Science Advances
AAAS

Supplementary Materials for

Cross-species identification of cancer resistance–associated genes that may mediate human cancer risk

Nishanth Ulhas Nair *et al.*

Corresponding author: Nishanth Ulhas Nair, nishanth.nair@nih.gov; Kuoyuan Cheng, kycheng@terpmail.umd.edu; Yuval Tabach, yuvaltab@ekmd.huji.ac.il; Eytan Ruppin, eytan.ruppin@nih.gov

Sci. Adv. **8**, eabj7176 (2022)
DOI: 10.1126/sciadv.abj7176

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

SUPPLEMENTARY FIGURES

Figure S1: *Summary of the top significantly enriched pathways (adjusted P<0.1) by the genes whose conservation scores are correlated with cancer-resistance estimates (MLTAW), using gene set enrichment analysis (GSEA) with gene set annotations from the Reactome database.*

Figure S2: *Scatter plots showing the correlation between the predicted cancer resistance (CR) scores computed based on gene conservation and for the cancer-resistance estimate MLTAW or '(Maximum longevity)^6 x (adult weight)', with leave-one-out cross-validation, for all species.*

Figure S3: *Scatter plots showing the correlation between the predicted cancer resistance (CR) scores computed based on gene conservation and for the cancer-resistance estimate MLTAW for Aves (birds).*

Figure S4: *Random controls experiments for predicting cancer resistance using all species.*

Figure S5: *Instead of human reference to compute gene conservation scores, we use Mus musculus (house mouse) and thirteen-lined ground squirrel (Ictidomys tridecemlineatus) as reference and predict cancer resistance in the all-species analysis.*

Figure S6: *Robustness analysis using 12 non-human species as reference for computing the gene conservation scores.*

Figure S7: *Plots show the distribution/frequency of Spearman's ρ between the predicted cancer resistance (CR) scores computed based on gene conservation and each of the two cancer-resistance estimates, using two-fold cross-validation (instead of LOOCV).*

Figure S8: *Predicting cancer-resistance (CR) scores by altering various parameters.*

Figure S9: *Predicting cancer-resistance (CR) scores by using both PC and NC genes (PC & NC genes); PC genes only; NC genes only.*

Figure S10: *Cancer resistance predictions on the entire mammalian species.*

Figure S11: *Predicting cancer resistance (CR) scores by identifying PC/NC genes by leaving out one class and testing on that left-out class (cross-validation; all-species analysis).*

Figure S12: *Predicting cancer resistance (CR) scores by identifying PC/NC genes (using mammalian data) by leaving out one order of mammals and testing on that left-out order (cross-validation; mammals-only analysis).*

Figure S13: *Gene set enrichment analysis (GSEA) of the correlation between the gene conservation scores and cancer-resistance estimates including (maximum longevity)^6 x (adult weight) (MLTAW), or the residue of maximum longevity after regressing out adult weight (MLCAW), for three classes of species: Mammalia (mammals), Aves (birds), and Teleostei (fish).*

Figure S14: *Spearman's correlation (ρ) in predicting cancer resistance (MLCAW) in all species using only TSGs, only oncogenes, both TSGs and oncogenes, using PC and NC genes in cross validation, using PC and NC genes after removing TSGs and oncogenes in cross validation is shown.*

Figure S15: *Looking at loss-of-function genes observed in canine transmissible venereal tumors.*

Figure S16: *The cancer resistance (CR) scores predicted on all species (in leave-one-out cross-validation) analysis are individually tested on different classes of species: Mammalia (mammals), Aves (birds), Teleostei (fish), Reptilia (reptiles).*

Figure S17: Gene conservation profiles for all genes and species, before (left sub-figure) and after (right sub-figure) rank-normalization.

SUPPLEMENTARY TABLES

Supplementary Table S1: The predicted cancer resistance (CR) scores are provided for various species.

Supplementary Table S2: Table contains Pearson correlation value of gene conservation and cancer resistance measures for all genes

Supplementary Table S3: Table contains gene set enrichment analysis (GSEA) using the Pearson correlation values of gene conservation and cancer resistance measures for all genes (Reactome pathways are used for the enrichment analysis).

Supplementary Table S4: Table contains Pearson correlation value of gene conservation and cancer resistance measures for all genes for various orders of mammals.

Supplementary Table S5: Table contains gene set enrichment analysis (GSEA) using the Pearson correlation values of gene conservation and cancer resistance measures (for all genes), for various orders of mammals (Reactome pathways are used for the enrichment analysis).

Supplementary Table S6: LOEUF and GSEA of PC and NC genes

Supplementary Table S7: Pathway enrichment of the known human cancer genes that overlap with PC genes, or NC genes, or that do not overlap with either PC or NC genes.

Supplementary Table S8: Table contains the analysis results on the mutated genes detected in a single phylogeny of a mouse melanoma tumor, their enrichment for PC/NC genes, and the pathway enrichment of the overlapping PC genes

Supplementary Table S9: In this table, we summarize different lines of evidence supporting PC genes.

Supplementary Table S10: In this table, we rank the PC genes based on the percentage of non-silent mutations in the TCGA pan-cancer dataset.

Supplementary Table S11: Table contains copy number of different genes in mammals from EggNOG database.

Supplementary Table S12: Summary of PC and NC genes within each pathway.

Supplementary Table S13: Genes (and their enriched pathways) that are specifically associated with cancer resistance estimates. All-species analysis.

Supplementary Table S14: Genes (and their enriched pathways) that are specifically associated with cancer resistance estimates. Mammals-only analysis.

Supplementary Table S15: Pearson correlation values of gene conservation and true cancer mortality risk (CMR) for all genes for 39 mammals. Gene set enrichment analysis (GSEA) using Reactome pathways are shown. PC and NC genes (from all-species and mammals-only analysis) that intersect with NCMR and PCMR genes respectively are also shown.

SUPPLEMENTARY NOTES

1. Short review of cancer resistance mechanisms in different species

Different species have independently evolved unique cancer resistance mechanisms. Repression of somatic telomerase activity, and replicative senescence are important tumor-suppressing mechanisms evolved in species greater than approximately 10 kilograms (2). Cells of smaller but relatively long-lived animals are reported to have slower proliferation in culture (*77*). There have been a few reports of more efficient DNA repair in cancer-resistant and long-lived animals (3*,* 9). African elephants, the largest land mammals, have 19 extra retrogene copies of the tumor suppressor gene *TP53* and are more sensitive to *TP53*-mediated DNA damage response (*78, 79*). The remarkable cancer resistance of naked mole rat has been previously partly attributed to the production of high molecular mass hyaluronan (HMM-HA) (*80, 81*) but this mechanism has been questioned in a recent study (*82*). Blind mole rats also have abundant HMM-HA and increased interferon-β expression that contribute to cancer resistance (3*, 81, 83*). Various large-bodied whales do not have additional *TP53* copies, and their cancer resistance mechanisms are not clearly understood (3*, 13*).

2. Summary of the numbers of PC and NC genes

The table below summarizes the numbers of PC and NC genes identified (analysis done with humans as the reference for computing the gene conservation matrix) for the cancer resistance predictors: all-species, mammals, birds, and teleost fish, at FDR < 0.1. The table below summarizes the number of PC/NC genes at FDR<0.1 for the cancer resistance predictors (all-species, mammals, birds) using the different MLTAW/MLCAW cancer resistance estimates (robustness was also shown for other FDR thresholds). Note that for birds with MLCAW and Teleost fish, no genes passed the FDR 0.1 cutoff.

	MLTAW (all-species) (n=193 species)	MLCAW (all-species) (n=193 species)	MLTAW (Mammallia) (n=108 species)	MLCAW (Mammallia) (n=108 species)	MLTAW (Aves) (n=55 species)
No. of PC genes at FDR < 0.1	3506	2797	2211	1912	428
No. of NC genes at FDR < 0.1	4108	2184	2079	1524	132

For the LOEUF score analysis, we enclose a summary of the resulting numbers of PC and NC genes in the table below (we only used PC/NC genes identified from all species or mammals in these analyses):

	MLTAW (all-species) (n=193 species)	MLCAW (all-species) (n=193 species)	MLTAW (Mammallia) (n=108 species)	MLCAW (Mammallia) (n=108 species)
No. of PC genes	594	385	423	160
No. of NC genes	483	244	392	159

3. Cancer resistance prediction in birds and teleost fishes

Using leave-one-out cross-validation (LOOCV) we find a significant positive correlation between the predicted cancer resistance (CR) scores and MLTAW on all the bird species (Spearman's ρ =

0.43, P = 0.00094, Fig. S3A) though this is weaker than the corresponding predictions obtained by learning on all species (Spearman's ρ = 0.57, P = 6.4e-6, Fig. S16A, using LOOCV). The correlation is stronger for the order Passeriformes (Spearman's ρ = 0.79, P = 0.0012, Fig. S3B) for which we have the largest number of samples. Among Passeriformes, the highest CR scores are obtained for American crow (Fig. S3B). The MLCAW measure did not yield any PC/NC genes (FDR < 0.1) for birds, and hence no CR predictor could be built. We could not identify any PC/NC genes at FDR < 0.1 for teleost fishes (for both MLTAW and MLCAW measures) probably because of a small sample size (n=18); and hence we could not however build a cancer resistance predictor for them.

A detailed pathway enriched analysis for mammals, birds, and teleost fishes are provided in Fig. 3A, S12C. Many pathways show group-specific enrichment. For example, many cell cycle and DNA repair-related pathways are enriched by the PC genes in mammals, but not or to a much lesser extent in birds or teleost fishes; complement activation is enriched by PC genes in teleost fishes, but by NC genes in mammals or birds (Fig. 3A, S12C, Table S3). Bird PC genes are uniquely enriched for certain processes including fatty acid and amino acid metabolism and PI3K-AKT signaling pathway despite sharing interleukin, interferon signaling and mRNA transcription with mammals (Fig. 3A, S12C). GPCR signaling is commonly enriched by the NC genes based on MLCAW in all three groups (Fig. 3A, S12C, Table S3).

4. Control and robustness analysis

Random control experiments

Random controls experiments for predicting cancer resistance were done for using all species. We chose random PC/NC genes with the same size as the actual PC/NC genes identified from the all-species analysis at FDR < 0.1. We can predict cancer resistance using these genes. We do this for 1000 iterations and the empirical P-value is computed. We see that they are not correlated in comparison to the 'true' correlation obtained using the actual PC/NC genes (randomization test P < 0.001). The results for MLTAW/MLCAW are shown in Fig. S4.

PC/NC genes and cancer resistance predictors are robust to the method of correlation used.

To identify cancer resistance-associated genes (PC/NC genes) for all species, we used Pearson correlation between the conservation scores of each gene and the cancer-resistance estimates (MLTAW and MLCAW) across all species. Pearson correlation coefficient was used (instead of Spearman) in order to reduce the number of ties which will affect the gene set enrichment analysis (GSEA). We now show that the robust identification of PC/NC genes is possible even if we use Spearman's correlation instead of Pearson. To do this, we recomputed PC/NC genes using Spearman's correlation and compared it to those obtained using Pearson's correlation (FDR < 0.1). Using Fisher's exact test, we get a significant overlap in comparing cancer resistance-associated genes obtained using Pearman and Spearman's correlation (PC genes: Odds-ratio/OR = 111.64, P < 2.2e-16 for MLTAW and OR = 94.22, P < 2.2e-16 for MLCAW; NC genes: OR = 126.52, P < 2.2e-16 for MLTAW and OR = 184.78, P < 2.2e-16 for MLCAW). (Note: whenever the enrichment test software shows P = 0, we write as P < 2.2e-16.) Cancer-resistance (CR) predictors computed from PC/NC genes identified using Pearson or Spearman's correlation works very similar across all species (Pearson-based PC/NC identification: MLTAW ρ=0.44, P=1.32e-10, MLCAW ρ=0.51, P=2.31e-14; Spearman's-based PC/NC identification: MLTAW ρ=0.43, P=3.65e-10, MLCAW ρ=0.51, P=5.16e-14; using LOOCV or 'leave-one-out cross-validation').

PC/NC genes and cancer resistance predictors are robust to the choice of reference species

We used the human genome as a reference, for computing gene conservation matrix as most of our downstream analysis was on human genes and we aimed to identify genes relevant to cancer resistance in humans. To check if our analysis is robust to changes in reference, we recomputed the gene conservation matrix using a few different species genomes including some cancer prone species like the house mouse (Mus Musculus) (in comparison humans are known to be relatively cancer resistant). The thirteen-lined ground squirrel was also chosen as a reference as it is likely to be cancer prone given that it is a species with poor longevity (it is relatively low MLTAW and MLCAW estimates). In the all-species analysis, the PC/NC genes obtained using mouse or squirrel genome as reference are extremely similar to the PC/NC genes obtained using humans as reference. Using Fisher's exact test, we get a significant overlap in comparing the PC genes

obtained using humans and mouse/squirrel genomes as reference (mouse: OR = 14.67, P < 2.2e-16 for MLTAW and OR = 30.87, P < 2.2e-16 for MLCAW; squirrel: OR = 33.78, P < 2.2e-16 for MLTAW and OR = 55.31, P < 2.2e-16 for MLCAW). Similarly, we get a significant overlap in comparing the NC genes obtained using humans and mouse/squirrel genomes as reference (mouse: OR = 10.78, P < 2.2e-16 for MLTAW and OR = 27.73, P < 2.2e-16 for MLCAW; squirrel: OR = 24.29, P < 2.2e-16 for MLTAW; OR = 57.68, P < 2.2e-16 for MLCAW). Similarly, cancer-resistance (CR) predictors computed from a gene conservation matrix which uses mouse or squirrel genomes as reference, show good prediction results between CR scores and cancer-resistance estimates; similar to what was obtained using human as reference (Fig. S5; LOOCV). We also see that humans are predicted to be relatively cancer resistant as expected (Fig. S5).

Furthermore, for the sake of a more comprehensive analysis, we recomputed the gene conservation scores using 12 non-human species genomes as a reference including many evolutionarily distant species from humans (includes mammals, birds, fish, even plants). Those are Mus musculus (house mouse), Ictidomys tridecemlineatus (thirteen-lined ground squirrel), Heterocephalus glaber (naked-mole rat; known to be cancer-resistant), Physeter catodon (sperm whale; large animal), Gallus gallus (chicken), Cyanistes caeruleus (Eurasian blue tit), Struthio camelus australis (ostrich), Danio rerio (zebrafish), Arabidopsis thaliana (thale cress), Petunia axillaris (large white petunia), Zea mays (corn), Solanum lycopersicum (tomato). We then recomputed the PC and NC genes (all-species analysis) using the gene conservation scores obtained from the analyses of each of these 12 reference species genomes (both the MLTAW and MLCAW measures), and checked if they significantly overlapped with the corresponding PC/NC genes identified using human genome as a reference (overlap enrichment test, using Fisher exact test) -- we see a very significant overlap in all 12 cases (FDR < 0.0005, Fig. S6). As probably expected, we see the strongest overlap if we choose any mammal as a reference (irrespective of its size, longevity, or whether it is cancer-prone/resistant) in comparison to using birds, fish, or plants (Fig. S6).

Next, we compared the overlap between the PC and NC genes identified using a cancer-prone species like house mouse and a cancer-resistant species like naked mole rat as reference

genomes. Notably, the PC/NC genes obtained using mouse genome as reference are very similar to the PC/NC genes obtained using naked-mole rat as reference: Using a Fisher's exact test, we get a significant overlap in comparing the PC/NC genes obtained using mouse and naked mole rat genomes as reference (PC genes: OR = 15.06, P < 2.2e-16 for MLTAW and OR = 46.22, P < 2.2e-16 for MLCAW; NC genes: OR = 8.72, P < 2.2e-16 for MLTAW and OR = 63.22, P < 2.2e-16 for MLCAW). Similarly, we compared the overlap between the PC and NC genes identified using a short-living species like house mouse and a long-living species like sperm whale as reference genomes. Like before, in the all-species analysis, the PC/NC genes obtained using mouse genome as reference are very similar to the PC/NC genes obtained using sperm whale as reference (PC genes: OR = 7.55, P < 2.2e-16 for MLTAW and OR = 53.91, P < 2.2e-16 for MLCAW; and NC genes: OR = 4.44, P < 1.02e-170 for MLTAW and OR = 48.18, P < 2.2e-16 for MLCAW). (Note that whenever the enrichment test software shows P = 0, we write as P < 2.2e-16.) Based on these results we can see that our identification of PC/NC genes is quite robust to the choice of reference.

Using two-fold cross-validation instead of LOOCV

We also did a two-fold cross-validation (instead of LOOCV) for predicting cancer-resistance scores, i.e., identifying PC and NC genes in the training group and testing the accuracy of the CR predictions in the left-out group. We see that our results using two-fold cross-validation is similar to that obtained by LOOCV in the all-species analysis (Fig. S7 in comparison to Fig. 2A, S2).

Our results are robust to changes in FDR criteria and thresholds used

We predicted cancer-resistance scores (CR) by altering various parameters. Our original predictor as described in the manuscript uses PC genes and NC genes which are significantly associated with cancer resistance at FDR < 0.1. We now show that our CR predictor is robust to changes in FDR thresholds from 0.1 to 0.01 or 0.2 (Fig. S8A,B). The original predictor also computes the number of PC genes whose conservation score > median conservation score; and the number of NC genes whose conservation score < median conservation score. We also show that the CR predictor is robust to altering the thresholds from median conservation score to top and bottom

33 percentile of the conservation scores for PC and NC genes respectively (Fig. S8C). The same analysis was also done using the top and bottom 20 percentile (Fig. S8D).

Alternative predictors using either PC or NC genes

For the original predictor which uses both PC and NC genes, our cancer resistance (CR) score was measured using the following equation:

Original predictor: CR score = [(No. of PC genes > MCS) + (No. of NC genes < MCS)] / (Total no. of genes)

where MCS is the median conservation score of all genes in a species; PC and NC genes are chosen for FDR < 0.1.

Now to test the individual contribution of using PC-only and NC-only genes to predict a good cancer-resistance estimate, we build alternative predictors as follows:

PC-only predictor: CR score = (No. of PC genes > MCS) / (Total no. of genes)

NC-only predictor: CR score = (No. of NC genes < MCS) / (Total no. of genes)

We then compare the PC-only and NC-only predictor with the original predictor in Fig. S9 for the all-species, Mammalia (mammals), and Aves (birds) analysis for both MLTAW and MLCAW measures. We see that both PC and NC genes have significant individual and comparable contributions to predict cancer resistance.

5. Cancer resistance prediction within specific mammalian orders

Using the predicted CR scores learnt from all mammalian species (in LOOCV), we further tested its association with cancer-resistance estimates for various orders with the class Mammalia, Rodentia, Primates, Carnivora, Artiodactyla, Cetacea, Chiroptera. We are able to predict at least one of the two cancer-resistance estimates effectively in Rodentia, Primates, Carnivora, and Chiroptera; but not for Artiodactyla, Cetacea (Fig. S10A,B). Just looking at rodents, we are able

to get good CR scores in known cancer resistant species like the naked mole rat and low CR scores for cancer prone species like the house mouse (Fig. S10C). Among primates, we see that animals like chimpanzees and gorillas are predicted to be cancer resistant (Fig. S10D). Among carnivores, we see that Steller sea lion and California sea lion have high CR scores (Fig. S10E). Among bats, we again see that species like the Brandt's bat which are known to live long for their body size are predicted to have high CR scores (Fig. 2D). Little brown bats are also seen to have high CR scores (Fig. 2D).

6. Copy numbers of top PC and NC genes across mammals

We have not explicitly considered the number of copies or paralogous genes in our analysis, and we have explained this as a limitation in the discussion section. To generally compute the copy numbers of genes in any species is not a trivial problem, requiring the careful choice of thresholds to find orthologues and paralogues. The appropriate thresholds depend on the phylogenetic distance between each pair of species and on sequence lengths. After determining the clusters of orthologous groups based on the normalized blast bit scores of each possible protein pair, gene copy numbers can be determined. A thorough investigation of the number of copies may be pursued in a future study. However, here we performed a smaller-scale analysis of top PC/NC gene copy numbers in mammalian species, as follows: From the EggNOG database (*84*), we obtained the gene orthologous group across mammals for each of the 5 top PC and NC genes correlated with MLTAW/MLCAW in the mammal-specific analysis. We found only one copy (a single orthologous gene per species) of most of these PC/NC genes in most species, and accordingly we do not see a significant correlation between copy-number and MLTAW/MLCAW measures. However, some of these PC/NC genes do have more than one copy (few paralogous) in well-known specific cancer resistance species (note that the term "single copy" is used here to describe the case where a gene is present on both chromosomes in a diploid genome). For example, we see that ZBED9 (a PC gene correlated with MLTAW) has 3 paralogues in the African elephant (the rest of the species have one copy) and KRBA2 (another PC gene correlated with

MLTAW) has 2 copies in the naked mole rat. This suggests that some of the PC/NC genes we identified may indeed undergo copy number alterations. As a sanity check, we also checked the number of copies for p53, and confirmed that there are multiple copies of p53 (most of them pseudo genes) in the African elephant. However, in most of the other species p53 has no paralogous, consistent with previous literatures (*78, 79*). These gene copy number results are given in Table S11.

7. Adaptation to different oxygen levels and cancer resistance

In our study, we have not explicitly considered factors potentially linked to variations in cancer resistance that are not reflected through body size and lifespan. These factors are more difficult to quantify with limited data available. One example is the adaptation to different oxygen concentrations and oxidative stress levels. Reactive oxygen species (ROS) levels have a complicated role in cancers, although one of the effects of ROS is DNA damage, which is linked to cancer development (*85*), and tolerance to hypoxia is also associated with cancer resistance, as evident in several well-known cancer resistant species including the naked mole rat and certain bats (*68, 86*). This was not considered in our analysis due to the challenge to quantitate hypoxia resistance for each species, which may underlie some notable outliers in our cancer resistance predictions. For example, the predicted gene conservation-based CR score was high for the small and short-lived star nosed mole (55 grams, 2.5 years; Fig. 2A, S2), which largely lives underground and is hypoxia-tolerant (*87*). When more phenotypic data across species become available in the future, further studies are required to refine and update our findings here.

8. Enrichment of PC and NC genes on the sex chromosomes of mammals and birds

To investigate whether the abundances of PC/NC genes on the sex chromosomes differ between mammals and birds, we downloaded the Mammalia genes that exist in X and Y mammalian-sex-chromosomes and birds' genes that exist in Z and W bird-sex-chromosomes. We then computed

the abundance of both PC and NC genes in sex chromosomes of birds and mammals. For each species in mammals or birds, we computed the total number of sex genes the species has and the number of sex genes that are present in PC/NC genes, and measured the enrichment of PC/NC genes in these sex chromosomes by odds ratios (ORs). The ORs for each set of PC or NC genes in the sex chromosomes of mammals (XY) were then compared to those for the birds (ZW) with Wilcoxon rank-sum tests. A summary of the distributions of the ORs and the two-sided Wilcoxon rank-sum test P-values are summarized in the table below. We see that the PC genes tend to be over-represented on the sex chromosomes of both mammals and birds (i.e., OR>1), while the NC genes tend to be under-represented (OR<1) on the sex chromosomes. However, the extent of under-representation for NC genes is significantly stronger in birds than in mammals. The extent of over-representation for PC genes are also different between birds and mammals, however, here the direction of the difference depends on whether MLTAW or MLCAW was used to identify the PC genes. MLCAW PC genes are more strongly over-represented on the sex chromosomes in birds than in mammals, while MLTAW PC genes exhibit the opposite trend.

genes	species	OR Min.	OR 1st Qu.	OR Median	OR Mean	OR 3rd Qu.	OR Max.	P-value
MLCAW PC	mammals (XY)	0.8807208336	1.116401695	1.165427249	1.193734882	1.189015498	2.355205658	3.48E-05
MLCAW PC	birds (ZW)	1.133449703	1.249361475	1.259597386	1.272658813	1.307364157	1.408332487	3.48E-05
MLCAW NC	mammals (XY)	0.7985345012	0.9139980765	0.9612346404	1.006265209	1.007656418	1.643790751	2.39E-07
MLCAW NC	birds (ZW)	0.338397378	0.3688231375	0.3904344464	0.3935403498	0.4063078041	0.4923647457	2.39E-07
MLTAW PC	mammals (XY)	0.5215046227	1.4511353	1.49583152	1.442961816	1.50863158	1.591965573	8.42E-08
MLTAW PC	birds (ZW)	0.9190058898	0.9963760673	1.030014239	1.031453072	1.072284008	1.127438833	8.42E-08
MLTAW NC	mammals (XY)	0.4492870386	0.4860804856	0.5077636619	0.5602773981	0.5707946462	1.057688892	6.36E-08
MLTAW NC	birds (ZW)	0.356980346	0.3734042482	0.3816719458	0.3965563195	0.3974321429	0.5778710576	6.36E-08

9. Rank-normalization

In general, a higher gene conservation score (for most genes) would be expected in species that are phylogenetically closer to humans than more distant species (as we use humans as a reference to compute gene conservation scores). To confirm this, we checked if the gene conservation scores are significantly higher in mammals in comparison to birds and teleost

fish and found that to be true (one-sided Wilcoxon rank-sum test, P < 2.2e-16; Fig. S17). However, reassuringly, after rank-normalization, we do not see any overall differences between the conservation profiles of mammals, birds, and teleost fish (one-sided Wilcoxon rank-sum test, P = 0.5; Fig. S17).

As we already noted, the gene conservation scores were obtained by rank-normalizing the protein length normalized bit scores across genes within each species, to control for the evolutionary distance between human and each species. These rank-normalized values range from 0 to 1, with higher values corresponding to higher levels of conservation. For ranking ties, we used the 'rank' function in R (https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/rank). Although we used ties.method = first for ranking ties, we could have also used ties.method = min. The overall results for identifying PC and NC genes are robust irrespective of whether we use ties.method = first or ties.method = min.

10. Cancer resistance estimates and Peto's paradox

As explained in the main manuscript, since the strength of intrinsic cancer resistance mechanisms of a species is a "latent" property that is not directly observable, we used two proxy cancer-resistance estimates that have been proposed in the literature – *MLTAW* and *MLCAW*. As per Peto's paradox cancer incidence within the normal lifespan of a species appears to have comparable orders of magnitudes across large or small, and long-lived or short-lived species. It follows that the intrinsic level of cancer resistance in a given species needs to roughly counteract its risk of cancer development due to cell division, which accordingly to a simple cancer development model is proportional to $ML^6 \times AW$, where ML denotes the species maximum longevity and AW denotes its adult weight. This is our MLTAW measure.

Recently, Vincze et al. (*20*) published cancer-related mortality (cancer mortality risk or CMR) of 191 mammalian species using data on adult zoo mammals (110,148 individuals). They also published "Adult life expectancy" and "Species body mass (kg)" for each of these

species (there was no maximum longevity information provided). So, we recomputed MLTAW and MLCAW measures for these 39 species "Adult life expectancy" instead of "Maximum longevity", and "Species body mass (kg)" instead of "Adult weight" (referred to as MLTAW' and MLCAW' respectively). As expected, we do not see a correlation between MLTAW' and CMR (Spearman's ρ=0.04, P=0.58) showing that Peto's paradox is correct, that is, cancer mortality risk is not very dependent on body mass or adult weight (as already discussed in Vincze et al.). This justifies the use of MLTAW as an intrinsic level of cancer resistance. There was however a small positive correlation between MLCAW' and CMR (Spearman's ρ=0.24, P=0.00078).

SUPPLEMENTARY FIGURES

Figure S1: Summary of the top significantly enriched pathways (adjusted P<0.1) by the genes whose conservation scores are correlated with cancer-resistance estimates (MLTAW), using gene set enrichment analysis (GSEA) with gene set annotations from the Reactome database. *The cancer-resistance estimate used is MLTAW or '(Maximum longevity)^6 x (adult weight)'. The normalized enrichment score is plotted on the Y-axis, where positive values correspond to enrichment by the positively correlated (PC) genes and negative values correspond to enrichment by the negatively correlated (NC) genes. The dot color represents the significance of the enrichment (negative log10 GSEA P-value), and the dot size represents the number of genes in the "leading edge", i.e., the set of genes that are enriched in a pathway. For the sake of clarity, only*

a subset of the enriched pathways (FDR<0.1) are shown and long pathway names have been shortened (using "..."). The complete pathway enrichment results are given in Table S3A.

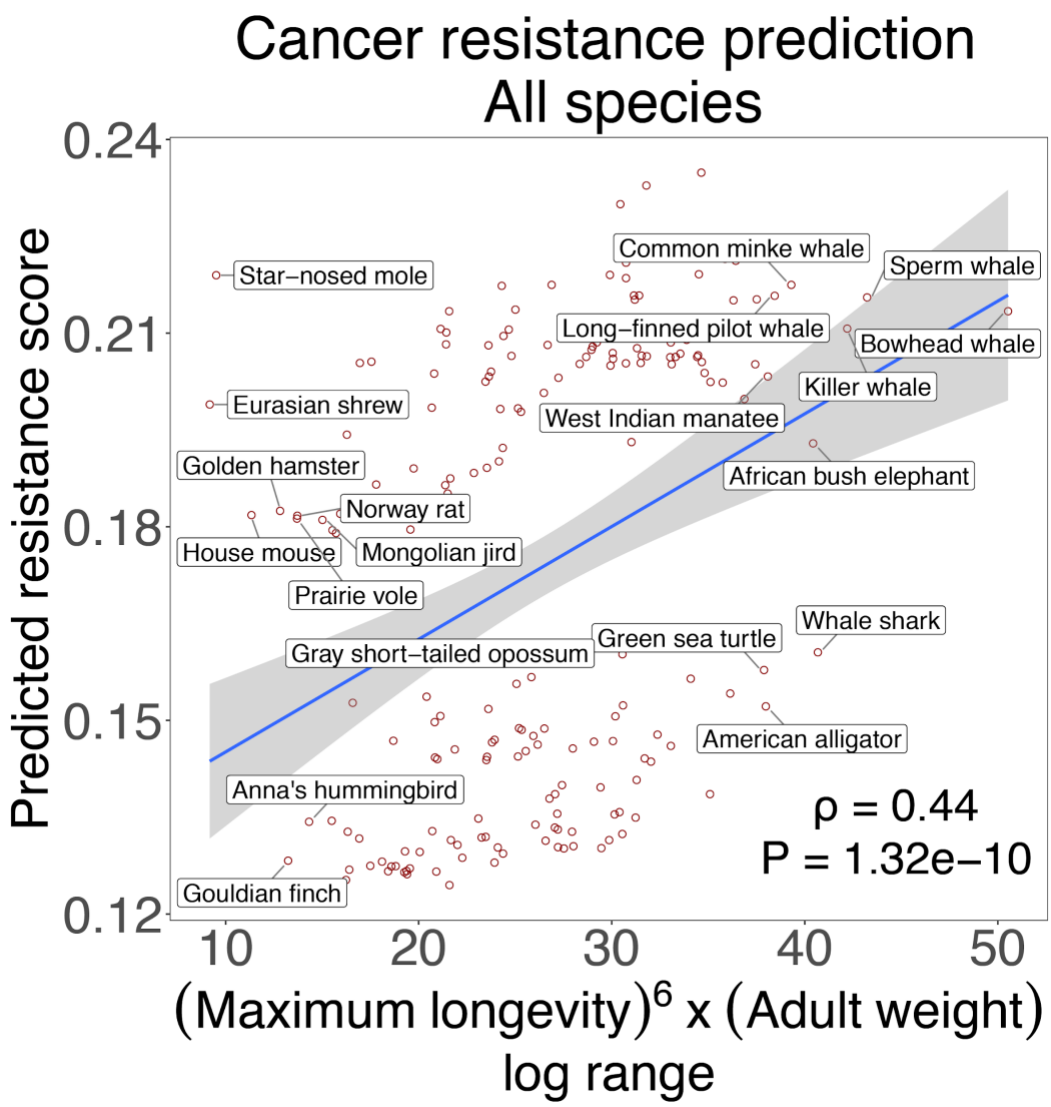


Figure S2: Scatter plots showing the correlation between the predicted cancer resistance (CR) scores computed based on gene conservation and for the cancer-resistance estimate MLTAW or '(Maximum longevity)^6 x (adult weight)', with leave-one-out cross-validation, for all species. *Species with the top and bottom 10% MLTAW values are labeled by their common names for the sake of clarity. Spearman's ρ and p-values (P) are shown.*

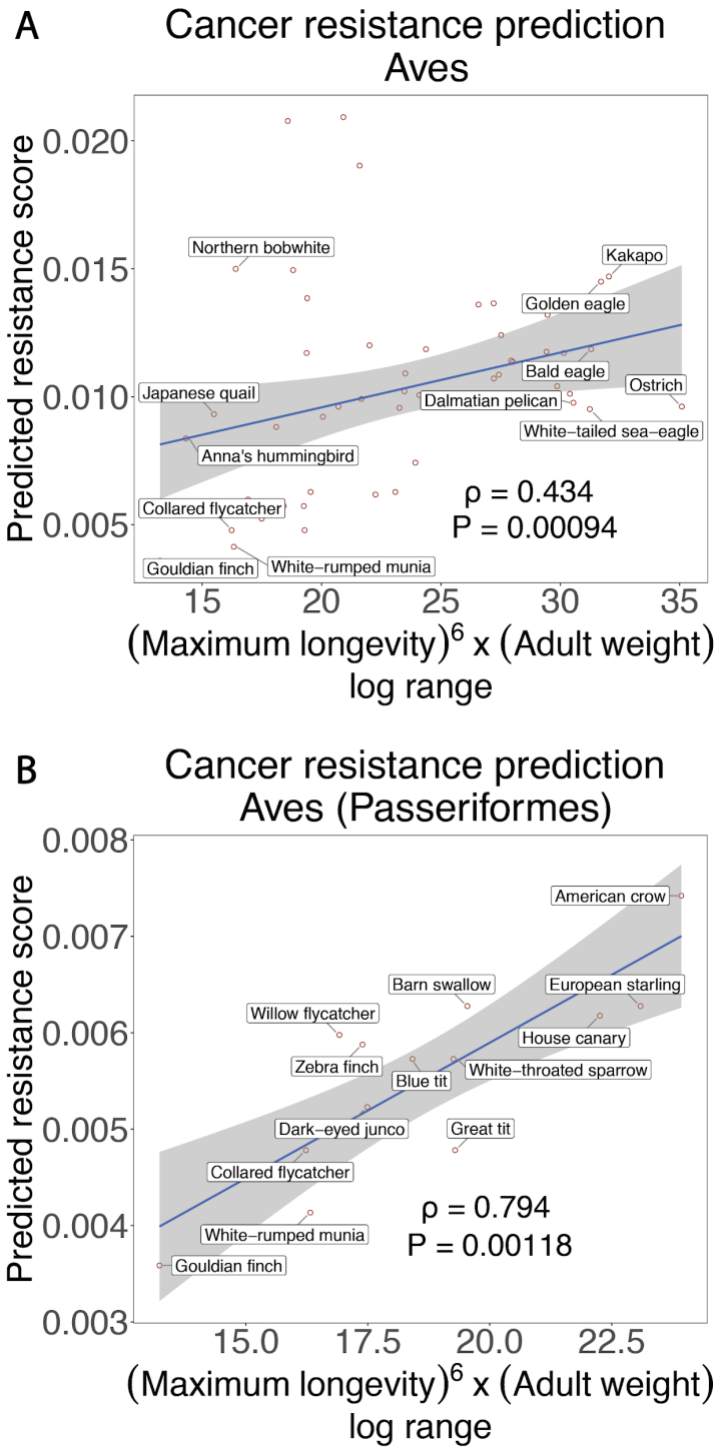


Figure S3: Scatter plots showing the correlation between the predicted cancer resistance (CR) scores computed based on gene conservation and for the cancer-resistance estimate MLTAW

for Aves (birds). (A) *Cancer resistance (CR) predictions are done by identifying PC/NC genes using only Aves (bird) species (in cross validation). Scatter plots showing the Spearman's correlation between the predicted cancer-resistance estimates and '(Maximum longevity)^6 x (adult weight)' or MLTAW is shown. Only species names for the top and bottom 10 percentile of the MLTAW measure are labeled for display clarity.* ***(B)*** *Scatter plots using the predicted scores in (A) are shown for only order Passeriformes (n=14) within the bird species. Spearman's ρ and p-values (P) are reported for (E, F). No CR predictor was built for birds using the MLCAW measure as we did not identify any PC/NC genes at (FDR < 0.1) for this measure.*

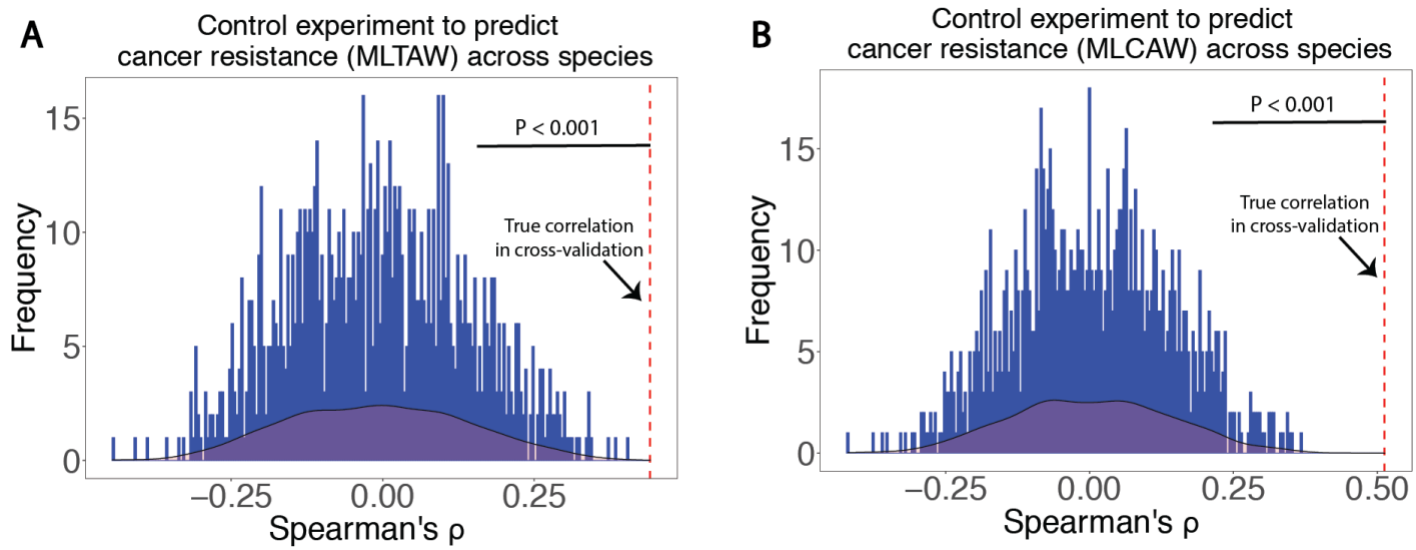


Figure S4: Random controls experiments for predicting cancer resistance using all species. *We chose random PC/NC genes of the same size as the actual PC/NC genes identified from the all-species analysis at FDR < 0.1. We can predict cancer resistance using these genes. We do this for 1000 iterations and the empirical P-value is computed. We see that they are not correlated in comparison to the 'true' correlation obtained using the actual PC/NC genes (randomization test P < 0.001). Cancer-resistance estimates used are: (A) MLTAW; (B) MLCAW.*

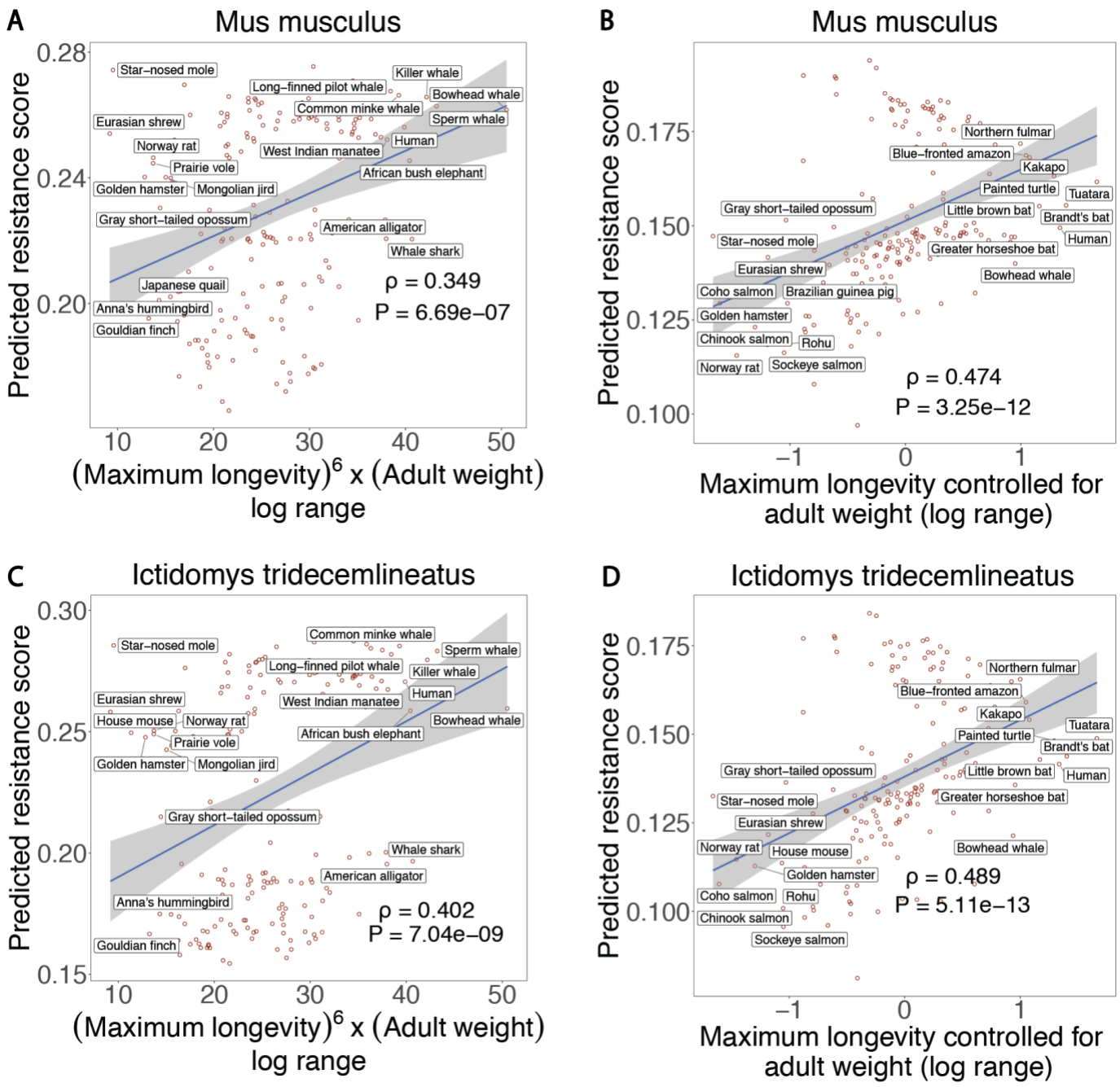


Figure S5: Instead of human reference to compute gene conservation scores, we use Mus musculus (house mouse) and thirteen-lined ground squirrel (Ictidomys tridecemlineatus) as references and predict cancer resistance in the all-species analysis. *Scatter plots along with the*

Spearman's correlation between the predicted cancer-resistance estimates (in cross-validation, LOOCV) and the cancer-resistance estimates like ***(A)*** *MLTAW or '(Maximum longevity)^6 x (adult weight)' for mouse;* ***(B)*** *MLCAW or 'Maximum longevity controlled for adult weight' for mouse;* ***(C)*** *MLTAW for squirrel;* ***(D)*** *MLCAW for squirrel, are shown. Both Spearman's ρ and p-values (P) are reported. Only species names for the top and bottom 5 percentile of the MLTAW/MLCAW measures are labeled for display clarity. We see that humans are predicted to be relatively cancer resistant as expected. The results obtained are quite similar to the corresponding results using human reference.*

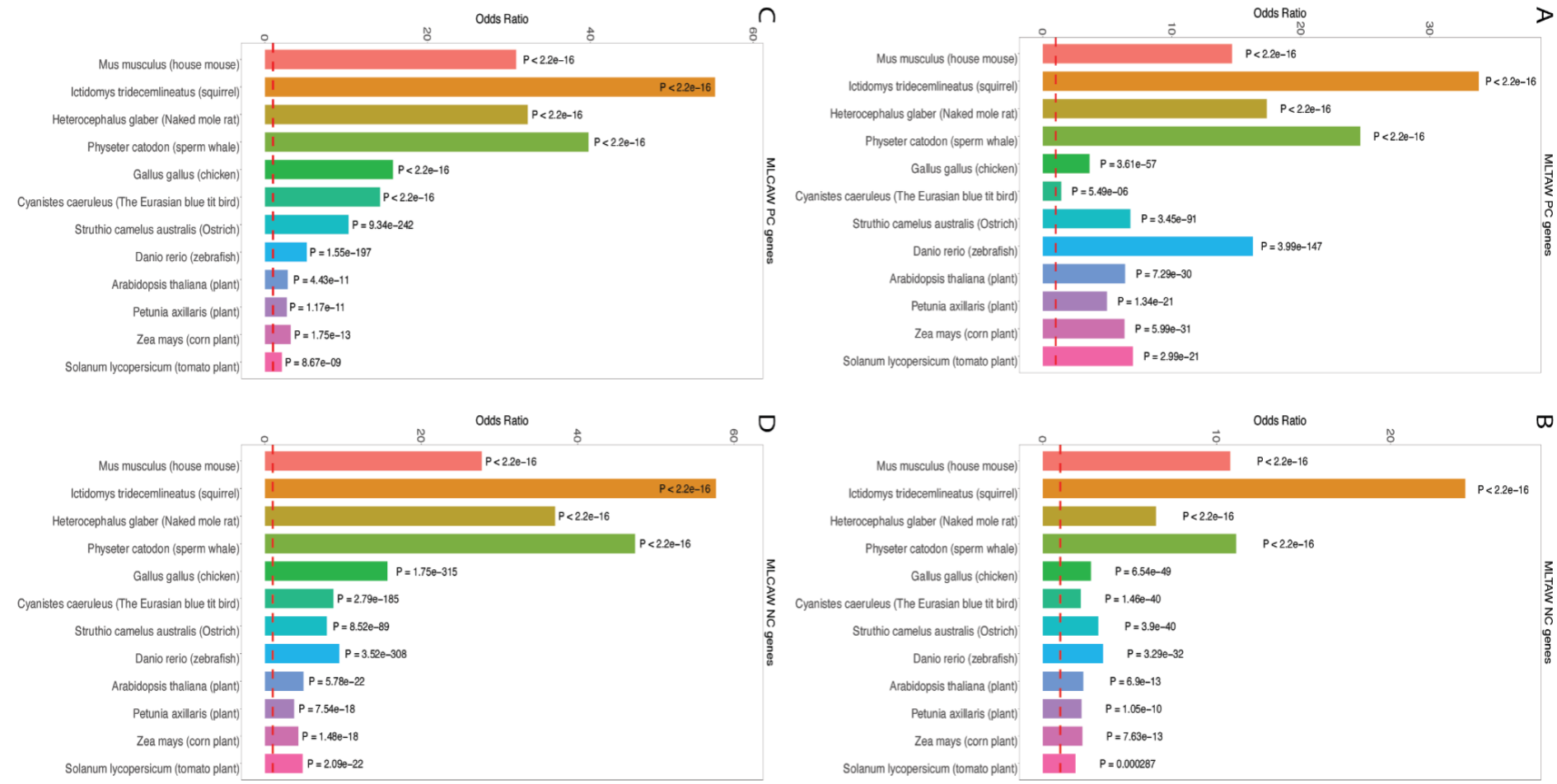

A
Odds Ratio
0
10
20
30
Mus musculus (house mouse)
Ictidomys tridecemlineatus (squirrel)
Heterocephalus glaber (Naked mole rat)
Physeter catodon (sperm whale)
Gallus gallus (chicken)
Cyanistes caeruleus (The Eurasian blue tit bird)
Struthio camelus australis (Ostrich)
Danio rerio (zebrafish)
Arabidopsis thaliana (plant)
Petunia axillaris (plant)
Zea mays (corn plant)
Solanum lycopersicum (tomato plant)
P < 2.2e−16
P < 2.2e−16
P < 2.2e−16
P < 2.2e−16
P = 3.61e−57
P = 5.49e−06
P = 3.45e−91
P = 3.99e−147
P = 7.29e−30
P = 1.34e−21
P = 5.99e−31
P = 2.99e−21
MLTAW PC genes
B
Odds Ratio
0
10
20
Mus musculus (house mouse)
Ictidomys tridecemlineatus (squirrel)
Heterocephalus glaber (Naked mole rat)
Physeter catodon (sperm whale)
Gallus gallus (chicken)
Cyanistes caeruleus (The Eurasian blue tit bird)
Struthio camelus australis (Ostrich)
Danio rerio (zebrafish)
Arabidopsis thaliana (plant)
Petunia axillaris (plant)
Zea mays (corn plant)
Solanum lycopersicum (tomato plant)
P < 2.2e−16
P < 2.2e−16
P < 2.2e−16
P < 2.2e−16
P = 6.54e−49
P = 1.46e−40
P = 3.9e−40
P = 3.29e−32
P = 6.9e−13
P = 1.05e−10
P = 7.63e−13
P = 0.000287
MLTAW NC genes
C
Odds Ratio
0
20
40
60
Mus musculus (house mouse)
Ictidomys tridecemlineatus (squirrel)
Heterocephalus glaber (Naked mole rat)
Physeter catodon (sperm whale)
Gallus gallus (chicken)
Cyanistes caeruleus (The Eurasian blue tit bird)
Struthio camelus australis (Ostrich)
Danio rerio (zebrafish)
Arabidopsis thaliana (plant)
Petunia axillaris (plant)
Zea mays (corn plant)
Solanum lycopersicum (tomato plant)
P < 2.2e−16
P < 2.2e−16
P < 2.2e−16
P < 2.2e−16
P < 2.2e−16
P < 2.2e−16
P = 9.34e−242
P = 1.55e−197
P = 4.43e−11
P = 1.17e−11
P = 1.75e−13
P = 8.67e−09
MLCAW PC genes
D
Odds Ratio
0
20
40
60
Mus musculus (house mouse)
Ictidomys tridecemlineatus (squirrel)
Heterocephalus glaber (Naked mole rat)
Physeter catodon (sperm whale)
Gallus gallus (chicken)
Cyanistes caeruleus (The Eurasian blue tit bird)
Struthio camelus australis (Ostrich)
Danio rerio (zebrafish)
Arabidopsis thaliana (plant)
Petunia axillaris (plant)
Zea mays (corn plant)
Solanum lycopersicum (tomato plant)
P < 2.2e−16
P < 2.2e−16
P < 2.2e−16
P < 2.2e−16
P = 1.75e−315
P = 2.79e−185
P = 8.52e−89
P = 3.52e−308
P = 5.78e−22
P = 7.54e−18
P = 1.48e−18
P = 2.09e−22
MLCAW NC genes

Figure S6: Robustness analysis using 12 non-human species as reference for computing the gene conservation scores. *Bar plots showing the odds ratio (OR) and p-values for this gene overlap enrichment test (Fisher exact test) for: (a) all-species PC genes for MLTAW, (b) all-species NC genes for MLTAW, (c) all-species PC genes for MLCAW, (d) all-species NC genes for MLCAW; identified from gene conservation scores computed using these 12 different references in comparison to identifying the corresponding PC/NC genes using humans as reference. We get significant overlap in all species (FDR < 0.0005). The dashed red line is at OR=1; OR > 1 signifies enrichment of gene sets.*

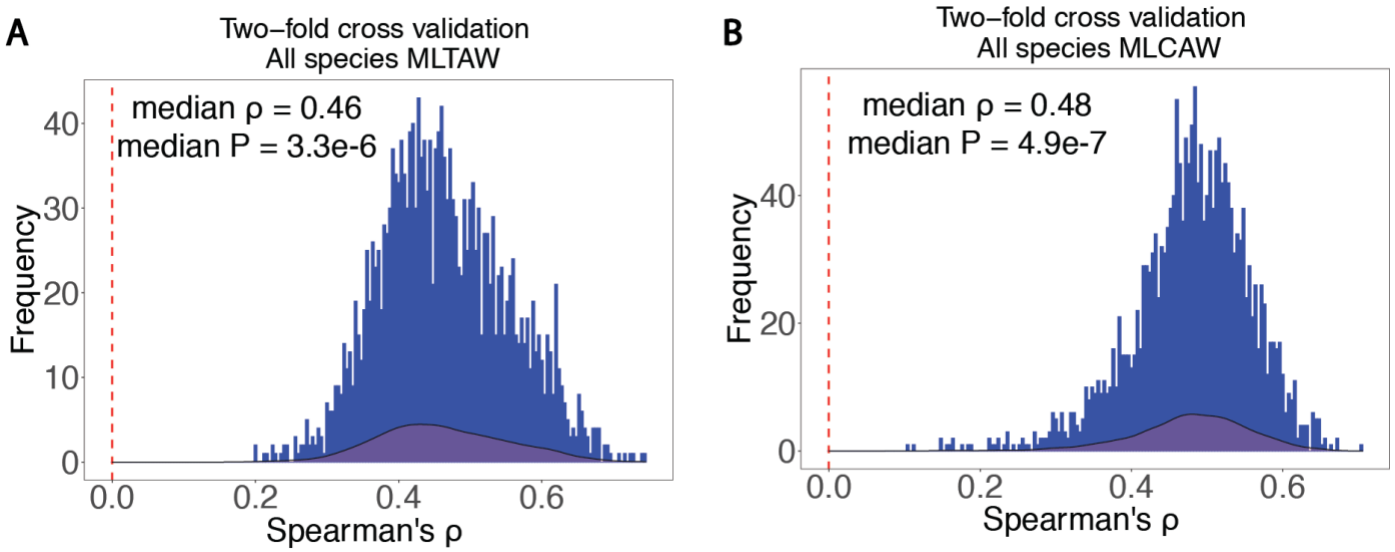


Figure S7: Plots show the distribution/frequency of Spearman's ρ between the predicted cancer resistance (CR) scores computed based on gene conservation and each of the two cancer-resistance estimates, using two-fold cross-validation (instead of LOOCV). *(A) MLTAW, i.e., (Maximum longevity)^6 x (adult weight); (B) MLCAW, i.e., maximum longevity controlled for adult weight). The two-fold cross validation was carried out 1000 times (2000 data points). Median Spearman's ρ and p-values (P) are mentioned.*

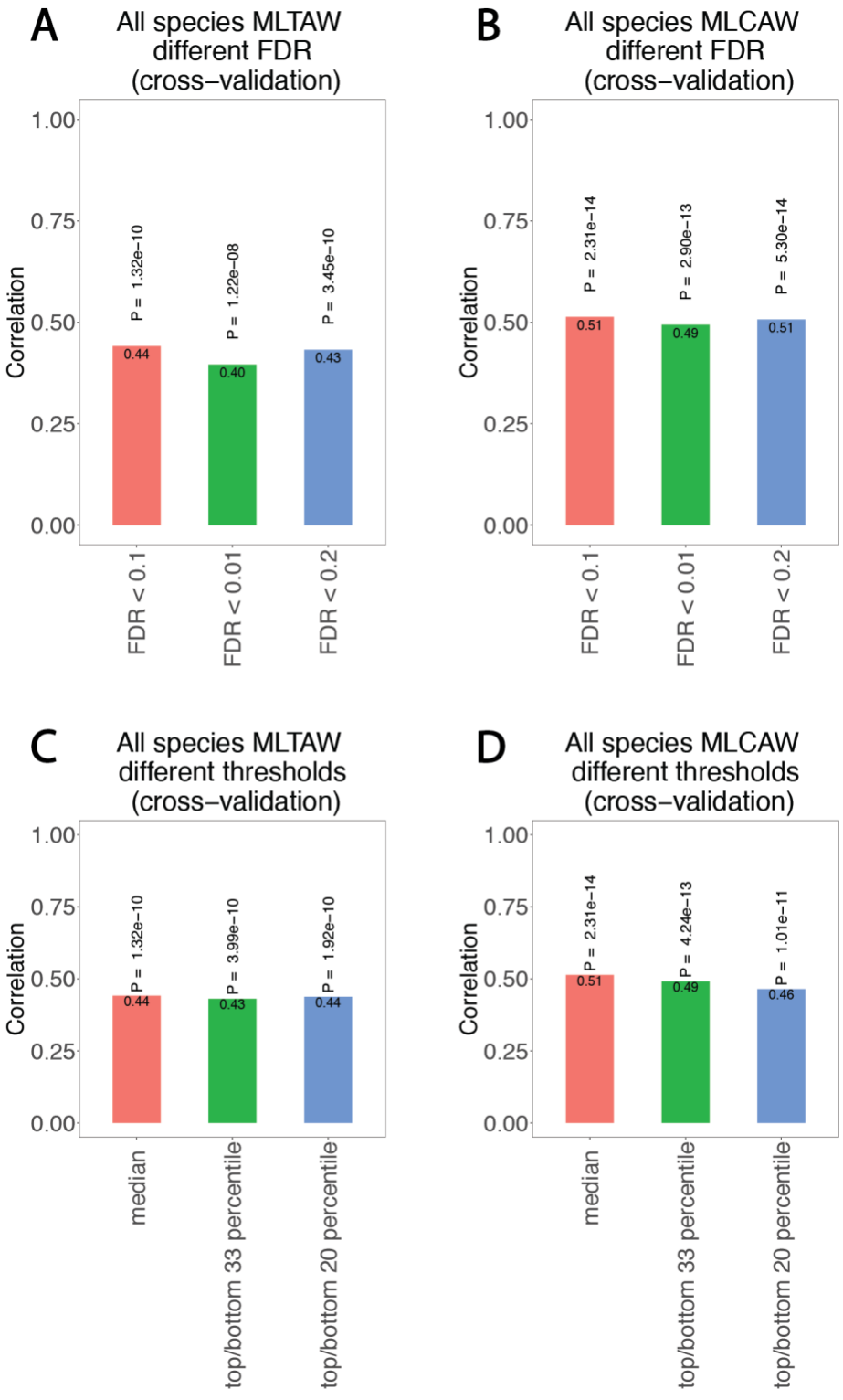

A
All species MLTAW different FDR (cross−validation)
Correlation
1.00
0.75
0.50
0.25
0.00
P = 1.32e−10
P = 1.22e−08
P = 3.45e−10
0.44
0.40
0.43
FDR < 0.1
FDR < 0.01
FDR < 0.2
B
All species MLCAW different FDR (cross−validation)
P = 2.31e−14
P = 2.90e−13
P = 5.30e−14
0.51
0.49
0.51
C
All species MLTAW different thresholds (cross−validation)
P = 1.32e−10
P = 3.99e−10
P = 1.92e−10
0.44
0.43
0.44
median
top/bottom 33 percentile
top/bottom 20 percentile
D
All species MLCAW different thresholds (cross−validation)
P = 2.31e−14
P = 4.24e−13
P = 1.01e−11
0.51
0.49
0.46

Figure S8: Predicting cancer-resistance (CR) scores by altering various parameters. *CR predictors for different FDR thresholds (0.01, 0.01, or 0.2) are shown using* ***(A)*** *MLTAW; and* ***(B)*** *MLCAW measures. The original predictor also computes the number of PC genes whose conservation score > median conservation score; and the number of NC genes whose conservation score < median conservation score. CR predictor results are shown to be robust by altering the thresholds from median conservation score to top and bottom 33/20 percentile of the conservation scores for PC and NC genes respectively (****C, D****). Both Spearman's ρ and p-values (P) are reported.*

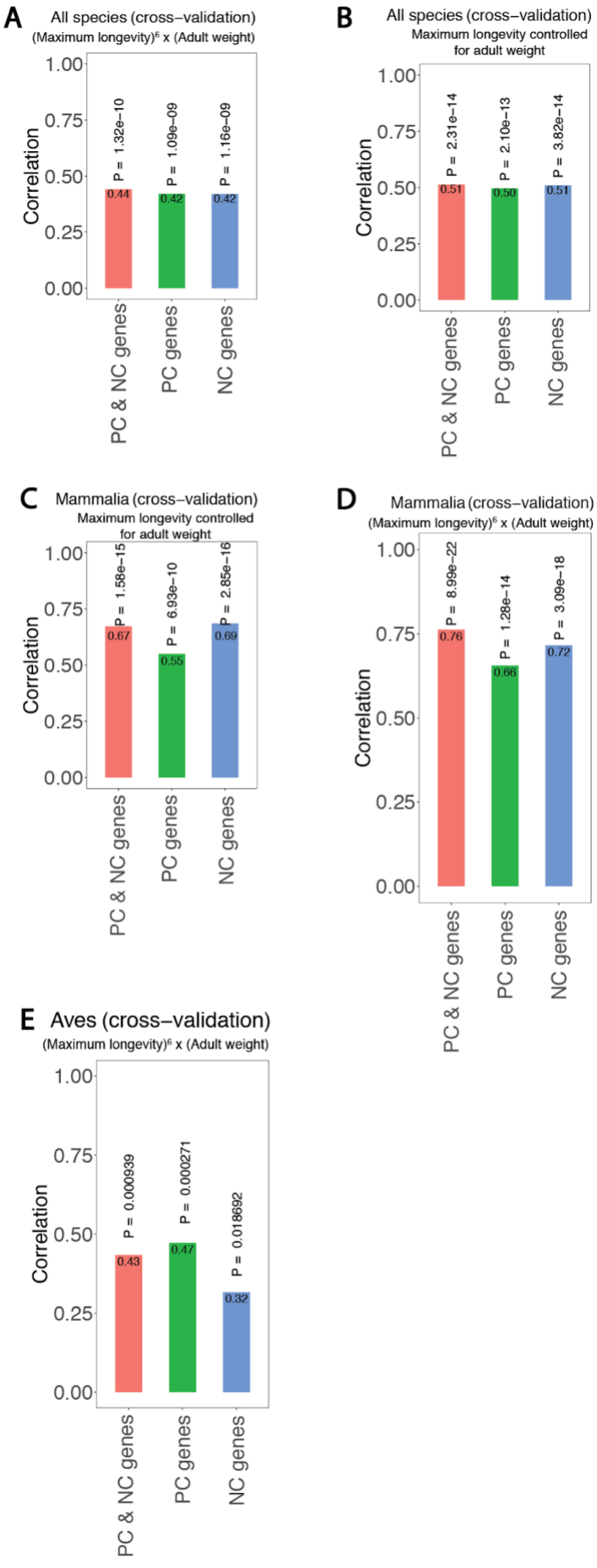
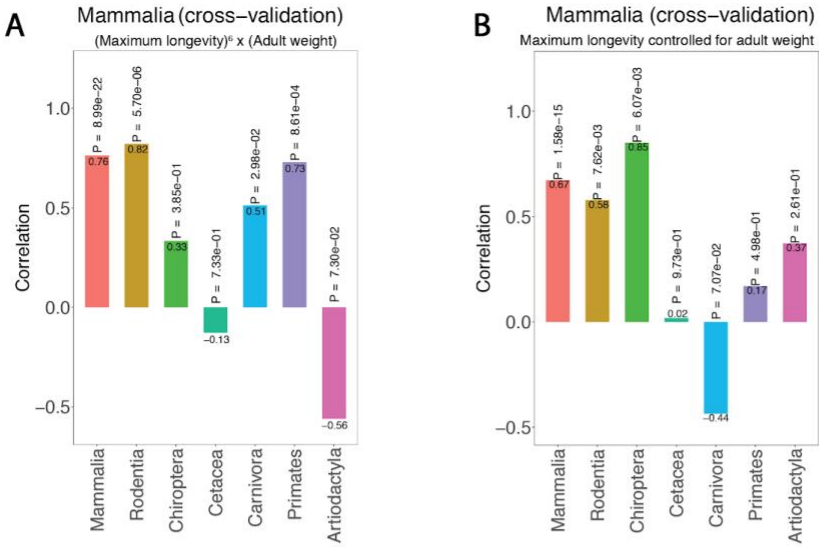
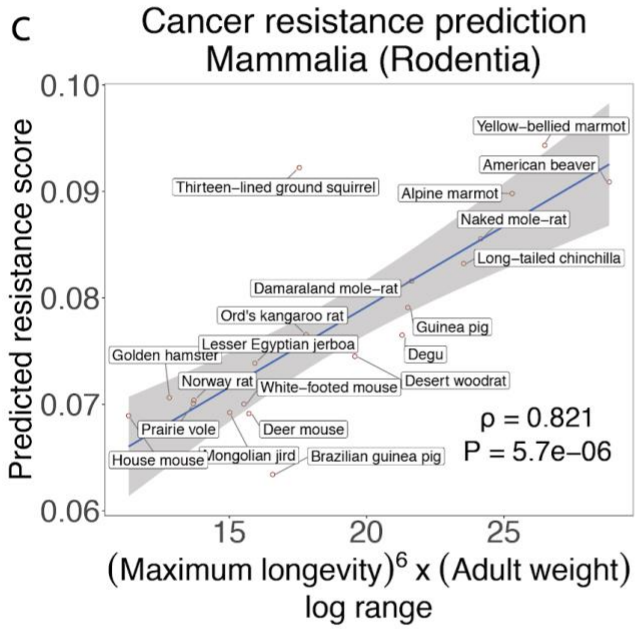
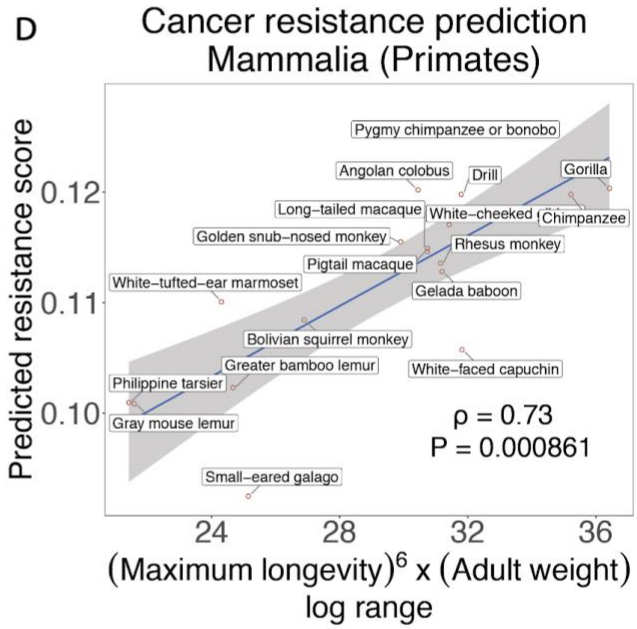

A
All species (cross−validation)
(Maximum longevity)6 x (Adult weight)
Correlation
1.00
0.75
0.50
0.25
0.00
P = 1.32e−10
P = 1.09e−09
P = 1.16e−09
0.44
0.42
0.42
PC & NC genes
PC genes
NC genes
B
All species (cross−validation)
Maximum longevity controlled
for adult weight
P = 2.31e−14
P = 2.10e−13
P = 3.82e−14
0.51
0.50
0.51
C
Mammalia (cross−validation)
Maximum longevity controlled
for adult weight
P = 1.58e−15
P = 6.93e−10
P = 2.85e−16
0.67
0.55
0.69
D
Mammalia (cross−validation)
(Maximum longevity)6 x (Adult weight)
P = 8.99e−22
P = 1.28e−14
P = 3.09e−18
0.76
0.66
0.72
E
Aves (cross−validation)
(Maximum longevity)6 x (Adult weight)
P = 0.000939
P = 0.000271
P = 0.018692
0.43
0.47
0.32

Figure S9: Predicting cancer-resistance (CR) scores by using both PC and NC genes (PC & NC genes); PC genes only; NC genes only. *Results for MLTAW or '(Maximum longevity)^6 x (adult weight)'; and MLCAW or 'Maximum longevity controlled for adult weight' are shown for the all-species analysis* ***(A,B)****, Mammalia-only analysis* ***(C,D)****, and Aves-only* ***(E)*** *analysis. MLCAW analysis is not shown for birds as we could not identify PC/NC genes at FDR < 0.1 and therefore could not build a CR predictor.*


A
Mammalia (cross−validation)
(Maximum longevity)^6 x (Adult weight)
Correlation
1.0
0.5
0.0
−0.5
P = 8.99e−22
0.76
P = 5.70e−06
0.82
P = 3.85e−01
0.33
P = 7.33e−01
−0.13
P = 2.98e−02
0.51
P = 8.61e−04
0.73
P = 7.30e−02
−0.56
Mammalia
Rodentia
Chiroptera
Cetacea
Carnivora
Primates
Artiodactyla
B
Mammalia (cross−validation)
Maximum longevity controlled for adult weight
Correlation
1.0
0.5
0.0
−0.5
P = 1.58e−15
0.67
P = 7.62e−03
0.58
P = 6.07e−03
0.85
P = 9.73e−01
0.02
P = 7.07e−02
−0.44
P = 4.98e−01
0.17
P = 2.61e−01
0.37
Mammalia
Rodentia
Chiroptera
Cetacea
Carnivora
Primates
Artiodactyla


C
Cancer resistance prediction
Mammalia (Rodentia)
Predicted resistance score
0.10
0.09
0.08
0.07
0.06
Yellow−bellied marmot
American beaver
Thirteen−lined ground squirrel
Alpine marmot
Naked mole−rat
Long−tailed chinchilla
Damaraland mole−rat
Ord's kangaroo rat
Guinea pig
Lesser Egyptian jerboa
Golden hamster
Degu
Norway rat
White−footed mouse
Desert woodrat
Prairie vole
Deer mouse
House mouse
Mongolian jird
Brazilian guinea pig
ρ = 0.821
P = 5.7e−06
15
20
25
(Maximum longevity)^6 x (Adult weight)
log range


D
Cancer resistance prediction
Mammalia (Primates)
Predicted resistance score
0.12
0.11
0.10
Pygmy chimpanzee or bonobo
Angolan colobus
Drill
Gorilla
Long−tailed macaque
White−cheeked (
Chimpanzee
Golden snub−nosed monkey
Rhesus monkey
Pigtail macaque
White−tufted−ear marmoset
Gelada baboon
Bolivian squirrel monkey
Greater bamboo lemur
White−faced capuchin
Philippine tarsier
Gray mouse lemur
Small−eared galago
ρ = 0.73
P = 0.000861
24
28
32
36
(Maximum longevity)^6 x (Adult weight)
log range

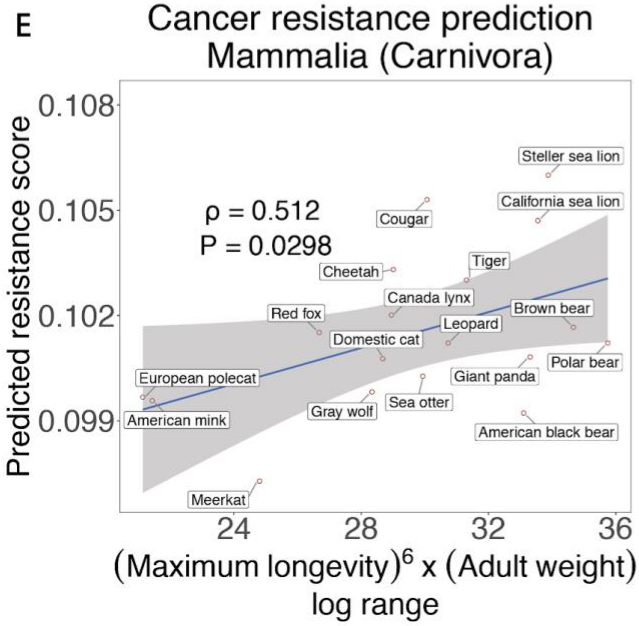

E
Cancer resistance prediction
Mammalia (Carnivora)
Predicted resistance score
0.108
0.105
0.102
0.099
Steller sea lion
California sea lion
Cougar
ρ = 0.512
P = 0.0298
Tiger
Cheetah
Canada lynx
Brown bear
Red fox
Leopard
Domestic cat
Polar bear
European polecat
Giant panda
Sea otter
Gray wolf
American mink
American black bear
Meerkat
24
28
32
36
(Maximum longevity)^6 x (Adult weight)
log range

Figure S10: Cancer resistance predictions on the entire mammalian species. *Cancer resistance predictions were done on the entire mammalian species (LOOCV, by learning PC/NC genes from mammals). Using these predictions, Spearman's correlation (ρ and p-values) for different orders (sub-groups) of mammals: Rodentia (rodents), Chiroptera (bats), Cetacea (aquatic mammals like whales), Carnivora (carnivores), Primates, Artiodactyla (even-toed hoofed mammals) are shown for (A) MLTAW or '(Maximum longevity)^6 x (adult weight)' and (B) MLCAW or 'Maximum longevity controlled for adult weight'. Scatter plots showing the Spearman's correlation between the predicted cancer-resistance estimates and the MLTAW cancer-resistance estimate for some of the orders are shown in (C-E). Spearman's ρ and p-values (P) are reported.*

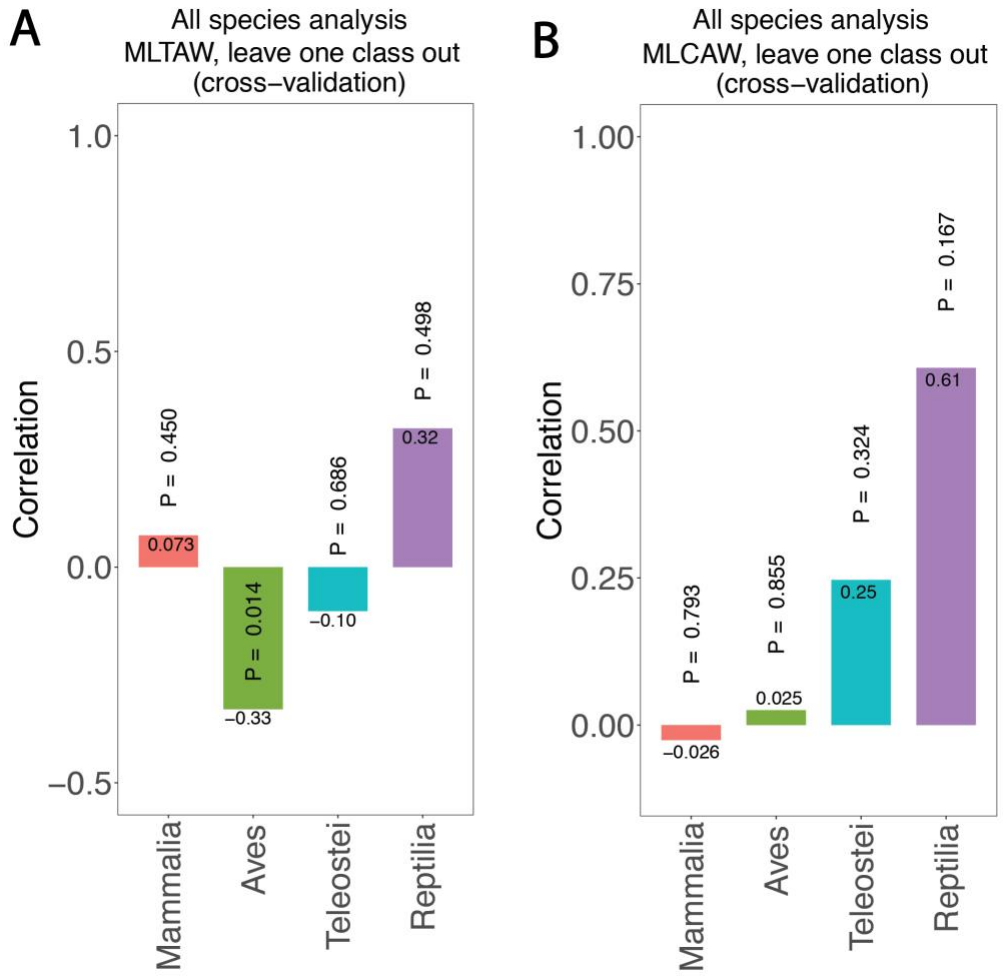


Figure S11: Predicting cancer resistance (CR) scores by identifying PC/NC genes by leaving out one class and testing on that left-out class (cross-validation; all-species analysis). *We show the accuracies for the following classes: Mammalia (mammals), Aves (birds), Teleostei (fish), and Reptilia (reptiles). Spearman's ρ and p-values (P) are reported using the two cancer-resistance estimates:* ***(A)*** *MLTAW; and* ***(B)*** *MLCAW.*

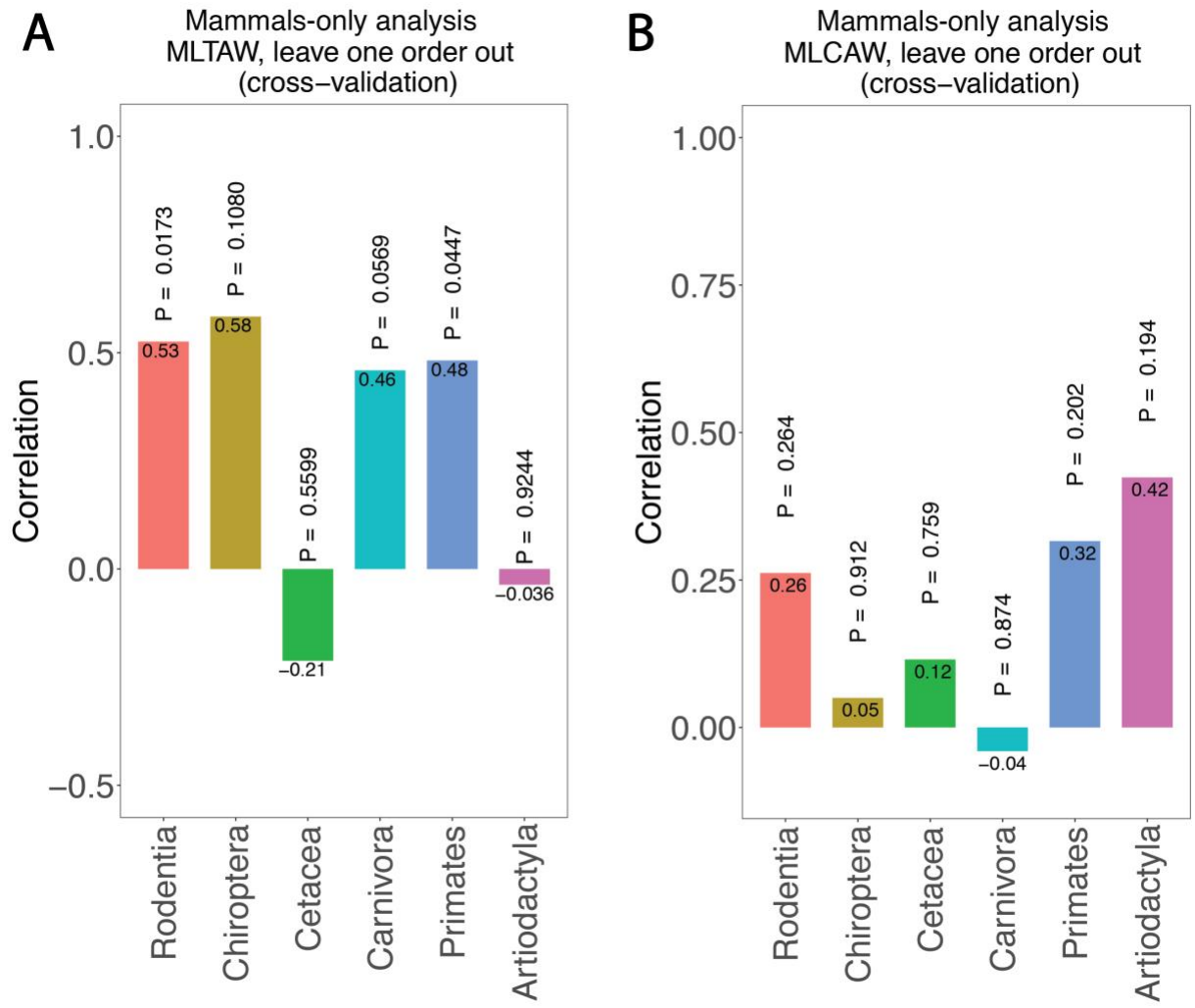
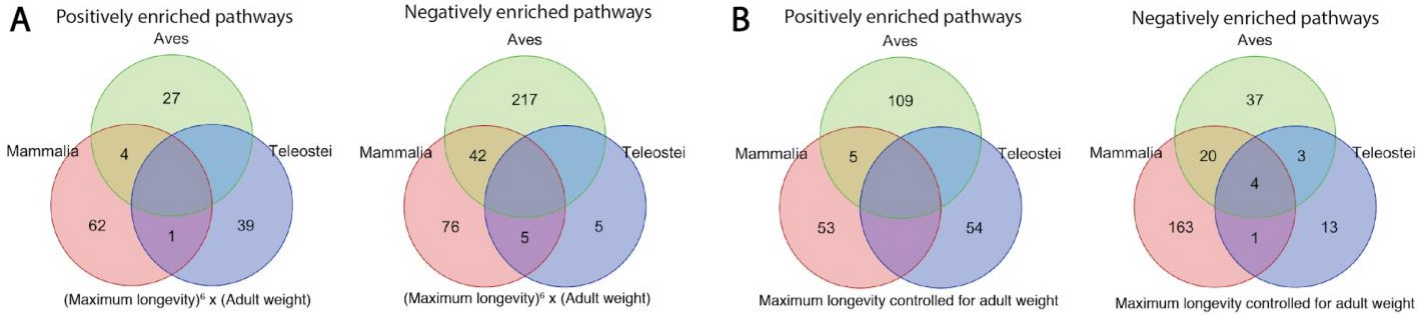
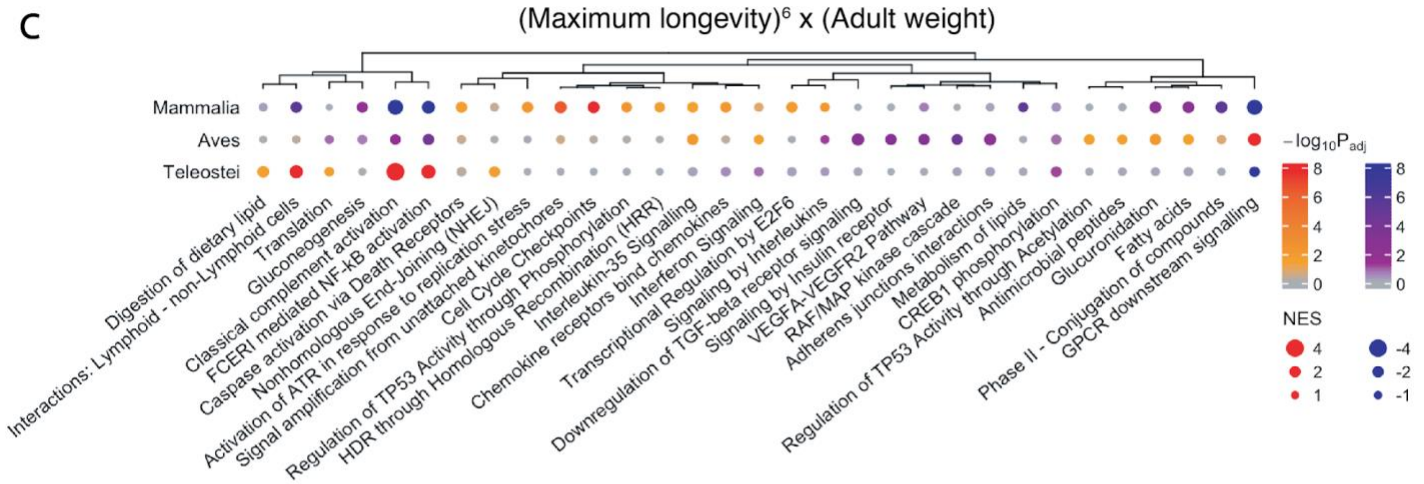
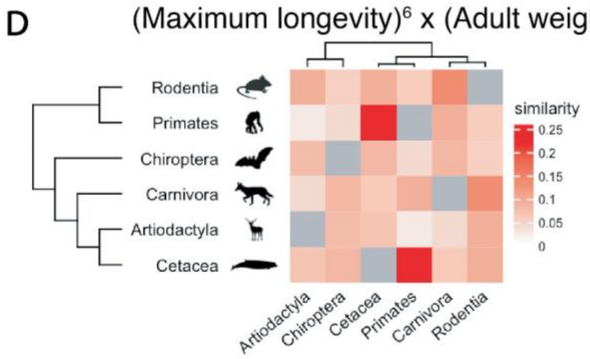


Figure S12: Predicting cancer resistance (CR) scores by identifying PC/NC genes (using mammalian data) by leaving out one order of mammals and testing on that left-out order (cross-validation; mammals-only analysis). *We show the accuracies for the following orders: Rodentia (rodents), Chiroptera (bats), Cetacea (aquatic mammals like whales), Carnivora (carnivores), Primates, Artiodactyla (even-toed hoofed mammals). Spearman's ρ and p-values (P) are reported using the two cancer-resistance estimates:* ***(A)*** *MLTAW; and* ***(B)*** *MLCAW.*


A
Positively enriched pathways
Aves
27
Mammalia
4
Teleostei
62
1
39
(Maximum longevity)⁶ x (Adult weight)
Negatively enriched pathways
Aves
217
Mammalia
42
Teleostei
76
5
5
(Maximum longevity)⁶ x (Adult weight)
B
Positively enriched pathways
Aves
109
Mammalia
5
Teleostei
53
54
Maximum longevity controlled for adult weight
Negatively enriched pathways
Aves
37
Mammalia
20
3
Teleostei
4
163
1
13
Maximum longevity controlled for adult weight


C
(Maximum longevity)⁶ x (Adult weight)
Mammalia
Aves
Teleostei
Digestion of dietary lipid
Interactions: Lymphoid - non-Lymphoid cells
Translation
Gluconeogenesis
Classical complement activation
FCERI mediated NF-kB activation
Caspase activation via Death Receptors
Nonhomologous End-Joining (NHEJ)
Activation of ATR in response to replication stress
Signal amplification from unattached kinetochores
Regulation of TP53 Activity through Phosphorylation
HDR through Homologous Recombination (HRR)
Cell Cycle Checkpoints
Interleukin-35 Signalling
Chemokine receptors bind chemokines
Interferon Signaling
Transcriptional Regulation by E2F6
Downregulation of TGF-beta receptor signaling
Signaling by Interleukins
Signaling by Insulin receptor
VEGFA-VEGFR2 Pathway
RAF/MAP kinase cascade
Adherens junctions interactions
Metabolism of lipids
CREB1 phosphorylation
Regulation of TP53 Activity through Acetylation
Antimicrobial peptides
Glucuronidation
Phase II - Conjugation of compounds
Fatty acids
GPCR downstream signalling
−log10Padj
8
6
4
2
0
NES
4
2
1
-4
-2
-1


D
(Maximum longevity)⁶ x (Adult weight)
Rodentia
Primates
Chiroptera
Carnivora
Artiodactyla
Cetacea
similarity
0.25
0.2
0.15
0.1
0.05
0
Artiodactyla
Chiroptera
Cetacea
Primates
Carnivora
Rodentia

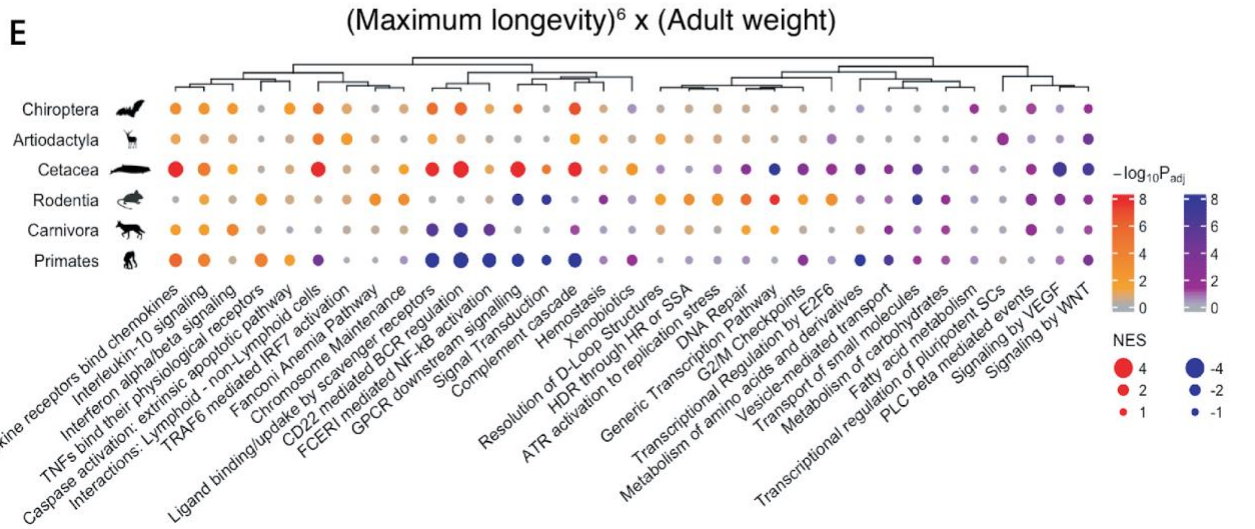

E
(Maximum longevity)⁶ x (Adult weight)
Chiroptera
Artiodactyla
Cetacea
Rodentia
Carnivora
Primates
Chemokine receptors bind chemokines
Interleukin-10 signaling
Interferon alpha/beta signaling
TNFs bind their physiological receptors
Caspase activation: extrinsic apoptotic pathway
Interactions: Lymphoid - non-Lymphoid cells
TRAF6 mediated IRF7 activation
Fanconi Anemia Pathway
Chromosome Maintenance
Ligand binding/updake by scavenger receptors
CD22 mediated BCR regulation
FCERI mediated NF-kB activation
GPCR downstream signalling
Signal Transduction
Complement cascade
Hemostasis
Xenobiotics
Resolution of D-Loop Structures
HDR through HR or SSA
ATR activation to replication stress
DNA Repair
Generic Transcription Pathway
G2/M Checkpoints
Transcriptional Regulation by E2F6
Metabolism of amino acids and derivatives
Vesicle-mediated transport
Transport of small molecules
Metabolism of carbohydrates
Fatty acid metabolism
Transcriptional regulation of pluripotent SCs
Signaling by VEGF
PLC beta mediated events
Signaling by WNT
−log10Padj
8
6
4
2
0
NES
4
2
1
-4
-2
-1

Figure S13: Gene set enrichment analysis (GSEA) of the correlation between the gene conservation scores and cancer-resistance estimates including (maximum longevity)^6 x (adult weight) (MLTAW), or the residue of maximum longevity after regressing out adult weight (MLCAW), for three classes of species: Mammalia (mammals), Aves (birds), and Teleostei (fish). ***(A,B)*** *Venn diagram showing the number of positively and negatively enriched gene sets in the three classes based on correlations with: (A) MLTAW and (B) MLCAW.* ***(C)*** *Summary visualization of the GSEA result for the top significantly enriched gene sets in the three classes (Mammalia, Aves, Teleostei) based on correlations with MLCAW. A selected subset of top gene sets are shown to save space, all with adjusted P<0.1 in at least one of the classes. GSEA significance (negative log10 adjusted P-values) is encoded by dot color, with two sets of colors (red-orange and blue-purple) representing positive or negative enrichment, respectively; grey color means adjusted P>=0.1. Dot size represents the absolute value of normalized enrichment scores (NES) measuring the effect size of enrichment. The complete GSEA results are given in Table S3.* ***(D)*** *GSEA analysis of the correlation between the gene conservation scores and cancer-resistance estimates such as MLTAW were performed for different orders of mammalian species including Rodentia (rodents), Primates (primates), Chiroptera (bats), Carnivora (carnivores), Artiodactyla (even-toed hoofed animals), and Cetacea (whales). A heatmap showing the similarity (Jaccard index) between the significantly enriched gene sets (FDR<0.1) from each pair of mammalian orders, based on the MLTAW correlation. The dendrogram on the left is the phylogenetic tree of the mammalian orders, and the rows of the heatmap are arranged accordingly. The dendrogram on the top represents the hierarchical clustering of the orders based on their similarities in the GSEA results.* ***(E)*** *Summary visualization of the GSEA result for the top significantly enriched gene sets in the mammalian orders based on MLTAW correlation. A selected subset of top gene sets are shown to save space, all with adjusted P<0.1 in at least one of the orders (complete results in Table S5). The color code and dot size are as described in (C).*

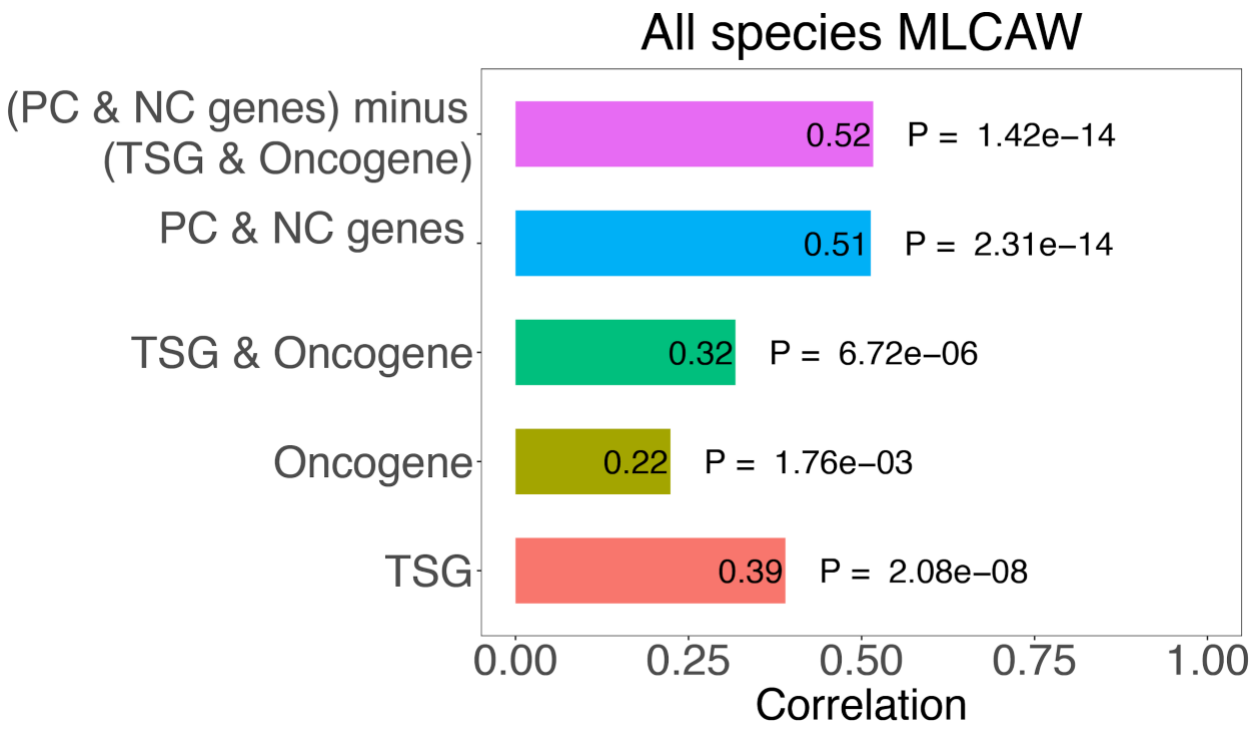


Figure S14: Spearman's correlation (ρ) in predicting cancer resistance (MLCAW) in all species using only TSGs, only oncogenes, both TSGs and oncogenes, using PC and NC genes in cross validation, using PC and NC genes after removing TSGs and oncogenes in cross validation is shown.

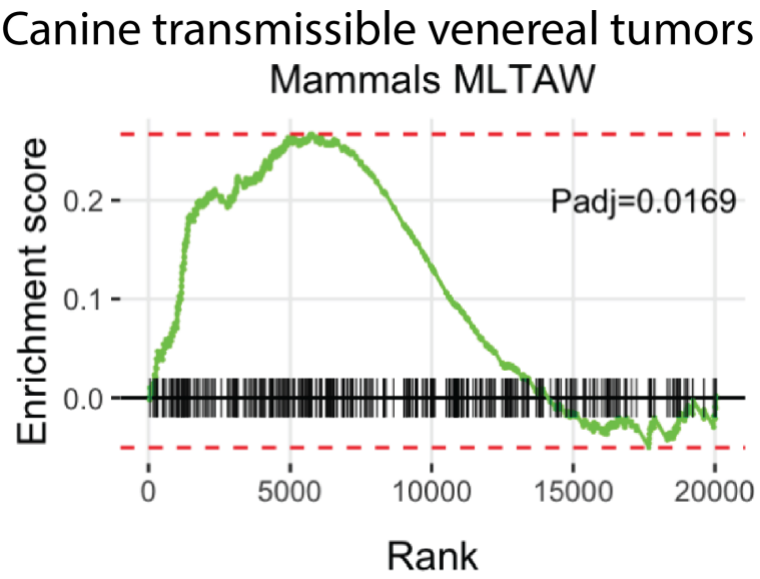


Figure S15: Looking at loss-of-function genes observed in canine transmissible venereal tumors. *Gene set enrichment analysis (GSEA) plot showing a significant enrichment of the PC genes in mammals (using MLTAW measure) for the loss-of-function genes observed in canine transmissible venereal tumors.*

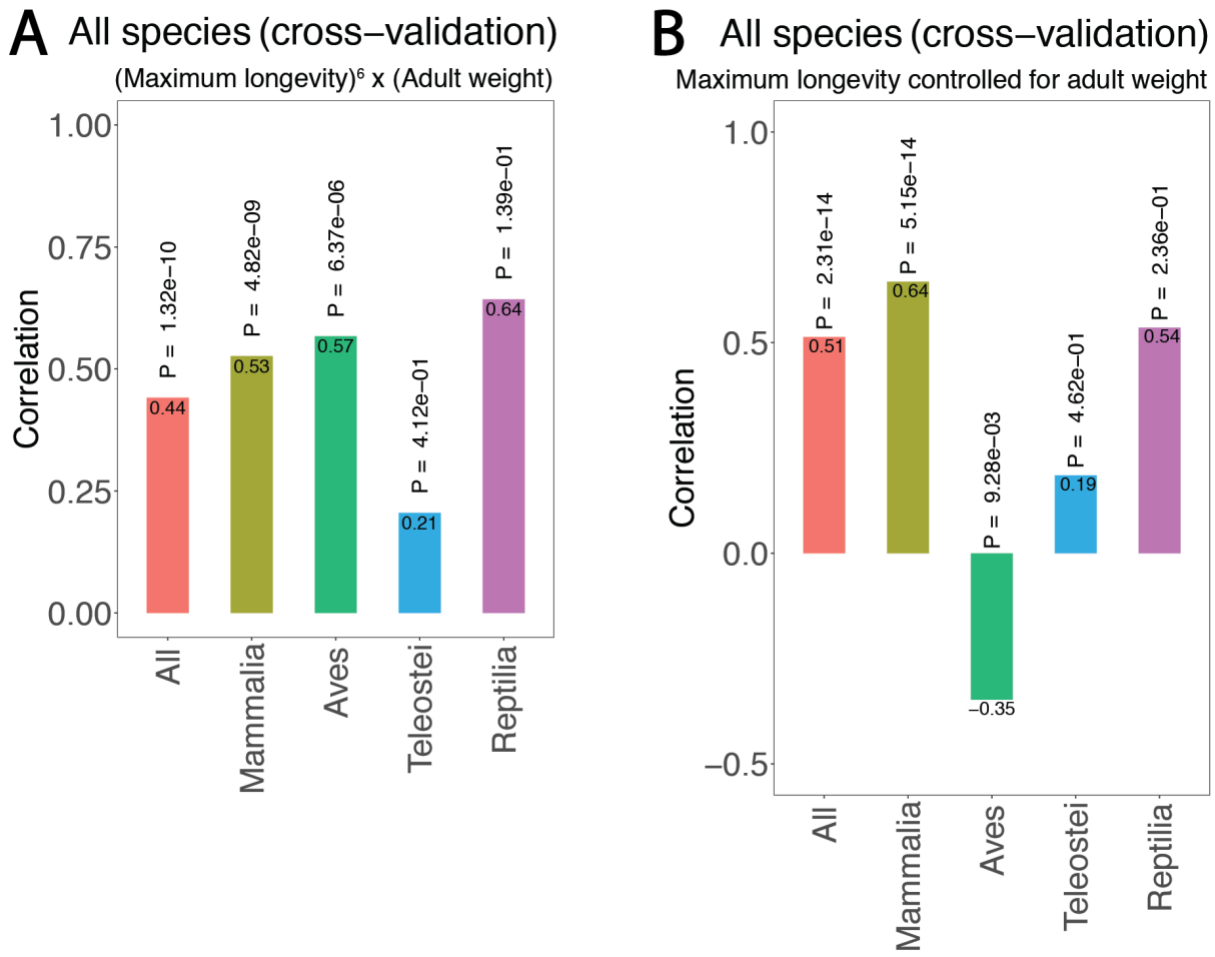


Figure S16: The cancer resistance (CR) scores predicted on all species (in leave-one-out cross-validation) analysis are individually tested on different classes of species: Mammalia (mammals), Aves (birds), Teleostei (fish), Reptilia (reptiles). *Spearman's ρ and p-values (P) are reported using the two cancer-resistance estimates:* ***(A)*** *MLTAW or '(Maximum longevity)^6 x (adult weight)'; and* ***(B)*** *MLCAW or 'Maximum longevity controlled for adult weight'.*

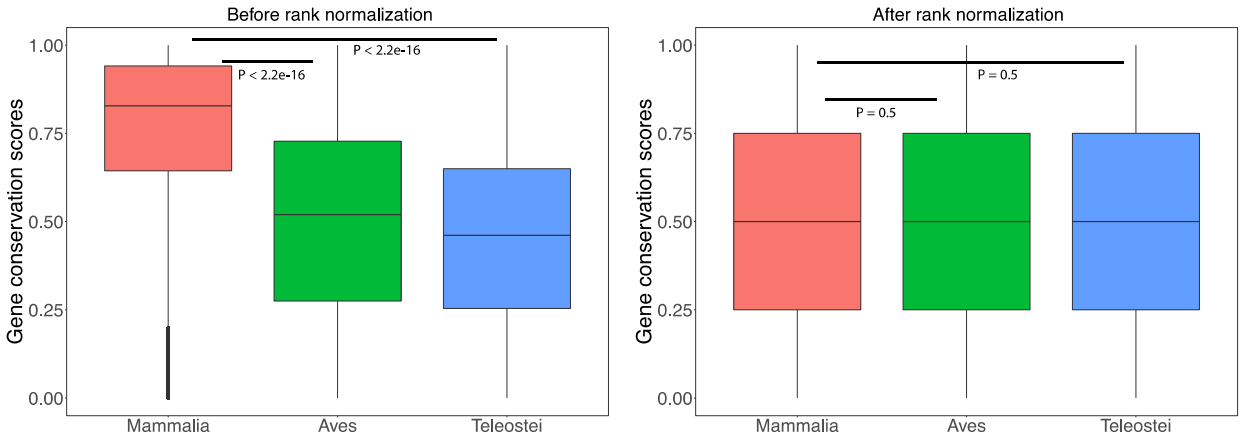


Figure S17: Gene conservation profiles for all genes and species, before (left sub-figure) and after (right sub-figure) rank-normalization. *Box plots show the range of gene conservation scores (obtained using humans as reference species) for mammals (Mammalia), birds (Aves), and fish (Teleostei). One-sided Wilcoxon rank-sum test comparing mammals to the other two classes are shown (whenever the p-value was computed as 0 by the Wilcoxon rank-sum test software, we wrote P < 2.2e-16 in the left figure).*

REFERENCES AND NOTES

1. T. A. F. Albuquerque, L. Drummond do Val, A. Doherty, J. P. de Magalhães, From humans to hydra: Patterns of cancer across the tree of life. *Biol. Rev. Camb. Philos. Soc.* **93**, 1715–1734 (2018).

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

3. A. Seluanov, V. N. Gladyshev, J. Vijg, V. Gorbunova, Mechanisms of cancer resistance in long-lived mammals. *Nat. Rev. Cancer* **18**, 433–441 (2018).

4. C. O. Nordling, A new theory on the cancer-inducing mechanism. *Br. J. Cancer* **7**, 68–72 (1953).

5. C. Tomasetti, L. Li, B. Vogelstein, Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355**, 1330–1334 (2017).

6. C. Tomasetti, B. Vogelstein, Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).

7. N. K. Khankari, X-O Shu, W. Wen, P. Kraft, S. Lindström, U. Peters, J. Schildkraut, F. Schumacher, P. Bofetta, A. Risch, H. Bickeböller, C. I. Amos, D. Easton, R. A. Eeles, S. B. Gruber, C. A. Haiman, D. J. Hunter, S. J. Chanock, B. L. Pierce, W. Zheng; Colorectal Transdisciplinary Study (CORECT); Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE); Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE); Transdisciplinary Research in Cancer of the Lung (TRICL), Association between adult height and risk of colorectal, lung, and prostate cancer: Results from meta-analyses of prospective studies and mendelian randomization analyses. *PLOS Med.* **13**, e1002118 (2016).

8. R. Peto, *Origins of Human Cancer* (Cold Spring Harbor Publications, 1977), pp. 1403–1428.

9. M. Tollis, A. M. Boddy, C. C. Maley, Peto's Paradox: How has evolution solved the problem of cancer prevention? *BMC Biol.* **15**, 60 (2017).

10. Y. Ikeno, G. B. Hubbard, S. Lee, L. A. Cortez, C. M. Lew, C. R. Webb, D. E. Berryman, E. O. List, J. J. Kopchick, A. Bartke, Reduced incidence and delayed occurrence of fatal neoplastic diseases in growth

hormone receptor/binding protein knockout mice. *J. Gerontol. A Biol. Sci. Med. Sci.* **64**, 522–529 (2009).

11. R. Lipman, A. Galecki, D. T. Burke, R. A. Miller, Genetic loci that influence cause of death in a heterogeneous mouse stock. *J. Gerontol. A Biol. Sci. Med. Sci.* **59**, 977–983 (2004).

12. H. Szymanska, J. Lechowska-Piskorowska, E. Krysiak, A. Strzalkowska, K. Unrug-Bielawska, B. Grygalewicz, H. M. Skurzak, B. Pienkowska-Grela, M. Gajewska, Neoplastic and nonneoplastic lesions in aging mice of unique and common inbred strains contribution to modeling of human neoplastic diseases. *Vet. Pathol.* **51**, 663–679 (2014).

13. M. Keane, J. Semeiks, A. E. Webb, Y. I. Li, V. Quesada, T. Craig, L. B. Madsen, S. van Dam, D. Brawand, P. I. Marques, P. Michalak, L. Kang, J. Bhak, H.-S. Yim, N. V. Grishin, N. H. Nielsen, M. P. Heide-Jørgensen, E. M. Oziolor, C. W. Matson, G. M. Church, G. W. Stuart, J. C. Patton, J. C. George, R. Suydam, K. Larsen, C. López-Otín, M. J. O'Connell, J. W. Bickham, B. Thomsen, J. P. de Magalhães, Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep.* **10**, 112–122 (2015).

14. J. C. George, J. Bada, J. Zeh, L. Scott, S. E. Brown, T. O'Hara, R. Suydam, Age and growth estimates of bowhead whales (Balaena mysticetus) via aspartic acid racemization. *Can. J. Zool.* **77**, 571–580 (1999).

15. A. Vicens, D. Posada, Selective pressures on human cancer genes along the evolution of mammals. *Genes* **9**, 582 (2018).

16. M. Tollis, A. K. Schneider-Utaka, C. C. Maley, The evolution of human cancer gene duplications across mammals. *Mol. Biol. Evol.* **37**, 2875–2886 (2020).

17. J. M. Vazquez, V. J. Lynch, Pervasive duplication of tumor suppressors in Afrotherians during the evolution of large bodies and reduced cancer risk. *eLife* **10**, e65041 (2021).

18. A. Kowalczyk, R. Partha, N. L. Clark, M. Chikina, Pan-mammalian analysis of molecular constraints underlying extended lifespan. *eLife* **9**, e51089 (2020).

19. E. Ferris, L. M. Abegglen, J. D. Schiffman, C. Gregg, Accelerated evolution in distinctive species reveals candidate elements for clinically relevant traits, including mutation and cancer resistance. *Cell Rep.* **22**, 2742–2755 (2018).

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

21. Y. Tabach, A. C. Billi, G. D. Hayes, M. A. Newman, O. Zuk, H. Gabel, R. Kamath, K. Yacoby, B. Chapman, S. M. Garcia, M. Borowsky, J. K. Kim, G. Ruvkun, Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. *Nature* **493**, 694–698 (2013).

22. Y. Tabach, T. Golan, A. Hernández-Hernández, A. R. Messer, T. Fukuda, A. Kouznetsova, J.-G. Liu, I. Lilienthal, C. Levy, G. Ruvkun, Human disease locus discovery and mapping to molecular pathways through phylogenetic profiling. *Mol. Syst. Biol.* **9**, 692 (2013).

23. UniProt Consortium, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

24. N. A O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, K. D. Pruitt, Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

25. K. L. Howe, P. Achuthan, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, M. Charkhchi, C. Cummins, L. D. R. Fioretto, C. Davidson, K. Dodiya, B. E. Houdaigui, R. Fatima, A. Gall, C. G. Giron, T. Grego, C. Guijarro-Clarke,

L. Haggerty, A. Hemrom, T. Hourlier, O. G. Izuogu, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, J. C. Marugán, T. Maurel, A. C. McMahon, S. Mohanan, B. Moore, M. Muffato, D. N. Oheh, D. Paraschas, A. Parker, A. Parton, I. Prosovetskaia, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, E. Steed, M. Szpak, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, B. Walts, A. Winterbottom, M. Chakiachvili, A. Chaubal, N. D. Silva, B. Flint, A. Frankish, S. E. Hunt, G. R. IIsley, N. Langridge, J. E. Loveland, F. J. Martin, J. M. Mudge, J. Morales, E. Perry, M. Ruffier, J. Tate, D. Thybert, S. J. Trevanion, F. Cunningham, A. D. Yates, D. R. Zerbino, P. Flicek, Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).

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

27. M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, T. O. Yeates, Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4285–4288 (1999).

28. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

29. D. Sherill-Rofe, D. Rahat, S. Findlay, A. Mellul, I. Guberman, M. Braun, I. Bloch, A. Lalezari, A. Samiei, R. Sadreyev, M. Goldberg, A. Orthwein, A. Zick, Y. Tabach, Mapping global and local coevolution across 600 species to identify novel homologous recombination repair genes. *Genome Res.* **29**, 439–448 (2019).

30. M. Braun, E. Sharon, I. Unterman, M. Miller, A. M. Shtern, S. Benenson, A. Vainstein, Y. Tabach, ACE2 co-evolutionary pattern suggests targets for pharmaceutical intervention in the COVID-19 pandemic. *iScience* **23**, 101384 (2020).

31. R. Peto, Quantitative implications of the approximate irrelevance of mammalian body size and lifespan to lifelong cancer risk. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **370**, 20150198 (2015).

32. J. R. Speakman, Body size, energy metabolism and lifespan. *J. Exp. Biol.* **208**, 1717–1730 (2005).

33. N. M. Varki, A. Varki, On the apparent rarity of epithelial cancers in captive chimpanzees. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **370**, 20140225 (2015).

34. G. S. Wilkinson, D. M. Adams, Recurrent evolution of extreme longevity in bats. *Biol. Lett.* **15**, 20180860 (2019).

35. S. A. Forbes, D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, M. Ding, S. Bamford, C. Cole, S. Ward, C. Y. Kok, M. Jia, T. De, J. W. Teague, M. R. Stratton, U. McDermott, P. J. Campbell, COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).

36. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski; Genome Aggregation Database Consortium, B. M. Neale, M. J. Daly, D. G. MacArthur, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

37. A. Balmain, C. C. Harris, Carcinogenesis in mouse and human cells: Parallels and paradoxes. *Carcinogenesis* **21**, 371–377 (2000).

38. E. SEER. Surveillance, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database Incidence, SEER 9 RegsResearch Data, Nov 2017 Sub (1973–2015) - Linked To County Attributes - Total U.S., 1969–2016 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2018, based on the November 2017 submission. (2018).

39. GTEx Consortium, The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

40. C. J. Bult, J. A. Blake, C. L. Smith, J. A. Kadin, J. E. Richardson; Mouse Genome Database Group, Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).

41. E. Pérez-Guijarro, H. H. Yang, R. E. Araya, R. E. Meskini, H. T. Michael, S. K. Vodnala, K. L. Marie, C. Smith, S. Chin, K. C. Lam, A. Thorkelsson, A. J. Iacovelli, A. Kulaga, A. Fon, A. M. Michalowski, W. Hugo, R. S. Lo, N. P. Restifo, S. K. Sharan, T. V. Dyke, R. S. Goldszmid, Z. W. Ohler, M. P. Lee, C.-P. Day, G. Merlino, Multimodel preclinical platform predicts clinical response of melanoma to immunotherapy. *Nat. Med.* **26**, 781−791 (2020).

42. K. L. Marie, E. Pérez-Guijarro, S. Malikić, E. S. Azer, H. H. Yang, C. Kızılkale, C. Gruen, W. Robinson, H. Liu, M. C. Kelly, C. Marcelus, S. Burkett, A. Buluç, F. Ergün, M. P. Lee, G. Merlino, C.-P. Day, S. C. Sahinalp, Profiles of expressed mutations in single cells reveal subclonal expansion patterns and therapeutic impact of intratumor heterogeneity. bioRxiv 2021.03.26.437185 [Preprint]. 23 May 2021. https://doi.org/10.1101/2021.03.26.437185.

43. E. P. Murchison, D. C. Wedge, L. B. Alexandrov, B. Fu, I. Martincorena, Z. Ning, J. M. C. Tubio, E. I. Werner, J. Allen, A. B. De Nardi, E. M. Donelan, G. Marino, A. Fassati, P. J. Campbell, F. Yang, A. Burt, R. A. Weiss, M. R. Stratton, Transmissible dog cancer genome reveals the origin and history of an ancient cell lineage. *Science* **343**, 437–440 (2014).

44. S. Chen, G. Parmigiani, Meta-analysis of BRCA1 and BRCA2 penetrance. *J. Clin. Oncol.* **25**, 1329–1333 (2007).

45. K. B. Kuchenbaecker, J. L. Hopper, D. R. Barnes, K.-A. Phillips, T. M. Mooij, M.-J. Roos-Blom, S. Jervis, F. E. van Leeuwen, R. L. Milne, N. Andrieu, D. E. Goldgar, M. B. Terry, M. A. Rookus, D. F. Easton, A. C. Antoniou; the BRCA1 and BRCA2 Cohort Consortium, L. M. Guffog, D Gareth Evans, D. Barrowdale, D. Frost, J. Adlard, K.-R. Ong, L. Izatt, M. Tischkowitz, R. Eeles, R. Davidson, S. Hodgson, S. Ellis, C. Nogues, C. Lasset, D. Stoppa-Lyonnet, J.-P. Fricker, L. Faivre, P. Berthet, M. J. Hooning, L. E. van der Kolk, C. M. Kets, M. A. Adank, E. M. John, W. K. Chung, I. L. Andrulis, M. Southey, M. B. Daly, S. S. Buys, A. Osorio, C. Engel, K. Kast, R. K. Schmutzler, T. Caldes, A. Jakubowska, J. Simard, M. L. Friedlander, S.-A. McLachlan, E. Machackova, L. Foretova, Y. Y. Tan, C.

F. Singer, E. Olah, A.-M. Gerdes, B. Arver, H. Olsson, Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA* **317**, 2402–2416 (2017).

46. C. Cybulski, D. Wokołorczyk, A. Jakubowska, T. Huzarski, T. Byrski, J. G. Masojć, T. Dębniak, B. Górski, P. Blecharz, S. A. Narod, J. Lubiński, Risk of breast cancer in women with a CHEK2 mutation with and without a family history of breast cancer. *J. Clin. Oncol.* **29**, 3747–3752 (2011).

47. S. J. Ramus, H. Song, E. Dicks, J. P. Tyrer, A. N. Rosenthal, M. P. Intermaggio, L. Fraser, A. Gentry-Maharaj, J. Hayward, S. Philpott, C. Anderson, C. K. Edlund, D. Conti, P. Harrington, D. Barrowdale, D. D. Bowtell, K. Alsop, G. Mitchell; AOCS Study Group, M. S. Cicek, J. M. Cunningham, B. L. Fridley, J. Alsop, M. Jimenez-Linan, S. Poblete, S. Lele, L. Sucheston-Campbell, K. B. Moysich, W. Sieh, V. M. Guire, J. Lester, N. Bogdanova, M. Dürst, P. Hillemanns; Ovarian Cancer Association Consortium, K. Odunsi, A. S. Whittemore, B. Y. Karlan, T. Dörk, E. L. Goode, U. Menon, I. J. Jacobs, A. C. Antoniou, P. D. P. Pharoah, S. A. Gayther, Germline mutations in the BRIP1, BARD1, PALB2, and NBN genes in women with ovarian cancer. *J. Natl. Cancer Inst.* **107**, dvj214 (2015).

48. A. W. Kurian, E. Hughes, E. A. Handorf, A. Gutin, B. Allen, A.-R. Hartman, M. J. Hall, Breast and ovarian cancer penetrance estimates derived from germline multiple-gene sequencing results in women. *JCO Precis. Oncol.* **1**, 1–12 (2017).

49. B. Zhang, A. Beeghly-Fadiel, J. R. Long, W. Zheng, Genetic variants associated with breast-cancer risk: Comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol.* **12**, 477–488 (2011).

50. M.-M. Hahn, L. Vreede, S. A. S. A Bemelmans, E. van der Looij, A. G. van Kessel, H. K. Schackert, M. J. L. Ligtenberg, N. Hoogerbrugge, R. P. Kuiper, R. M. de Voer, Prevalence of germline mutations in the spindle assembly checkpoint gene BUB1B in individuals with early-onset colorectal cancer. *Gene Chromosome Cancer* **55**, 855–863 (2016).

51. W.-C. Chou, S.-C. Chou, C.-Y. Liu, C.-Y. Chen, H.-A. Hou, Y.-Y. Kuo, M.-C. Lee, B.-S. Ko, J.-L. Tang, M. Yao, W. Tsay, S.-J. Wu, S.-Y. Huang, S.-C. Hsu, Y.-C. Chen, Y.-C. Chang, Y.-Y. Kuo, K.-T. Kuo, F.-Y. Lee, M.-C. Liu, C.-W. Liu, M.-H. Tseng, C.-F. Huang, H.-F. Tien, TET2 mutation is an

unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood* **118**, 3803–3810 (2011).

52. R. G. W. Verhaak, C. S. Goudswaard, W. van Putten, M. A. Bijl, M. A. Sanders, W. Hugens, A. G. Uitterlinden, C. A. J. Erpelinck, R. Delwel, B. Löwenberg, P. J. M. Valk, Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): Association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. *Blood* **106**, 3747–3754 (2005).

53. O. Fuge, N. Vasdev, P. Allchorne, J. S. Green, Immunotherapy for bladder cancer. *Res. Rep. Urol.* **7**, 65–79 (2015).

54. Z. Urban-Wojciuk, M. M. Khan, B. L. Oyler, R. Fåhraeus, N. Marek-Trzonkowska, A. Nita-Lazar, T. R. Hupp, D. R. Goodlett, The role of TLRs in anti-cancer immunity and tumor rejection. *Front. Immunol.* **10**, 2388 (2019).

55. L. Ni, J. Lu, Interferon gamma in cancer immunotherapy. *Cancer Med.* **7**, 4509–4516 (2018).

56. C. Sanden, U. Gullberg, The DEK oncoprotein and its emerging roles in gene regulation. *Leukemia* **29**, 1632–1636 (2015).

57. K. Ellrott, M. H. Bailey, G. Saksena, K. R. Covington, C. Kandoth, C. Stewart, J. Hess, S. Ma, K. E. Chiotti, M. McLellan, H. J. Sofia, C. Hutter, G. Getz, D. Wheeler, L. Ding; MC3 Working Group; Cancer Genome Atlas Research Network, Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst.* **6**, 271–281.e7 (2018).

58. K. Gala, Q. Li, A. Sinha, P. Razavi, M. Dorso, F. Sanchez-Vega, Y. R. Chung, R. Hendrickson, J. J. Hsieh, M. Berger, N. Schultz, A. Pastore, O. Abdel-Wahab, S. Chandarlapaty, KMT2C mediates the estrogen dependence of breast cancer through regulation of ERα enhancer function. *Oncogene* **37**, 4692–4710 (2018).

59. Z. Guo, X. Yan, C. Song, Q. Wang, Y. Wang, X.-P. Liu, J. Huang, S. Li, W. Hu, FAT3 mutation is associated with tumor mutation burden and poor prognosis in esophageal cancer. *Front. Oncol.* **11**, 603660 (2021).

60. C. Zhu, Q. Yang, J. Xu, W. Zhao, Z. Zhang, D. Xu, Y. Zhang, E. Zhao, G. Zhao, Somatic mutation of DNAH genes implicated higher chemotherapy response rate in gastric adenocarcinoma patients. *J. Transl. Med.* **17**, 109 (2019).

61. A. Cagan, A. Baez-Ortega, N. Brzozowska, F. Abascal, T. H. H. Coorens, M. A. Sanders, A. R. J. Lawson, L. M. R. Harvey, S. Bhosle, D. Jones, R. E. Alcantara, T. M. Butler, Y. Hooks, K. Roberts, E. Anderson, S. Lunn, E. Flach, S. Spiro, I. Januszczak, E. Wrigglesworth, H. Jenkins, T. Dallas, N. Masters, M. W. Perkins, R. Deaville, M. Druce, R. Bogeska, M. D. Milsom, B. Neumann, F. Gorman, F. Constantino-Casas, L. Peachey, D. Bochynska, E. S. J. Smith, M. Gerstung, P. J. Campbell, E. P. Murchison, M. R. Stratton, I. Martincorena. Somatic mutation rates scale with lifespan across mammals. *Nature* **604**, 517–524 (2022).

62. R. W. Hart, R. B. Setlow, Correlation between deoxyribonucleic acid excision-repair and life-span in a number of mammalian species. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2169–2173 (1974).

63. L. Zhang, X. Dong, X. Tian, M. Lee, J. Ablaeva, D. Firsanov, S.-G. Lee, A. Y. Maslov, V. N. Gladyshev, A. Seluanov, V. Gorbunova, J. Vijg, Maintenance of genome sequence integrity in long- and short-lived rodent species. *Sci. Adv.* **7**, eabj3284 (2021).

64. S. R. Woo, L. Corrales, T. F. Gajewski, Innate immune recognition of cancer. *Annu. Rev. Immunol.* **33**, 445–474 (2015).

65. A. Evdokimov, M. Kutuzov, I. Petruseva, N. Lukjanchikova, E. Kashina, E. Kolova, T. Zemerova, S. Romanenko, P. Perelman, D. Prokopov, A. Seluanov, V. Gorbunova, A. Graphodatsky, V. Trifonov, S. Khodyreva, O. Lavrik, Naked mole rat cells display more efficient excision repair than mouse cells. *Aging* **10**, 1454–1473 (2018).

66. J. Mellors, T. Tipton, S. Longet, M. Carroll, Viral evasion of the complement system and its importance for vaccines and therapeutics. *Front. Immunol.* **11**, 1450 (2020).

67. M. S. Moore, J. D. Reichard, T. D. Murtha, B. Zahedi, R. M. Fallier, T. H. Kunz, Specific alterations in complement protein activity of little brown Myotis (Myotis lucifugus) Hibernating in white-nose syndrome affected sites. *PLOS ONE* **6**, e27430 (2011).

68. E. U. Hammarlund, K. von Stedingk, S. Pahlman, Refined control of cell stemness allowed animal evolution in the oxic realm. *Nat. Ecol. Evol.* **2**, 220–228 (2018).

69. I. Bloch, D. Sherill-Rofe, D. Stupp, I. Unterman, H. Beer, E. Sharon, Y. Tabach, Optimization of co-evolution analysis through phylogenetic profiling reveals pathway-specific signals. *Bioinformatics* **36**, 4116–4125 (2020).

70. I. Omar, G. Guterman-Ram, D. Rahat, Y. Tabach, M. Berger, N. Levaot, Schlafen2 mutation in mice causes an osteopetrotic phenotype due to a decrease in the number of osteoclast progenitors. *Sci. Rep.* **8**, 13005 (2018).

71. T. Tsaban, D. Stupp, D. Sherill-Rofe, I. Bloch, E. Sharon, O. Schueler-Furman, R. Wiener, Y. Tabach, CladeOScope: Functional interactions through the prism of clade-wise co-evolution. *NAR Genom. Bioinform.* **3**, lqab024 (2021).

72. B. Jassal, L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, R. Haw, F. Loney, B. May, M. Milacic, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).

73. A. Buniello, J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, C. Malangone, A. McMahon, J. Morales, E. Mountjoy, E. Sollis, D. Suveges, O. Vrousgou, P. L. Whetzel, R. Amode, J. A. Guillen, H. S. Riat, S. J. Trevanion, P. Hall, H. Junkins, P. Flicek, T. Burdett, L. A. Hindorff, F. Cunningham, H. Parkinson, The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

74. T. A. Myers, S. J. Chanock, M. J. Machiela, LDlinkR: An R package for rapidly calculating linkage disequilibrium statistics in diverse populations. *Front. Genet.* **11**, 157 (2020).

75. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).

76. A. Gordon, G. Glazko, X. Qiu, A. Yakovlev, Control of the mean number of false discoveries, Bonferroni and stability of multiple testing. *Ann. Appl. Stat.* **1**, 179–190 (2007).

77. X. Tian, K. Doerig, R. Park, A. C. R. Qin, C. Hwang, A. Neary, M. Gilbert, A. Seluanov, V. Gorbunova, Evolution of telomere maintenance and tumour suppressor mechanisms across mammals. *Philos. Trans. R Soc. B Biol. Sci.* **373**, 20160443 (2018).

78. L. M. Abegglen, A. F. Caulin, A. Chan, K. Lee, R. Robinson, M. S. Campbell, W. K. Kiso, D. L. Schmitt, P. J. Waddell, S. Bhaskara, S. T. Jensen, C. C. Maley, J. D. Schiffman, Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *JAMA* **314**, 1850–1860 (2015).

79. M. Sulak, L. Fong, K. Mika, S. Chigurupati, L. Yon, N. P. Mongan, R. D. Emes, V. J. Lynch, TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *eLife* **5**, e11994 (2016).

80. R. Buffenstein, Negligible senescence in the longest living rodent, the naked mole-rat: Insights from a successfully aging species. *J. Comp. Physiol. B* **178**, 439–445 (2008).

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

82. F. Hadi, Y. Kulaberoglu, K. A. Lazarus, K. Bach, R. Ugur, P. Beattie, E. S. J. Smith, W. T. Khaled, Transformation of naked mole-rat cells. *Nature* **583**, E1–E7 (2020).

83. V. Gorbunova, C. Hine, X. Tian, J. Ablaeva, A. V. Gudkov, E. Nevo, A. Seluanov, Cancer resistance in the blind mole rat is mediated by concerted necrotic cell death mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19392–19396 (2012).

84. J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S. K. Forslund, H. Cook, D. R. Mende, I. Letunic, T. Rattei, L. J. Jensen, C. von Mering, P. Bork, eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).

85. G. Waris, H. Ahsan, Reactive oxygen species: Role in the development of cancer and various chronic conditions. *J. Carcinog.* **5**, 14 (2006).

86. A. N. Ilacqua, A. M. Kirby, M. E. Pamenter, Behavioural responses of naked mole rats to acute hypoxia and anoxia. *Biol. Lett.* **13**, 20170545 (2017).

87. I. W. McIntyre, K. L. Campbell, R. A. MacArthur, Body oxygen stores, aerobic dive limits and diving behaviour of the star-nosed mole (Condylura cristata) and comparisons with non-aquatic talpids. *J. Exp. Biol.* **205**, 45–54 (2002).