# A *CDKN2C* retroduplication in Bowhead whales is associated with the evolution of extremely long lifespans and alerted cell cycle dynamics

Juan M. Vazquez[1], Morgan Kraft[2], and Vincent J. Lynch[2]

[1] Department of Integrative Biology, University of California – Berkeley, 4111 Valley Life Sciences MC3140, Berkeley, CA 94720-3140, USA.

[2] Department of Biological Sciences, University at Buffalo, SUNY, 551 Cooke Hall, Buffalo, NY, 14260, USA.
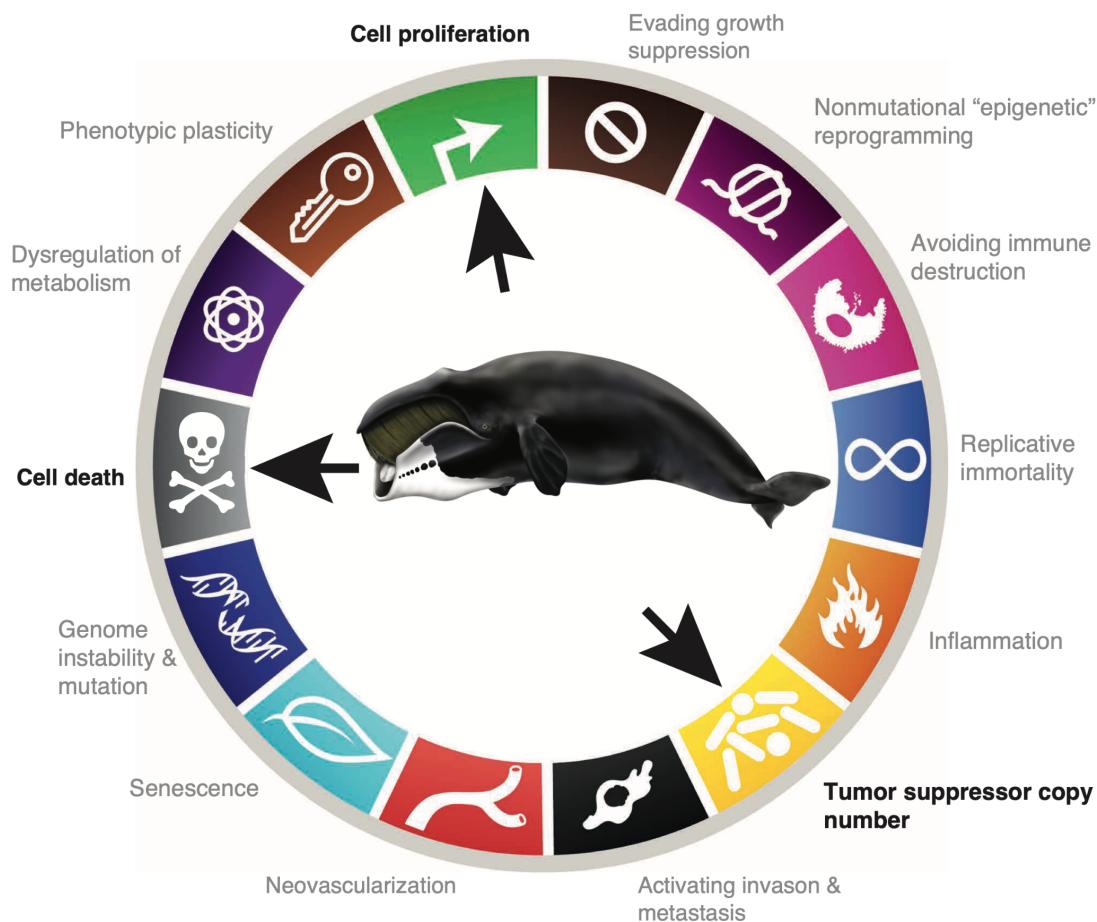
Abstract

Among the constraints on the evolution of remarkably long lifespans is an increased risk of developing cancer because organisms with long lifespans have more time to accumulate cancer-causing mutations than organisms with shorter lifespans. Indeed, while there is a strong correlation between lifespan and cancer risk within species, there is no correlation between maximum lifespan and cancer risk across species ('Peto's Paradox'). Here we use evolutionary genomics and comparative experimental biology to explore the mechanisms by which Bowhead whales (*Balaena mysticetus*), which can live at least 211 years, evolved their extremely long lifespans. We found that the Bowhead whale genome encodes a species-specific retroduplicated *CDKN2C* (p18^{INK4C}) gene (*CDKN2CRTG*). The *CDKN2CRTG* gene is embedded within a Cetacean-specific LINE L1 element, and is highly expressed in Bowhead whale tissues likely because it coopted an L1 promoter to drive constitutive expression. Furthermore we use a series of gain of function experiments to show how the duplicate *CDKN2CRTG* gene may influence cellular phenotypes such as cell cycle progression and DNA damage repair in ways that are beneficial for aging and cancer resistance. Remarkably, Bowhead and Right whales only diverged ~4-5 million years ago, suggesting the long lifespan of Bowheads may have evolved relatively recently and coincident with the origin of *CDKN2CRTG*.

Introduction

A major constraint on the evolution of long lifespans is an increased risk of developing cancer because organisms with long lifespans have more time to accumulate cancer-causing mutations than organisms with shorter lifespans (Armitage and Doll, 1954; Caulin et al., 2015; Caulin and Maley, 2011; Doll, 1971; Peto, 1975, 2015). The maximum lifespan among vertebrates, for example, ranges from over 400 years in the Greenland shark (Nielsen et al., 2016) to only 59 days in the pygmy goby (Depczynski and Bellwood, 2005). Thus, Greenland sharks have more time to acquire cancer causing mutations than pygmy gobies and should therefore have a higher risk of developing cancer than pygmy gobies. Indeed there is a strong correlation between lifespan and cancer risk within species, for example, cancer prevalence increases with age in humans and rodents (Anisimov et al., 2005) and human genetic variants that predispose to cancer are associated with decreased lifespan (Liu et al., 2022). Conversely, offspring of longer lived parents have a substantially lower incidence of cancer, as well as other common diseases,

2

than offspring of shorter lived parents (Dutta et al., 2013). There is no correlation, however, between maximum lifespan and cancer risk across species (Abegglen et al., 2015; Boddy et al., 2020; Bulls et al., 2022; Vincze et al., 2022) – this lack of correlation is often referred to as 'Peto's Paradox' (Peto, 1975, 2015).

Among the mechanisms long lived animals may have evolved that resolve Peto's paradox are a decrease in the copy number or functional activity of oncogenes, an increase in the copy number or functional activity of tumor suppressor genes (Caulin et al., 2015; Leroi et al., 2003; Sulak et al., 2016a; Vazquez et al., 2018; Vazquez and Lynch, 2021), the origin of novel tumor suppressors, or increased checks on cell growth, among many others (Gorbunova et al., 2012; Katzourakis et al., 2014; Nagy et al., 2007; Tian et al., 2019, 2013). Mice with extra copies of the master tumor suppressor *TP53* ('super p53' mice) or the cell cycle checkpoint regulator *CDKN2A* ('super Ink4a/Arf' mice), for example, are cancer resistant and age normally, whereas mice with extra copies of *TP53* and *CDKN2A* are cancer resistant and have extended lifespans (Garcia-Cao et al., 2002; Matheu et al., 2007, 2004). Aging is also suppressed in mice that overexpress the tumor suppressor *Klotho*, a peptide hormone that antagonizes the IGF-1 signaling pathway (Kurosu et al., 2005; Xie et al., 2013). Similarly while transgenic mice with an increased dosage of *SIRT1*, a protein deacetylase that contributes to telomere maintenance (Palacios et al., 2010), have normal lifespans they are protected from the development of aging-associated diseases such as glucose intolerance, osteoporosis and cancer (Herranz et al., 2010). These data indicate that mutations which increase copy number or functional dosage can inhibit tumor growth, promote healthy aging, and increase lifespan (Pinkston et al., 2006).

In this study we use evolutionary genomics and comparative experimental biology to explore the mechanisms by which Bowhead whales (*Balaena mysticetus*), which can live which live at least 211 years unlike all other Cetaceans which live 50-70 years (George et al., 2011, 1999; O'Hara et al., 2002; Tacutu et al., 2013), evolved their extremely long lifespans. We found that the Bowhead whale genome encodes a recently duplicated *CDKN2C* (p18/INK4C) gene, unlike all other Sarcopterygians including their closest living relative the Right whale (*Eubalaena sp.*). The duplicate *CDKN2C* is a retrogene (*CDKN2CRTG*) embedded within a Cetacean-specific LINE L1 element, and is highly expressed in Bowhead whale tissues likely because it coopted a L1 promoter to drive constitutive expression. Furthermore we use a series of gain of function experiments to show how the duplicate *CDKN2CRTG* gene may influence cellular phenotypes

3

such as the cell cycle and DNA damage repair in ways that are beneficial for aging and cancer resistance. Remarkably, Bowhead and Right whales only diverged ~4-5 million years ago (Árnason et al., 2018; Dornburg et al., 2012; McGowen et al., 2009), indicating the long lifespan of Bowheads evolved relatively recently and coincident with the origin of *CDKN2CRTG*.

Results

Extremely low cancer prevalence in Cetaceans

Sporadic cases of cancer have been reported in many Cetaceans, including several dolphin species, sperm, sei, humpback, beluga, and fin wales, and even remarkably large blue whales and long-lived bowhead whales, (Geraci et al., 1987; Uys and Best, 1966)(Migaki and Albert, 1982; Stimmelmayr et al., 2017)(Geraci et al., 1987; Uys and Best, 1966). Unfortunately, however, while several studies have recently reported cancer prevalence data from diverse mammals these studies were based mostly on zoo animals and therefore did not include Cetaceans. To explore cancer prevalence in Cetacea, we gathered data from the literature and found several published studies with multiple necropsy reports for individual species as well as studies that combined data from multiple species (**Figure 1A**; **Figure 1 – source data 1**). Unexpectedly, among these were two reports that described neoplasias from 7/211 (Migaki and Albert, 1982; Stimmelmayr et al., 2017)) and 18/382 (R. et al., 2020) subsistence-harvested Bowhead whales conservatively suggesting cancer prevalence in Bowheads to be 3.3% (95% CI: 1.3–6.7%) and 4.7% (95% CI: 2.8–7.3%), respectively; By far the most common tumor reported was hepatic lipoma (84%). Other studies, which had very large samples but that combined data from multiple species, reported cancer in only 14/1800 (0.78%, 95% CI: 0.04-0.13%) and 2/2000 (0.01%, 95% CI: 0.00–0.04%%) necropsies (**Figure 1A**; **Figure 1 – source data 1**). Next, we compared neoplasia prevalence in Cetacea to other Therian mammals from three published studies that included pathology reports from 37 (Boddy et al., 2020) and 191 (Vincze et al., 2022) species of Therian mammals, and a large dataset of Xenarthrans compiled from multiple published sources (Vazquez et al., 2022); The dataset includes cancer prevalence data from 244 species. As a group the mean prevalence of neoplasia in Cetacea was 1.4% (95% CI: 1.07–1.70%) when aggregated across all datasets and excluding St. Lawrence estuary belugas (**Figure 1B**; **Figure 1 – source data 2**). By rank, Cetaceans are the second most cancer resistant mammalian order (**Figure 1B**; **Figure 1 – source data 2**). Thus, cancer occurs but is also likely extremely uncommon in Cetaceans including Bowhead whales.

The Bowhead whale genome encodes a duplicate *CDKN2C* retrogene (*CDKN2CRTG*)

Whales, particularly long-lived Bowheads, must have evolved cancer protection mechanisms that reduce the enhanced risk of cancer imposed by their large bodies and long lifespans. Previous studies, for example, identified several gene duplications in the Bowhead whale genome associated with cancer biology (Keane et al., 2015; Tejada-Martinez et al., 2021). These studies, however, used methods that may not be able to identify gene duplications that have very low sequence divergence and did not include Right whales as an outgroup. Thus, they were unable to identify very recent duplications, such as those that occurred since the Bowhead-Right whale divergence ~4-5 MYA, or Bowhead whale specific duplications. To identify duplicates with low sequence divergence, we used reciprocal best BLAT (RBB) and refined our copy number estimates based on total copy coverage (ECNC, (Vazquez and Lynch, 2021)) to and duplications fragmented in lower quality genomes (Vazquez and Lynch, 2021) to identified gene duplications in the genome of a Bowhead whale, a South Atlantic Right whale, a North Atlantic Right whale, and a sperm whale (**Figure 2A**; **Figure 2 – source data 1**). Intersection of these duplicate gene sets identified a single duplication unique to the Bowhead whale genome: *Cyclin Dependent Kinase Inhibitor 2C* (*CDKN2C*; **Figure 2B**), which encodes a member of the INK4 family of cyclin-dependent kinase inhibitors that regulate the G_1/S checkpoint of the cell cycle (Cui et al., 2015).

To verify that the *CDKN2C* gene is uniquely duplicated in Bowhead whales, we manually verified *CDKN2C* copy number in the genomes of 80 Sarcopterygians including 19 Cetaceans. We found that all Cetacean (**Figure 2C**) and Sarcopterygian genomes (**Figure 2D**; **Figure 2 – source data 2**) encoded a single *CDKN2C* gene except the Bowhead whale, which we confirmed encoded a typical *CDKN2C* gene as well as a *CDKN2C* retrogene (*CDKN2CRTG*), i.e. an intronless *CDKN2C* gene, and the allotetraploid African clawed frog (*Xenopus laevis*) which has previously been reported to have two *CDKN2C* genes (CDKn2c.L and CDKn2c.S) (Tanaka 2017). *CDKN2C* encodes two transcripts that initiate from alternative promoters (**Figure 3A**). While both transcripts contain coding exons 2 and 3, transcripts initiated from the distal promoter contain an alternative noncoding exon (exon 1) that results in ~1-2kb 5'-UTR whereas transcripts generated from the proximal promoter have a ~400bp 5'-UTR (Blais et al., 1998; Phelps et al., 1998). The long transcript is predominantly expressed in germ cells (Buchold et al., 2007; Franklin and Xiong, 1996), suggesting the long transcript is more likely to be retroduplicated than the short one. Indeed, the *CDKN2CRTG* gene includes non-coding exon 1, indicating it was reverse transcribed from the long *CDKN2C* transcript (**Figure 3B**). As expected for a recent LINE L1 mediated

5

retrotransposition event, the *CDKN2CRTG* gene includes remnants of a long poly-A tail, and is flanked by a LINE1 target site (GT/AAAA) and direct repeats (**Figure 3B**).

The *CDKN2CRTG* gene is embedded within the 5'-UTR of a Cetacean-specific LINE L1 (L1-2_Ttr) that is flanked by a unique combination of transposable elements that allow us to annotate the locus in other Cetacean genomes (**Figure 3B**). The orthologous locus in other whale genomes, including North and South Atlantic Right whales, Minke whale, and Dolphin do not contain *CDKN2CRTG* (**Figure 3B**). The Bowhead whale *CDKN2C* and *CDKN2CRTG* genes are 100% identical across the 5'-UTR, coding region, and 3'-UTR. These data indicate that the retroduplication event occurred very recently in the Bowhead-lineage, coincident with the evolution of a long lifespan, and may have generated a functional *CDKN2C* gene.

CDKN2CRTG is transcribed in multiple tissues

If retoduplication of the *CDKN2C* gene played a role in the evolution of long lifespans in Bowhead whales, then *CDKN2CRTG* should be transcribed. Therefore, we used previously published RNA-Seq data from seven Bowhead whale tissues (Keane et al., 2015), five Minke whale tissues, (Yim et al., 2014), one Humpback whale tissue, and one Sperm whale tissue, and 454 GS FLX data from five dolphin tissues (Westbury et al., 2019) to determine if *CDKN2CRTG* was transcribed and compared with *CDKN2C* expression levels (**Figure 4 – source data 1**). We found the *CDKN2CRTG* transcripts from Bowhead tissues spanned the junction between the *CDKN2CRTG* 3'-end and the L1-2_Ttr LINE element (**Figure 4A**), allowing us to discriminate between *CDKN2C* and *CDKN2CRTG* transcripts. *CDKN2CRTG* transcripts were on average ~27x more abundant than *CDKN2C* transcripts (**Figure 4B**), which were expressed at relatively low levels or not at all (**Figure 4B**). We note that cDNA libraries from Bowhead tissues were constructed from poly-A captured RNA, indicating the *CDKN2CRTG* transcript is likely poly-adenylated.

A Coopted LINE L1 promoter drives *CDKN2CRTG* transcription

Most retrogenes lack native regulatory elements such as promoters and enhancers that direct their expression, thus expression of *CDKN2CRTG* is either regulated by existing regulatory elements or ones that evolved *de novo*. Our observation that the *CDKN2CRTG* gene is embedded within a LINE L1 element suggests that *CDKN2CRTG* transcription may be regulated by the endogenous LINE L1 promoter/enhancer located within the LINE L1 5'-UTR (Lavie et al., 2004;

Swergold, 1990). To test if this region has regulatory abilities, we synthesized ten 301bp fragments including 51bp 5'- and 3'- overlaps that tile the region from upstream of the L1-2_Ttr element to downstream of the *CDKN2CRTG* initiation codon (**Figure 4C**) and cloned each fragment into the pGL4.14[*luc2*/Hygro] and pGL4.26[*luc2*/minP/Hygro] luciferase reporter vectors; the pGL4.14[*luc2*/Hygro] vector lacks an enhancer and a promoter whereas the pGL4.26[*luc2*/minP/Hygro] vector contains a minimal promoter allowing us to test the cloned fragments for promoter and enhancer functions. Next, we transiently transfected Chinese Hamster Ovary (CHO) cells with each reporter construct and used a dual luciferase reporter assay to quantify luciferase expression. Fragments from the 5'-UTR of the L1-2_Ttr element significantly increased luciferase expression from the pGL4.14[*luc2*/minP/Hygro] reporter vector peaking with fragment four, which increased luciferase expression 87-fold (95.0%CI 70.7-107 fold) (**Figure 4C** and **Figure 4 – figure supplement 1**). In contrast, these fragments did not increase luciferase expression when cloned into the pGL4.26[*luc2*/minP/Hygro] vector (**Figure 4C** and **Figure 4 – figure supplement 2**), suggesting this region can function as a promoter but likely not an enhancer. Fragments five and six, however, which include the target site and the most proximal region of the *CDKN2CRTG* 5'-UTR, did increase luciferase expression from the pGL4.26[*luc2*/minP/Hygro] vector (**Figure 3C** and **Figure 3 – figure supplement 2**) indicating this region has enhancer/promoter potential. These data suggest that expression of *CDKN2CRTG* is regulated by an enhancer/promoter located within the LINE L1-2_Ttr element immediately upstream *CDKN2CRTG* 5'-UTR.

CDKN2CRTG overexpression induces cell-cycle delay at the G_1 and G_1/S checkpoints

CDKN2C expression is cell-cycle dependent with transcription initiating around the late G_1/S checkpoint, peaking in the S and G_2/M phases before falling again in early G_1 (Hirai et al., 1995) and acts as a G_1/S checkpoint (Cui et al., 2015). However, the high level of *CDKN2CRTG* expression and the usage of LINE L1 derived promoter to drive expression suggests constitutive expression of *CDKN2CRTG* throughout the cell-cycle. To explore the consequences of constitutively high *CDKN2CRTG* expression on cell cycle dynamics, we first generated a stable Fucci-2a reporter CHO cell line (FuCHO). The Fucci-2a system consists of a monomeric Cherry (mCherry) tagged fragment of human CDT1 (hCdt1) gene and a monomeric Venus (mVenus) tagged fragment of human GMNN (hGem) gene joined by the *Thosea asigna* virus 2A (T2A) self-cleaving peptide (Mort et al., 2014; Sakaue-Sawano et al., 2008) (**Figure 5A**). After transcription,

the T2A peptide cleaves generating an mCherry-hCdt1(30/120) probe that accumulates during G_1 phase and is degraded at the G_1-S transition and a mVenus-hGem probe that accumulates during S/G_2/M phases and is rapidly degraded prior to cytokinesis (**Figure 5B**). Next we synthesized and cloned the *CDKN2CRTG* gene into the mammalian pcDNA3.1-C-eGFP expression vector , which adds a C-terminal eGFP tag, transiently transfected this vector into FuCHO cells, and quantified the cell-cycle stage using flow cytometry. Consistent with previous studies of CDKN2C, we found that overexpression of *CDKN2CRTG* induced an accumulation of cells in G_1 (Hirai et al., 1995) and at the G_1/S checkpoint (Cui et al., 2015), which was particularly pronounced in the GFP+ population of FuCHO cells (**Figure 5C-E**). These results indicate that the bowhead whale CDKN2CRTG protein is functionally equivalent to CDKN2C with respect to its ability to negatively regulate the cell cycle during the G_1 phase. .

CDKN2CRTG overexpression protects cells from DNA damage induced apoptosis

Ectopic overexpression of *CDKN2C* can induce cell cell arrest and apoptosis in normal and cancerous mouse and human cells (Komata et al., 2003a; Kulkarni et al., 2002; Schreiber et al., 1999), suggesting that a duplicate *CDKN2C* gene expressed at high levels may also affect DNA damage processes. Therefore, we used an ApoTox-Glo assay to explore the effects of CDKN2CRTG-eGFP overexpression in FuCHO cells treated with doxorubicin for 18 hours. Doxorubicin, a topoisomerase poison, induces DNA double stranded breaks by intercalating with DNA and inhibiting the progression of topoisomerase II complex during transcription (Pommier et al., 2010) and during DNA replication in S phase (Tacar et al., 2013). As expected, doxorubicin treatment alone induced apoptosis (**Figure 6A**) and decreased cell viability (**Figure 6B**) whereas CDKN2CRTG overexpression protected cells from apoptosis and dramatically preserved cell viability (**Figure 6A/B**). Thus, CDKN2CRTG expression can prevent the accumulation of DNA damage.

Bowhead whale cells have longer doubling times than other whale cells

The role of CDKN2C in regulating the G_1 and G_1/S cell cycle checkpoints suggest that Bowhead whale cells may have different cell cycle dynamics than other species. While measurements of cell cycle phase length or overall duration are not available for most species, many studies report population doubling times from mammalian cells in culture. Population doubling time is proportional to the cumulative length of individual cell cycle phases, particularly

8

the length of G_1 (Blank et al., 2018). Therefore, we performed a thorough literature review and assembled a dataset of population doubling times from 60 species of Eutherian mammals (Vazquez et al., 2022). While there was considerable variation in population doubling times between species, whale cells had similar doubling times as other species of mammals (**Figure 7A**). Bowhead whale cell lines (n=4), however, had significantly longer doubling times than cell lines from other whales (n=13) with an unpaired mean difference in doubling time of 15.4 hours (95.0%CI 9.46-24.9) and a two-sided permutation t-test P-value=0.0104 (**Figure 7B**). Bowhead whale cell lines also had longer doubling times than cell lines from Right whales (n=3), with an unpaired mean difference in doubling time of 14.2 hours (95.0%CI 7.58-25.9) and a two-sided permutation t-test P-value=0.0 (**Figure 7C**). We note that these differences in population doubling times may reflect biological differences between species or technical variation not related to species-specific differences in doubling time or cell cycle length.

Discussion

Among the challenges for discovering the genetic and molecular changes associated with the evolution of remarkably long lifespans is identifying long-lived species that are closely related to species with 'normal' body sizes and lifespans. The ideal study system for exploring how animals evolved long healthy lifespans, for example, would be one in which an exceptionally long-lived lineage was deeply nested within a clade with short lifespans, for which genomes and potentially other biological materials are available for study. The long-lived species should also be very closely related to short-lived species, allowing us to identify derived genetic changes that are phylogenetically associated with lifespan extension rather than other traits or neutral genetic changes. Bowhead whales are an excellent system in which to study lifespan evolution because they can live over 200 years, unlike other whales which live ~60 years, and diverged from Right whales only ~4-5 million years ago (Árnason et al., 2018; Dornburg et al., 2012; McGowen et al., 2009). Thus, Bowhead whales evolved their extreme lifespans relatively recently allowing us to identify potentially causal mutations from a sea of neutral ones.

Extremely low cancer prevalence in Cetaceans

Previous studies have reported sporadic cases of cancer in many Cetacean species, including several dolphin species, sperm, sei, humpback, beluga, fin, blue, right, and bowhead whales, (Geraci et al., 1987; Migaki and Albert, 1982; Stimmelmayr et al., 2017; Uys and Best,

1966), but there have been few systematic studies of cancer prevalence in whales. To estimate cancer prevalence in Cetaceans, we gathered mortality data from previously published studies and found that in general, the prevalence of cancer among Cetaceans was low – only 1.4% (95% CI: 1.07–1.70%) when aggregated across all datasets and excluding St. Lawrence estuary belugas (**Figure 1**). Remarkably, among these published studies were two that reported neoplasias from 7/211 (Stimmelmayr et al., 2017) and 18/382 (R. et al., 2020) subsistence-harvested Bowhead whales, suggesting that cancer prevalence in Bowheads to be 3.3% (95% CI: 1.3–6.7%) and 4.7% (95% CI: 2.8–7.3%), respectively. Thus, despite their remarkably large sizes, and in the case of Bowhead whales long lifespans, cancer prevalence in Cetaceans is generally very low. Indeed, Cetaceans are the second most cancer resistant mammalian order (**Figure 1**). We note, however, that these estimates are likely to be very dependent on the sample size of necropsy reports. For example, cancer was detected in at least one individual in almost all species with more than 82 individual pathological records illustrating that cancer is likely to be detected in all mammals with adequate sample sizes (Vincze et al., 2022). The number of post-mortem pathology reports was below 82 for most Ceteceans suggesting caution when interpreting the low cancer prevalence in these species, yet these data indicate that whales must have evolved enhanced cancer suppression mechanisms to reduce their cancer risk because in aggregate cancer is relatively rare in Cetaceans.

What might be the consequences of *CDKN2C* duplication? A longer cell cycle

CDKN2C inhibits cyclin-CDK4/6 complexes (Guan et al., 1994; Noh et al., 1999) that are required for G_0/G_1 progression and entry in S phase (Robitaille et al., 2016), thereby acting as a G_1/S checkpoint (Cui et al., 2015) and regulating the length of G_1 (Dong et al., 2018). Ectopic overexpression of *CDKN2C,* for example, reduces cell growth by preventing CDK4/6 activation, lengthening G_1, and delaying progression through the G_1/S checkpoint (Bench et al., 1996; Komata et al., 2003b; Schreiber et al., 1999; Solomon et al., 2008). In contrast, pharmacological inhibition or knockdown of *CDKN2C* expression promotes cell cycle progression and proliferation (Gao et al., 2015; McClendon et al., 2012). Indeed, we found that overexpression of $CDKN2C^{RTG}$ increased the proportion of FuCHO cells in G_1 and G_1/S. These data are broadly congruent with the "textbook" model of the eukaryotic cell cycle (Lubischer, 2007), in which the length of the G_1 phase accounts for most of the variation in cell cycle length between cell-types and species (Blank et al., 2018; Brauer et al., 2008; Fisher, 2016; Johnston et al., 1977; Lubischer, 2007; Wheals,

1985). These data suggest that cell cycle length, particularly the length of the G_1 phase, may be longer in Bowhead whale cells than cells from other species.

Consistent with this hypothesis, Bowhead whale cells grow extremely slowly (Jarrell, 1979; Smith et al., 1987) prompting us to perform a literature review to compare growth rates of mammalian cells. The most commonly reported measurement for the growth rate of cells in culture is the population doubling time, i.e. the time it takes for a population to double its size (Lindström and Friedman, 2020); while not a direct estimate of cell cycle length, the population doubling time is proportional to cell cycle length and largely determined by the duration of G_1 (Blank et al., 2018). We found that there was significant variation in doubling time between mammalian cells, and that the mean population doubling time of multiple Bowhead whale cell-types was 14.2 hours (95.0%CI 7.58-25.9) longer than the doubling times of Right whale cells and 15.4 hours (95.0%CI 9.46-24.9) longer than the doubling times of other whale cells. We note that these differences in population doubling times may reflect real biological differences in cell cycle length between species, as well as differences between cell-types, technical differences in how cells were cultured, or how doubling times were calculated, among other coundounfers not directly related to species-specific differences in doubling time or cell cycle length. Yet, these data are consistent with a model in which a consequence of duplicate *CDKN2C* gene, particularly one that is highly expressed, is a decrease in cell growth rates.

What might be the consequences of *CDKN2C* duplication? Enhanced DNA damage repair

Activation of cell cycle checkpoints in response to DNA damage pauses cell cycle progression, which give cells time to repair damage or induce apoptosis thereby preventing transmission of damaged or incompletely replicated chromosomes (Bartek and Lukas, 2001; O'Connell and Cimprich, 2004; O'Connor, 1997). Previous studies have also found that the INK4 family of cyclin-dependent kinase inhibitors directly enhance DNA damage repair independent of their role in cell cycle regulation (Cánepa et al., 2007). *CDKN2A* (p16[INK4a]), for example, protects cells from UV-induced apoptosis by downregulating the pro-apoptotic BAX protein (Al-Mohanna et al., 2004). After DNA damage, *CDKN1A* (p21[Cip1]) or *CDKN2A* overexpression induces cell cycle arrest and inhibits DNA damage induced apoptotic events such as cytochrome c release, mitochondrial membrane depolarization, and activation of the caspase cascade (Le et al., 2005). *CDKN2D* (p19[INK4d]) also plays a crucial role in regulating genomic stability and cell viability in

11

response to genotoxic stress (Ceruti et al., 2005; Scassa et al., 2007; Tavera-Mendoza et al., 2006; Tavera-Mendoza et al., 2006). *CDKN2D* overexpression, for example, enhanced DNA damage repair whereas *CDKN2D*-deficient cells had diminished ability to repair DNA damage; Enhancement of DNA repair inversely correlated with apoptosis and was independent of its role in cell cycle regulation (Cánepa et al., 2007; Ceruti et al., 2005; Scassa et al., 2007).

Our finding that *CDKN2CRTG* overexpression protected Chinese hamster ovary cells from doxorubicin induced apoptosis likely extends the observation that the INK4 family of cyclin-dependent kinase inhibitors can enhance DNA damage repair to *CDKN2C*. While we have not explored the mechanisms by which *CDKN2CRTG* overexpression protects from apoptosis induced cell death, it is may be by slowing the progression of the cell cycle at the G_1 and G_1/S checkpoints and preventing rapid progression into G_2, allowing cells more time to repair DNA damage or mechanisms independent of its role in cell cycle regulation like *CDKN2D*. *CDKN2CRTG* overexpression may also slow cell cycle progression into G_2 when doxorubicin induces most DNA damage (Potter et al., 2002), thereby limiting the level of DNA damage and/or allowing cells to repair damage before DNA damage can initiate apoptosis. Thus, a longer cell cycle might allow Bowhead whale cells more time to correct DNA damage and other genomic instabilities such as aneuploidy but additional studies are needed to determine if the mechanism is direct, i.e., by allowing cells more time to repair DNA damage, indirect, i.e., by limiting the accumulation of DNA damage, or some combination of direct and indirect effects. Both indirect and direct effects models are consistent with observations that whale cells generally have more efficient DNA damage repair mechanisms than cells from some other mammals (Browning et al., 2017; Chen et al., 2009; Meaza et al., 2020; Wise et al., 2010, 2008).

What might be the consequences of *CDKN2C* duplication? Reduced cancer risk

CDKN2C is a tumor suppressor for a variety of cancer types (Franklin et al., 2000), which may be related to both its role in cell cycle regulation and a direct role DNA damage repair, and this function is dosage sensitive. *CDKN2C* has a dosage dependent inhibitory effect on teratoma formation in NOD/SCID mice, such that tumor volume from *CDKN2C* overexpressing teratomas was significantly reduced compared to control and *CDKN2C*$^{-/-}$ teratomas. Similarly, *CDKN2C* overexpression in the nasopharyngeal carcinoma epithelial tumor cell line HONE-1 prevented tumor growth when implanted to a nude mouse xenograft model (Li 2021,Lin 2018). *CDKN2C*

12

haploinsufficiency also sensitizes mice to carcinogen-induced tumorigenesis (Bai et al., 2003). Remarkably, direct administration of a cell-permeable CDKN2C protein can inhibit tumor growth by 86-98% in a HCT116 tumor xenograft model (Lim et al., 2012) and *CDKN2C* copy number loss was associated with enhanced metastasis, metastasis to more distant sites, and decreased overall survival in patients with medullary thyroid carcinoma (Grubbs et al., 2016). These data indicate that *CDKN2C* dosage has functional effects on cancer initiation and progression and suggests that an additional copy of *CDKN2C* in Bowhead whales might be protective against the kinds of unregulated cell proliferation and genotoxic stress associated with cancer, particularly metastasis and survival once cancer develops.

What might be the consequences of *CDKN2C* duplication? An aging, cancer resistance, and reproduction life history trade-off

While *CDKN2C* duplication may have played a role in the evolution of a remarkably long lifespan in Bowhead whales, our survey of mammalian *CDKN2C* copy number variation suggest that *CDKN2C* duplication is likely to be uncommon and thus may be associated with antagonistic pleiotropic effects that impose a constraint (cost) on increased dosage. Life history trade-offs between body size, lifespan, cancer resistance, and other organismal traits may be common (Garland, 2014; Stearns, 1989), particularly for reproductive (Boddy et al., 2015; Charnov and Ernest, 2006; Walker et al., 2008) and cellular phenotypes (Aktipis et al., 2013). For example, missense alleles in *ARHGAP27* are associated with increased number of children born but reduce reproductive lifespan (Mathieson et al., 2020), while missense alleles in *BRCA1* and *BRCA2* associated with increased lifetime risk of developing breast or ovarian cancer are also associated with an increased number of children born (Smith et al., 2012). Similarly, genetic variants that are associated with *HLA-F* and *TAP2* expression in humans increase fecundability but also predispose autoimmune disorders (Burrows et al., 2016; Mika et al., 2018; Mika and Lynch, 2016). Thus while additional copies of *CDKN2C* may provide cells with added robustness against the kinds of genotoxic stress associated with cancer progression, it may also be associated with antagonistic pleiotropic effects that constrain *CDKN2C* copy number changes.

CDKN2C, for example, is expressed in adipose tissue where it regulates adipocyte differentiation and lipid storage and *CDKN2C* expression levels have been associated with type-2 diabetes and central obesity (Pereira et al., 2022). Genome-wide association studies in the

13

diversity outbred mouse population have also implicated *CDKN2C* in testis weight variability (Yuan et al., 2018). Consistent with this genetic association, *CDKN2C* is expressed in the seminiferous tubules of postnatal mice including post-mitotic spermatocytes undergoing meiosis (Yuan et al., 2018; Zindy et al., 2001) and *CDKN2C* knockout mice have generalized organomegaly including testicular enlargement and hyperplasia of interstitial testicular Leydig cells, which produce reduced levels of testosterone (Franklin et al., 2000, 1998; Zindy et al., 2001). While loss of *CDKN2C* (or *CDKN2D*) alone had no observable effect on male fertility, spermatogonia do not differentiate properly in *CDKN2C* and *CDKN2D* double-knockout mice and they produce few mature sperm; The residual spermatozoa have reduced motility and decreased viability (Yuan et al., 2018; Zindy et al., 2001). Thus, *CDKN2C* (*CDKN2D*) dosage is essential for testis and sperm development and male fertility.

In contrast to the testis phenotypes in *CDKN2C* in knockout mice, Bowhead whales have relatively small testes that are much smaller than than Right whales (~200kg vs >1,000kg), and "cryptic" Leydig (interstitial) cells that are difficult to identify by histological examination (O'Hara et al., 2002). Remarkably, several male pseudohermaphrodite Bowheads have also been described (O'Hara et al., 2002; Tarpley et al., 1995). While the external phenotype male pseudohermaphrodite Bowheads was female, the gonads were underdeveloped testes (O'Hara et al., 2002; Tarpley et al., 1995). In the most thoroughly studied individual, all derivatives of the mesonephric (Wolffian) and paramesonephric (Müllerian) ducts, the embryonic structures that form the internal male and female reproductive tract respectively, were absent (Tarpley et al., 1995). Male Bowhead whales also may have the oldest age at sexual maturation of any Cetacean (George et al., 1999). These phenotypes are consistent with Bowhead whales having an additional copy of *CDKN2C* and suggest tradeoff between aging and cancer resistance and male reproduction.

Ideas and speculations: On the rapid evolution of cancer resistance

The increased longevity of Bowhead evolved very rapidly, within a few million years after their divergence with Right whales, which likely limited the kinds of genetic variation that could contribute to enhanced lifespan. For example, the average mammalian gene duplication rate, both segmental and retroduplication, is estimated to be $0.5–20×10^{-3}$ per gene per million years (Lynch and Conery, 2000; Pan and Zhang, 2007), suggesting that there should be about one gene duplication unique to Bowhead whales given Bowhead and Right whales diverged only ~4-

5 million years ago (Árnason et al., 2018; Dornburg et al., 2012; McGowen et al., 2009). These data indicate that gene duplication has likely played a minor, even if important, role in the evolution of long lifespans and reduced cancer risk in Bowhead whales unlike the pervasive role gene duplication appears to have played in the evolution of reduced cancer risk in elephants, Xenarthrans, and Galapagos tortoises (Glaberman et al., 2021; Vazquez et al., 2022; Vazquez and Lynch, 2021). Thus, other mechanisms must have played a more numerous role than copy number gains and losses, such as gene regulatory evolution.

The evolutionary possibilities for solutions to Peto's Paradox to arise is likely also constrained not just by the short time during which Bowheads evolved long lifesans, but also by the fundamental nature of developmental processes such as embryogenesis and growth rates that require cellular proliferation. We have shown that by insertion into a LINE1 element with a strong promoter, *CDKN2CRTG* achieves an extraordinarily high level of expression. This simple event likely led to an immediate phenotypic consequence in individuals with the *CDKN2CRTG* duplication, and possibly immediate reduction in cancer risk, that may have facilitated any ongoing increases in body size and lifespan in the Bowhead Lineage. Of course, to prove such a hypothesis requires the discovery of additional Mystecini fossils to reconstruct the history of body size expansions within this clade, which could then be correlated with the gene duplications described herein with functional validation; A developmental and evolutionary proof of the importance of the *CDKN2CRTG* duplication for body size and lifespan evolution in Bowheads would also require the kinds forward and reverse genetic experiments that are possible in model organisms but impossible in whales.

Caveats and limitations

Our gain of function experiments indicate that *CDKN2CRTG* expression may affect the cell cycle length and the response of cells to DNA damage. A limitation of these experiments, however, is that we have not directly quantified DNA damage and repair, therefore it is unclear if the protective effects of *CDKN2CRTG* overexpression we observed is because of direct effects on DNA damage and repair rates or indirect effects based on cell cycle slowing. To disentangle these possibilities, we must quantify DNA damage repair rates between cells in different cycle phases and with *CDKN2CRTG* overexpression. We must also perform similar gain and loss of function experiments in Bowhead whale cells to establish a causal connection between *CDKN2CRTG* and cellular phenotypes related to healthy aging and lifetime cancer resistance.

Finally, we are unable to determine causality, i.e. that a specific genetic change, such as the retroduplication of *CDKN2C*, is necessary and sufficient for the realization of phenotypes that contribute to the evolution of long lifespans in Bowhead whales using traditional forward and reverse genetic techniques. While this constrains our ability to demonstrate proximate causation within an organismal context, we can construct strong circumstantial arguments in favor of causation using methods like those used to associate genetic variation with phenotypes in genome-wide association studies.

Conclusions

Bowhead whales are an excellent model system in which to explore the evolution of long lifespans because they are closely related to species with much shorter lifespans and evolved that long lifespan since their divergence with Right whales ~4-5 million years ago. In this study we used comparative genomics to identify gene duplications that are phylogenetically associated with the evolution of long lifespans in Bowhead whales. We identified a single gene duplication, a retroduplication of the cyclin dependent kinase inhibitor *CDKN2C* (*CDKN2CRTG)*, unique to Bowhead whales. Remarkably, the putative phenotypic consequences of a duplicate *CDKN2C* gene are relatively straightforward given its role in cell cycle regulation and DNA repair. Our experimental data indicate that *CDKN2CRTG* may regulate the cell cycle and contribute to enhanced DNA damage repair. These data suggest that Bowhead whales evolved their extremely long lifespans at least in part through duplication of the *CDKN2C* gene, which may reduce their lifetime risk of developing cancer.

Materials and methods

Identification of gene duplications in the Bowhead, Right, and Sperm whale genomes

Approach: We identified putative gene duplicates in the genome of a Bowhead whale (PRJNA194091), North Atlantic (https://www.dnazoo.org/assemblies/Eubalaena_glacialis) and South Atlantic (https://www.dnazoo.org/assemblies/Eubalaena_australis) Right whales, and Sperm whale (assembly ASM283717v2) using a combination of Reciprocal Best Hit BLAT (RBHB) and Effective Copy Number By Read Coverage (ECNC), which we have previously described (Vazquez and Lynch, 2021). Note that the draft Right whale assemblies was generated by the DNA Zoo team and Mark Yandell's lab at the University of Utah with contributions from

16

Reciprocal best hit BLAT: We previously developed a reciprocal best hit BLAT pipeline to identify putative homologs and estimate gene copy number across species (Vazquez and Lynch, 2021). The Reciprocal Best Hit (RBH) search strategy is conceptually straightforward: 1) Given a gene of interest G_A in a query genome AA, one searches a target genome B for all possible matches to G_A G_A; 2) For each of these hits, one then performs the reciprocal search in the original query genome to identify the highest-scoring hit; 3) A hit in genome B is defined as a homolog of gene G_A if and only if the original gene G_A G_A is the top reciprocal search hit in genome AA. We selected BLAT as our algorithm of choice, as this algorithm is sensitive to highly similar (>90% identity) sequences, thus identifying the highest-confidence homologs while minimizing many-to-one mapping problems when searching for multiple genes. RBH performs similar to other more complex methods of orthology prediction and is particularly good at identifying incomplete genes that may be fragmented in low quality/poorly assembled regions of the genome.

Estimated copy number by coverage: In low-quality genomes, many genes are fragmented across multiple scaffolds, which results in BLA(S)T-like methods calling multiple hits when in reality there is only one gene. To compensate for this, we developed a novel statistic, Estimated Copy Number by Coverage (ECNC), which averages the number of times we hit each nucleotide of a query sequence in a target genome over the total number of nucleotides of the query sequence found overall in each target genome (Vazquez and Lynch, 2021). This allows us to correct for genes that have been fragmented across incomplete genomes, while accounting for missing sequences from the human query in the target genome. Details of the ECNC method can be found in (Vazquez and Lynch, 2021).

RecSearch pipeline: We created a custom Python pipeline for automating RBHB searches between a single reference genome and multiple target genomes using a list of query sequences from the reference genome. For the query sequences in our search, we used the hg38 UniProt proteome, which is a comprehensive set of protein sequences curated from a combination of predicted and validated protein sequences generated by the UniProt Consortium. Next, we

excluded genes from downstream analyses from large multigene families for which assignment of homology using the RBBH approach was uncertain, including uncharacterized ORFs, LOCs, HLA genes, replication dependent histones, odorant receptors, ribosomal proteins, zinc finger transcription factors, Coiled Coil Domain Containing proteins, Family With Sequence Similarity genes, Mitochondrial Ribosomal Proteins, NADH:Ubiquinone Oxidoreductase, Solute Carrier Family genes, TMEM Transmembrane Proteins, DnaJ Heat Shock Protein Family genes, Dual Specificity Phosphatases, Eukaryotic Translation Initiation Factors, F-Box Proteins, Polypeptide N-acetylgalactosaminyltransferases, Ring Finger Proteins, Sterile Alpha Motif Domain Containing proteins, Serpins, SEC14 And Spectrin Domain Containing proteins, Small Nuclear Ribonucleoproteins, Serine/Threonine Kinases, TBC1 Domain Family Members, Tubulins, Ubiquitin Conjugating Enzymes, WD domain containing proteins, viral and repetitive-element-associated proteins, "Uncharacterized," "Putative," and "Fragment" proteins.

Duplicate gene inclusion criteria: In order to condense transcript-level hits into single gene loci, and to resolve many-to-one genome mappings, we removed exons where transcripts from different genes overlapped, and merged overlapping transcripts of the same gene into a single gene locus call. The resulting gene-level copy number table was then combined with the maximum ECNC values observed for each gene in order to call gene duplications. We called a gene duplicated if its copy number was two or more, and if the maximum ECNC value of all the gene transcripts searched was 1.5 or greater; We have shown that incomplete duplications can encode functional genes (Sulak et al., 2016b; Vazquez et al., 2018), therefore partial gene duplications were included provided they passed additional inclusion criteria (see below). The ECNC cut-off of 1.5 was selected empirically, as this value minimized the number of false positives seen in a test set of genes and genomes. We thus identified 1710 duplicate genes in the Bowhead whale genome. We note that: 1) The RBBH/ECNC approach is not exhaustive and other gene duplications that are 100% identical between query and target sequences within the Bowhead whale genome may be collapsed into a single hit, undercounting gene duplicates; 2) These duplicate genes may not be specific to Bowhead whales and represent a mixture of both ancient and recent gene duplications; and 3) We have not filtered these duplicates to exclude those with signatures of pseudogenes such as inframe stop codons, retrogenes, or non-expressed genes. Thus this duplicate gene set will include a mixture of functional and non-functional genes, but is a useful starting point to more detailed studies such as manual verification of duplicate genes.

18

Identification of the *CDKN2C* gene in additional Sarcopterygian and Cetacean genomes

We also used BLAT to verify the copy number of CDKN2C in 61 Sarcopterygian genomes (**Figure 1 – source data 2**). After identifying the canonical gene from each species, we used the corresponding nucleotide sequences as the query sequence for additional BLAT searches within that species' genome. To further confirm the orthology of each gene we used a reciprocal best BLAT approach, sequentially using the putative CDS of each gene as a query against the human genome.

RNA-Seq transcript abundance estimation

We used previously published RNA-Seq data from seven Bowhead whale tissues (Keane et al., 2015); five Minke whale tissues (Yim et al., 2014); one Humpback whale tissue; one Sperm whale tissue; and data from 454 GS FLX pyrosequencing of five dolphin tissues (Westbury et al., 2019) to estimate transcript abundances (**Figure 3 – source data 1**). Individual datasets were aligned to the Bowhead whale, Minke whale, Humpback whale, Sperm whale, or dolphin reference genome on Galaxy with HISAT2 (Galaxy Version 2.2.1+galaxy0) using default settings and reporting alignments tailored for transcript assemblers including StringTie, and transcripts assembled with StringTie (Galaxy Version 2.1.7+galaxy1) without a reference to guide assembly. Assembled transcripts from each dataset were merged into a single gtf file with StringTie merge transcripts (Galaxy Version 2.1.7+galaxy1) using default settings, except that minimum input transcript abundances (FPKM and TPM) to include in the merge was set to 0. Sequences for each transcript were extracted from the merged gtf file with gffread (Galaxy Version 2.2.1.3+galaxy0) and used as the reference transcriptome to estimate transcript abundances with Kallisto and bias correction (Bray et al., 2016).

Fucci2a construct synthesis and stable cell line generation

We generated a Fucci2a reporter vector by synthesizing the mCherry-hCdt1(30/120) (GenBank accession: AB512478), *Thosea asigna* virus 2A (T2A) self-cleaving peptide (GSGEGRGSLLTCGDVEENPGP), and mVenus-hGeminin(1/110) (GenBank accession: AB512479) reporters as a single gene with mouse codon usage (GeneScript);. The mCherry-hCdt1(30/120)-T2A-mVenus-hGeminin(1/110) fusion gene was then cloned into the pcDNA5/FRT vector (GeneScript). We generated a stable Fucci-2a reporter Chinese Hamster Ovary (FuCHO) cell line using the Flp-In System (Thermo Fisher K601001) and the Flp-In-CHO cell line (invitrogen R758-07)). Flp-In-CHO Chinese Hamster ovary cells were maintained in T75 culture flasks at

19

37ºC, 5% CO2, and 90-90% humidity in a culture medium consisting of Ham's F12, 10% FBS, 2mM L-glutamine, 1% Pen-Strep, and 100ug/mL of Zeocin (to maintain stable integration of the FRT landing site). To integrate the Fucci-2a reporter into the Flp-In-CHO cell line, 1x10^6 cells were transiently with the Fucci-2a pcDNA5/FRT vector using Lipofectamine LTX (Invitrogen), the CHO-K1 transfection protocol, and the recommended ratio of insert plasmid and flippase plasmid. 24 hours after transfection the cells were selected using 800ug/mL hygromycin (Gibco) for 5 days. Cells were propagated in media containing 200ug/mL Hygromycin. Stability of the knock-in was confirmed by estimating the abundance of fluorescent cells in culture over several passages, both of the original lines as well as from cryopreserved aliquots.

Luciferase reporter vector construction and Dual-Luciferase Reporter assay

To test if the region 5' to the *CDKN2CRTG* gene regulatory abilities, we divided it into ten 301bp fragments including 51bp 5'- and 3'- overlaps that tile the region from upstream of the L1-2_Ttr element to downstream of the *CDKN2CRTG* initiation codon using OVERFRAG (http://143.107.58.249/cgi-bin/overfrag.cgi) (Ferreira-Junior and Digiampietri, 2019). Each fragment was synthesized and cloned (GeneScript) into the pGL4.14[*luc2*/Hygro] and pGL4.26[*luc2*/minP/Hygro] firefly luciferase reporter vectors (Promega); the pGL4.14[*luc2*/Hygro] vector lacks an enhancer and a promoter whereas the pGL4.26[*luc2*/minP/Hygro] vector contains a minimal promoter. Luciferase assays were performed in FuCHO cells, which were maintained in T75 culture flasks at 37ºC, 5% CO2, and 90-98% humidity in a culture medium consisting of Ham's F12, 10% FBS, 2mM L-glutamine, 200ug/mL Hygromycin, and 1% Pen-Strep. 10^4 cells were harvested and seeded into 96-well white culture plates. 24 hours later, cells were transfected using Lipofectamine LTX (Invitrogen) and either 100ug of the pGL4.14[*luc2*/Hygro] or pGL4.26[*luc2*/minP/Hygro] vectors and 1ng of the pGL4.74[hRluc/TK] *Renilla* control reporter vector according the standard protocol, with 0.5 mL/well of Lipofectamine LTX reagent and 0.1 mL/well of PLUS reagent. Luminescence was quantified 48 hours after transfection using a Dual-Luciferase Reporter Assay System (Promega) in a GloMax-Multi+ Reader (Promega). For all experiments, firefly luminescence was standardized to *Renilla* luminescence to control for differences in transfection efficiency across samples, and firefly/*Renilla* luminescence values of firefly luciferase vectors with tiling inserts were standardized to firefly/*Renilla* luminescence values of firefly luciferase vectors without tiling inserts.

CDKN2CRTG-eGFP construct synthesis and flow cytometry methods

We synthesized the Bowhead whale *CDKN2CRTG* protein with mouse codon usage (GeneScript) and directly cloned the synthesized gene into the pcDNA3.1-C-eGFP vector (Life Technologies), which added a C-terminal eGFP tag. Cell cycle analyses were performed in FuCHO cells, which were maintained in T75 culture flasks as described. FUCHO cells were transfected with either the CDKN2CRTG-eGFP transgene plasmid, a GFP-control plasmid, or a DNA-free mock transfection. Cells were grown for 48 hours as described in the prior section. Furthermore, FLP-IN CHO cells were transfected with either a mock solution or the eGFP plasmid as a no-color and GFP controls. Cells were harvested using 0.05% Trypsin-EDTA and washed three times with DPBS. Cells were resuspended at a concentration of 1 million cells/mL in DPBS, and then fixed in 1% paraformadehyde for 15 minutes at room temperature. Cells were then washed twice with DPBS and resuspended at a concentration of 1-2 million cells/mL in 10% BSA in DPBS for flow cytometry. Cells were analyzed using a LSR Fortessa (BD Biosciences) configured with 488nm, 640nm, 561nm and 405nm lasers using FCS Express. Due to the high transfection efficiency, the eGFP sample was spiked with 10% mock transfected FLP-IN CHO control as an internal negative control. Voltages, gates, and compensation for eGFP (520/50), mCherry (610/20), and mVenus (520/50) were calibrated using the eGFP-only and mock transfected FuCHO cells. 10,000 events were captured for each sample after calibration. All analyses were done using the FACS Diva software.

ApoTox-Glo Triplex Assay

ApoTox-Glo Triplex Assays (Promega) were performed in FuCHO cells, which were maintained in T75 culture flasks at 37ºC, 5% CO2, and 90-90% humidity in a culture medium consisting of Ham's F12, 10% FBS, 2mM L-glutamine, and 1% Pen-Strep. 5×10^4 cells were harvested and seeded into 96-well white culture plates and reverse-transfected with 100ng of the CDKN2CRTG-eGFP reporter vector using Lipofectamine LTX (Invitrogen) according the standard protocol, with 0.5 mL/well of Lipofectamine LTX reagent and 0.1 mL/well of PLUS reagent. 24 hours after transfection cells were treated with 0, 0.5, 1, 5, 10, or 50uM doxorubicin for 6 or 18 hours. The ApoTox-Glo Triplex Assay (Promega) was performed using the default ApoTox-Glo Triplex Assay in a GloMax-Multi+ Reader (Promega).

Mammalian population doubling time estimation

We gathered population doubling times from previously published studies (Burkard et al., 2019, 2015; Carvan et al., 1994; Chen et al., 2009; Gomes et al., 2011; Mancia et al., 2012; Rajput et al., 2018; Seluanov et al., 2008; Wang et al., 2011, 2021; Wise et al., 2011, 2008; Yajing et al., 2018). Some studies did not report doubling times, but showed graphs of doubling times. For these datasets, doubling times were extracted from figures using WebPlotDigitizer Version 4.5 (Rohatgi, 2021).

Bowhead whale cell culture and doubling time estimation

Bowhead whale primary kidney fibroblast cells were maintained in T75 culture flasks at 37°C/5% CO2 in a culture medium consisting of fibroblast growth medium (FGM)/Eagle's minimum essential medium (EMEM) (1:1) supplemented with insulin, fibroblast growth factor (FGF), 6% FBS, and gentamicin/amphotericin B (FGM-2; singlequots; Clonetics/Lonza). At 80% confluency, cells were harvested and seeded into six-well culture plates at 10,000 cells/well. Once cells recovered to 80% confluency, they were harvested by trypsinization. Cells were manually counted using trypan blue staining and a hemocytometer.

Cancer data collection from previous studies

Published cancer prevalence data for *Delphinapterus leucas (Martineau et al., 2002, 1988)*, *Globicephala melaena (Albuquerque et al., 2018)*, *Tursiops truncatus (Albuquerque et al., 2018)*, *Phocoena phocoena (Albuquerque et al., 2018)*, *Balaena mysticetus (R. et al., 2020; Stimmelmayr et al., 2017)*, *Eubalaena glacialis (Sharp et al., 2019)*, *Ziphius cavirostris* (Rogan et al., 1997) was collected from published surveys of morbidity and mortality in these species. In addition, we also collected data from several studies reporting cancer prevalence aggregated across multiple species (Díaz-Delgado et al., 2018; Geraci et al., 1987; Uys and Best, 1966). Neoplasia prevalence in these Cetaceans were compared to to other Therian mammals using data from two published studies that included pathology reports from 37 (Boddy et al., 2020) and 191 (Vincze et al., 2022) species of Therian mammals, as well as several species of Xenarthrans we also collected from the literature, including giant anteater (*Myrmecophaga tridactyla*), southern tamandua (*Tamandua tetradactyla*), silky anteater (*Cyclopes* sp.), maned three-toed sloths (*Bradypus torquatus*), brown-throated sloths (*Bradypus variegatus*), pale-throated sloth (*Bradypus tridactylus*), Linné's two-toed sloth (*Choloepus didactylus*), three-toed sloths

(*Bradypus sp.*), and two-toed slots (*Choloepus sp.*) (Arenales et al., 2020a, 2020b; Diniz and Oliveira, 1999; Diniz et al., 1995); The total dataset includes cancer prevalence data from 221 species. Neoplasia reports from *Dasypus novemcinctus* from Boddy et al. were combined with cancer prevalence data from the NHDP colony (Vazquez et al., 2022) and *Choloepus didactylus* data from Vincze et al were combined with *Choloepus didactylus* data from (Arenales et al., 2020b). Confidence intervals (95%) on lifetime neoplasia prevalence were estimated in PropCI package (https://github.com/shearer/PropCIs) in R using the Clopper-Pearson exact CI function: exactci(x, n, conf.level=0.95), where x is the number of successes (necropsies with neoplasia), n is the total sample size (number of necropsies), and conf.level=0.95 is the 95% lower and upper confidence intervals.

Data analysis

All data analysis was performed using Python version 3.8 and R version 4.0.2 (2020-06-22). Scripts for performing the reciprocal best hit BLAT and ECNC correction were described in Vazquez and Lynch (2021, https://github.com/docmanny/atlantogenataGeneDuplication) copy archived at swh:1:rev:6bc68ac31ef148131480710e50b0b75d06077db2. All files necessary to reproduce the data in this manuscript are provided in Source data 1.

Cell lines

Flp-In-CHO Chinese Hamster ovary cells, which are derived from the CHO-K1 cell line (ATCC CCL-61), were purchased from ThermoFisher (invitrogen R758-07); their identity has been authenticated by Invitrogen. Bowhead whale primary kidney fibroblast cells were provided by the T.C. Hsu Cryo-Zoo at the University of Texas MD Anderson Cancer Center. Both cell lines were determined to be mycoplasma free.

Data availability

All genome data analyzed during this study are publicly available (**Figure 1 - source data 1**) all other data generated or analyzed in this study are included in the manuscript and supporting files. A supplemental R markdown file is provided that includes raw data and scripts to reproduce the luciferase assay data shown in Figure 3, Figure 3 – figure supplement 1, and Figure 3 – figure supplement 2, the flow data shown in Figure 4, and the ApoTox-Glo data shown Figure 5.

Acknowledgements

Competing interests

The Authors have no conflicts of interest to report.

Funding Source

References

Abegglen LM, Caulin AF, Chan A, Lee K, Robinson R, Campbell MS, Kiso WK, Schmitt DL, Waddell PJ, Bhaskara S, Jensen ST, Maley CC, Schiffman JD. 2015. Potential Mechanisms for *Cancer Res*istance in Elephants and Comparative *Cell*ular Response to DNA Damage in Humans. *JAMA* 314:1850–1860.

Aktipis CA, Boddy AM, Gatenby RA, Brown JS, Maley CC. 2013. Life history trade-offs in cancer evolution. *Nat Rev Cancer* 13:883–892. doi:10.1038/nrc3606

Albuquerque TAF, Val LD do, Doherty A, Magalhães JP de. 2018. From humans to hydra: patterns of cancer across the tree of life. *Biol Rev* 93:1715–1734. doi:10.1111/brv.12415

Al-Mohanna MA, Manogaran PS, Al-Mukhalafi Z, Al-Hussein KA, Aboussekhra A. 2004. The tumor suppressor p16INK4a gene is a regulator of apoptosis induced by ultraviolet light and cisplatin. *Oncogene* 23:201–212. doi:10.1038/sj.onc.1206927

Anisimov VN, Ukraintseva SV, Yashin AI. 2005. Cancer in rodents: does it tell us about cancer in humans? Nat Rev Cancer 5:807–819. doi:10.1038/nrc1715

Arenales A, Gardiner CH, Miranda FR, Dutra KS, Oliveira AR, Mol JP, Costa MET da, Tinoco HP, Coelho CM, Silva RO, Pinto HA, Hoppe EG, Werther K, Santos RL. 2020a. Pathology of Free-Ranging and Captive Brazilian Anteaters. *J Comp Pathol* 180:55–68. doi:10.1016/j.jcpa.2020.08.007

Arenales A, Silva FL, Miranda F, Guedes PEB, Werther K, Costa MELT da, Tinoco HP, Coelho CM, Santos RL. 2020b. Pathologic findings IN 36 sloths from Brazil. *J Zoo Wildlife Med* 51:672–677. doi:10.1638/2020-0002

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

Bartek J, Lukas J. 2001. Mammalian G1- and S-phase checkpoints in response to DNA damage. *Curr Opin Cell Biol* 13:738–747. doi:10.1016/s0955-0674(00)00280-5

Bench GS, Friz AM, Corzett MH, Morse DH, Balhorn R. 1996. DNA and total protamine masses in individual sperm from fertile mammalian subjects. *Cytometry* 23:263–271. doi:10.1002/(sici)1097-0320(19960401)23:4<263::aid-cyto1>3.0.co;2-i

Blais A, Labrie Y, Pouliot F, Lachance Y, Labrie C. 1998. Structure of the Gene Encoding the Human Cyclin-Dependent Kinase Inhibitor p18 and Mutational Analysis in Breast Cancer. *Biochem Bioph Res Co* 247:146–153. doi:10.1006/bbrc.1998.8497

Blank HM, Callahan M, Pistikopoulos IPE, Polymenis AO, Polymenis M. 2018. Scaling of G1

25

Duration with Population Doubling Time by a Cyclin in Saccharomyces cerevisiae. *Genetics* 210:895–906. doi:10.1534/genetics.118.301507

Boddy AM, Abegglen LM, Pessier AP, Schiffman JD, Maley CC, Witte C. 2020. {Lifetime cancer prevalence and life history traits in mammals}. *Evol Medicine Public Heal*. doi:10.1093/emph/eoaa015

Boddy AM, Kokko H, Breden F, Wilkinson GS, Aktipis CA. 2015. Cancer susceptibility and reproductive trade-offs: a model of the evolution of cancer defences. *Philosophical Transactions Royal Soc B Biological Sci* 370:20140220. doi:10.1098/rstb.2014.0220

Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D. 2008. Coordination of Growth Rate, *Cell Cycle*, Stress Response, and Metabolic Activity in Yeast. *Mol Biol Cell* 19:352–367. doi:10.1091/mbc.e07-08-0779

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527. doi:10.1038/nbt.3519

Browning CL, Wise CF, Wise JP. 2017. Prolonged particulate chromate exposure does not inhibit homologous recombination repair in North Atlantic right whale (Eubalaena glacialis) lung cells. *Toxicol Appl Pharm* 331:18–23. doi:10.1016/j.taap.2017.04.006

Buchold GM, Magyar PL, Arumugam R, Lee MM, O'Brien DA. 2007. p19Ink4d and p18Ink4c cyclin-dependent kinase inhibitors in the male reproductive axis. *Mol Reprod Dev* 74:997–1007. doi:10.1002/mrd.20716

Bulls S, Platner L, Ayub W, Moreno N, Arditi J-P, McCain S, Wagner P, Sonsbeek LGRB- van, Lynch VJ, Claude J, Glaberman S, Chiari Y. 2022. Cancer prevalence is remarkably low in turtles and is correlated with life history traits in birds, mammals, and squamates. doi:10.1101/2022.07.12.499088

Burkard M, Nash SB, 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 Biol Toxicol* 35:387–398. doi:10.1007/s10565-018-09457-1

Burkard M, Whitworth D, Schirmer K, Nash SB. 2015. Establishment of the first humpback whale fibroblast cell lines and their application in chemical risk assessment. *Aquat Toxicol* 167:240–247. doi:10.1016/j.aquatox.2015.08.005

Burrows CK, Kosova G, Herman C, Patterson K, Hartmann KE, Edwards DRV, Stephenson MD, Lynch VJ, Ober C. 2016. Expression quantitative trait locus mapping studies in mid-

secretory phase endometrial cells identifies HLA-F and TAP2 as fecundability-associated genes. *PLoS genetics* 12:e1005858.

Cánepa ET, Scassa ME, Ceruti JM, Marazita MC, Carcagno AL, Sirkin PF, Ogara MF. 2007. INK4 proteins, a family of mammalian CDK inhibitors with novel biological functions. *Iubmb Life* 59:419–426. doi:10.1080/15216540701488358

Carvan MJ, SANTOSTEFANO M, Safe S, Busbee D. 1994. Characterization of a bottlenose dolphin (tursiops truncatus) kidney epithelial cell line. *Mar Mammal Sci* 10:52–69. doi:10.1111/j.1748-7692.1994.tb00389.x

Caulin AF, Graham TA, Wang L-S, Maley CC. 2015. Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 370.

Caulin AF, Maley CC. 2011. Peto's Paradox: evolution's prescription for cancer prevention. *Trends Ecol Evol* 26:175–182. doi:10.1016/j.tree.2011.01.002

Ceruti JM, Scassa ME, Fló JM, Varone CL, Cánepa ET. 2005. Induction of p19INK4d in response to ultraviolet light improves DNA repair and confers resistance to apoptosis in neuroblastoma cells. Oncogene 24:4065–4080. doi:10.1038/sj.onc.1208570

Charnov EL, Ernest SKM. 2006. The Offspring-Size/Clutch-Size Trade-Off in Mammals. *Am Nat* 167:578–582. doi:10.1086/501141

Chen TL, Wise SS, Kraus S, Shaffiey F, Levine KM, Thompson WD, Romano T, O'Hara T, Wise JP. 2009. Particulate hexavalent chromium is cytotoxic and genotoxic to the North Atlantic right whale (Eubalaena glacialis) lung and skin fibroblasts. *Environ Mol Mutagen* 50:387–393. doi:10.1002/em.20471

Cui H, Zhao C, Gong P, Wang Ling, Wu H, Zhang K, Zhou R, Wang Li, Zhang T, Zhong S, Fan H. 2015. DNA methyltransferase 3A promotes cell proliferation by silencing CDK inhibitor p18INK4C in gastric carcinogenesis. *Sci Rep-uk* 5:13781. doi:10.1038/srep13781

Díaz-Delgado J, Fernández A, Sierra E, Sacchini S, Andrada M, Vela AI, Quesada-Canales Ó, Paz Y, Zucca D, Groch K, Arbelo M. 2018. Pathologic findings and causes of death of stranded cetaceans in the Canary Islands (2006-2012). *Plos One* 13:e0204444. doi:10.1371/journal.pone.0204444

Diniz LS, Oliveira PM. 1999. Clinical problems of sloths (Bradypus sp. and Choloepus sp.) in captivity. *J Zoo Wildl Medicine Official Publ Am Assoc Zoo Vet* 30:76–80.

Diniz LSM, Costa EO, Oliveira PMA. 1995. Clinical disorders observed in anteaters

(Myrmecophagidae, Edentata) in captivity. *Vet Res Commun* 19:409–415. doi:10.1007/bf01839320

Doll R. 1971. The age distribution of cancer: implications for models of carcinogenesis. *Journal of the Royal Statistical Society Series A (łdots*.

Dong P, Zhang C, Parker B-T, You L, Mathey-Prevot B. 2018. Cyclin D/CDK4/6 activity controls G1 length in mammalian cells. Plos One 13:e0185637. doi:10.1371/journal.pone.0185637

Dutta A, Henley W, Robine J-M, Langa KM, Wallace RB, Melzer D. 2013. Longer Lived Parents: Protective Associations With Cancer Incidence and Overall Mortality. *Journals Gerontology Ser* 68:1409–1418. doi:10.1093/gerona/glt061

Ferreira-Junior JR, Digiampietri LA. 2019. OVERFRAG: an overlapping DNA fragments generator for molecular cloning and synthetic biology. *Data Brief* 23:103806. doi:10.1016/j.dib.2019.103806

Fisher RP. 2016. Getting to S: CDK functions and targets on the path to cell-cycle commitment. *F1000research* 5:2374. doi:10.12688/f1000research.9463.1

Franklin DS, Godfrey VL, Lee H, Kovalev GI, Schoonhoven R, Chen-Kiang S, Su L, Xiong Y. 1998. CDK inhibitors p18INK4c and p27Kip1 mediate two separate pathways to collaboratively suppress pituitary tumorigenesis. *Gene Dev* 12:2899–2911. doi:10.1101/gad.12.18.2899

Franklin DS, Godfrey VL, O'Brien DA, Deng C, Xiong Y. 2000. Functional Collaboration between Different Cyclin-Dependent Kinase Inhibitors Suppresses Tumor Growth with Distinct Tissue Specificity. *Mol Cell Biol* 20:6147–6158. doi:10.1128/mcb.20.16.6147-6158.2000

Franklin DS, Xiong Y. 1996. Induction of p18INK4c and its predominant association with CDK4 and CDK6 during myogenic differentiation. Mol Biol Cell 7:1587–1599. doi:10.1091/mbc.7.10.1587

Gao Y, Yang P, Shen H, Yu H, Song X, Zhang L, Zhang P, Cheng Haizi, Xie Z, Hao S, Dong F, Ma S, Ji Q, Bartlow P, Ding Y, Wang L, Liu H, Li Y, Cheng Hui, Miao W, Yuan W, Yuan Y, Cheng T, Xie X-Q. 2015. Small-molecule inhibitors targeting INK4 protein p18(INK4C) enhance ex vivo expansion of haematopoietic stem cells. *Nature communications* 6:6328.

Garcia-Cao I, García-Cao M, Marťın-Caballero J, Criado LM, Klatt P, Flores JM, Weill J-C, Blasco MA, Serrano M. 2002. "Super p53" mice exhibit enhanced DNA damage response, are tumor resistant and age normally. *EMBO Journal* 21:6225–6235.

Garland T. 2014. Trade-offs. *Curr Biol* 24:R60–R61. doi:10.1016/j.cub.2013.11.036

George JC, Bada J, Zeh J, Scott L. 1999. Age and growth estimates of bowhead whales

(Balaena mysticetus) via aspartic acid racemization. *Journal of Zoology*.

George JC, Follmann E, Zeh J. 2011. A new way to estimate the age of bowhead whales (Balaena mysticetus) using ovarian corpora counts. Journal of Zoology.

Geraci JR, Palmer NC, Aubin DJSt. 1987. Tumors in Cetaceans: Analysis and New Findings. *Can J Fish Aquat Sci* 44:1289–1300. doi:10.1139/f87-152

Glaberman S, Bulls SE, Vazquez JM, Chiari Y, Lynch VJ. 2021. Concurrent evolution of anti-aging gene duplications and cellular phenotypes in long-lived turtles. *Genome Biol Evol* 13:evab244-. doi:10.1093/gbe/evab244

Gomes NMV, Ryder OA, Houck ML, Charter SJ, Walker W, Forsyth NR, Austad SN, Venditti C, Pagel M, Shay JW, Wright WE. 2011. Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging cell* 10:761–768.

Gorbunova V, Hine C, Tian X, Ablaeva J, Gudkov AV, Nevo E, Seluanov A. 2012. Cancer resistance in the blind mole rat is mediated by concerted necrotic cell death mechanism. *Proc National Acad Sci*, Proceedings of the National Academy of *Science*s of the United States of America 109:19392–19396. doi:10.1073/pnas.1217211109

Grubbs EG, Williams MD, Scheet P, Vattathil S, Perrier ND, Lee JE, Gagel RF, Hai T, Feng L, Cabanillas ME, Cote GJ. 2016. Role of CDKN2C Copy Number in Sporadic Medullary *Thyroid* Carcinoma. Thyroid 26:1553–1562. doi:10.1089/thy.2016.0224

Guan KL, Jenkins CW, Li Y, Nichols MA. 1994. Growth suppression by p18, a p16INK4/MTS1- and p14INK4B/MTS2-related CDK6 inhibitor, correlates with wild-type pRb function. *Genes & ….*

Herranz D, Muñoz-Martin M, Cãnamero M, Mulero F, Martinez-Pastor B, Fernandez-Capetillo O, Serrano M. 2010. Sirt1 improves healthy ageing and protects from metabolic syndrome-associated cancer. Nature communications 1:3.

Hirai H, Roussel MF, Kato JY, Ashmun RA, Sherr CJ. 1995. Novel INK4 proteins, p19 and p18, are specific inhibitors of the cyclin D-dependent kinases CDK4 and CDK6. Mol Cell Biol 15:2672–2681.

Jarrell GH. 1979. Karyotype of the Bowhead Whale (Balaena mysticetus). *J Mammal* 60:607–610. doi:10.2307/1380102

Johnston GC, Pringle JR, Hartwell LH. 1977. Coordination of growth with cell division in the yeast Saccharomyces cerevisiae. *Exp Cell Res* 105:79–98. doi:10.1016/0014-4827(77)90154-9

Katzourakis A, Magiorkinis G, Lim AG, Gupta S, Belshaw R, Gifford R. 2014. Larger mammalian body size leads to lower retroviral activity. *Plos Pathog* 10:e1004214. doi:10.1371/journal.ppat.1004214

Keane M, Semeiks J, Webb AE, Li YI, Quesada V, Craig T, Madsen LB, Dam S van, Brawand D, Marques PI, Michalak P, Kang L, Bhak J, Yim H-S, Grishin NV, Nielsen NH, Heide-Jørgensen MP, Oziolor EM, Matson CW, Church GM, Stuart GW, Patton JC, George JC, Suydam R, Larsen K, López-Otín C, O'Connell MJ, Bickham JW, Thomsen B, Magalhães JP de. 2015. Insights into the evolution of longevity from the bowhead whale genome. *Cell Reports* 10:112–122. doi:10.1016/j.celrep.2014.12.008

Komata T, Kanzawa T, Takeuchi H, Germano IM, Schreiber M, Kondo Y, Kondo S. 2003a. Antitumour effect of cyclin-dependent kinase inhibitors (p16(INK4A), p18(INK4C), p19(INK4D), p21(WAF1/CIP1) and p27(KIP1)) on malignant glioma cells. British Journal of Cancer 88:1277–1280.

Komata T, Kanzawa T, Takeuchi H, Germano IM, Schreiber M, Kondo Y, Kondo S. 2003b. Antitumour effect of cyclin-dependent kinase inhibitors (p16(INK4A), p18(INK4C), p19(INK4D), p21(WAF1/CIP1) and p27(KIP1)) on malignant glioma cells. *Brit J Cancer* 88:1277–1280. doi:10.1038/sj.bjc.6600862

Kulkarni MS, Daggett JL, Bender TP, Kuehl WM, Bergsagel PL, Williams ME. 2002. Frequent inactivation of the cyclin-dependent kinase inhibitor p18 by homozygous deletion in multiple myeloma cell lines: ectopic p18 expression inhibits growth and induces apoptosis. *Leukemia* 16:127–134. doi:10.1038/sj.leu.2402328

Kurosu H, Yamamoto M, Clark JD, Pastor JV, Nandi A, Gurnani P, McGuinness OP, Chikuda H, Yamaguchi M, Kawaguchi H, Shimomura I, Takayama Y, Herz J, Kahn CR, Rosenblatt KP, Kuro-o M. 2005. Suppression of aging in mice by the hormone Klotho. Science 309:1829–1833.

Lavie L, Maldener E, Brouha B, Meese EU, Mayer J. 2004. The human L1 promoter: Variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res* 14:2253–2260. doi:10.1101/gr.2745804

Le H-V, Minn AJ, Massagué J. 2005. Cyclin-dependent Kinase Inhibitors Uncouple Cell Cycle Progression from Mitochondrial Apoptotic Functions in DNA-damaged Cancer Cells*. *J Biol Chem* 280:32018–32025. doi:10.1074/jbc.m504689200

Leroi AM, Koufopanou V, Burt A. 2003. Cancer selection. Nat Rev Cancer 3:226–231. doi:10.1038/nrc1016

Lim J, Kim Jungeun, Duong T, Lee G, Kim Junghee, Yoon J, Kim Jaetaek, Kim H, Ruley HE, El-Rifai W, Jo D. 2012. Antitumor Activity of Cell-Permeable p18(INK4c) With Enhanced Membrane and Tissue Penetration. *Mol Ther* 20:1540–1549. doi:10.1038/mt.2012.102

Lindström HJG, Friedman R. 2020. Inferring time-dependent population growth rates in cell cultures undergoing adaptation. *Bmc Bioinformatics* 21:583. doi:10.1186/s12859-020-03887-7

Liu JZ, Chen C-Y, Tsai EA, Whelan CD, Sexton D, John S, Runz H. 2022. The burden of rare protein-truncating genetic variants on human lifespan. *Nat Aging* 1–6. doi:10.1038/s43587-022-00182-3

Lubischer JL. 2007. The Cell Cycle, Principles of Control. David O. Morgan. *Integr Comp Biol* 47:794–795. doi:10.1093/icb/icm066

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155.

Mancia A, Spyropoulos DD, McFee WE, Newton DA, Baatz JE. 2012. Cryopreservation and in vitro culture of primary cell types from lung tissue of a stranded pygmy sperm whale (Kogia breviceps). *Comp Biochem Physiology Part C Toxicol Pharmacol* 155:136–142. doi:10.1016/j.cbpc.2011.04.002

Martineau D, Lagacé A, Béland P, Higgins R, Armstrong D, Shugart LR. 1988. Pathology of stranded beluga whales (Delphinapterus leucas) from the St. Lawrence Estuary, Québec, Canada. J Comp Pathol 98:287–310. doi:10.1016/0021-9975(88)90038-2

Martineau D, Lemberger K, Dallaire A, Labelle P, Lipscomb TP, Michel P, Mikaelian I. 2002. Cancer in wildlife, a case study: beluga from the St. Lawrence estuary, Québec, Canada. *Environmental health perspectives* 110:285–292.

Matheu A, Maraver A, Klatt P, Flores I, Garcia-Cao I, Borras C, Flores JM, Viña J, Blasco MA, Serrano M. 2007. Delayed ageing through damage protection by the Arf/p53 pathway. Nature 448:375–379. doi:10.1038/nature05949

Matheu A, Pantoja C, Efeyan A, Criado LM, Martín-Caballero J, Flores JM, Klatt P, Serrano M. 2004. Increased gene dosage of Ink4a/Arf results in cancer resistance and normal aging. *Genes and Development* 18:2736–2746.

Mathieson I, Day FR, Barban N, Tropf FC, Brazel DM, Consortium B, Vaez A, Zuydam N van, Bitarello BD, Snieder H, Hoed M den, Ong KK, Mills MC, Perry JRB. 2020. Genome-wide analysis identifies genetic effects on reproductive success and ongoing natural selection at the FADS locus. *Biorxiv* 2020.05.19.104455. doi:10.1101/2020.05.19.104455

31

McClendon AK, Dean JL, Rivadeneira DB, Yu JE, Reed CA, Gao E, Farber JL, Force T, Koch WJ, Knudsen ES. 2012. CDK4/6 inhibition antagonizes the cytotoxic response to anthracycline therapy. Cell Cycle 11:2747–55. doi:10.4161/cc.21127

Meaza I, Speer RM, Toyoda JH, Lu H, Wise SS, Croom-Perez TJ, Aboueissa AE-M, Wise JP. 2020. Prolonged exposure to particulate Cr(VI) is cytotoxic and genotoxic to fin whale cells. *J Trace Elem Med Bio* 62:126562. doi:10.1016/j.jtemb.2020.126562

Migaki G, Albert TF. 1982. Lipoma of the Liver in a Bowhead Whale (Balaena mysticetus). *Vet Pathol* 19:329–331. doi:10.1177/030098588201900314

Mika KM, Li X, DeMayo FJ, Lynch VJ. 2018. An Ancient Fecundability-Associated Polymorphism Creates a GATA2 Binding Site in a Distal Enhancer of HLA-F. *Am J Hum Genetics* 103:509–521. doi:10.1016/j.ajhg.2018.08.009

Mika KM, Lynch VJ. 2016. An Ancient Fecundability-Associated Polymorphism Switches a Repressor into an Enhancer of Endometrial TAP2 Expression. Am J Hum Genetics 99:1059–1071. doi:10.1016/j.ajhg.2016.09.002

Mort RL, Ford MJ, Sakaue-Sawano A, Lindstrom NO, Casadio A, Douglas AT, Keighren MA, Hohenstein P, Miyawaki A, Jackson IJ. 2014. Fucci2a: A bicistronic cell cycle reporter that allows Cre mediated tissue specific expression in mice. Cell Cycle 13:2681–2696. doi:10.4161/15384101.2015.945381

Nagy JD, Victor EM, Cropper JH. 2007. Why don't all whales have cancer? A novel hypothesis resolving Peto's paradox. Integr Comp Biol 47:317–328.

Nielsen J, Hedeholm RB, Heinemeier J, Bushnell PG, Christiansen JS, Olsen J, Ramsey CB, Brill RW, Simon M, Steffensen KF, Steffensen JF. 2016. Eye lens radiocarbon reveals centuries of longevity in the Greenland shark (Somniosus microcephalus). Science 353:702–704.

Noh SJ, Li Y, Xiong Y, Guan KL. 1999. Identification of functional elements of p18INK4C essential for binding and inhibition of cyclin-dependent kinase (CDK) 4 and CDK6. Cancer Res 59:558–64.

O'Connell MJ, Cimprich KA. 2004. G2 damage checkpoints: what is the turn-on? *J Cell Sci* 118:1–6. doi:10.1242/jcs.01626

O'Connor PM. 1997. Mammalian G1 and G2 phase checkpoints. *Cancer Surv* 29:151–82.

O'Hara TM, George JC, Tarpley RJ, Burek K. 2002. Sexual maturation in male bowhead whales (Balaena mysticetus) of the Bering-Chukchi-Beaufort Seas stock. *Journal of Cetacean łdots*.

Palacios JA, Herranz D, Bonis MLD, Velasco S, Serrano M, Blasco MA. 2010. SIRT1

contributes to telomere maintenance and augments global homologous recombination. *The Journal of cell biology* 191:1299–1313.

Pan D, Zhang L. 2007. Quantifying the major mechanisms of recent gene duplications in the human and mouse genomes: a novel strategy to estimate gene duplication rates. Genome Biol 8:R158. doi:10.1186/gb-2007-8-8-r158

Pereira MJ, VRANIC M, KAMBLE PG, JERNOW H, KRISTÓFI R, HOLBIKOVA E, SKRTIC S, KULLBERG J, SVENSSON MK, HETTY S, ERIKSSON JW. 2022. CDKN2C expression in adipose tissue is reduced in type II diabetes and central obesity: impact on adipocyte differentiation and lipid storage? *Transl Res* 242:105–121. doi:10.1016/j.trsl.2021.12.003

Peto FJRPNLLLJCR. 1975. Cancer and ageing in mice and men. British Journal of Cancer 32:411.

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 of London Series B, Biological sciences 370:20150198.

Phelps DE, Hsiao K-M, Li Y, Hu N, Franklin DS, Westphal E, Lee EY-HP, Xiong Y. 1998. Coupled Transcriptional and Translational Control of Cyclin-Dependent Kinase Inhibitor p18 INK4c Expression during Myogenesis. Mol Cell Biol 18:2334–2343. doi:10.1128/mcb.18.4.2334

Pinkston JM, Garigan D, Hansen M, Kenyon C. 2006. Mutations that increase the life span of C. elegans inhibit tumor growth. Science 313:971–975.

Pommier Y, Leo E, Zhang H, Marchand C. 2010. DNA Topoisomerases and Their Poisoning by Anticancer and Antibacterial Drugs. *Chem Biol* 17:421–433. doi:10.1016/j.chembiol.2010.04.012

Potter AJ, Gollahon KA, Palanca BJA, Harbert MJ, Choi YM, Moskovitz AH, Potter JD, Rabinovitch PS. 2002. Flow cytometric analysis of the cell cycle phase specificity of DNA damage induced by radiation, hydrogen peroxide and doxorubicin. *Carcinogenesis* 23:389–401. doi:10.1093/carcin/23.3.389

R. S, D. R, Gay S, H.K. B, J.C. G. 2020. Diseases and parasites 471–498.

Rajput IR, Xiao Z, Yajing S, Yaqoob S, Sanganyado E, Ying H, Fei Y, Liu W. 2018. Establishment of pantropic spotted dolphin (Stenella attenuata) fibroblast cell line and potential influence of polybrominated diphenyl ethers (PBDEs) on cytokines response. Aquat Toxicol 203:1–9. doi:10.1016/j.aquatox.2018.07.017

Robitaille K, Rourke JL, McBane JE, Fu A, Baird S, Du Q, Kin T, Shapiro AMJame, Screaton RA. 2016. High-throughput Functional Genomics Identifies Regulators of Primary Human Beta

Cell Proliferation*. J Biol Chem 291:4614–4625. doi:10.1074/jbc.m115.683912

Rogan E, Baker JR, Jepson PD, Berrow S, Kiely O. 1997. A mass stranding of white-sided dolphins (Lagenorhynchus acutus) in Ireland: biological and pathological studies. *J Zool* 242:217–227. doi:10.1111/j.1469-7998.1997.tb05798.x

Rohatgi A. 2021. Webplotdigitizer: Version 4.5. https://automeris.io/WebPlotDigitizer

Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, Osawa H, Kashiwagi S, Fukami K, Miyata T, Miyoshi H, Imamura T, Ogawa M, Masai H, Miyawaki A. 2008. Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. Cell 132:487–498. doi:10.1016/j.cell.2007.12.033

Scassa ME, Marazita MC, Ceruti JM, Carcagno AL, Sirkin PF, González-Cid M, Pignataro OP, Cánepa ET. 2007. Cell cycle inhibitor, p19INK4d, promotes cell survival and decreases chromosomal aberrations after genotoxic insult due to enhanced DNA repair. *Dna Repair* 6:626–638. doi:10.1016/j.dnarep.2006.12.003

Schreiber M, Muller WJ, Singh G, Graham FL. 1999. Comparison of the effectiveness of adenovirus vectors expressing cyclin kinase inhibitors p16INK4A, p18INK4C, p19INK4D, p21WAF1/CIP1 and p27KIP1 in inducing cell cycle arrest, apoptosis and inhibition of tumorigenicity. Oncogene 18:1663–1676. doi:10.1038/sj.onc.1202466

Seluanov A, Hine C, Bozzella M, Hall A, Sasahara THC, Ribeiro AACM, Catania KC, Presgraves DC, Gorbunova V. 2008. Distinct tumor suppressor mechanisms evolve in rodent species that differ in size and lifespan. Aging cell 7:813–823.

Sharp S, McLellan W, Rotstein D, Costidis A, Barco S, Durham K, Pitchford T, Jackson K, Daoust P, Wimmer T, Couture E, Bourque L, Frasier T, Frasier B, Fauquier D, Rowles T, Hamilton P, Pettis H, Moore M. 2019. Gross and histopathologic diagnoses from North Atlantic right whale Eubalaena glacialis mortalities between 2003 and 2018. *Dis Aquat Organ* 135:1–31. doi:10.3354/dao03376

Smith AW, Skilling DE, Benirschke K, Albert TF, Barlough JE. 1987. SEROLOGY AND VIROLOGY OF THE BOWHEAD WHALE (BALAENA MYSTICETUS L.). *J Wildlife Dis* 23:92–98. doi:10.7589/0090-3558-23.1.92

Smith KR, Hanson HA, Mineau GP, Buys SS. 2012. Effects of BRCA1 and BRCA2 mutations on female fertility. *Proc Royal Soc B Biological Sci* 279:1389–1395. doi:10.1098/rspb.2011.1697

Solomon DA, Kim J-S, Jenkins S, Ressom H, Huang M, Coppa N, Mabanta L, Bigner D, Yan H, Jean W, Waldman T. 2008. Identification of p18 INK4c as a tumor suppressor gene in

glioblastoma multiforme. Cancer Res 68:2564–9. doi:10.1158/0008-5472.can-07-6388

Stearns SC. 1989. Trade-Offs in Life-History Evolution. *Funct Ecol* 3:259. doi:10.2307/2389364

Stimmelmayr R, Rotstein D, Seguel M, Gottdenker N. 2017. Hepatic lipomas and myelolipomas in subsistence-harvested bowhead whales Balaena mysticetus, Alaska (USA): a case review 1980-2016. Dis Aquat Organ 127:71–74. doi:10.3354/dao03186

Sulak M, Fong L, Mika K, Chigurupati S, Yon L, Mongan NP, Emes RD, Lynch VJ. 2016a. TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *eLife* 5.

Sulak M, Fong L, Mika K, Chigurupati S, Yon L, Mongan NP, Emes RD, Lynch VJ. 2016b. TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. eLife 5.

Swergold GD. 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol 10:6718–6729. doi:10.1128/mcb.10.12.6718

Tacar O, Sriamornsak P, Dass CR. 2013. Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. *J Pharm Pharmacol* 65:157–170. doi:10.1111/j.2042-7158.2012.01567.x

Tacutu R, Craig T, Budovsky A, Wuttke D, Lehmann G, Taranukha D, Costa J, Fraifeld VE, Magalhães JP de. 2013. Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Research* 41:D1027-33.

Tarpley RJ, Jarrell GH, George JC, Cubbage J, Stott GG. 1995. Male Pseudohermaphroditism in the Bowhead Whale, Balaena mysticetus. J Mammal 76:1267–1275. doi:10.2307/1382622

Tavera-Mendoza L, Wang T, Lallemant B, Zhang R, Nagai Y, Bourdeau V, Ramirez-Calderon M, Desbarats J, Mader S, White JH. 2006. Convergence of vitamin D and retinoic acid signalling at a common hormone response element. *Embo Rep* 7:180–185. doi:10.1038/sj.embor.7400594

Tavera-Mendoza LE, Wang T-T, White JH. 2006. P19INK4D and Cell Death. Cell Cycle 5:596–598. doi:10.4161/cc.5.6.2585

Tejada-Martinez D, Magalhães JP de, Opazo JC. 2021. Positive selection and gene duplications in tumour suppressor genes reveal clues about how cetaceans resist cancer. Proc Royal Soc B 288:20202592. doi:10.1098/rspb.2020.2592

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 X, Firsanov D, Zhang Zhihui, Cheng Y, Luo L, Tombline G, Tan R, Simon M, Henderson S, Steffan J, Goldfarb A, Tam J, Zheng K, Cornwell A, Johnson A, Yang J-N, Mao Z, Manta B, Dang W, Zhang Zhengdong, Vijg J, Wolfe A, Moody K, Kennedy BK, Bohmann D, Gladyshev VN, Seluanov A, Gorbunova V. 2019. SIRT6 Is Responsible for More Efficient DNA Double-Strand Break Repair in Long-Lived Species. Cell 177:622-638.e22. doi:10.1016/j.cell.2019.03.043

Uys CJ, Best PB. 1966. Pathology of lesions observed in whales flensed at Saldanha Bay, South Africa. J Comp Pathol 76:407-IN16. doi:10.1016/0021-9975(66)90061-2

Vazquez JM, Lynch VJ. 2021. Pervasive duplication of tumor suppressors in Afrotherians during the evolution of large bodies and reduced cancer risk. *Elife* 10:e65041. doi:10.7554/elife.65041

Vazquez JM, Pena MT, Muhammad B, Kraft M, Adams LB, Lynch VJ. 2022. Parallel evolution of reduced cancer risk and tumor suppressor duplications in Xenarthra. doi:10.1101/2022.08.04.502824

Vazquez JM, Sulak M, Chigurupati S, Lynch VJ. 2018. A Zombie LIF Gene in Elephants Is Upregulated by TP53 to Induce Apoptosis in Response to DNA Damage. *Cell reports* 24:1765–1776.

Vincze O, Colchero F, Lemaître J-F, Conde DA, Pavard S, Bieuville M, Urrutia AO, Ujvari B, Boddy AM, Maley CC, Thomas F, Giraudeau M. 2022. Cancer risk across mammals. Nature 601:263–267. doi:10.1038/s41586-021-04224-5

Walker RS, Gurven M, Burger O, Hamilton MJ. 2008. The trade-off between number and size of offspring in humans and other primates. Proc Royal Soc B Biological Sci 275:827–834. doi:10.1098/rspb.2007.1511

Wang Jingzhen, Su W, Nie W, Wang Jinhuan, Xiao W, Wang D. 2011. Establishment and characterization of fibroblast cell lines from the skin of the Yangtze finless porpoise. *Vitro Cell Dev Biology Animal* 47:618–30. doi:10.1007/s11626-011-9448-x

Wang T, Li Z, Wei J, Zheng D, Wang C, Xu C, Chen W, Wang B. 2021. Establishment and characterization of fibroblast cultures derived from a female common hippopotamus (Hippopotamus amphibius) skin biopsy. *Cell Biol Int* 45:1571–1578. doi:10.1002/cbin.11596

Westbury MV, Petersen B, Lorenzen ED. 2019. Genomic analyses reveal an absence of contemporary introgressive admixture between fin whales and blue whales, despite known hybrids. Plos One 14:e0222004. doi:10.1371/journal.pone.0222004

36

Wheals AE. 1985. The biology of cell reproduction. *Trends Biochem Sci* 10:371. doi:10.1016/0968-0004(85)90123-9

Wise JP, Wise SS, Holmes AL, LaCerte C, Shaffiey F, Aboueissa A-M. 2010. The cytotoxicity and genotoxicity of hexavalent chromium in Steller sea lion lung fibroblasts compared to human lung fibroblasts. Comp Biochem Physiology Part C Toxicol Pharmacol 152:91–98. doi:10.1016/j.cbpc.2010.03.001

Wise JP, Wise SS, Kraus S, Shaffiey F, Grau M, Chen TL, Perkins C, Thompson WD, Zheng T, Zhang Y, Romano T, O'Hara T. 2008. Hexavalent chromium is cytotoxic and genotoxic to the North Atlantic right whale (Eubalaena glacialis) lung and testes fibroblasts. *Mutat Res Genetic Toxicol Environ Mutagen* 650:30–38. doi:10.1016/j.mrgentox.2007.09.007

Wise JP, Wise SS, LaCerte C, Wise JP, Aboueissa A. 2011. The genotoxicity of particulate and soluble chromate in sperm whale (physeter macrocephalus) skin fibroblasts. Environ Mol Mutagen 52:43–49. doi:10.1002/em.20579

Xie B, Chen J, Liu B, Zhan J. 2013. Klotho acts as a tumor suppressor in cancers. *Pathology & Oncology Research*.

Yajing S, Rajput IR, Ying H, Fei Y, Sanganyado E, Ping L, Jingzhen W, Wenhua L. 2018. Establishment and characterization of pygmy killer whale (Feresa attenuata) dermal fibroblast cell line. Plos One 13:e0195128. doi:10.1371/journal.pone.0195128

Yim H-S, Cho YS, Guang X, Kang SG, Jeong J-Y, Cha S-S, Oh H-M, Lee Jae-Hak, Yang EC, Kwon KK, Kim YJ, Kim TW, Kim W, Jeon JH, Kim S-J, Choi DH, Jho S, Kim H-M, Ko J, Kim H, Shin Y-A, Jung H-J, Zheng Y, Wang Z, Chen Y, Chen M, Jiang A, Li E, Zhang S, Hou H, Kim TH, Yu L, Liu S, Ahn K, Cooper J, Park S-G, Hong CP, Jin W, Kim H-S, Park C, Lee K, Chun S, Morin PA, O'Brien SJ, Lee H, Kimura J, Moon DY, Manica A, Edwards J, Kim BC, Kim S, Wang J, Bhak J, Lee HS, Lee Jung-Hyun. 2014. Minke whale genome and aquatic adaptation in cetaceans. *Nat Genet* 46:88–92. doi:10.1038/ng.2835

Yuan JT, Gatti DM, Philip VM, Kasparek S, Kreuzman AM, Mansky B, Sharif K, Taterra D, Taylor WM, Thomas M, Ward JO, Holmes A, Chesler EJ, Parker CC. 2018. Genome-wide association for testis weight in the diversity outbred mouse population. *Mamm Genome* 29:310–324. doi:10.1007/s00335-018-9745-8

Zindy F, Besten W den, Chen B, Rehg JE, Latres E, Barbacid M, Pollard JW, Sherr CJ, Cohen PE, Roussel MF. 2001. Control of Spermatogenesis in Mice by the Cyclin D-Dependent Kinase Inhibitors p18 Ink4c and p19 Ink4d. Mol Cell Biol 21:3244–3255. doi:10.1128/mcb.21.9.3244-
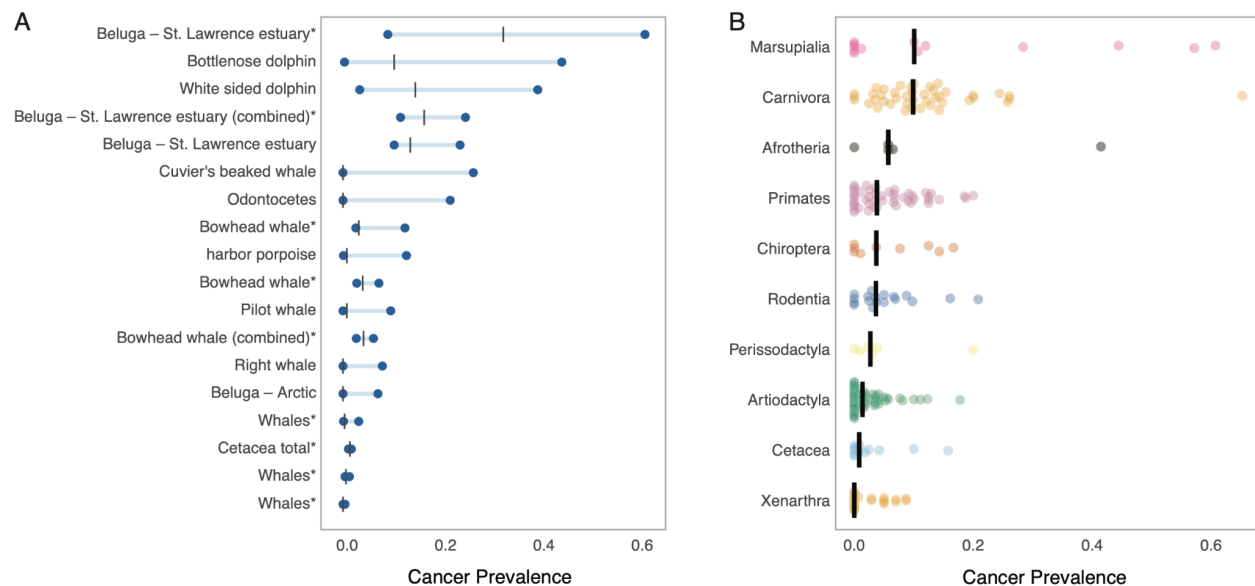
37

3255.2001

Figure 1. Extremely low cancer prevalence in Cetaceans.

(A) Lolliplot plot showing the mean (black bar) and Clopper-Pearson exact 95% confidence intervals of cancer prevalence in each group, asterisks (*) indicate taxa with more than 82 pathology reports. Data from more than two studies are indicated with "(combined)".

(B) Cancer prevalence data for each species (dot) grouped by clade and shown as jittered stripchart. The number of species per clade is: Afrotheria=5, Artiodactyla=68, Carnivora=43, Cetacea=4, Chiroptera=9, Marsupialia=13, Perissodactyla=6, Primates=47, Rodentia=18, and Xenarthra=14. Medians for each clade are shown with vertical black bars.

Figure 1 – source data 1. Cancer prevalence shown in panel A.

Figure 1 – source data 2. Cancer prevalence shown in panel B.
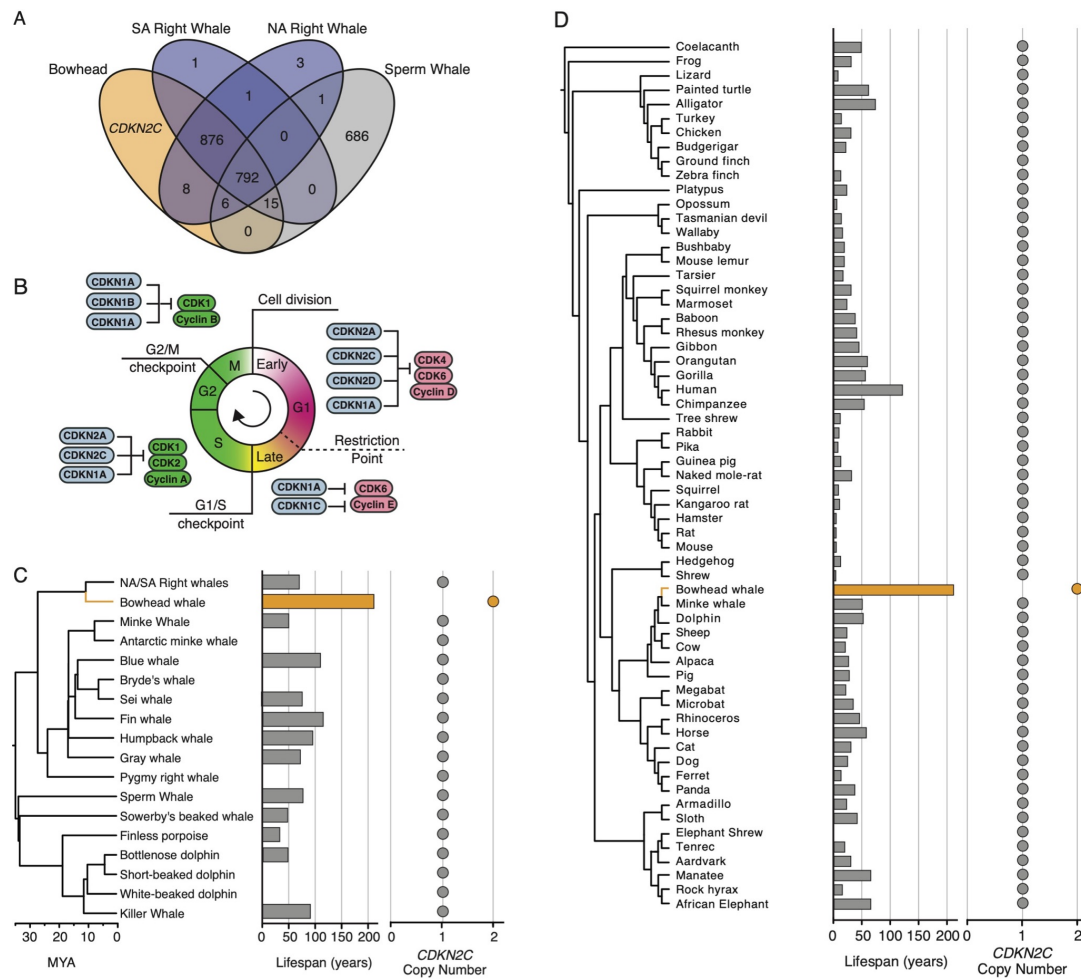
Figure 2. Duplication of the cell-cycle regulator *CDKN2C* in Bowhead whales.

(A) Venn diagram showing intersection of gene duplicated between Bowhead whale, South Atlantic (SA) Right whale, North Atlantic (NA) Right whale, and Sperm whale. Note that only 1 gene duplicate is unique to Bowhead whales: *CDKN2C*.

(B) Cartoon of the cell cycle, with major regulatory genes for each phase.

(C) Lifespan and *CDKN2C* copy number in 18 Cetaceans. Cetacean phylogenetic tree is scaled to time. Lifespan data is from (de Magalhães and Costa, 2009).

(D) Lifespan and *CDKN2C* copy number in 61 Sarcopterygians. Lifespan data is from (de Magalhães and Costa, 2009).

Figure 2 – source data 1. Duplicate genes identified in the Bowhead whale, North and South Atlantic Right whales, and Sperm whale.

Figure 2 – source data 2. Sarcopterygian genomes characterized for *CDKN2C* copy number.
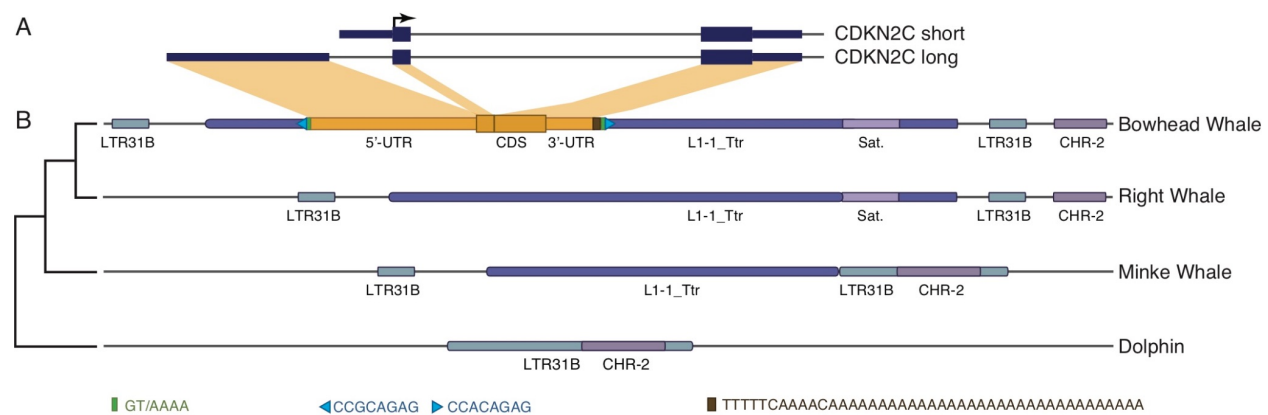
Figure 3. The Bowhead whale duplicate *CDKN2C* is a retrogene (*CDKN2CRTG*) embedded within a LINE L1 element.

(A) The canonical *CDK2NC* gene encodes two isoforms that use alternative 3'-UTR exons.

(B) Structure of the locus encoding *CDK2NCRTG* gene in Bowhead whale, and the homologous locus in Right whale, Minke whale, and Dolphin (phylogeny shown on left). The *CDK2NCRTG* gene is embedded within a Cetacean-specific LINE L1 element (L-1_Ttr) and nearby other transposable elements. The location of the poly-A tail (brown rectangle), LINE1 target site (green rectangle), and direct repeats (blue triangle) are shown. The Regions of the *CDK2NC* long transcript that contribute to the *CDK2NCRTG* gene are highlighted in yellow.
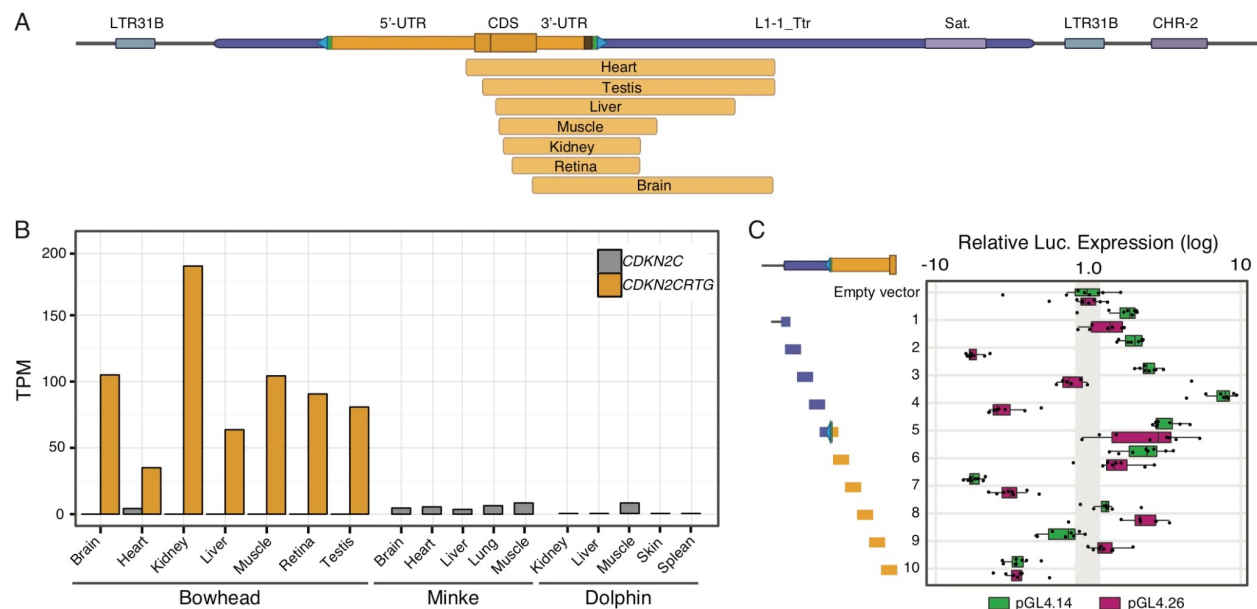
Figure 4. *CDKN2CRTG* is expressed and from a coopted LINE L1 promoter.

(A) Upper, *CDKN2CRTG* locus. Lower, *CDKN2CRTG* transcripts reconstructed with HISAT2 and StringeTie. Note that all transcripts extend downstream into the LINE L1 L2-1_Ttr and include the junction between the *CDKN2CRTG* gene and the ctranposable element which allows for discrimination between *CDKN2CRTG* and *CDKN2C* transcripts.

(B) Expression of *CDKN2CRTG* (yellow) and *CDKN2C* (grey) in Bowhead whale, Minke whale, and Dolphin tissues. Transcript abundances in transcripts per million (TPM) were inferred with Kallisto.

(C) Dual luciferase reporter assay in transiently transfected mouse 3T3 cells from ten 301bp overlapping fragments that tile the region from upstream of the L1-2_Ttr element to downstream of the *CDKN2CRTG* initiation codon. Each fragment was cloned into luciferase reporter vector pGL4.14[*luc2*/Hygro], which lacks both an enhancer and a promoter, and pGL4.26[*luc2*/minP/Hygro], which lacks an enhancer but has a minimal promoter. Relative luciferase expression (standardized to *Renilla* and empty vector control), n=8.

Figure 4 – source data 1. Transcriptome sample information.

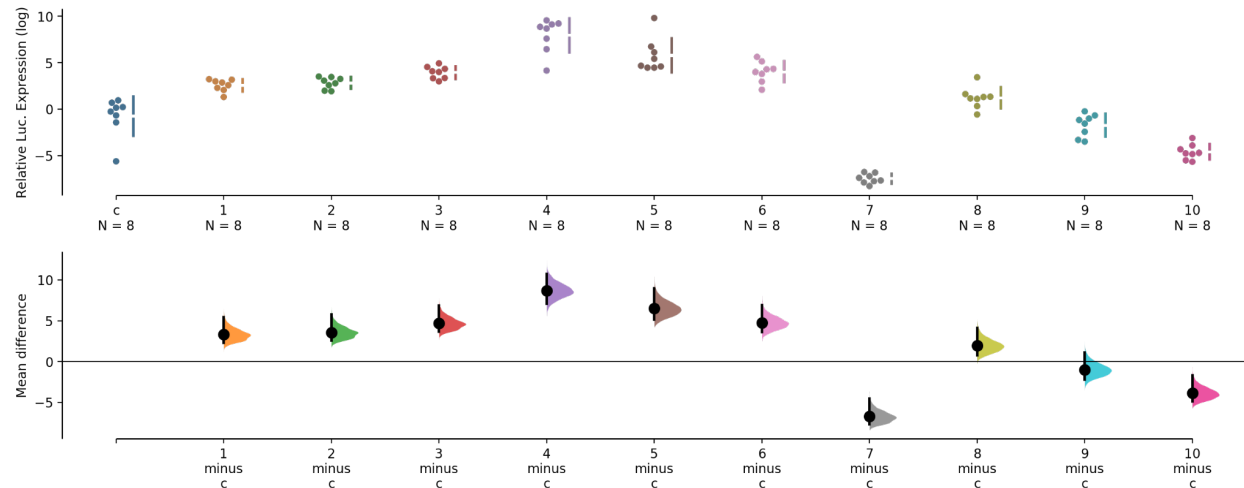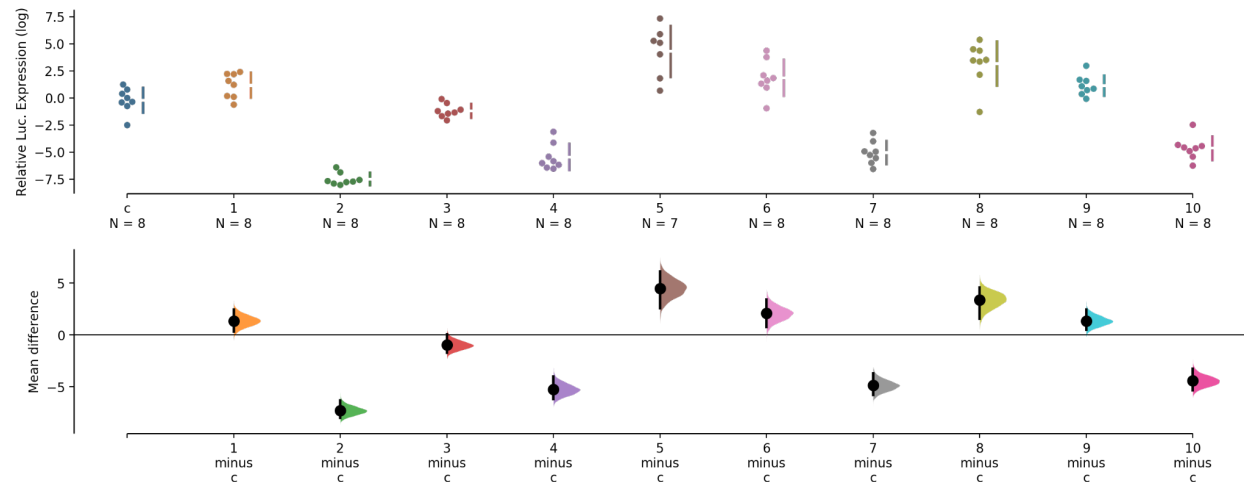**Figure 4 – source data 2. Data shown in panel C.**

Figure 4 – figure supplement 1. pGL4.14[*luc2*/Hygro] dual luciferase reporter assay results.
Ten 301bp overlapping fragments that tile the region from upstream of the L1-2_Ttr element to downstream of the *CDKN2CRTG* initiation codon were cloned into luciferase reporter vector pGL4.14[*luc2*/Hygro], which lacks both an enhancer and a promoter, and in transiently transfected mouse 3T3 cells. Relative luciferase expression (standardized to *Renilla* and empty vector control), n=8. The mean difference for 10 comparisons against the shared empty vector control (c) are shown as a Cumming estimation plot. The raw data is plotted on the upper axes. On the lower axes, mean differences are plotted as bootstrap sampling distributions. Each mean difference is depicted as a dot. Each 95% confidence interval is indicated by the ends of the vertical error bars. The effect sizes and CIs are reported above as: effect size [CI width   lower bound; upper bound]. 5000 bootstrap samples were taken; the confidence interval is bias-corrected and accelerated. The P value(s) reported are the likelihood(s) of observing the effect size(s), if the null hypothesis of zero difference is true. Results: The unpaired mean difference between c and fragment 1 is 3.31 [95.0%CI 2.31, 5.43]. The P value of the two-sided permutation t-test is 0.0. The unpaired mean difference between c and fragment 2 is 3.57 [95.0%CI 2.57, 5.77]. The P value of the two-sided permutation t-test is 0.0. The unpaired mean difference between c and fragment 3 is 4.69 [95.0%CI 3.69, 6.89]. The P value of the two-sided permutation t-test is 0.0. The unpaired mean difference between c and fragment 4 is 8.7 [95.0%CI 7.07, 10.7]. The P value of the two-sided permutation t-test is 0.0. The unpaired mean difference between c and fragment 5 is 6.53 [95.0%CI 5.16, 9.0]. The P value of the two-sided permutation t-test is 0.0002. The unpaired mean difference between c and fragment 6 is 4.77 [95.0%CI 3.58, 6.89]. The P value of the two-sided permutation t-test is 0.0. The unpaired mean difference between c

and fragment 7 is -6.73 [95.0%CI -7.7, -4.55]. The P value of the two-sided permutation t-test is 0.0. The unpaired mean difference between c and fragment 8 is 1.96 [95.0%CI 0.779, 4.15]. The P value of the two-sided permutation t-test is 0.0116. The unpaired mean difference between c and fragment 9 is -0.991 [95.0%CI -2.2, 1.09]. The P value of the two-sided permutation t-test is 0.257. The unpaired mean difference between c and fragment 10 is -3.86 [95.0%CI -4.91, -1.7]. The P value of the two-sided permutation t-test is 0.0004.

Figure 4 – figure supplement 2. pGL4.26[*luc2*/minP/Hygro] dual luciferase reporter assay results. Ten 301bp overlapping fragments that tile the region from upstream of the L1-2_Ttr element to downstream of the *CDKN2CRTG* initiation codon were cloned into luciferase reporter vector pGL4.26[*luc2*/minP/Hygro], which lacks an enhancer but has a minimal promoter, and in transiently transfected mouse 3T3 cells. Relative luciferase expression (standardized to *Renilla* and empty vector control), n=8. The mean difference for 10 comparisons against the shared empty vector control (c) are shown as a Cumming estimation plot. The raw data is plotted on the upper axes. On the lower axes, mean differences are plotted as bootstrap sampling distributions. Each mean difference is depicted as a dot. Each 95% confidence interval is indicated by the ends of the vertical error bars. The effect sizes and CIs are reported above as: effect size [CI width lower bound; upper bound]. 5000 bootstrap samples were taken; the confidence interval is bias-corrected and accelerated. The P value(s) reported are the likelihood(s) of observing the effect size(s), if the null hypothesis of zero difference is true. Results: The unpaired mean difference between c and 1 is 1.36 [95.0%CI 0.326, 2.45]. The P value of the two-sided permutation t-test is 0.0292. The unpaired mean difference between c and 2 is -7.29 [95.0%CI -8.0, -6.32]. The P value of the two-sided permutation t-test is 0.0. The unpaired mean difference between c and 3 is -0.972 [95.0%CI -1.7, 0.042]. The P value of the two-sided permutation t-test is 0.0462. The unpaired mean difference between c and 4 is -5.26 [95.0%CI -6.19, -4.01]. The P value of the two-sided permutation t-test is 0.0. The unpaired mean difference between c and 5 is 4.5 [95.0%CI 2.6, 6.12]. The P value of the two-sided permutation t-test is 0.0002. The unpaired mean difference between c and 6 is 2.08 [95.0%CI 0.749, 3.44]. The P value of the two-sided permutation t-test is 0.013. The unpaired mean difference between c and 7 is -4.86 [95.0%CI -

5.78, -3.69]. The P value of the two-sided permutation t-test is 0.0. The unpaired mean difference between c and 8 is 3.38 [95.0%CI 1.55, 4.59]. The P value of the two-sided permutation t-test is 0.002. The unpaired mean difference between c and 9 is 1.35 [95.0%CI 0.504, 2.44]. The P value of the two-sided permutation t-test is 0.0168. The unpaired mean difference between c and 10 is -4.43 [95.0%CI -5.32, -3.25]. The P value of the two-sided permutation t-test is 0.0002.
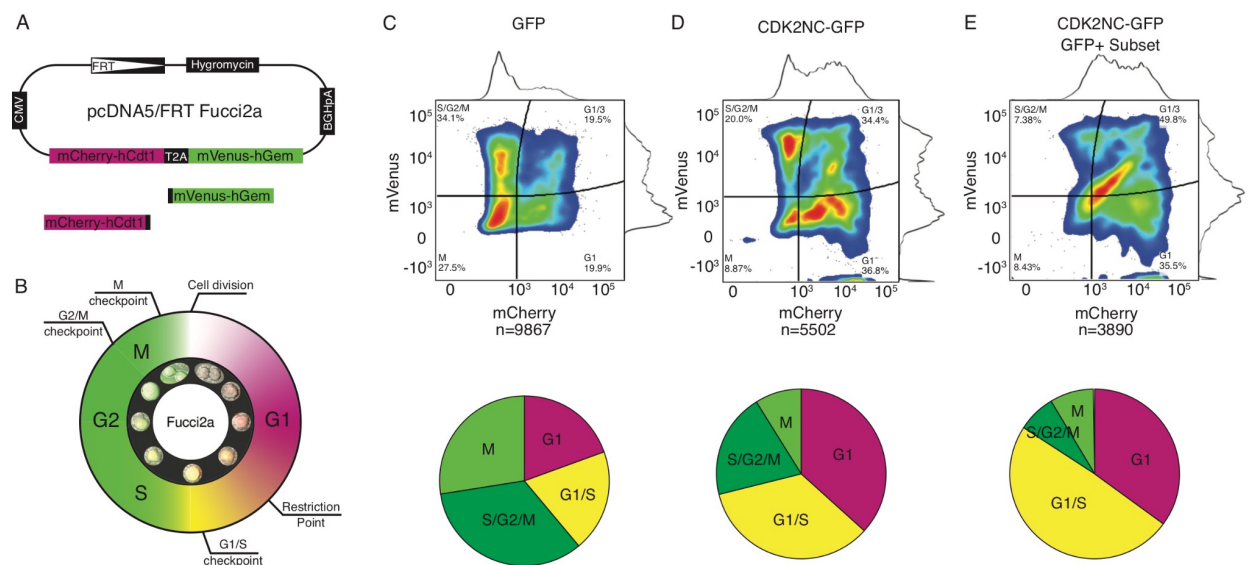
Figure 5. CDK2CRTG overexpression induces cell cycle delay at G_1 and G_1/S.

(A) Fucci2a Flp-In reporter used to generate the FuCHO cell line.

(B) Schematic of the cell cycle as visualized with the mCherry-hCdt1 and mVenus-hGem fluorescent reporters. Inset shows fluorescence intensity of single FuCHO cell. Cell cycle stage in FuCHO cells either mock transfected or transfected with a GFP-tagged copy of the Bowhead *CDKN2CRTG* was assayed after 20 hours by flow cytometry.

(C) The distribution of cell cycle states in mock transfected FuCHO cells (upper; inset percentages indicate the frequency of cells in each stage), pie chart (lower) shows the percentage of cells in each cell cycle stage.

(D) The distribution of cell cycle states in *CDKN2CRTG* transfected FuCHO cells (upper; inset percentages indicate the frequency of cells in each stage), pie chart (lower) shows the percentage of cells in each cell cycle stage. Note the shift in cells towards G1 and G1/S phase compared to mock transfected cells (shown in panel C).

(E) The distribution of cell cycle states GFP gated *CDKN2CRTG* transfected FuCHO cells. Note the shift in cells towards G1 and G1/S phase compared to mock transfected cells (shown in panel C). The shift in the population to G1 and G1/S is particularly pronounced compared to D.
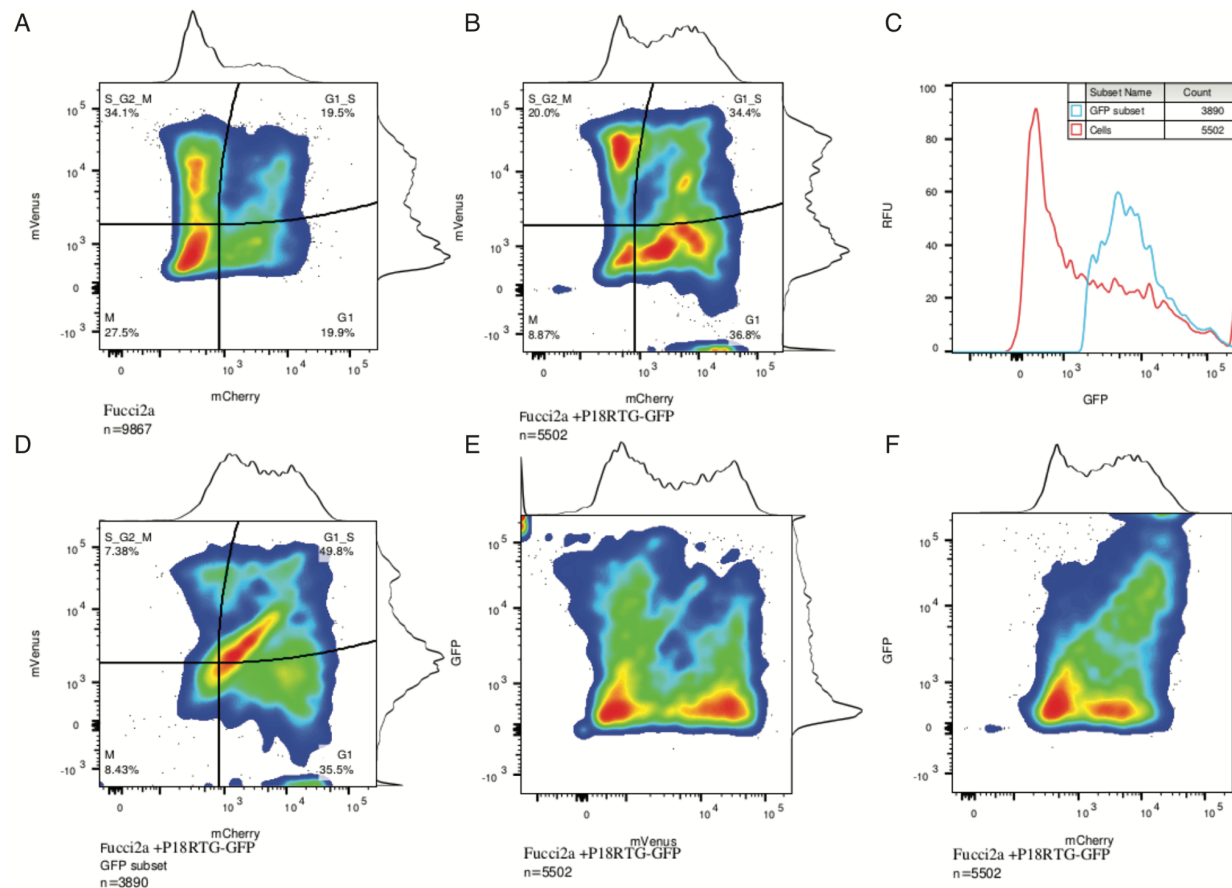
Figure 5 – figure supplement 1.

(A) The distribution of cell cycle states in mock transfected Fucci2a CHO cells, 19.5% of cells are in G1/S.

(B) The distribution of cell cycle states in *CDKN2CRTG* transfected Fucci2a CHO cells. Note the shift in cells towards G1 and G1/S phase from 19.5% to 34.4%.

(C) Gated subset of GFP-cells shown in F.

(D) The distribution of cell cycle states GFP gated *CDKN2CRTG* transfected Fucci2a CHO cells. Note the shift in cells towards G1 and G1/S phase from 7.38% to 49.8%. The shift in the population to G1 and especially G1/S is clear.

(E) GFP levels as a function of mVenus in the ungated populations.

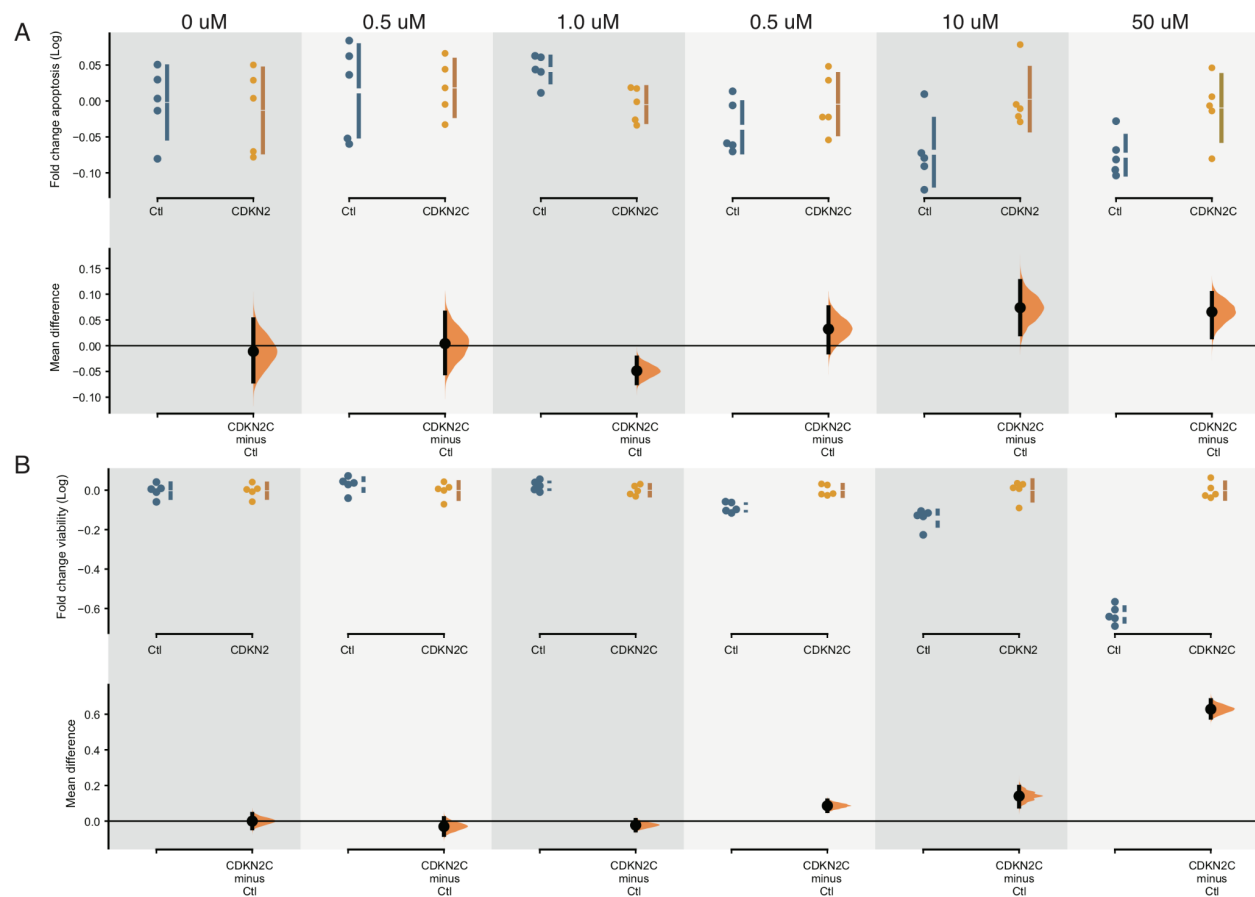(F) GFP levels as a function of mCherry in the ungated populations.

Figure 6. CDK2CRTG overexpression protects FuCHO cells from doxorubicin induced apoptosis and cell death. ApoTox-Glo results for control (Ctl) and CDKN2C transfected FuCHO cells treated with increasing doses of doxorubicin (0-50 uM) for 18 hours. Fold change in apoptosis (A) and viability (B) are shown on the upper axes of each plot. On the lower axes, mean differences are plotted as bootstrap sampling distributions. Each mean difference is depicted as a dot. Each 95% confidence interval (CI) is indicated by the ends of the vertical error bars. The effect sizes and CIs are reported above as: effect size [CI width lower bound; upper bound]. 5000 bootstrap samples were taken; the confidence interval is bias-corrected and accelerated. The P-value(s) reported are the likelihood(s) of observing the effect size(s), if the null hypothesis of zero difference is true from a two-sided permutation t-test.

(A) Apoptosis results: The unpaired mean difference between Ctl-0 and CDKN2C-0 is -0.0111 [95.0%CI -0.0695, 0.051], P=0.8. The unpaired mean difference between Ctl-0.5 and CDKN2C-0.5 is 0.00405 [95.0%CI -0.0535, 0.0643], P=0.878. The unpaired mean difference between Ctl-1 and CDKN2C-1 is -0.0488 [95.0%CI -0.0727, -0.0231],

P=0.0218. The unpaired mean difference between Ctl-5 and CDKN2C-5 is 0.0322 [95.0%CI -0.013, 0.0746], P=0.24. The unpaired mean difference between Ctl-10 and CDKN2C-10 is 0.0738 [95.0%CI 0.0222, 0.125], P=0.0292. The unpaired mean difference between Ctl-50 and CDKN2C-50 is 0.0657 [95.0%CI 0.0164, 0.102], P=0.0188.

(B) Cell viability results: The unpaired mean difference between Ctl-0 and CDKN2C-0 is -0.000142 [95.0%CI -0.0401, 0.04], P=0.984. The unpaired mean difference between Ctl-0.5 and CDKN2C-0.5 is -0.0301 [95.0%CI -0.0769, 0.0166], P=0.294. The unpaired mean difference between Ctl-1 and CDKN2C-1 is -0.0227 [95.0%CI -0.0518, 0.00616], P=0.192. The unpaired mean difference between Ctl-5 and CDKN2C-5 is 0.0863 [95.0%CI 0.0569, 0.116], P=0.0014. The unpaired mean difference between Ctl-10 and CDKN2C-10 is 0.141 [95.0%CI 0.0823, 0.192], P=0.0002. The unpaired mean difference between Ctl-50 and CDKN2C-50 is 0.628 [95.0%CI 0.581, 0.679], P=0.0.

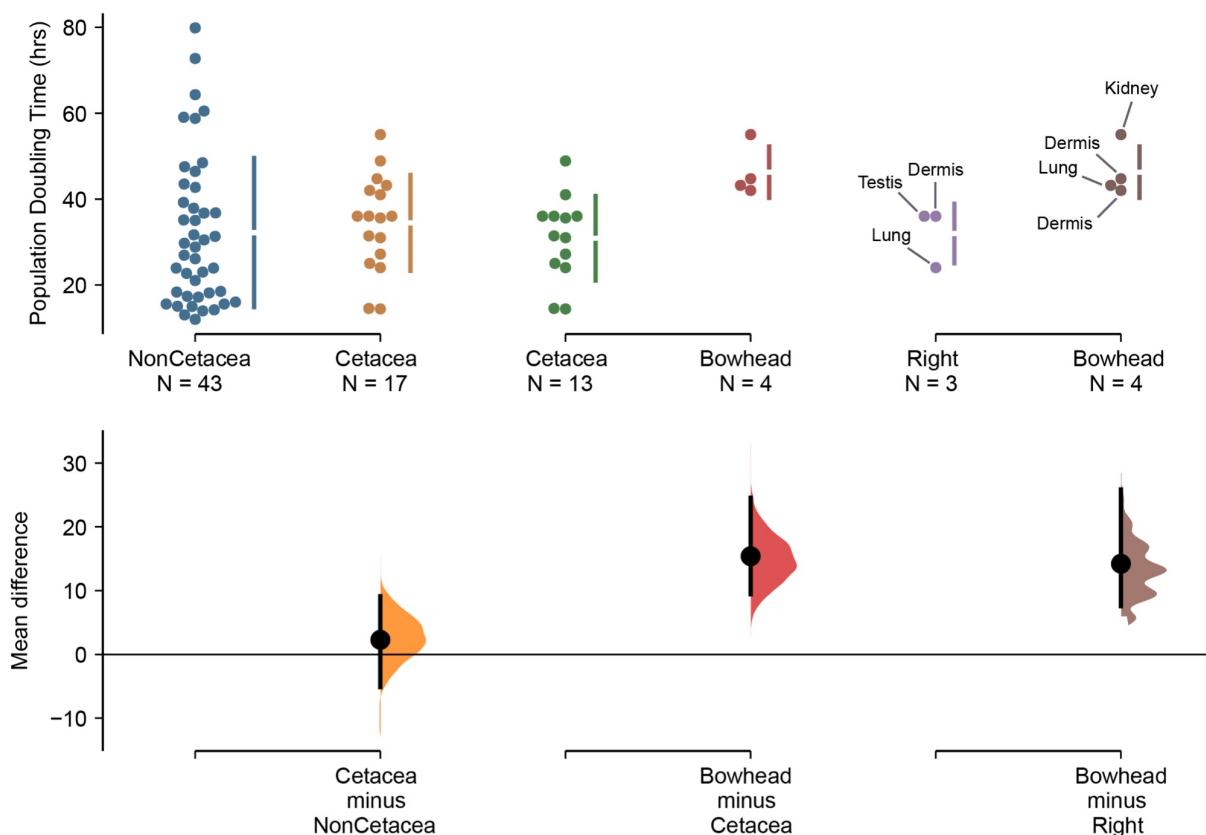**Figure 6 – source data 1. Data shown in panels A and B.**

Figure 7. Bowhead whale cells have longer doubling times than cells from other whales.

Population doubling times from 60 species of Eutherian mammals collected from the literature. Population doubling times for each group is shown on the upper axis (along with cell-types from which doubling times were estimated for Right and Bowhead whales). On the lower axes, mean differences are plotted as bootstrap sampling distributions. Each mean difference is depicted as a dot. Each 95% confidence interval (CI) is indicated by the ends of the vertical error bars. The effect sizes and CIs are reported above as: effect size [CI width lower bound; upper bound]. 5000 bootstrap samples were taken; the confidence interval is bias-corrected and accelerated. The P-value(s) reported are the likelihood(s) of observing the effect size(s), if the null hypothesis of zero difference is true from a two-sided permutation t-test. The unpaired mean difference between NonCetacea and Cetacea is 2.28 [95.0%CI -5.12, 9.14], P=0.616. The unpaired mean difference between Cetacea and Bowhead is 15.4 [95.0%CI 9.46, 24.6], P=0.0104. The unpaired mean difference between Right and Bowhead is 14.2 [95.0%CI 7.58, 25.9], P=0.0. Cell-types are shown for Bowhead and Right Whale cell lines.

Figure 7 – source data 1. Data shown in Figure 7.