bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Positive selection and fast turnover rate in tumor suppressor genes reveal how cetaceans resist cancer

Daniela Tejada-Martinez[1,2,3*], João Pedro de Magalhães[3*] and Juan C. Opazo[2,4*]

[1*]Programa de Doctorado en Ciencias mención Ecología y Evolución, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile.

[2*]Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile.

[3*]Integrative Genomics of Ageing Group, Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool L7 8TX, UK.

[4*]Millennium Nucleus of Ion Channels-Associated Diseases (MiNICAD).

Corresponding authors: dtejadamartinez@gmail.com*; jp@senescence.info*; jopazo@gmail.com*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Cetaceans are the longest-lived species of mammals and the largest in the history of the planet. They have developed mechanisms against diseases like cancer, however their underlying molecular and genetic basis remain unknown. The goal of this study was to investigate the role of natural selection in the evolution of tumor suppressor genes in cetaceans. We found signal of positive selection 29 tumor suppressor genes and duplications in 197 genes. The turnover rate of tumor suppressor genes was almost 6 times faster in cetaceans when compared to other mammals. Those genes with duplications and with positive selection are involved in important cancer regulation mechanisms (e.g. chromosome break, DNA repair and biosynthesis of fatty acids). They are also related with multiple ageing and neurological disorders in humans (e.g. Alzheimer, Nijmegen breakage syndrome, and schizophrenia). These results provide evolutionary evidence that natural selection in tumor suppressor genes could act on species with large body sizes and extended life span, providing insights into the genetic basis of disease resistance. We propose that the cetaceans are an important model in cancer, ageing and neuronal, motor and behavior disorders.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Why large organisms do not always develop more cancer than smaller ones? This is an unresolved question that was proposed by Richard Peto in 1975 and it is known as "Peto's paradox" [1]. Peto's paradox is based on the observation that the risk of developing cancer should increase with the number of cells. Higher number of cells is associated with more cell divisions, that is one of the main causes of errors that may result in DNA damage transforming a normal cell into a cancer cell [2,3]. According to the Peto's paradox, we should expect a higher probability of developing cancer in large organisms in comparison to small ones, however, this does not occur [4–6]. Recently, the availability of genomes of a wide variety of species have stimulated the search for genes that could explain Peto's paradox. In particular, understanding the molecular basis of the anticancer mechanisms in big and long-lived species will shed light into fundamental areas like medicine, behavior and ecology.

The ways in which organisms reach and maintain larger body sizes and/or increased longevity, have puzzled scientists for decades [7]. For example, the naked mole rat (*Heterocephalus glaber*) has an average body mass of 35g and a lifespan of 35 years [8]. Long term studies suggest that their mortality rate [9] and risk of cancer does not increase with age [10]. In this species special modifications in telomerase activity and sensitive contact inhibition mechanisms would be, in part, responsible for their extended lifespan [11]. Among vertebrates other anti cancer-ageing mechanisms are described that could explain the maintenance of large body sizes and/or increased longevity. For example, in the little brown bat (*Myotis lucifugus*) it is described that telomere dynamics and changes in genes associated with growth factors are related to repair mechanisms that prevent the DNA damage with age [12,13]. In the yellowmouth rockfish (*Sebastes reedi*) there are no signs of aging in replicative senescence [14] and in the African elephant (*Loxodonta africana*) it is described variation in gene copy number in TP53 and LIF genes [15]. Molecular variation in tumor suppressor genes could work like a compensatory mechanism, by providing additional protection against DNA damage that could result in less probability of developing cancer [2]. However, the evolutionary history of tumor suppressor genes remains unknown.

Cetaceans evolved from a terrestrial ancestor and re-entered into the oceans around 50 million years ago [16]. Their adaptation to an aquatic lifestyle required large amount of morphological, physiological and behavior changes. The sub-order Cetacea is composed of two main lineages; Odontoceti (or toothed whales) and Mysticeti (or baleen whales) [16].

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Baleen whales and toothed whales display a huge range of body mass and maximum lifespan, ranging from the 17 years in the Pygmy sperm whale (*Kogia breviceps*) to 211 years in the bowhead whale (*Balaena mysticetus*) and from 50 kg in the Maui's dolphin (*Cephalorhynchus hectori maui*) to 175 tons in the blue whale (*Balaenoptera musculus*) [8]. Because there are species that live more than one hundred years, during their lifetime, they are exposed for a much longer time to the appearance of harmful mutations, pathogens and diseases. Then, it is expected that cetaceans developed improvements in the immune systems and DNA repair that could act favoring multiple traits. [17]. In agreement with that, in the longest-living mammal, the bowhead whale (*Balaena mysticetus,* lifespan ≈ 211 years, weight ≈ 100 tons) and the Humpback whale, the signal of positive selection were identified in genes involved in DNA repair, cell cycle regulation, resistance to ageing and cancer [18,19]. Further, recent investigations have found the signal of positive selection in genes that could be related to their variation of body size [17]. Although for mammals the evidence shows that tumor suppressor genes play a key role in the evolution of body size and longevity, through decreasing the incidence of diseases such as cancer [7,20], it is still an open question which molecular variants allow cetaceans to have evolved a lower incidence of cancer [21–23].

The goal of this study was to investigate the evolution of tumor suppressor genes in cetaceans, evaluating two forms of molecular variation (d_N/d_S and gene copy number variation). We report a signal of positive selection in 29 tumor suppressor genes involved in multiple processes that control the cancer onset and progression. The turnover rate of tumor suppressor genes was faster in cetaceans in comparison to other mammals. We report almost two hundred genes with duplications in one or more species of cetaceans involved in ageing, motor and neuronal disorders, deficits in learning and memory and developmental disorders. This approach highlights how studying the role of natural selection in genes associated with human health could lead to advances in our understanding of the genetic basis of disease resistance.

Methods

DNA sequences and taxonomic sampling

To study the evolution of tumor suppressor genes (TSGs) in cetaceans we implemented a phylogenetic design including 15 mammalian species. Our taxonomic sampling included five Odontocetes (bottlenose dolphin, *Tursiops truncatus*; orca, *Orcinus orca*; beluga,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Delphinapterus leucas; yangtze river dolphin, *Lipotes vexillifer*; and the sperm whale, *Physeter catodon*), two Mysticetes (common minke whale, *Balaenoptera acutorostrata*; bowhead whale, *Balaena mysticetus*), five other members of the superorder Laurasiatheria (cow, *Bos taurus*; pig, *Sus scrofa*; dog, *Canis familiaris*; horse, *Equus caballus*; microbat *Myotis lucifugus*), two Euarchontoglires (human, *Homo sapiens*; mouse, *Mus musculus*) and one Atlantogenata (African elephant, *Loxodonta africana*). The coding sequences of each species were downloaded from Ensembl v.96 ((http://www.ensembl.org), NCBI database [24] and the Bowhead whale genome project (http://www.bowhead-whale.org/) (supplementary table 1). To remove low quality records, sequences were clustered using CD-HITest *v*.4.6 [25] with a sequence identity threshold of 90% and an alignment coverage control of 80%. After that, the longest open reading frame was kept using TransDecoder LongOrfs and TransDecoder-predicted in TransDecoder v3.0.1 (https://github.com/TransDecoder/TransDecoder/).

Homology inference

We inferred homologous relationships between the 1088 tumor suppressor genes described for humans, available in public databases (Tumor Suppressor Gene Database, https://bioinfo.uth.edu/TSGene/ and the Tumor Associate Gene, http://www.binfo.ncku.edu.tw/TAG/GeneDoc.php), and the other 14 species included in our study using the program OMA standalone *v*.2.3.1 [26] . We inferred two types of groupings 1) OMA Groups (OG), containing the sets of orthologous genes and 2) Hierarchical Orthologous Groups (HOGs), which are all genes that have descended from a common ancestral gene. The amino acid sequences were aligned using the L-INS-i algorithm from MAFFT v.7 [27]. Nucleotide alignments were generated using the amino acid alignments as a template using the function pxaa2cdn in phyx [28]. Finally, to reduce the chance of false positives given for low quality alignment regions, we made a cleaning step with the codon.clean.msa algorithm of the rphast package [29], with the associated human tumor suppressor gene as reference sequence.

Natural selection analysis

To evaluate the role of natural selection in the evolution of tumor suppressor genes, we used the codon-based model in a maximum likelihood framework using the program PAML v4.9

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[30], as is implemented in the ETE-toolkit with the *ete-evol* function [31]. We used the branch-site model to estimate the signature of positive selection in the last common ancestor of Cetacea, Mysticeti (baleen whales) and Odontoceti (toothed whales). We compared the null model, where the value of ω in the foreground branch was set to 1, with the model in which the omega value was estimated from the data using the likelihood ratio tests (LRT) [32] (supplementary table 2).

Copy number variation analysis

To estimate the gene turnover rate of the tumor suppressor genes we used the software CAFE, version 4.2.1 (Computational Analysis of Gene Family Evolution) [33]. Briefly, CAFE uses a stochastic birth and death model to estimate the expansions and contractions of genes using as a frame of reference the sister group relationships among species and divergence times. Using this approach we can infer the rate of evolution (λ) and the direction of the change in the size of the gene families in different lineages. We implemented two models. In the first, one λ was estimated for cetaceans as total group and other for the outgroup. In the second model, we estimated four λ values (i) the stem cetacea, (ii) the total group of Mysticeti, (iii) the total group of Odontoceti and (iv) the outgroup. All the comparisons were calculated using $p<0.01$ and the CAFE correction for the genome assembly and annotation random error. The divergence time between species was obtained from the TimeTree database (http://www.timetree.org/) [34] . However, Odontoceti shows a polytomy, then we re-calibrate the tree using the packages APE and Phytools [33,35] in R [36], and the time-data of the most actualized time calibrated phylogeny of Odontoceti [37].

Enrichment analysis

To gain insight into the specific functions associated to the TSGs with signal of positive selection (d_N/d_S or CNV) we performed enrichment analyses using the gene ontology classification through DAVID bioinformatics resources - DAVID 6.7 [38].

Results

Homology inference

Whales are the longest-lived mammals and the largest in the history of our planet. During their evolutionary history they developed mechanisms against ageing and diseases, however, most of them remain unknown. In this study we performed an evolutionary analysis to

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

understand the role of natural selection, by estimating d_N/d_S and gene copy number variation, in the evolution of tumor suppressor genes in cetaceans. From the 1088 TSGs described for humans, we obtained 364 OrthoGroups (OG) containing the 15 mammalian species included in this study (supplementary file3) and 1044 hierarchical orthologous groups (HOG) containing two or more species (supplementary file 4, supplementary table 6).

TSGs with positive selection in cetaceans are related with multiple human disorders

According to our analyses, carried out in a set of 364 orthologous genes, we found the signal of positive selection in 29 tumor suppressor genes: 4 in the stem Cetacea, 4 in the stem of toothed whales (Odontoceti) and 21 in the stem of baleen whales (Mysticeti) (Fig.1, supplementary table 3). The TSGs with signal of positive selection are involved in multiple types of cancer and other human disorders (e.g., Fanconi anemia, Alzheimer, Nijmegen breakage syndrome and Coffin-Siris syndrome) (Table 1). The genes with a signature of positive selection were significantly enriched in 33 biological processes (supplementary table 4). In the last common ancestor of cetaceans we found enriched categories related with cancer regulation like chromosome break (BRCA2) and chemotaxis (CXCR2), but also involved in brain development that are linked with nervous system disorders in humans (PALB2 and BRCA2). In toothed whales, we found enriched categories related with the regulation of the cell cycle arrest (MYBBP1A), an important anti-cancer ageing mechanism reported before in other long-lived species (Huang et al., 2019), and also related with nervous system development (SMARCA4 and ACHE). In the last common ancestor of baleen whales, the lineage that includes the largest and long-lived species of mammals, the most represented functional categories found were related with aging and ADN repair (PRKAA1 and PRKAA2), apoptosis (SERPINB5), chromosome break (BRCA1), biosynthesis of fatty acids and cholesterol and regulation of TOR signaling (PRKAA1 and PRKAA2).

In summary, we report the signature of natural selection in TSGs involved in multiple human disorders. Since TSGs with the signature of positive selection in baleen whales and toothed whales are categorized into different biological processes, we suggest that these lineages could have evolved similar anticancer/ageing phenotype independently, reducing the risk of cancer, and favoring longevity and body mass.

Cetaceans have an accelerated gene turnover rate in comparison to other mammals

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Our results revealed that the turnover rate of TSGs of cetaceans, as a total group, is more than five times faster ($\lambda c = 0.0037$) in comparison to the rest of the tree ($\lambda o = 0.0007$) (Fig. 2a). In the second model, in which we specified four rates of gene family evolution, we found that the turnover rate values estimated for the last common ancestor of baleen whales (λ_{my}= 0.0036) and the last common ancestor of cetaceans (λ_C= 0.0037) are similar, but lower in comparison to the estimate of the last common ancestor of toothed whales (λ_{od}= 0.0046) (Fig. 2b). In all cases the gene turnover rate is almost 6 times faster in comparison to the outgroup (Fig. 2b).

To gain insights into the mechanisms related to ageing and disease resistance that are related to the tumor suppressor gene families, we identified the TSGs with specific duplications in cetaceans. According to our analyses we identified 197 TSGs with specific duplications in one or more cetacean species (Fig 3a, supplementary table 7). These duplicate genes are related with ageing (21 TSGs, supplementary table 8), immune system, vision and neurological, cardiovascular, metabolic and developmental disorders in humans (Fig 3b). Also they are involved in multiple biological processes (supplementary table 9) which are directly associated with the regulation of cancer, like cell proliferation (DAB2, FES, YAP1, BIN1), cell migration (EFNA1, PTPRK), apoptosis (LITAF, EPB41L3. KIF1B) and metabolism pathways (NEO1, PAX6, PTPRD and PTPRK). In our analysis, we also report extensive variation in the expansion and contraction of TSGs families in cetaceans and in the African elephant (*Loxodonta africana*) and the little brown bat (*Myotis lucifugus*), mammalian species that are well known for being cancer resistant (Gorbunova et al., 2014) (Fig. 4).

In summary, we found an accelerated rate of evolution of TSGs in cetaceans in comparison to other mammals. The gene families with an accelerated rate of evolution are related with important anticancer process that has also been reported in other long-lived species, like DNA repair, cell cycle and replicative senescence. Finally, we report almost two hundred genes duplicated just in one or more cetaceans, that are involved in multiple human disorders that comes from immune system and ageing to neurological disorders.

Discussion

TSGs with the signature of positive selection are related to cancer and Falcony anemia

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To understand the evolution of tumor suppressor genes in cetaceans, we studied the role of positive selection in genes involved in detection and repair of genetic damage, because this could be an important process in preventing the appearance of mutations that promote the spread of cancer cells in the organism. In the ancestor of cetaceans we report the signature of positive selection in the CXCR2 gene, that is a chemotaxis receptor, that regulates the recruitment of leukocytes during the inflammation processes and has also been described as a crucial factor in the tumor cell dissemination [39]. The MYBBP1A gene had the signature of positive selection in the ancestor of toothed whales, and is involved in p53 activation playing a fundamental role in programmed cell death [40]. In baleen whales we also found genes that could protect against the tumor growth, like the SERPINB5 gene, that has an important role in angiogenesis and metastasis [41]. Considering the evolution of the enormous body size in cetaceans, molecular variants in CXCR2, MYBBP1A and SERPINB5 could arise as a protective mechanism against the development of tumors.

We also found the signature of positive selection in the BRCA2 gene (marginally significant p-value of 0.06) in the ancestor of cetaceans and BRCA1 and PALB2 in the ancestor of baleen whales. BRCA1-2 genes are strongly associated with breast, ovarian and prostate cancer, the most common malignancy among the human population [42]. These genes have a central role in the maintenance of genomic stability and are also involved with chromosome breakage and cell cycle control [43]. Previous studies have also shown the signature of positive selection in BRCA2 in long-lived species of bats [13], suggesting that BRCA2, in association with their paralog BRCA1, could improve the DNA repair process protecting against the accumulation of DNA damage in cetaceans.

In humans, the BRCA1-2 and PALB2 genes are also related with Fanconi anemia, a genetic disease that is involved in bone-marrow failure and congenital malformation, which made patients more susceptible to cancer at younger ages. Damages in those genes cause an increased cancer predisposition by DNA repair deficiency [44]. Patients with Fanconi anemia also have accelerated telomere shortening, product of the DNA damage and oxidative stress [45], which are determinants of ageing and cell senescence. In cetaceans, the combination of new molecular variants (e.g. BRCA1/2 and PALB2) could reduce the incidence of cancer and lead us to understand new routes that could improve the mechanisms of DNA damage, cancer and ageing.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TSGs with the signature of positive selection are related with ageing

Ageing is the biggest risk factor for developing cancer and is characterized by the progressive accumulation of cellular damage, but the mechanisms linking these two processes -ageing and cancer- remain unclear [46]. One of the mechanisms that define how well we age, is the regulation of energy flow in the cells [47]. The genes PRKAA1 and PRKAA2 had signature of positive selection in baleen whales, and are part of the AMP- activated protein kinase (AMPK) complex, one of the primary regulators of energy homeostasis in eukaryotic cells and is considered as "a metabolic master-switch" [48]. The AMPK modulators have an important role in cardiac disorders (e.g. Wolff-Parkinson-White) [49], metabolic diseases (e.g. diabetes type II and obesity) [50] and also preventing the proliferation of cancer cells [51]. This pro-longevity mechanism has been used as therapeutic targets for age-related disorders in mammals [52,53] and it could be an important ageing mechanism in the evolution of baleen whales, who can live more than two hundred years. In baleen whales, the mechanisms associated with cancer resistance may have responded to selective pressures related to the increase in body size. Then, new pathways that improved their repair mechanisms of DNA damage, product of the metabolism of fatty acids and cholesterol, could arise in response to variants of these genes.

Previous studies with the bowhead whale genome (the longest-lived mammal) found a link between genes involved in metabolism and the evolution of longevity [19]. A number of studies have supported associations between sindroms related with metabolism and cardiovascular diseases, diabetes and schizophrenia [54]. For example, Hansen et al., (2011) studied 410 Danish patients with type II diabetes, 4089 with schizophrenia and others 17,597 european patients as controls. They found that the patients with type II diabetes increases the risk of developing schizophrenia and the genetic risk factor is associated mutations in the gene TCF7L2 [55]. In our study, we found ten copies of this gene in the killer whale and seven in beluga, however, the physiological consequences of having an increased number of copies and its relation with diabetes and schizophrenia remains unknown. Comparative studies in genes related with ageing and cancer, will provide insights into the evolution of physiological, morphological and behavioural traits.

Another anticancer and ageing mechanism is replicative senescence. This process is in charge of stopping cell proliferation [56] and in species that evolved large body sizes is considered as an important tumour suppressor mechanism [11]. In our study we report four

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

copies of the IRF5 gene in the orca and three copies of the PTPRD gene in the orca and in the sperm whale, both of them promoters of replicative senescence. The PTPRD gene has been associated with different types of cancer, including laryngeal, head and neck squamous cell carcinomas [57]. The fact that cases of oral squamous cell carcinoma have been diagnosed in dolphins of the genus *Tursiops* [58], coincides with the fact that they only have one copy of the PTPRD gene. In contrast, we report five copies of the PTPRD gene in beluga and minke whale and two copies in the killer whale and the river dolphin. An increase in the copy number of TSGs (e.g PTPRD and IRF5) could work as an extra protection against the development of squamous cell carcinomas in cetaceans, since it has been suggested that increasing the number of copies would act as "guardians" to prevent somatic mutations spreading in the cell population [4,59]. This result is in agreement with previous studies, where it has been shown that longevity-associated gene families evolve faster (increase the number of gene family members) in long-lived species [60].

TSGs with the signature of positive selection are related with neurological disorders

Given that whales, porpoises and dolphins have developed complex neurological systems and have the greatest absolute brain size among animals [61], it is expectable that they possess molecular variants to reduce the development of nervous system disorders (e.g. Alzheimer's, Nijmegen breakage syndrome, Coffin-Siris syndrome and spatial learning and motor coordination diseases). For example, the EPB41L3 gene is involved in meningiomas development [62], that is the most common central nervous system tumors in humans [63]. In our study we found 13 copies in the minke whale, eight in the beluga, four in the river dolphin and two in the killer whale and the sperm whale. On the other hand, BRCA2 gene (with positive selection in the stem of cetaceans) has also an important role in maintaining neural homeostasis in the central nervous system [64]. Experiments in mice have shown that the deletion of the BRCA2 gene affect their embryonic and postnatal neural development [65], while in humans, mutations in this gene could cause microcephaly [66]. Other duplicated gene in cetaceans related with neurological disorders, like depression and schizophrenia, is the DLG1 gene [67], and we identified three to five copies in five cetaceans species. Thus, amino acid changes in the BRCA2 gene in addition extra copies of the EPB41L3 and DLG1 genes, could be seen as an adaptive process improving the DNA repair system and providing an

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

improved protection against cancer and neurological disorders, a hypothesis that needs to be tested.

Alzheimer's disease is another neurodegenerative disorder developed mainly in old ages that affects the cognitive functions and generating slow memory loss [68]. In cetaceans, we report multiple copies of genes that has been recognized as a genetic risk factors in human Alzheimer's. Some of those genes are EPHA1, that is a positive regulator of angiogenesis, and metastases [69]; the ACHE gene, that encodes for the enzyme acetylcholinesterase, which is a catabolic enzyme for the neurotransmitter acetylcholine that is involved in the nervous system development pathways [70] and the BIN1 gene, that is a neuroinflammation related gene [71]. Particularly, for the BIN1 gene we found eleven copies in the beluga, eight in the river dolphin and four in the killer whale. The extra copies could serve as a complement to prevent the emergence of diseases that can cause memory loss that could affect the social abilities of the group. Then, in toothed whales, the anticancer mechanism could have evolved in response to the cognitive demands associated with complex cognitive system, sociability and use of echolocation. Thus TSGs may have been positively selected by natural selection by improving the mechanisms that lead to the appearance of diseases such as Alzheimer's.

In the ancestor of toothed whales, we found TSGs with positive selection involved in neurodevelopmental disorders like Nijmegen breakage syndrome (NBS) and Coffin-Siris syndrome (CSS). NBS is a genetic disorder characterized by severe microcephaly leaded by growth retardation, short stature, and malfunctioning of the immune system, making the individuals with more predisposition to cancer [72]. The NBS is associated with mutations in the gene MDC1, that is involved in the signalling, detection and reparation of DNA damage [73]. The gene MDC1 could be an important anticancer mechanism in toothed whales. On the other hand, Coffin-Siris syndrome (CSS) is a rare neurodevelopmental disorder, characterized by cognitive and developmental disability [74]. CSS is related with mutations in the SMARCA4 gene in the 11% of patients [75]. SMARCA4 is part of the ATP-dependent chromatin remodelling complex BAF (or SWI\SNF complex) [76]. This complex is in charge of the regulation of gene expression and cell differentiation and maintenance of stem cell pluripotency [77]. Particularly, SMARCA4 has been also associated with small-cell carcinoma of the ovary hypercalcemic type (SCCOHT), a malignant tumour with poor response to chemotherapy [78]. Overall, these genes which are related to neurodevelopmental disorders in

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

humans, in species that developed complex cognitive systems, like the toothed whales, could underlie new discoveries and gene pathways that could have an impact in human medicine.

TGSs duplications in cetaceans are related with spatial learning and motor disorders in humans

As in humans, cetaceans possess a fascinating evolution of behavior, with a complex social structure, developinging language comprehension and self-recognition ability [79]. Dolphins and whales are characterized for having few fully developed young, which develop a faster self-recognition ability and are capable of following the mother shortly after birth in a behavior that is known as a "infant carrying" [80]. During lactancy and infant carrying, the morphological and physiological hydrodynamic development of the young impacts in the swimming performance [81]. Newborns with diseases, motor malformations or poorly self-recognition ability usually are killed or left to die as they do not have enough skills to survive in nature [82]. In our study we identify duplicated genes in cetaceans that are related with spatial learning and motor disorders in humans. For example, EPB41 gene is involved in involved in the actomyosin structure organization that is a machinery that works in the contractile apparatus in muscle cells and neuronal membrane receptor complexes [83]. Experiments with EPB41 knockout mice also revealed erythroid disorders and others neuronal defects related with spatial learning and motor coordination [84]. In our study we found four copies of the EPB41 gene in the minke whale and beluga and five copies in the river dolphin. An increase in the copy number of EPB41 could represent a way to develop new pathways to regulate the machinerie that controls the muscle contractions in an aquatic environment and also control and/or avoid neuronal disorders.

Another gene duplicated in cetaceans is the PAX6 gene, that is involved in the central nervous system and eye development across the tree of life [85,86]. Mutations in the PAX6 gene cause ocular pathologies, like aniridia in humans [87] or malformed retina and lens in frogs (*Xenopus tropicalis*) [88]. Also, it has been related with neurological phenotypes (e.g autism and mental retardation) [89,90]. In zebrafish (*Danio rerio*) has been identified two PAX6 genes, as a consequence of the teleost-specific genome duplication, suggesting a division of labor between the two copies during the development of brain and eye structures [91]. In our study we report four copies in beluga and four copies in the common minke whale. Studying human ageing under a comparative evolutionary perspective offers an avenue that could provide

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

information about the role of natural selection in the evolution of species that found another way to avoid ageing diseases like cancer.

In the same direction, ABI2 gene is a regulator of cell migration and when is mutated generates aberrant dendritic spine morphogenesis and deficits in learning and memory [92]. ABI2 has a paralog, ABI1, which positively regulates lung metastasis of aggressive breast cancer [93]. Regua et al., (2008) suggest that in the absence of ABI1, its paralog could play a compensatory role that may support primary tumor growth. In our study, we report seven copies of ABI2 in the killer whale, four in the beluga, six in the river dolphin, two in the sperm whale, and two in the minke whale. Charcot Marie Tooth disease (CMT) is a hereditary motor and sensory neuropathy. The CMT type 1C cause abnormalities in myelin and affect the health of the nerve fiber, this disorder is associated with mutations in the LITAF gene [94]. In CMT type 2A the sensory peripheral nerves are malfunction, developing alterations in the sensorial capacity, atrophy and muscle weakness. The CMT type 2A? has been related related with mutations in the KIF1B gene that is a precursor in the axon from the cell body to the synapse [95]. However, even when affect an important morphological and physiological traits, it seems that the CMT disease do not decrease the lifespan of the individuals [96]. In our study, we report four copies of the LITAF gene in the river dolphin and three copies in the bowhead whale found an increased number of copies of the KIF1B gene in four cetaceans species: six in the minke whale, four in beluga, four in the sperm whale and two in the river dolphin. The role of the expanded repertoire of LITAF and KIF1B in cetaceans remain unknown, however, they could play a role at neuronal and motor level, taking into account that the organs of senses (including the sonar) and the development of hydrodynamic body shape were ones of the most remarkable morphological modifications in the evolution of the cetaceans.

In big, social and long-lived organisms, natural selection could favor disease resistant mechanisms for the improvement of the immune system, DNA repair mechanisms and metabolic pathway, but could also select positively molecular variants associated to avoid neuronal and motor pathologies.

Conclusions

The main goal of this study was to shed light into the evolution of tumor suppressor genes in cetaceans. We reported signal of positive selection in 29 TSGs and an accelerated gene

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

turnover rate in comparison to other mammals, with duplications in 197 genes in one or more cetacean species. These genes were involved in important process in ageing (e.g. PTPRD, PRKAA1 and PRKAA2) and cancer regulation like chromosome break, chemotaxis, programmed cell death and metabolism of fatty acids and cholesterol. In cetaceans, natural selection could favor new molecular variants that could improve the mechanism of DNA repair, energy consumption and ageing. We also report TSGs with positive selection and/or gene duplications related with multiple human disorders like Fanconi anemia (BRCA1-2 and PALB2), Alzheimer (EPHA1, ACHE), diabetes and schizophrenia (TCF7L2, DLG1), Nijmegen breakage syndrome (MDC1), Coffin-Siris syndrome (SMARCA4). Other TGSs with duplications in cetaceans were related with spatial learning and motor disorders in humans (EPB41, ABI2, KIF1B, LITAF). Studying genes that are related to human malignancies from a comparative perspective offers a novel avenue that could provide clues about the role of natural selection in the evolution of species that found a way to beat diseases. This study provide evolutionary evidence that natural selection in tumor suppressor genes could act on species with large body sizes and extended life span, providing insights into the genetic basis associated with the evolution of disease resistance. We propose that the cetaceans are an important model to understand longevity, neuronal, motor and behavioral disorders in humans and other animals.

Author contributions

DTM and JCO conceived the project; DTM performed the bioinformatic analysis. DTM wrote the manuscript; JCO and JPM edited the manuscript and were the project advisors.

Acknowledgements

Members of the Integrative Genomics of Ageing Group at the University of Liverpool, for all the support and advice during DTM's internship. To Richard Gregory in the Centre for Genomic Research (CGR) and Ian C. Smith in the Advanced Research Computing at the University of Liverpool for their help with computing resources. This work was supported by Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) - Chile through the doctoral studentship number N°21170433 and the scholarship from MECESUP AUS 2003 to DTM, Millennium Nucleus of Ion Channels Associated Diseases (MiNICAD), Iniciativa Científica Milenio, Ministry of Economy, Development and Tourism, Chile to

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

JCO. Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) grant 1160627 to JCO. To Dr. Andres Parada, for the academic discussions and programming advice. To Alejandra Tejada-Martinez for the graphic designs. Finally to Jesus Tejada and Gloria Martinez for their unconditional support. AnAge is supported by funding from the Biotechnology and Biological Sciences Research Council (BB/R014949/1) to JPM.

References

1. Peto, R., Roe, F. J., Lee, P. N., Levy, L. & Clack, J. Cancer and ageing in mice and men. *Br. J. Cancer* **32**, 411–426 (1975).
2. Tollis, M., Boddy, A. M. & Maley, C. C. Peto's Paradox: how has evolution solved the problem of cancer prevention? *BMC Biol.* **15**, 60 (2017).
3. Tomasetti, C. *et al.* Role of stem-cell divisions in cancer risk. *Nature* **548**, E13–E14 (2017).
4. Caulin, A. F. & Maley, C. C. Peto's Paradox: evolution's prescription for cancer prevention. *Trends Ecol. Evol.* **26**, 175–182 (2011).
5. Green, J. *et al.* 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* vol. 12 785–794 (2011).
6. Nagy, J. D., Victor, E. M. & Cropper, J. H. Why don't all whales have cancer? A novel hypothesis resolving Peto's paradox. *Integrative and Comparative Biology* vol. 47 317–328 (2007).
7. Tollis, M., Schiffman, J. D. & Boddy, A. M. Evolution of cancer suppression as revealed by mammalian comparative genomics. *Curr. Opin. Genet. Dev.* **42**, 40–47 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

8. Tacutu, R. *et al.* Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).

9. Ruby, J. G., Smith, M. & Buffenstein, R. Naked Mole-Rat mortality rates defy gompertzian laws by not increasing with age. *Elife* **7**, (2018).

10. Liang, S., Mele, J., Wu, Y., Buffenstein, R. & Hornsby, P. J. Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (Heterocephalus glaber). *Aging Cell* vol. 9 626–635 (2010).

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

12. Foley, N. M. *et al.* Growing old, yet staying young: The role of telomeres in bats' exceptional longevity. *Sci Adv* **4**, eaao0926 (2018).

13. Zhang, G. *et al.* Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* **339**, 456–460 (2013).

14. Cailliet, G. M. *et al.* Age determination and validation studies of marine fishes: do deep-dwellers live longer? *Experimental Gerontology* vol. 36 739–764 (2001).

15. Sulak, M. *et al.* TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *eLife* vol. 5 (2016).

16. Mancia, A. On the revolution of cetacean evolution. *Mar. Genomics* **41**, 1–5 (2018).

17. Sun, Y. *et al.* Insights into body size variation in cetaceans from the evolution of body-size-related genes. *BMC Evol. Biol.* **19**, 157 (2019).

18. Tollis, M. *et al.* Return to the Sea, Get Huge, Beat Cancer: An Analysis of Cetacean Genomes Including an Assembly for the Humpback Whale (Megaptera novaeangliae). *Mol. Biol. Evol.* **36**, 1746–1763 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

19. Keane, M. *et al.* Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep.* **10**, 112–122 (2015).

20. Tian, X., Seluanov, A. & Gorbunova, V. Molecular Mechanisms Determining Lifespan in Short- and Long-Lived Species. *Trends Endocrinol. Metab.* **28**, 722–734 (2017).

21. Magalhães, J. P. de, de Magalhães, J. P. & Kean, M. Endless paces of degeneration—applying comparative genomics to study evolution's moulding of longevity. *EMBO reports* vol. 14 661–662 (2013).

22. Nunney, L. The real war on cancer: the evolutionary dynamics of cancer suppression. *Evol. Appl.* **6**, 11–19 (2013).

23. Gorbunova, V., Seluanov, A., Zhang, Z., Gladyshev, V. N. & Vijg, J. Comparative genetics of longevity and cancer: insights from long-lived rodents. *Nat. Rev. Genet.* **15**, 531–540 (2014).

24. NCBI Resource Coordinators & NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research* vol. 45 D12–D17 (2017).

25. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

26. Altenhoff, A. M. *et al.* OMA standalone: orthology inference among public and custom genomes and transcriptomes. doi:10.1101/397752.

27. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

28. Brown, J. W., Walker, J. F. & Smith, S. A. Phyx: phylogenetic tools for unix. *Bioinformatics* vol. 33 1886–1888 (2017).

29. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).

30. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* vol. 24 1586–1591 (2007).
31. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
32. Zhang, J. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. *Molecular Biology and Evolution* vol. 22 2472–2479 (2005).
33. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
34. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
35. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* vol. 3 217–223 (2012).
36. Team, R. Core. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing 2012." URL https://www. R-project. org. (2019).
37. Zhang, X., Kong, L., Gao, Y. & Ma, H. Odontoceti phylogeny and divergence data resolved: evidence from nuclear genes and complete mitochondrial genomes of Neophocaena phocaenoides. *Conservation Genetics Resources* vol. 10 47–49 (2018).
38. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* vol. 4 44–57 (2009).
39. Roussos, E. T., Condeelis, J. S. & Patsialou, A. Chemotaxis in cancer. *Nat. Rev. Cancer*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

11, 573–587 (2011).

40. Akaogi, K., Ono, W., Hayashi, Y., Kishimoto, H. & Yanagisawa, J. MYBBP1A suppresses breast cancer tumorigenesis by enhancing the p53 dependent anoikis. *BMC Cancer* **13**, 65 (2013).
41. Teoh, S. S. Y., Wang, H., Risbridger, G. P., Whisstock, J. C. & Bird, P. I. A Versatile Monoclonal Antibody Specific to Human SERPINB5. *Hybridoma* vol. 31 333–339 (2012).
42. King, M.-C. & -C. King, M. Breast and Ovarian Cancer Risks Due to Inherited Mutations in BRCA1 and BRCA2. *Science* vol. 302 643–646 (2003).
43. Narod, S. A. & Foulkes, W. D. BRCA1 and BRCA2: 1994 and beyond. *Nat. Rev. Cancer* **4**, 665–676 (2004).
44. D'Andrea, A. D. & Grompe, M. The Fanconi anaemia/BRCA pathway. *Nature Reviews Cancer* vol. 3 23–34 (2003).
45. Adelfalk, C. *et al.* Accelerated telomere shortening in Fanconi anemia fibroblasts - a longitudinal study. *FEBS Letters* vol. 506 22–26 (2001).
46. Chatsirisupachai, K., Palmer, D., Ferreira, S. & de Magalhães, J. P. A human tissue-specific transcriptomic analysis reveals a complex relationship between aging, cancer, and cellular senescence. *Aging Cell* **18**, e13041 (2019).
47. Burkewitz, K., Weir, H. J. M. & Mair, W. B. AMPK as a Pro-longevity Target. *Experientia Suppl.* **107**, 227–256 (2016).
48. Frøsig, C. *et al.* AMPK and insulin action--responses to ageing and high fat diet. *PLoS One* **8**, e62338 (2013).
49. Davies, J. K. *et al.* Characterization of the role of gamma2 R531G mutation in AMP-activated protein kinase in cardiac hypertrophy and Wolff-Parkinson-White

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

syndrome. *Am. J. Physiol. Heart Circ. Physiol.* **290**, H1942–51 (2006).

50. O'Neill, H. M., Holloway, G. P. & Steinberg, G. R. AMPK regulation of fatty acid metabolism and mitochondrial biogenesis: implications for obesity. *Mol. Cell. Endocrinol.* **366**, 135–151 (2013).
51. Motoshima, H., Goldstein, B. J., Igata, M. & Araki, E. AMPK and cell proliferation--AMPK as a therapeutic target for atherosclerosis and cancer. *J. Physiol.* **574**, 63–71 (2006).
52. Steinberg, G. R. & Kemp, B. E. AMPK in Health and Disease. *Physiol. Rev.* **89**, 1025–1078 (2009).
53. Supnet, C. & Bezprozvanny, I. The dysregulation of intracellular calcium in Alzheimer disease. *Cell Calcium* vol. 47 183–189 (2010).
54. Mukherjee, S., Schnur, D. B. & Reddy, R. Family history of type 2 diabetes in schizophrenic patients. *Lancet* **1**, 495 (1989).
55. Hansen, T. *et al.* At-risk variant in TCF7L2 for type II diabetes increases risk of schizophrenia. *Biol. Psychiatry* **70**, 59–63 (2011).
56. Campisi, J. Senescence, Cellular. *Encyclopedia of Cancer* 205–211 (2002) doi:10.1016/b0-12-227555-1/00221-5.
57. Szaumkessel, M. *et al.* Recurrent epigenetic silencing of the PTPRD tumor suppressor in laryngeal squamous cell carcinoma. *Tumour Biol.* **39**, 1010428317691427 (2017).
58. Bossart, G. D. *et al.* Hematological, Biochemical, and Immunological Findings in Atlantic Bottlenose Dolphins (Tursiops truncatus) with Orogenital Papillomas. *Aquatic Mammals* vol. 34 166–177 (2008).
59. Leroi, A. M., Koufopanou, V. & Burt, A. Cancer selection. *Nat. Rev. Cancer* **3**, 226–231 (2003).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

60. Doherty, A. & de Magalhães, J. P. Has gene duplication impacted the evolution of Eutherian longevity? *Aging Cell* **15**, 978–980 (2016).

61. McGowen, M. R., Montgomery, S. H., Clark, C. & Gatesy, J. Phylogeny and adaptive evolution of the brain-development gene microcephalin (MCPH1) in cetaceans. *BMC Evol. Biol.* **11**, 98 (2011).

62. Smith, M. J. Germline and somatic mutations in meningiomas. *Cancer Genet.* **208**, 107–114 (2015).

63. Ostrom, Q. T. *et al.* CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007-2011. *Neuro. Oncol.* **16 Suppl 4**, iv1–63 (2014).

64. Johnson-Tesch, B. A., Gawande, R. S., Zhang, L., MacMillan, M. L. & Nascene, D. R. Fanconi anemia: correlating central nervous system malformations and genetic complementation groups. *Pediatr. Radiol.* **47**, 868–876 (2017).

65. Frappart, P.-O., Lee, Y., Lamont, J. & McKinnon, P. J. BRCA2 is required for neurogenesis and suppression of medulloblastoma. *EMBO J.* **26**, 2732–2742 (2007).

66. Rump, P. *et al.* Whole-exome sequencing is a powerful approach for establishing the etiological diagnosis in patients with intellectual disability and microcephaly. *BMC Med. Genomics* **9**, 7 (2016).

67. Uezato, A. *et al.* Reduced cortical expression of a newly identified splicing variant of the DLG1 gene in patients with early-onset schizophrenia. *Transl. Psychiatry* **5**, e654 (2015).

68. Masters, M. C., Morris, J. C. & Roe, C. M. 'Noncognitive' symptoms of early Alzheimer disease: A longitudinal analysis. *Neurology* vol. 84 617–622 (2015).

69. Naj, A. C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

associated with late-onset Alzheimer's disease. *Nat. Genet.* **43**, 436–441 (2011).

70. Paxinos, G. & Ashwell, K. W. S. *Atlas of the Developing Rat Nervous System: Fourth Edition*. (Academic Press, 2018).
71. Saito, T. & Saido, T. C. Neuroinflammation in mouse models of Alzheimer's disease. *Clinical and Experimental Neuroimmunology* vol. 9 211–218 (2018).
72. Bogdanova, N. *et al.* Nijmegen Breakage Syndrome mutations and risk of breast cancer. *Int. J. Cancer* **122**, 802–806 (2008).
73. Xu, C. *et al.* Structure of a second BRCT domain identified in the nijmegen breakage syndrome protein Nbs1 and its function in an MDC1-dependent localization of Nbs1 to DNA damage sites. *J. Mol. Biol.* **381**, 361–372 (2008).
74. Vergano, S. S. & Deardorff, M. A. Clinical features, diagnostic criteria, and management of Coffin-Siris syndrome. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* vol. 166 252–256 (2014).
75. Tsurusaki, Y. *et al.* Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat. Genet.* **44**, 376–378 (2012).
76. Errichiello, E. *et al.* SMARCA4 inactivating mutations cause concomitant Coffin-Siris syndrome, microphthalmia and small-cell carcinoma of the ovary hypercalcaemic type. *J. Pathol.* **243**, 9–15 (2017).
77. Alfert, A., Moreno, N. & Kerl, K. The BAF complex in development and disease. *Epigenetics & Chromatin* vol. 12 (2019).
78. Clarke, B. A. *et al.* Loss of SMARCA4 (BRG1) protein expression as determined by immunohistochemistry in small-cell carcinoma of the ovary, hypercalcaemic type distinguishes these tumours from their mimics. *Histopathology* **69**, 727–738 (2016).
79. Marino, L. Convergence of complex cognitive abilities in cetaceans and primates. *Brain*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Behav. Evol. **59**, 21–32 (2002).

80. Noren, S. R. Infant carrying behaviour in dolphins: costly parental care in an aquatic environment. *Functional Ecology* vol. 22 284–288 (2008).
81. Noren, S. R., Biedenbach, G. & Edwards, E. F. Ontogeny of swim performance and mechanics in bottlenose dolphins (Tursiops truncatus). *J. Exp. Biol.* **209**, 4724–4731 (2006).
82. Foote, A. D. Mortality rate acceleration and post-reproductive lifespan in matrilineal whale species. *Biol. Lett.* **4**, 189–191 (2008).
83. Parra, M. *et al.* Differential domain evolution and complex RNA processing in a family of paralogous EPB41 (protein 4.1) genes facilitate expression of diverse tissue-specific isoforms. *Genomics* **84**, 637–646 (2004).
84. Walensky, L. D. *et al.* The 13-kD FK506 Binding Protein, FKBP13, Interacts with a Novel Homologue of the Erythrocyte Membrane Cytoskeletal Protein 4.1. *The Journal of Cell Biology* vol. 141 143–153 (1998).
85. Chow, R. L., Altmann, C. R., Lang, R. A. & Hemmati-Brivanlou, A. Pax6 induces ectopic eyes in a vertebrate. *Development* **126**, 4213–4222 (1999).
86. Ashery-Padan, R. & Farhy, C. The multiple roles of Pax6 in mammalian retinogenesis. *Neuroscience Research* vol. 68 e23 (2010).
87. Kozmik, Z. Pax genes in eye development and evolution. *Curr. Opin. Genet. Dev.* **15**, 430–438 (2005).
88. Nakayama, T. *et al.* Xenopus pax6 mutants affect eye development and other organ systems, and have phenotypic similarities to human aniridia patients. *Dev. Biol.* **408**, 328–344 (2015).
89. Zhang, X. *et al.* Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7**,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

90–100 (2010).

90. Umeda, T. *et al.* Evaluation of Pax6 mutant rat as a model for autism. *PLoS One* **5**, e15500 (2010).

91. Nornes, S. *et al.* Zebrafish contains two pax6 genes involved in eye development. *Mech. Dev.* **77**, 185–196 (1998).

92. Grove, M. *et al.* ABI2-deficient mice exhibit defective cell migration, aberrant dendritic spine morphogenesis, and deficits in learning and memory. *Mol. Cell. Biol.* **24**, 10905–10922 (2004).

93. Regua, A. *et al.* Abstract 108: Abi1 positively regulates lung metastasis of aggressive breast cancer in PyMT mouse model. *Tumor Biology* (2018) doi:10.1158/1538-7445.am2018-108.

94. Gerding, W. M., Koetting, J., Epplen, J. T. & Neusch, C. Hereditary motor and sensory neuropathy caused by a novel mutation in LITAF. *Neuromuscul. Disord.* **19**, 701–703 (2009).

95. Azzedine, H., Senderek, J., Rivolta, C. & Chrast, R. Molecular Genetics of Charcot-Marie-Tooth Disease: From Genes to Genomes. *Molecular Syndromology* (2012) doi:10.1159/000343487.

96. Bird, T. D., Kraft, G. H., Lipe, H. P., Kenney, K. L. & Sumi, S. M. Clinical and pathological phenotype of the original family with Charcot-Marie-Tooth type 1B: a 20-year study. *Ann. Neurol.* **41**, 463–469 (1997).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure legends

Figure 1. Tumor suppressor genes with the signature of positive selection in different branches of the cetacean tree of life.

Figure 2. Turnover rate of tumor suppressor genes (TSGs) in cetaceans. a) The first model represent the rate of evolution (λ) of TSGs for the total group of cetacea (branches in orange) and the rate of evolution of TGSs for the outgroup (branches in grey). The λ values show that the gene turnover rate of TSGs in cetaceans is almost 6 times more accelerated in comparison to other placental mammals. b) In the second model, the λ values represents the rate of evolution of the total group of baleen whales (branches in green), the total group of toothed whales (branches in light blue) and ancestor of cetaceans (branch in orange).

Figure 3. Tumor suppressor genes (TSGs) with duplications in Cetaceans. a) The heat map represents the copy number variation of TSGs duplicated in one or more cetacean species. The color code correspond to the number of copies of each gene per species. The symbols on the branches of the tree represents the name of the species: Tt -*Tursiops truncatus*, Oc-Orcinus Orca, Dl-*Delphinapterus leucas*, Lv-*Lipotes vexillifer*, Pc-*Physeter catodon (*synonym name of *Physeter macrocephalus),* Ba-*Balaenoptera acutorostrata,* Bm-*Balaena mysticetus,* SS-*Sus scrofa*, Ec-*Equus caballus*, Ml-*Myotis lucifugus,* Mm-*Mus musculus*, Hs-*Homo sapiens,* La-*Loxodonta africana.* b) Functional classification of the TSGs duplicated in cetaceans in relation with the Gene Associated Disease (GAD) class given by the DAVID enrichment analysis.

Figure 4. Expansion and Contraction of Tumor suppressor gene families in placental mammals.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tables

Table 1. Human diseases related with the genes with the signature of positive selection in the ancestor of cetaceans, baleen whales and toothed whales and their relation with ageing. The information was obtained from GeneCards (https://www.genecards.org/), GeneAge database (https://genomics.senescence.info/genes/) and the KEGG enrichment analysis (supplementary table 5).

Group	Genes with positive selection signal	GeneCards - Diseases associated	GenAge genes and longevity pathways
Stem Cetacea	BRCA2	Breast, ovarian, prostate, gastric and pancreatic cancer - Fanconi anemia	No
	CDH1	Blepharo Cheilo Dontic syndrome 1 - Gastric cancer – Bladder cancer	No
	EPHA1	Placenta praevia – Ovarian serous adenocarcinoma	No
	CXCR2	Autosomal recessive severe congenital neutropenia – Neutrophil migration - Immune system	No
Stem toothed whales	ACHE	Yt Blood group antigen – Colonic pseudo-obstruction	No
	MYBBP1A	Cerebrovascular benign neoplasm – Circadian rhythm	No
	MDC1	Nijmegen breakage syndrome - Brachydactyly, Type C	No
	SMARCA4	Coffin-Siris syndrome 4 - Rhabdoid tumor predisposition syndrome 2	Yes
Stem baleen whales	ADAMTS8	Diseases of glycosylation – Lung neoplasm	No
	AKAP12	Myasthenia Gravis - Juvenile myelomonocytic leukemia	No

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

	BRCA1	Breast, ovarian and gastric cancer - Fanconi Anemia	Yes
	CHFR	Cell cycle progression and tumorigenesis	No
	DAB2	Malignant epithelial mesothelioma – Hypercholesterolemia autosomal recessive	No
	DSC3	Hypotrichosis - Recurrent skin vesicles - Subcorneal pustular dermatosis	No
	DSP	Skin Fragility-Woolly Hair syndrome – Cardiomyopathy – Immune system	No
	EPHA2	Cataract 6, multiple types - Early-Onset posterior subcapsular cataract	No
	IGF2R	Hepatocellular carcinoma - Inclusion-Cell disease	No
	INTS6	Expression suppressing in tumor cell growth	No
	KL	Tumoral calcinosis hypophosphatemic familial 1 and 3	Yes
	PALB2	Pancreatic cancer – Fanconi Anemia	No
	PLCB3	Multiple endocrine neoplasia, Type I	No
	PLCE1	Nephrotic syndrome, type 3 - Sporadic idiopathic steroid-resistant nephrotic syndrome with diffuse mesangial sclerosis	No
	PRKAA1	Wolff-Parkinson-White Syndrome and Body Mass Index	Yes
	PRKAA2	Peutz-Jeghers Syndrome and Wolff-Parkinson-White Syndrome	Yes
	RECK	Ovarian Cancer - Middle ear squamous cell carcinoma	No
	SAFB2	Regulation of immune genes	No

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.15.908244; this version posted January 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

	SERPINB5	Breast cancer	No
	TMPRSS11A	Iron-Refractory iron deficiency anemia – Ichthyosis follicular	No
	UNC5A	Axon guidance – Downregulated in several cancers	No

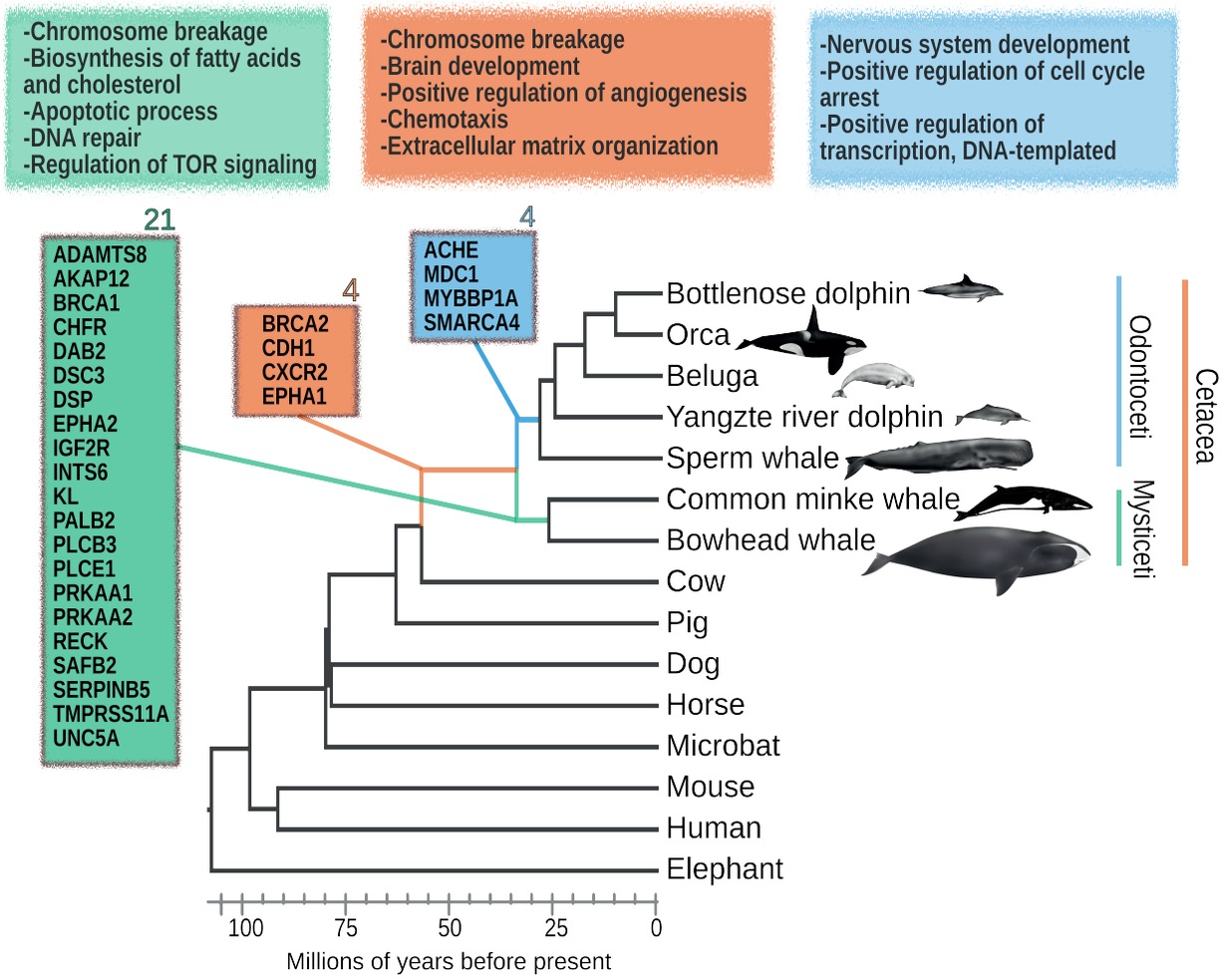
-Chromosome breakage
-Biosynthesis of fatty acids and cholesterol
-Apoptotic process
-DNA repair
-Regulation of TOR signaling
-Chromosome breakage
-Brain development
-Positive regulation of angiogenesis
-Chemotaxis
-Extracellular matrix organization
-Nervous system development
-Positive regulation of cell cycle arrest
-Positive regulation of transcription, DNA-templated
21
ADAMTS8
AKAP12
BRCA1
CHFR
DAB2
DSC3
DSP
EPHA2
IGF2R
INTS6
KL
PALB2
PLCB3
PLCE1
PRKAA1
PRKAA2
RECK
SAFB2
SERPINB5
TMPRSS11A
UNC5A
4
BRCA2
CDH1
CXCR2
EPHA1
4
ACHE
MDC1
MYBBP1A
SMARCA4
Bottlenose dolphin
Orca
Beluga
Yangzte river dolphin
Sperm whale
Common minke whale
Bowhead whale
Cow
Pig
Dog
Horse
Microbat
Mouse
Human
Elephant
Odontoceti
Mysticeti
Cetacea
100
75
50
25
0
Millions of years before present

Figure 1

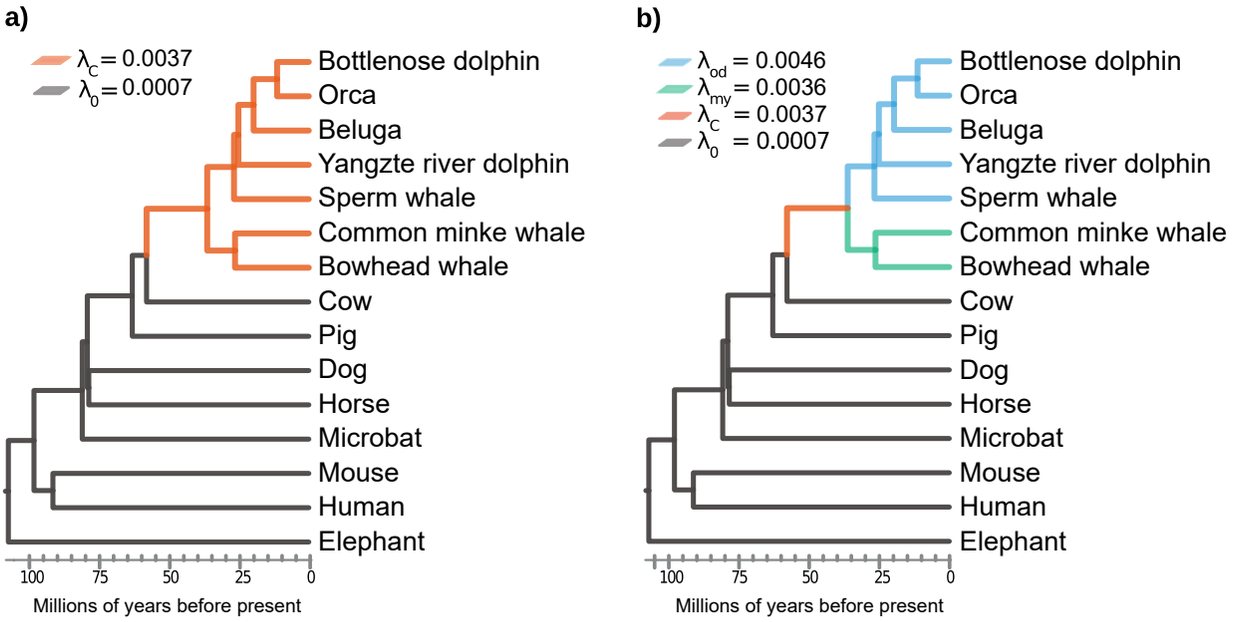
a)
λC = 0.0037
λ0 = 0.0007
Bottlenose dolphin
Orca
Beluga
Yangzte river dolphin
Sperm whale
Common minke whale
Bowhead whale
Cow
Pig
Dog
Horse
Microbat
Mouse
Human
Elephant
100
75
50
25
0
Millions of years before present
b)
λod = 0.0046
λmy = 0.0036
λC = 0.0037
λ0 = 0.0007
Bottlenose dolphin
Orca
Beluga
Yangzte river dolphin
Sperm whale
Common minke whale
Bowhead whale
Cow
Pig
Dog
Horse
Microbat
Mouse
Human
Elephant
100
75
50
25
0
Millions of years before present

Figure 2

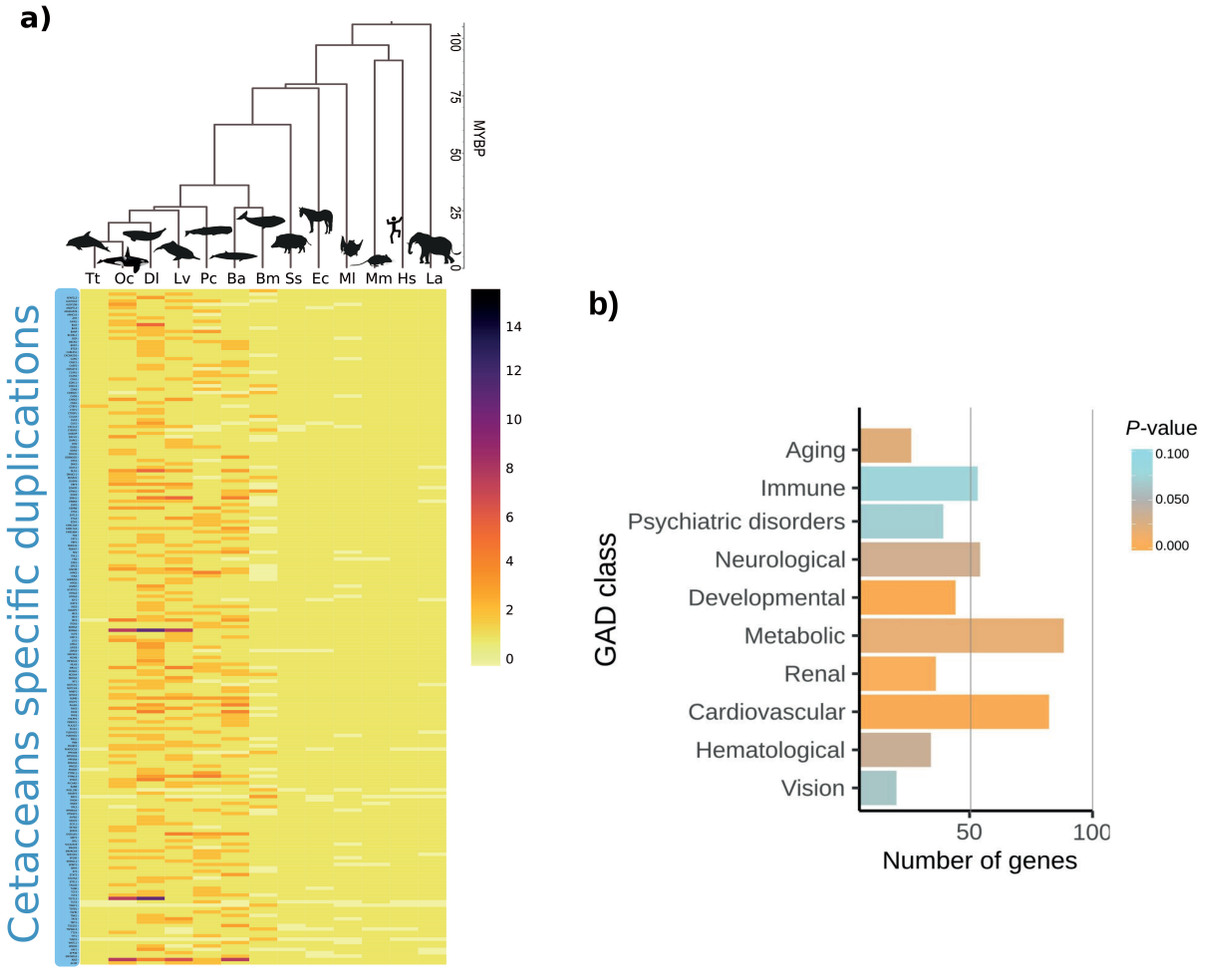
a)
100
75
50
25
0
MYBP
Tt Oc Dl Lv Pc Ba Bm Ss Ec Ml Mm Hs La
Cetaceans specific duplications
14
12
10
8
6
4
2
0
b)
GAD class
Aging
Immune
Psychiatric disorders
Neurological
Developmental
Metabolic
Renal
Cardiovascular
Hematological
Vision
50
100
Number of genes
P-value
0.100
0.050
0.000

Figure 3

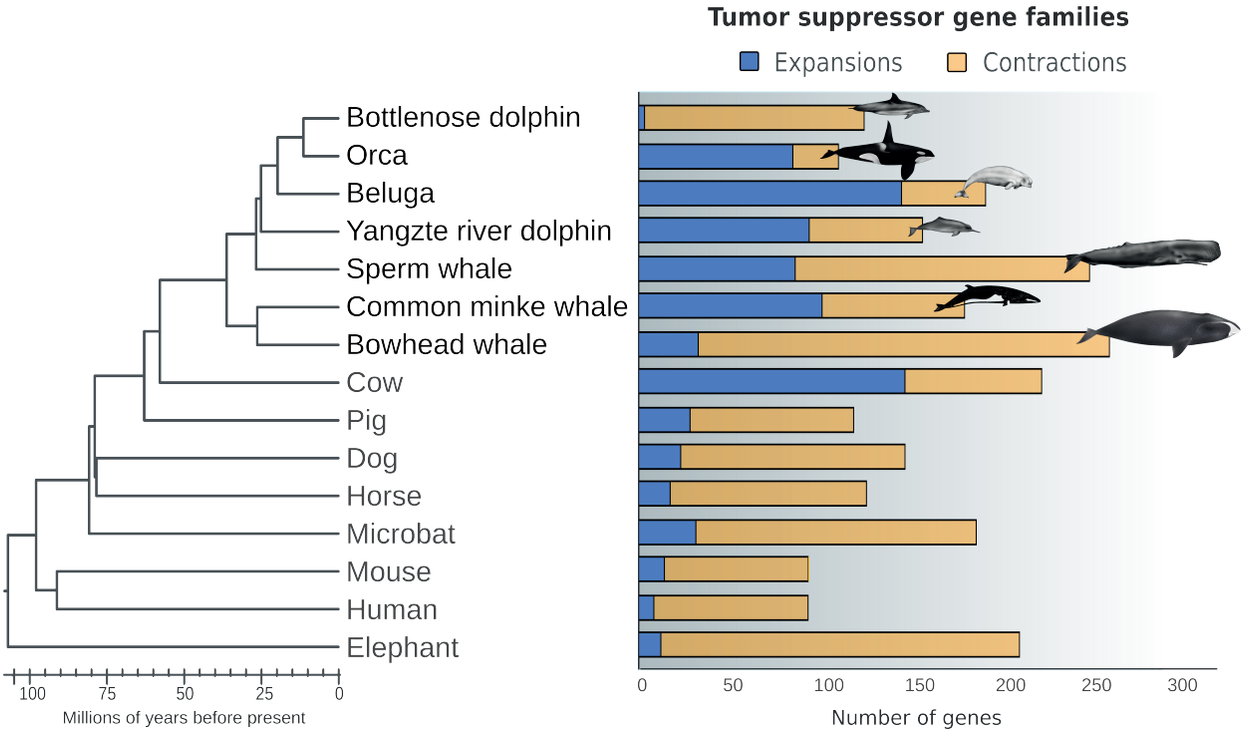
Tumor suppressor gene families
Expansions
Contractions
Bottlenose dolphin
Orca
Beluga
Yangzte river dolphin
Sperm whale
Common minke whale
Bowhead whale
Cow
Pig
Dog
Horse
Microbat
Mouse
Human
Elephant
100
75
50
25
0
Millions of years before present
0
50
100
150
200
250
300
Number of genes

Figure 4