25% in humans (Jeukendrup & Gleeson, 2009). This type of divergence was corrected for to some degree when modelling cancer progression in Chapter 4 by focusing on specific cancer types (lung and colorectal adenocarcinomas), however even then assumptions included a linear scaling exponent to approximate intestinal length and therefore crypt (and cell) number in each species. This was particularly inappropriate applied speculatively cross-taxa - ungulates, for example, are likely to have much longer gut lengths than comparatively sized carnivores. Otherwise, in all other chapters, accounting for heterogeneity in organismal tissue architecture was not undertaken. A more appropriate approach would have to control for the varying proportions of tissue type between organisms (and also, as above, potentially between sexes), or only compare pathology limited to particular tissue type - little to no reliable data has yet been gathered in order to achieve this.

A potential way to acquire such data would be to focus on skin tissue and melanomas. Taking advantage of the availability of 3D models representing hundreds of species, it ought to be relatively simple to calculate surface areas and thereby reliably parameterise models of melanoma progression across various species.

2.3. Cetacean Genomics

Current evidence suggests a considerable number of TSG duplication events have occurred within the cetacean clade, at a disproportionately higher rate compared to terrestrial mammals (at least, averaged). Though we cannot rule out the hypothesis that such genomic changes have been pushed by other factors associated with the clade's unique and dramatic adaptations to aquatic life, the fact that the highest rates of TSG duplication events and genes exhibiting size-associated positive selection occurred within the baleen whales (the largest, and longest-living mammals to ever exist) suggests such genetic changes are adaptively associated with size expansion (and therefore cancer suppression) as opposed to other broader cetacean-associated adaptation (i.e. aquatic life). Further evidence in support of this could be undertaken by cross-comparing genomic and functional molecular changes

between cetacea and other mammal taxa - investigating influence of aquatic adaptation with the Sirenia (manatees and dugongs) and Pinnipeds (seals and sealions), for example.

As discussed in Chapter 5 (3.1), cetacean genomes present unique challenges, and only two - the vaquita (*Phocoena sinus*) and blue whale (*Balaenoptera musculus*) - are considered of high enough quality to be considered representative per the standards set and used by the Vertebrate Genome Project (VGP) and Darwin Tree of Life sequencing projects (Morin *et al.*, 2020). Though using the available draft genomes have proven useful in answering evolutionary questions, a more focused approach targeting specific cancer-associated genes or molecular pathways requires fully resolved reference genomes.

A major limitation when it comes to sequencing cetacean genomes is the availability of fresh tissue that can yield enough high-quality DNA for long-read sequencing (>40-300Kb) and intact 3D chromatin for long range Hi-C linking to appropriately build scaffolds. Morin *et al.* (2020) currently estimate such material is available for approximately 25% of cetacean species, either in living captive specimens or adequately deep-frozen (e.g. the FrozenZoo tissue culture collection). Otherwise, though ongoing projects continue to push in the right directions, it may well be many years until sufficiently robust assemblies will be available for much of the lineage.

2.3.1. Functional Homology

The work presented in this thesis assumes that homologs are functionally equivalent across taxa. We know however from our understanding of many pathways in model organism that this isn't always the case (see Gharib & Robinson-Rechavi, 2011). For example, the molecular interactions of both P- and E-selectins differ between human and mouse. The human ortholog of mouse P-selectin is unresponsive to TNF (tumor necrosis factor), a major inflammatory factor, which results in markedly diverging phenotypes downstream (Liu *et al.*, 2010). Transcriptomics likewise confirms difference in expression patterns in the thymus - expressed there in mouse, but not human, further suggesting roles are not inter-specifically conserved. A non-trivial proportion of the biomedical literature on thymus activity and disorder in humans is associated with P-selectin, but derived from mouse data.

In another example, intra-cellular localisation of *TDP1* expression is divergent between mouse and humans (in the nucleus and cytoplasm respectively); a single amino acid change (1478A¿G) in humans results in a neurological SCAN1 disorder; the equivalent substitution in mouse *TDP1* confers

no phenotype change (Hirano *et al*., 2007). There is no evidence of differential gene expression patterns nor sequence positive selection between the two species - to look at their genomic sequence, they ought to be completely equivalent; yet intracellular localisation results in different phenotypes.

Even within a single species, near-identical sequences can confer different functions. The highly sequence-conserved homologs *ASPP1* and *iASPP* encode proteins that exert entirely oppositional activities - *ASPP1* induces apoptosis, while *iASPP* inhibits it (Amir *et al*., 2015). This divergence in function results from only a small number of tightly-focused amino acid changes that alter the tertiary ligand interface structure enough to prevent co-compatible binding of their targets.

Neither functional sequence annotations nor experimental evidence supporting functional equivalence in proteins outside the standard suite of model organisms yet exists. Though methods nonetheless exist that lend high credence to the assumption that functional non-equivalence in highly similar sequences is rare (Sangar *et al*., 2007), incorrect annotation is still non-zero, and cannot be ruled out. Promising cancer-suppressing candidate genes in other species must be verified experimentally.

2.4. Immunocompetence

Immunologically-associated strategies of suppressing cancer were beyond the scope of this project - however, per the multistage framework, these strategies are of consideration.

As discussed in Chapter 2 (3.3), it is well established that increased immunocompetence results in greater neoplastic suppression. The recently identified 'immunoediting' process presents a mechanism by which cancer cell lineages select for mutations that promote immunological evasion. As reviewed by Dunn, Koebel & Schreiber (2006), a series of investigations demonstrated immunodeficient mice present higher rates of neoplasia than a control group, but transplantation experiments of tumours from the former into the latter results in higher rates of rejection (compared to tumours from similarly immunocompetent individuals). This can be interpreted via a disparity in selective pressure: the immune system in immunodeficient mice exerts less selective pressure on cancer cell lineages to develop effective immunoevasion, and vice-versa. Immune systems better able to effectively navigate this evolutionary conflict ought to result in higher rates of mutated cell elimination. But what are these rates? Though some information can be gleaned from targeted *in vitro* experimentation, given the decentralised, extra-cellular nature of the immune system, results gleaned *in vivo* are required. As such there is little known about the immunological abilities of species outside a few models, and are

likely to remain that way - performing *in vivo* immunocompromising experiments on elephants isn't likely to pass ethical review.

Computational modelling is more approachable, and modelling outcomes following introduction of a modifier of rate-limiting number of steps to carcinogeneis ("immunocompetence") to multistage models presented in Chapter 4 may prove insightful. Although difficult to parameterise beyond the broad 'success rate' of eliminating pre- or fully-transformed cells, parameter space estimation may prove useful in identifying expected comparative rates of immune responses - i.e. if, under reasonable biological assumptions, a blue whale needs an immune system with an infeasibly low policing error rate to bring incidence rates to expected levels, we can reject this hypothesis that improved immunocompetence has a plausible cancer suppressive effect. Like metabolism, immunocompetence is not likely to be as powerful a modifier as the number of rate-limiting steps (an exponent), but like metabolic scaling may provide some suppressive benefit nonetheless.

2.5. No Two Cancers are Equal

The most glaring elephant (whale?) in the room is the assumption made during analyses throughout this project is that all cancers are equivalent.

Though all cancers arise at the broadest level through similar trajectories of serial genetic mutation and clonal expansion driven by evolutionary forces, the myriad pathologies that develop are only grouped together as 'cancer' as an abstraction: it's a simple way to classify what are complex systems of cell behaviours that sample from some or all of a set of generalised hallmarks. In a not dissimilar way to the abstraction of 'species', there is no useful definition except only in the broadest of terms; focus in on the boundaries and they are revealed to be diffuse and arbitrary.

Every cancer is unique. With the recent explosion in single cell genomics and other methodologies, it has become clear that the vast majority of genes involved in driving neoplasia are mutated in a tissue- and type-specific manner only (see Bianchi *et al*., 2020 for review). In other words, a set of mutations in genes within one particular cell type will result in tumorigenesis; the same mutations in another cell elsewhere, in the same individual, will not - all dependent on a host of cell-intrinsic and extrinsic factors. This is likely to compound as one moves up the hierarchy of biological organisation - cell, to tissue, to individual, to species, through larger taxonomic groups - as the degree and number of confounding variables increases.

A brief qualitative overview of the cross-taxa incidence data collated in Chapter 3, as well as established reporting in the literature, hints at variable rates of different cancer types in different organisms. As such, as with the metrics of lifespan and body size, a more suitable cross-taxa comparative approach should focus on specific neoplastic cell or tissue types. The paucity of available data is a major limit in this regard; though as demonstrated in Chapter 3, data does exist. Now that the field of comparative oncology is growing rapidly, there is increasing interest and funding to establish robust datasets; efforts to develop standardised protocols of tumour classification and reporting should be encouraged. The increasing use of the ZIMS database software at zoological institutions is a good starting point. Though currently most institutional ZIMS data is partitioned locally with end users, a push to develop automatic collation of information relevant to research into an open-access database would be of enormous benefit.

To The Future...

The 'One World, One Health' concept, originating in 2004, recognises the close connection between human health and that of other species and our shared environment (Dujon *et al*., 2021). Though mainly focusing on zoonotic disease, increasing focus has begun to leverage the similarities and differences between species to provide fundamental insight into the development of other diseases affecting humans. Evolutionary forces have driven a diversity of phenotypes that present solutions to many pathological challenges we continue to face, not limited to cancer.

It is becoming increasingly understood that animals present Alzheimer's-like diseases (ALDs) with considerable overlapping etiology with human neurodegenerative disorders, such as increasing accumulation of amyloid plaques, cerebral amyloid angiopathy, and dysfunctional mitochondria with age (Golaszewska *et al*., 2019). Like cancer, might there be solutions to help prevent neurodegeneration out there in the animal kingdom?

Many long-lived mammals such as elephants and several whale species have, like humans, adaptive long post-reproductive lifespans, relying on high cognitive ability to survive well into old age. Might selective pressure be acting on pathways in those species to suppress neurodegeneration into advanced age? It's certainly hard to imagine how a bowhead whale could survive 200+ years with dementia. The field of comparative gerontology attempts to answer such questions; though, unlike cancer which is comparatively easy to diagnose, investigating neurodegenerative disorders in other animals is challenging: how does one begin to test the memory of a blue whale?

As we look forward towards the rest of the 21st century, no doubt many continued challenges will be overcome, and progress in comparative methods will be made. Though we are only beginning to resolve Peto's paradox, the hints of evolutionary insight being exposed are informative, and one hopes the answers we continue to uncover will one day be translated into human patients with cancer. If a whale can prevent tumours, why not us too?

This is the dream of comparative oncology.

Fin.

References

– Amir, A.I., Rosmalen, M.V., Mayer, G., Lebendiker, M., Danieli, T. & Friedler, A. (2015) Highly homologous proteins exert opposite biological activities by using different interaction interfaces. *Scientific Reports*. 5, 11,629

– Armitage, P. & Doll, R. (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer*. 8 (**1**), 1-12

– Bianchi, J.J., Zhao, X., Mays, J.C. & Davoli, T. (2020) Not all cancers are created equal: Tissue specificity in cancer genes and pathways. Current Opinion in Cell Biology. 63, 135-143

– Bowie, M.B., McKnight, K.D., Kent, D.G., McCaffrey, L., Hoodless, P.A. & Eaves, C.J. (2006) Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect. *The Journal of Clinical Investigation*. 116 (**10**), 2,808-2,816

– Dorak, M.T. & Karpuzoglu, E. (2012) Gender Differences in Cancer Susceptibility: An Inadequately Addressed Issue. *Frontiers in Genetics*. 3, 268

– Dujon, A.M., Brown, J.S., Destoumieux-Garzón, D., Vittecoq, M., Hamede, R., Tasiemski, A., Boutry, J., Tissot, S., Alix-Panabieres, C., Pujol, P., Renaud, F., Simard, F., Roche, B., Ujvari, B & Thomas, F. (2021) On the need for integrating cancer into the One Health perspective. *Evolutionary Applications* (early review).

– Dunn, G.P., Koebel, C.M. & Schreiber, R.D. (2006) Interferons, immunity and cancer immunoediting. *Nature Reviews Immunology*. 6, 836-848

– Gharib, W.H. & Robinson-Rechavi, M. (2011) When orthologs diverge between human and mouse. *Briefings in Bioinformatics*. 12 (5), 436-441

– Golaszewska, A., Bik, W., Motyl, T. & Orzechowski, A. (2019) Bridging the Gap between Alzheimer's Disease and Alzheimer's-like Diseases in Animals. *International Journal of Molecular Sciences*. 20 (**7**), 1664

– Hirano, R., Interthal, H., Huang, C., Nakamura, T., Deguchi, K., Choi, K., Bhattacharjee, M.B., Arimura, K., Umehara, F., Izumo, S., Northrop, J.L., Salih, M.A.M., Inoue, K., Armstrong, D.L., Champoux, J.J., Takashima, H. & Boerkoel, C.F. (2007) Spinocerebellar ataxia with axonal neuropathy: consequence of a Tdp1 recessive neomorphic mutation?. *EMBO J*. 26, 4,732-5,743

– Horvath, S. (2013) DNA methylation age of human tissues and cell types. *Genome Biology*. 14, 3156

– Li, C.H., Prokopec, S.D., Sun, R.X., Yousif, F., Schmitz, N., PCAWG Tumour Subtypes and Clinical Translation, Boutros, P.C. & PCAWG Consortium (2020) Sex differences in oncogenic mutational processes. *Nature Communications*. 11, 4330

– Liu, Z., Miner, J.J., Yago, T., Yao, L., Lupu, F., Xia, L. & McEver, R.P. (2010) Differential regulation of human and murine P-selectin expression and function in vivo. *Journal of Experimental Medicine*. 207 (**13**), 2,975-2,987

– Pack, A.A., Herman, L.M., Spitz, S.S., Craig, A.S., Hakala, S., Deakos, M.H., Herman, E.Y.K., Milette, A.J., Carroll, E., Levitt, S. & Lowe, C. (2012) Size-assortative pairing and discrimination of potential mates by humpback whales in the Hawaiian breeding grounds. *Animal Behaviour*. 84 (**4**), 983-993

– Peto, R., Roe, F.J., Lee, P.N. & Clack, J. (1975) Cancer and ageing in mice and men. *British Journal of Cancer*. 32, 411-426

– Poole, J.H. (1989) Mate guarding, reproductive success and female choice in African elephants. Animal Behaviour. 37 (5), 842-849

– Rozhoh, A.I. & DeGregori, J. (2017) The evolution of lifespan and age-dependent cancer risk. *Trends in Cancer*. 2 (**10**), 552-560

– Sangar, V., Blankenberg, D.J., Altman, N. & Lesk, A.M. (2007) Quantitative sequence-function relationships in proteins based on gene ontology. *BMC Bioinformatics*. 8, 294

– Vijg, J., Busuttil, R.A., Bahar, R. & Dollé, M.E.T. (2005) Aging and Genome Maintenance. *Annals of the New York Academy of Sciences*. 1055, 35-47