



Importance of metabolic rate to the relationship between the number of genes in a functional category and body size in Peto's paradox for cancer

著者	Takemoto Kazuhiro, Ii Masato, Nishizuka Satoshi S.
journal or publication title	Royal Society Open Science
volume	3
page range	160267-1-160267-11
year	2016-09-01
その他のタイトル	Importance of Metabolic Rate to the Relationship Between the Number of Genes in a Functional Category and Body Size in Peto's Paradox for Cancer
URL	http://hdl.handle.net/10228/00007270

doi: [info:doi/10.1098/rsos.160267](https://doi.org/10.1098/rsos.160267)



Cite this article: Takemoto K, Ii M, Nishizuka SS. 2016 Importance of metabolic rate to the relationship between the number of genes in a functional category and body size in Peto's paradox for cancer. *R. Soc. open sci.* **3**: 160267. <http://dx.doi.org/10.1098/rsos.160267>

Received: 18 April 2016

Accepted: 4 August 2016

Subject Category:

Cellular and molecular biology

Subject Areas:

bioinformatics/physiology/genomics

Keywords:

cancer, Peto's paradox, functional analysis of genome, metabolic rate, body size

Author for correspondence:

Kazuhiro Takemoto

e-mail: takemoto@bio.kyutech.ac.jp

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsos.160267> or via <http://rsos.royalsocietypublishing.org>.

Importance of metabolic rate to the relationship between the number of genes in a functional category and body size in Peto's paradox for cancer

Kazuhiro Takemoto¹, Masato Ii¹ and Satoshi S. Nishizuka²

¹Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka, Fukuoka 820-8502, Japan

²Molecular Therapeutics Laboratory, Department of Surgery, Iwate Medical University School of Medicine, Morioka, Iwate 020-8505, Japan

KT, 0000-0002-6355-1366

Elucidation of tumour suppression mechanisms is a major challenge in cancer biology. Therefore, Peto's paradox, or low cancer incidence in large animals, has attracted focus. According to the gene-abundance hypothesis, which considers the increase/decrease in cancer-related genes with body size, researchers evaluated the associations between gene abundance and body size. However, previous studies only focused on a few specific gene functions and have ignored the alternative hypothesis (metabolic rate hypothesis): in this hypothesis, the cellular metabolic rate and subsequent oxidative stress decreases with increasing body size. In this study, we have elected to explore the gene-abundance hypothesis taking into account the metabolic rate hypothesis. Thus, we comprehensively investigated the correlation between the number of genes in various functional categories and body size while at the same time correcting for the mass-specific metabolic rate (B_c). A number of gene functions that correlated with body size were initially identified, but they were found to be artefactual due to the decrease in B_c with increasing body size. By contrast, immune system-related genes were found to increase with increasing body size when the correlation included this correction for B_c . These findings support the gene-abundance hypothesis and emphasize the importance of also taking into account the metabolic rate

when evaluating gene abundance–body size relationships. This finding may be useful for understanding cancer evolution and tumour suppression mechanisms as well as for determining cancer-related genes and functions.

1. Introduction

Cancer is a complex and robust disease that remains difficult-to-treat despite the development of numerous anti-cancer therapies [1,2]. Elucidation of cellular tumour suppression mechanisms is a long-standing goal for cancer researchers. In recent years, Peto's paradox, first reported in 1975 [3], has attracted renewed attention to this topic [4]. This paradox is often explained as follows: assuming that the probability of normal cells transitioning to cancer cells is equivalent among all cells, larger animals (i.e. organisms with more cells) are expected to show a higher incidence of cancer. However, a positive correlation between body size (or mass) and cancer incidence has not been observed [4–6]. In fact, previous studies have reported no association between body size and cancer incidence (and also the lifespan-adjusted equivalent) [7,8]. These results indicate the existence of body-size-dependent mechanisms of tumour suppression.

Several hypotheses have been proposed to resolve Peto's paradox (or to explain body-size-dependent suppression mechanisms). The metabolic rate hypothesis [9,10] is particularly interesting because cancer cells have altered glucose metabolism in that they typically possess enhanced aerobic glycolysis (the so-called Warburg effect). This hypothesis states that larger animals have relatively lower cancer incidence because the metabolic (oxygen consumption) rate per unit body mass, which is roughly equal to the rate at the cellular level, decreases with body size. Oxygen radicals often enhance ontogenetic mutations and induce the transition from normal cells to tumour cells. In short, large animals have lower cancer rates, because they can avoid oxidative stress and subsequent genetic mutations. In fact, oxidative DNA damage is associated with metabolic rate [11], and the mutation rate decreases with body size because of the decrease in the mass-specific metabolic rate [12,13]. The negative correlation between mass-specific metabolic rate (B_c) and body mass (M) is well known in the context of Kleiber's Law (or the allometric scaling of metabolic rate) [14–16]; in particular, the relationship between metabolic rate (B) and body mass is generally described as a power-law function: $B \propto M^{3/4}$. Thus, mass-specific metabolic rate shows a power-law decay with body mass: $B_c = B/M \propto M^{-1/4}$. The metabolic rate hypothesis may show promise for resolving Peto's paradox; however, it is unsuitable for investigating the cellular (microscopic) mechanisms of tumour suppression, because it is based on physiology (macroscopic behaviour). In addition to this, the basis for allometric scaling of metabolic rate is still unclear even though several hypotheses, or models, have been proposed [17–21].

Genetic studies are useful in this context. In particular, a number of cancer-related genes (e.g. tumour suppressor genes (TSGs) and oncogenes) and pathways [22,23] have been identified by several new technologies including the use of high-throughput methods. Previous studies have proposed an alternative hypothesis: the *gene-abundance hypothesis* [5,24,25], which states that larger animals have more (and/or more highly expressed) genes that suppress cancer progression (e.g. TSGs) and fewer (and/or more lowly expressed) genes that induce cancer growth. Further theoretical and experimental studies have provided deeper insights into this hypothesis. For example, a theoretical study [24] showed that the number and/or expression levels of (proto-)oncogenes decreases with increasing body size, whereas the number and/or expression levels of TSGs are hardly correlated with body size because of energetic limitations. In particular, the increase of TSGs requires energetic costs resulting from repairing mutations and reproduction; thus, such an evolutionary strategy (i.e. the increase of TSGs) for cancer suppression is hardly adopted (see [24] for details). A comparative genomic study [26] demonstrated that bowhead whale-specific gene mutations were linked to cancer and ageing; moreover, it reported that duplications frequently occurred in genes related to DNA-repair, cell cycle and ageing. Another study [8] showed that elephants have a larger TP53 copy number, which is a well-known TSG, compared with other mammals, and also reported that elephant cells are more resistant to DNA damage, than human cells. Retroviral integration can also induce cancer. A previous study [27], based on bioinformatics and mathematical biology, has shown that the levels and activity of endogenous retroviruses (ERVs) acquired in the last 10 Myr decrease with body size; it was also reported that the mean age of ERVs increases with body size. Inspired by the fact that cancer cell mutations directly increase the number of microsatellites in tumour DNA [28], researchers have revealed that the number of microsatellites decreases with body size [29]. Taken together, these results support the gene-abundance hypothesis.

However, a more careful examination may be required to prove the validity of the gene-abundance hypothesis. Previous studies have only focused on the relationship between specific biological functions and body size, and they generally compared functions among a few mammals. The effect of metabolic rate (i.e. the metabolic rate hypothesis) also needs to be considered when investigating the gene abundance–body size relationships. The gene-abundance hypothesis may therefore be overlapping with the metabolic rate hypothesis. In particular, it remains possible that the gene abundance–body size relationships are spurious correlations that result from the body-size-dependent metabolic rate. For example, the number of genes in a functional category (NOGF) generally shows a positive correlation with the mass-specific metabolic rate [30] (i.e. a number of functional categories are associated with mass-specific metabolic rate). This fact predicts that NOGF will generally show a negative correlation with body size because of the negative association between mass-specific metabolic rate and body size. Moreover, oncogenes and tumour suppressors are known to regulate metabolism (reviewed in [9,10]). For example, Myc is an oncogene that is known to regulate glycolysis and to activate mitochondrial biogenesis, whereas the p53 protein, which is a tumour suppressor, is known to inhibit glycolysis. Glycolysis and biogenesis are strongly related to respiration (i.e. metabolic rate). Assuming that oncogenes generally increase the metabolic rate at the cellular level and that TSGs decrease it, the number of oncogenes and TSGs are expected to show negative and positive associations with body size, respectively. This expectation is consistent with the prediction from the gene-abundance hypothesis.

Therefore, we performed an integrated evaluation of the metabolic rate hypothesis and the gene-abundance hypothesis. In particular, genomic data, metabolic rate and body size for mammals were collected from public databases and the literature, and using statistical methods, the relationships between NOGF and body size were investigated by removing the effect of mass-specific metabolic rate.

2. Material and methods

2.1. Data on metabolic rate, body mass and genome

We used data on mammalian metabolic rates and body masses obtained in our previous study [30]. Additionally, we also collected data for those mammals whose genomes were available in a species-level database within the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [31]. The units of mass-specific metabolic rate and body mass were converted to watts per gram (W/g) and grams (g), respectively. To reduce the phylogenetic signals, one representative species was selected from each genus, according to the year in which the species genome sequence was first completed. Finally, we obtained data for 33 mammals (electronic supplementary material, table S1).

2.2. Functional categories of genes

The method we used is almost similar to that reported in our previous study [30]. We used the third level of KEGG BRITE Functional Hierarchy [32] in the KEGG metabolic map (www.genome.jp/kegg-bin/get_htext?br08901.keg) for measuring the number of genes in functional categories. We downloaded the data on functional category–gene identifier relationships of species *S* from the KEGG FTP site (<ftp://ftp.bioinformatics.jp/kegg/brite/organisms/S/S00001.keg>) on 9 May 2015, where *S* corresponds to the KEGG organism identifier (electronic supplementary material, table S1). In this study, we did not consider Gene Ontology (GO) [33] as an alternative definition of functional categories because there were fewer organisms whose annotations in the GO database were completed, compared to those in the KEGG database. In addition, we selected TSGs and oncogenes identified in the literature [22,23]. In total, we investigated 342 functional categories.

2.3. Statistical analysis

Statistical analysis was performed using R software (v. 3.2.4; www.r-project.org). The association between NOGF and body size/mass-specific rate was evaluated based on the Pearson's product-moment correlation coefficient r and its associated p -value. Mass-specific metabolic rate and body mass were log-transformed for all analyses. To evaluate NOGF–body size correlations when the mass-specific metabolic rate was kept constant, the partial correlation analysis was considered; in particular, we used the *pcor* function that is available in the R package *ppcor* (v. 1.1). Additionally, we also considered a multivariate analysis; in particular, the standardized partial regression coefficient for M in the formula $\text{NOGF} \sim M + B_c$ (i.e. the B_c -corrected estimate) was calculated for each functional category using the *lm*

function (electronic supplementary material, tables S2 and S3). We used the *lm* function to calculate the residuals for variables.

To find the functional categories that are specifically associated with body size, we used an enrichment analysis based on Fisher's exact test, inspired by gene set enrichment analysis [34]. In particular, for the KEGG BRITe Functional Hierarchy we focused on the second and third functional hierarchical levels and calculated the statistical significance, defined as $-\log_{10}(p\text{-value using Fisher's exact test})$, for the ratio of functional categories (i.e. the categories at the third level) correlated with body size ($p < 0.05$) to all the functional categories in the upper (i.e. second-level) category. The *fisher.test* function was used to perform Fisher's exact test. The changes in the p -value threshold for the NOGF–body size correlations did not significantly affect the conclusion.

2.4. Phylogenetically independent contrasts

Phylogenetically independent contrasts (PICs) [35] of the variables were evaluated in order to remove any possible phylogenetic effect on the association between biological variables in the context of the phylogenetic comparative analysis [36,37]. This method is similar to that used in our previous study [38]. In particular, the mammalian phylogenetic tree was constructed using the matrix extracellular phosphoglycoprotein precursor (MEPE) gene (electronic supplementary material, figure S1), downloaded from the KEGG database on 20 August 2015. Note that the platypus, *Ornithorhynchus anatinus*, was omitted from the tree because the MEPE gene for this species was not found. Based on the phylogenetic tree, the PICs were calculated using the function *pic* in the R package *ape* (v. 3.4).

3. Results

3.1. The number of genes in functional categories generally show a negative correlation with body size

We found that NOGF was generally associated with body size. Genome size and the total number of genes (proteins) did not affect the observed NOGF–body size correlations because body size did not correlate with either genome size ($r = -0.054$, $p = 0.77$) or number of proteins ($r = -0.17$, $p = 0.35$).

In general, the associations observed were negative (figure 1a). In particular, the mean of r was -0.14 (95% CI, from -0.16 to -0.12). About 13% (43/342) of the functional categories showed a statistically significant correlation with body size ($p < 0.05$; also the electronic supplementary material, table S2). For example, the number of oncogenes decreases with body size ($r = -0.38$, $p = 0.031$); by contrast, the number of TSGs shows no correlation with body size ($r = -0.28$, $p = 0.11$). This result supports the gene-abundance hypothesis and is consistent with the theoretical prediction [24] that the number of oncogenes decreases with body size, whereas the number of TSGs is almost independent of body size due to reproduction. In addition, disease incidence (and by inference disease-related genes) also decreases with increasing body size. For instance, body size was negatively associated with viral myocarditis ($r = -0.63$, $p = 8.0 \times 10^{-5}$), type I diabetes mellitus ($r = -0.57$, $p = 6.1 \times 10^{-4}$), and graft-versus-host disease ($r = -0.55$, $p = 3.7 \times 10^{-3}$). These observed negative associations could be considered reasonable because these diseases possibly arise due to cancer treatment and/or because they increase the incidence of cancer [39–41]. Therefore, this result also supports the gene-abundance hypothesis.

3.2. Body-size-dependent metabolic rate strongly influences correlations between the number of genes in functional categories and body size

A conclusion that the gene-abundance hypothesis is correct needs to be made with caution, because it remains possible that the observed associations are spurious. For some of the functional categories, the NOGF–body size relationships were found to be questionable. For example, the relationship between cancer and transduction of taste signals could not be explained biologically although a negative correlation between NOGF and body size was clearly observed ($r = -0.57$, $p = 6.9 \times 10^{-4}$; see also §3.3). Similarly, the association between serotonergic synapses and body size ($r = -0.48$, $p = 5.2 \times 10^{-3}$) could also not be readily explained in the context of tumour biology. We can, however, speculate that the observed association may suggest a role for serotonin in cancer, because low doses of serotonin are known to inhibit cancer growth by decreasing tumour blood supply [42]. It should be noted that blood

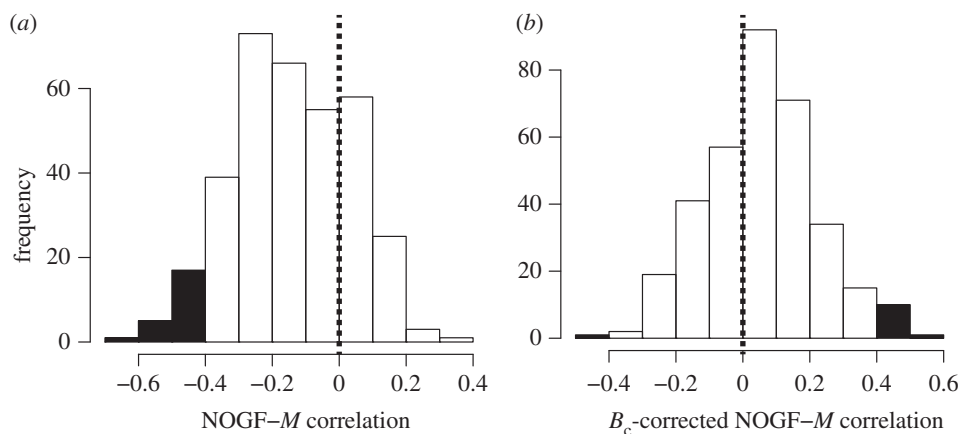


Figure 1. Distribution of the correlation coefficient between the NOGF and body mass (M). Mass-specific metabolic rate (B_c) correction was (a) not considered (b) considered. The filled area indicates the p -value of less than 0.05.

supply is positively correlated with metabolic rate [15,17]. In summary, these observations suggest the possibility that genetic changes may have an indirect effect on cancer. In particular, the genetic change does not contribute directly to the inhibition or enhancement of cancer growth; rather it implies that genetic changes (e.g. increase/decrease in NOGFs) lead to a change in metabolic rate, which then influences cancer growth. Therefore, the noted differences in metabolic rate according to body size should be taken into consideration when evaluating NOGF–body size associations. In particular, NOGFs are expected to show negative correlations with body size, because they are positively correlated with mass-specific metabolic rate which itself is negatively correlated with body size. In fact, all the statistically significant correlations that were observed, were negative (figure 1a); moreover, the observed NOGF–body mass negative correlation was not concluded when corrected with the mass-specific metabolic rate using the partial correlation analysis (figure 1b).

Additional results suggest the influence of metabolism on the NOGF–body size relationship. In particular, functional categories related to metabolism were frequently observed (see the following section for details). For example, body size was linked to lipid metabolism (e.g. fatty acid degradation (figure 2a; $r = -0.46$, $p = 6.7 \times 10^{-3}$) and fatty acid biosynthesis ($r = -0.50$, $p = 3.3 \times 10^{-3}$)) as well as nucleotide (pyrimidine) metabolism ($r = -0.52$, $p = 1.7 \times 10^{-3}$). Lipid metabolism [10,43] and nucleotide metabolism [10,44] are known to play important roles in cancer. However, their role in cancer is often considered in the context of energy regulation, which of course is related to metabolic rate [16]. In addition, fatty acids are believed to determine metabolic rate (the so-called *membrane-pacemaker hypothesis* [45–47]). In fact, the number of genes involved in lipid metabolism is known to be correlated with the mass-specific metabolic rate [30]. In our dataset, the number of genes involved in fatty acid degradation was also positively correlated with mass-specific metabolic rate (figure 2b; $r = 0.49$, $p = 3.4 \times 10^{-3}$).

For all functional categories, we calculated the correlation coefficient between the NOGF–body mass (M) correlation and the NOGF–mass-specific metabolic rate (B_c) correlation (figure 2c). Assuming that the mass-specific metabolic rate influences the NOGF– M correlation, a negative association between the NOGF– M correlation and the NOGF– B_c correlation was observed due to the observed negative correlation between mass-specific rate and body mass ($r = -0.95$, $p < 2.2 \times 10^{-16}$): $B_c \propto M^{-0.24 \pm 0.01}$. As expected, this negative association was confirmed (figure 2c; $r = -0.96$, $p < 2.2 \times 10^{-16}$). This result indicates that the observed NOGF– M correlations result from the B_c – M correlation.

3.3. Immune system-related genes increase with body size when the mass-specific metabolic rate is kept constant

We used a partial correlation analysis to eliminate the effect of the mass-specific metabolic rate on the NOGF– M correlations, and found that the NOGF– M associations for functional categories changed significantly (figure 3). In particular, the NOGF– M associations, identified by a simple correlation analysis, were not observed when the effect of metabolic rate was considered. For example, the observed NOGF– M correlation for taste transduction (figure 3a; $r = -0.57$, $p = 6.9 \times 10^{-4}$) was not observed when

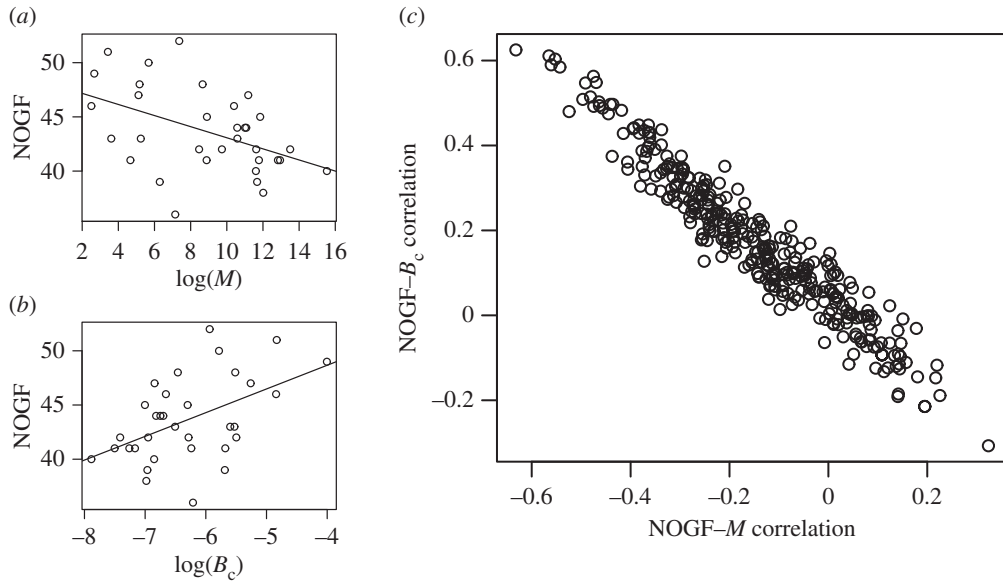


Figure 2. Relationship between the NOGF, body mass (M) and mass-specific metabolic rate (B_c). Correlations of NOGF for fatty acid degradation with (a) $\log(M)$ and (b) $\log(B_c)$. (c) Scatter plot of the correlation coefficient (r) between NOGF and M versus r between NOGF and B_c .

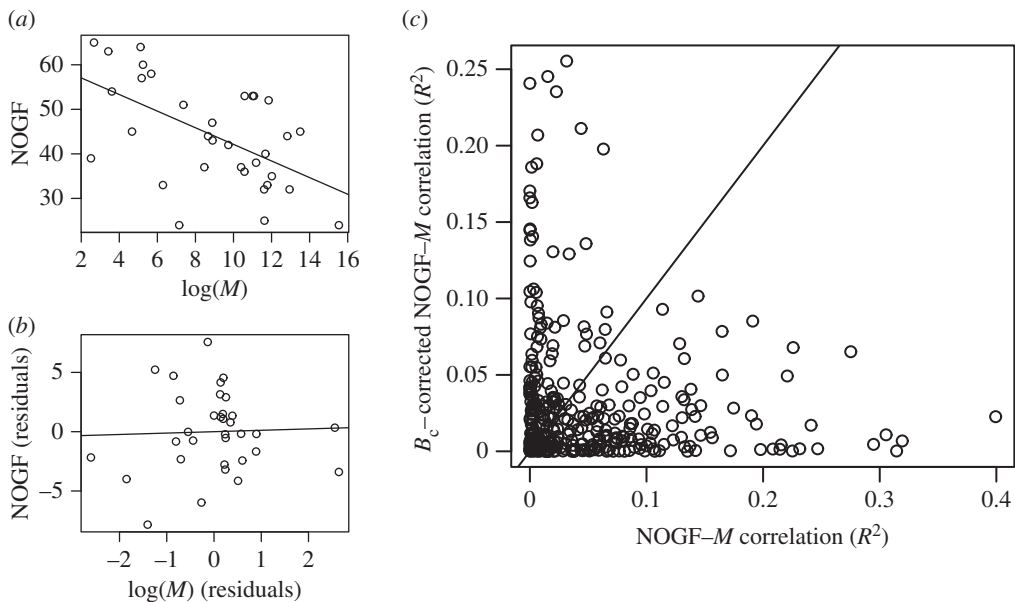


Figure 3. Change in the relationship between the NOGF and body mass (M). (a) NOGF- M correlation for taste transduction. (b) NOGF- M correlation for taste transduction corrected with the mass-specific metabolic rate (B_c). (c) Scatter plot for the determination of coefficient (R^2) of NOGF- M correlation versus R^2 of NOGF- M correlation corrected with B_c . The solid line indicates the diagonal line.

corrected for the mass-specific metabolic rate (figure 3b; $r_p = 0.013$, $p = 0.94$). This indicates that most of the observed correlations were artefacts likely occurring due to the body-size-dependent mass-specific metabolic rate. Moreover, the mean of r_p was positive (0.054 (95% CI, from 0.036 to 0.072); see also figure 1b), unlike in the case of simple correlation analysis where the metabolic rate effect was not considered (figure 1a). Around 6% (20/342) of the functional categories showed a statistically significant correlation with body size ($p < 0.05$; see also the electronic supplementary material, table S2). Representative examples of the functional categories that were correlated with body mass when the mass-specific metabolic rate was kept constant are shown in table 1. For example, the increase in

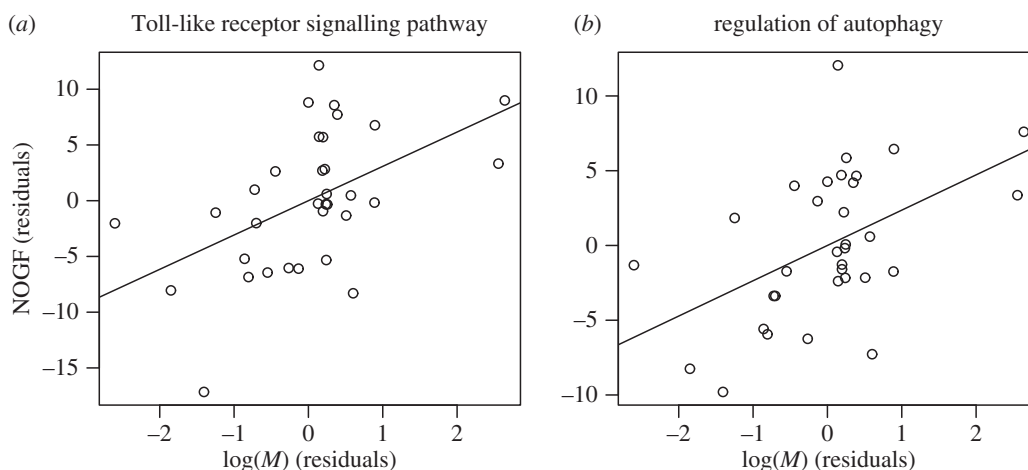


Figure 4. Scatter plot of the NOGF (residuals) versus log-transformed body mass (M) (residuals). (a) Toll-like receptor signalling pathway. (b) Regulation of autophagy.

Table 1. Correlations between the NOGF and body mass. The partial (mass-specific metabolic rate-corrected) correlation coefficients (r_p) and correlation coefficients (r) are shown. The functional categories are only represented when the p -value is less than 0.01 using the partial correlation analysis (see electronic supplementary material, table S2 for the full version), and they are displayed in descending order of r_p . Parenthetic values indicate the p -values.

upper category	functional category	B_c -corrected G - M correlation r_p (p -value)	G - M correlation r (p -value)
immune system	toll-like receptor signalling pathway	0.51 (3.2×10^{-3})	0.18 (0.32)
immune system	Fc ϵ RI signalling pathway	0.50 (4.0×10^{-3})	0.12 (0.49)
transport and catabolism	regulation of autophagy	0.49 (4.4×10^{-3})	0.0029 (0.99)
immune system	RIG-I-like receptor signalling pathway	0.49 (4.9×10^{-3})	0.15 (0.40)
immune system	natural killer cell-mediated cytotoxicity	0.46 (8.1×10^{-3})	-0.21 (0.24)
signalling molecules and interaction	cytokine-cytokine receptor interaction	0.45 (9.0×10^{-3})	0.082 (0.65)

the number of genes in the Toll-like receptor signalling pathway (figure 4a) may contribute to cancer resistance, because this pathway plays an important role in host defence from infection [48]. The other signalling pathways, identified based on NOGF-body size correlations, are also expected to contribute to the resistance against cancer. In particular, the IgE-mediated allergic reaction mediated by Fc ϵ RI signalling has been proposed as a target for cancer immunotherapy [49]. The RIG-I-like receptor signalling pathway has been shown to be activated in cancer therapies through endogenous non-coding RNAs [50]. The increase in the number of genes involved in the regulation of autophagy (figure 4b) may be explained in the context of the protective function of autophagy to limit tumour necrosis and inflammation [51]. Since natural killer cells contribute to the killing of tumour cells without a required immunization or deliberate activation [52], the increase in the number of natural killer cell genes that mediate cytotoxicity can be expected to enhance resistance against cancer. Cytokine interactions are also known to be important in the context of cancer pathogenesis and cancer therapy [53]. It should be noted that these functional categories were identified only when the correlation based on mass-specific metabolic rate was considered.

The functional categories that were correlated with body mass were enriched in the immune system (red bars in figure 5). This result indicates that the number of immune system-related genes generally increases with body size. On the contrary, no clear pattern was found when the effect of metabolic rate was ignored (blue bars in figure 5) even though such functional categories were enriched in lipid and nucleotide metabolism.

We did not consider PICs in the above analyses because of the alternative methodology we used (see §2.4 for details) and because there are limitations to phylogenetic comparative analysis. In particular,

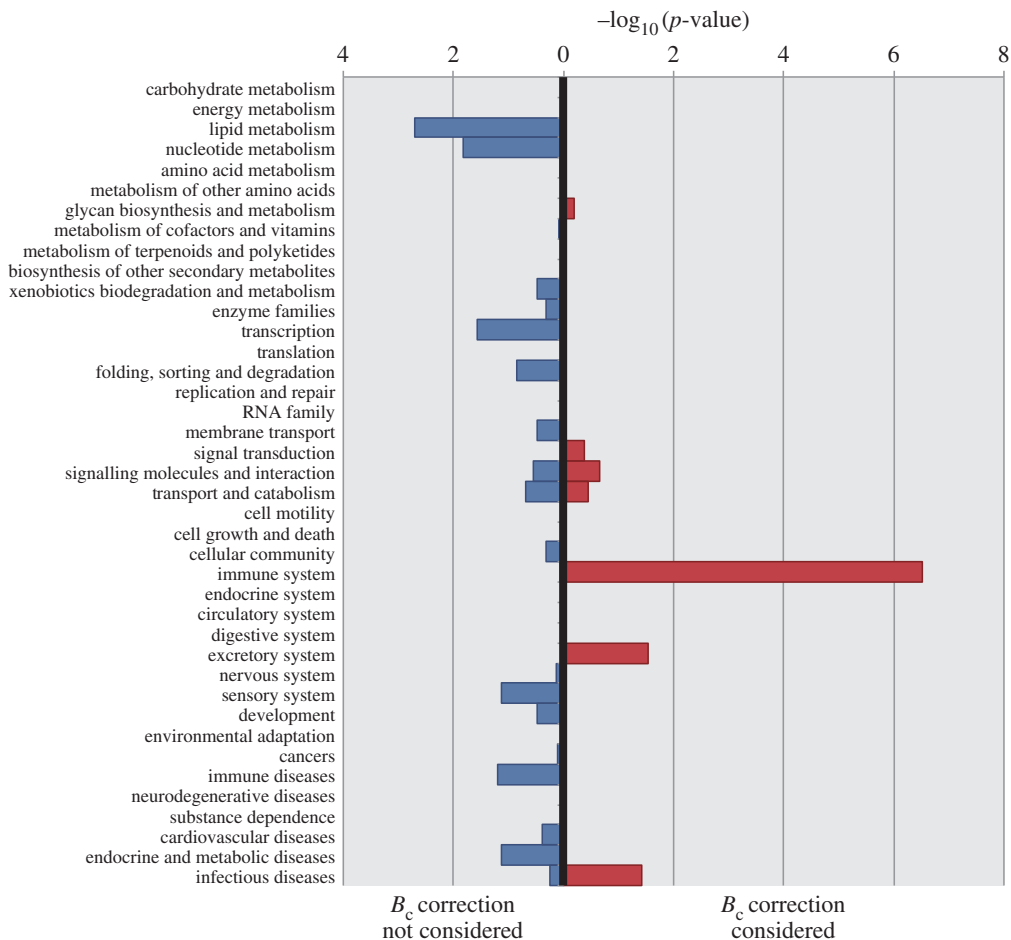


Figure 5. Enrichment of functional categories correlated with body size. The blue bars and red bars indicate the case that mass-specific metabolic rate (B_c) correction was not considered for evaluating the relationships between the NOGF and body mass (M) and the case that B_c correction was considered for evaluating NOGF– M correlations, respectively.

phylogenetic comparative analysis assumes a Brownian motion-like evolution of biological traits on a phylogenetic tree with accurate branch lengths and may thus result in misleading conclusions. In addition, statistical power decreases when a dataset is reduced in size following phylogenetic corrections [54]. Despite these caveats, almost identical conclusions were obtained when PICs were considered (electronic supplementary material, table S3). Therefore, phylogenetic effects did not influence the outcome of this study.

4. Discussion

We comprehensively evaluated the gene-abundance hypothesis in the context of the metabolic rate hypothesis, and found that NOGF generally showed a negative correlation with body size. However, such associations were found to be artefactual due to the relationship between metabolic rate and body mass. In particular, the result of the NOGF–body size correlations was significantly different when the analysis was corrected for metabolic rate. Moreover, NOGF tended to be related to metabolic rate rather than to cancer. When evaluating the NOGF–body size relationships corrected for metabolic rate, we found that immune system-related genes increased with body size. A number of previous studies support the notion that gene functions identified based on metabolic rate-corrected NOGF–body size relationships are reasonable in the context of tumour suppression and cancer therapies. This result was obtained only when the metabolic rate was kept constant, indicating that cancer-related genes/functions could easily be overlooked when the difference in metabolic rate according to body size is not considered. These results emphasize the importance of metabolic rate in evaluating gene abundance–body size relationships.

In recent years, researchers in this area have focused on the genomes of large animal (e.g. whales [26] and elephants [8]) to test the gene-abundance hypothesis. However, such studies did not consider the effect of metabolic rate and therefore may provide misleading conclusions. Our results do not contradict those from previous studies because of the strong relationship between metabolism and cancer [10]; however, the NOGF–body size associations identified when the effect of metabolic rate was ignored should be considered in the context of the metabolic rate hypothesis rather than the gene-abundance hypothesis. Our results provide significant support for the gene-abundance hypothesis. In particular, we have shown that the hypothesis is supported in a broad range of mammals, even though previous studies have generally focused on large animal to small animal comparisons.

However, more careful analyses may be required in further studies. For example, our conclusion (the importance of immune system, in particular) is still debatable because we only performed correlation analyses. In this context, a multivariate analysis would be more appropriate. However, in this study we could not perform such an analysis because of the multi-collinearity that arises from the fact that there was gene overlap among functional categories. In fact, we were unable to calculate partial regression coefficients and associated *p*-values because of singularities in the generalized linear model. As mentioned in our previous study [30], this is a particular problem in GO analysis. To resolve this problem, for example, we will need to define alternative (e.g. non-overlapping) functional categories (see below for details). Despite these limitations, the bulk of the experimental evidence supports our conclusions.

It should also be noted that we evaluated NOGFs–body size associations by assuming that the Peto’s paradox is correct. However, ideally, we need to consider the relationship with actual cancer incidence rates in animals. In this study, we did not consider the actual cancer incidence rates because only limited data are available. However, recent studies have begun to investigate cancer rates in wildlife (e.g. reviewed in [55]). The data on cancer risks obtained in these studies will be very useful in further studies. In addition to this, factors like lifespan, generation time and mutation rate are also different across species; however, because it is believed that these biological parameters are related to metabolic rate, a correction based on mass-specific metabolic rate may be the most useful one to remove most of the effects of these biological parameters [11,13,16,56]. A critical review [57] claims that Peto’s paradox relies on several questionable assumptions, and it emphasizes the importance of organ-level comparisons rather than species-level comparisons in addressing the variation in cancer risk across tissues [58]. The differences in risk may be explained in terms of differences in the number of tissue cell divisions [58] as well as differences in the robustness of cancer signalling networks [59,60].

Another consideration is that the definition of gene abundance is controversial. For simplicity, we have only considered the NOGF as an indicator of gene abundance. However, gene abundance could also be measured in terms of expression level, copy number and activity of functional genes. Indeed, a previous study reported that elephants show a higher expression level and copy number of TP53, compared to humans. It is also possible that the body-size-dependent metabolic rate can give rise to differences in gene expression and copy number, because the expression levels of (enzymatic) genes are also known to determine metabolic rate [61,62]. Thus, the correction for metabolic rate may also be required when considering the relationships between body size and expression level, copy number and/or activity. We attempted to consider expression levels using gene expression data from various mammalian organs [63]; however, a comprehensive analysis was not possible because of the limited amount of data available. The measurement of gene abundance at the organ level using high-throughput techniques (e.g. microarray and RNA-Seq [64]) should therefore be undertaken for a wide range of organisms in the future.

As is the case for other genome analysis studies, our study has some general limitations. For example, we only considered organisms for which genome sequences were completed and available; thus, our species selection was unintentionally biased. In this context, metagenomic techniques may help to complete the genomes of more organisms. Moreover, our findings depend significantly on the quality of genome annotation. The functional categories, defined by the KEGG BRITE Functional Hierarchy in this study, might be somewhat arbitrary (i.e. they depend on the opinions of database administrators). Computational frameworks may increase the quality of genome annotation in terms of protein function prediction (e.g. machine learning-based methods [65]); they may also provide alternative definitions of functional categories. For example, computational methods based on the following frameworks may be useful; graph clustering or community detection of networks [66], biosynthetic capability [67], gene clusters and chemical transformation patterns [68,69].

Despite the limitations of our data analysis, these findings enhance our understanding of cancer evolution and tumour suppression mechanisms with respect to body size. Furthermore, they may be usefully applied in future research for estimating cancer-related genes and functions.

Data accessibility. The datasets supporting this article have been uploaded as part of the electronic supplementary material. The datasets are also deposited at Dryad (<http://dx.doi.org/10.5061/dryad.8j0kf>).

Authors' contributions. K.T. and M.I. conceived and designed the study, performed data analysis and drafted the manuscript. S.S.N. helped draft the manuscript. All authors prepared the data, interpreted the results and gave final approval for publication.

Competing interests. We have no competing interests.

Funding. This study was supported by a Grant-in-Aid for Young Scientists (A) from the Japan Society for the Promotion of Science (no. 25700030).

References

1. Kitano H. 2004 Biological robustness. *Nat. Rev. Genet.* **5**, 826–837. (doi:10.1038/nrg1471)
2. Tian T, Olson S, Whitacre JM, Harding A. 2011 The origins of cancer robustness and evolvability. *Integr. Biol. (Camb.)* **3**, 17–30. (doi:10.1039/c0ib00046a)
3. Peto R, Roe FJ, Lee PN, Levy L, Clack J. 1975 Cancer and ageing in mice and men. *Br. J. Cancer* **32**, 411–426. (doi:10.1038/bjcr.1975.242)
4. Nunney L, Maley CC, Breen M, Hochberg ME, Schifman JD. 2015 Peto's paradox and the promise of comparative oncology. *Phil. Trans. R. Soc. B* **370**, 20140177. (doi:10.1098/rstb.2014.0177)
5. 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)
6. 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. (doi:10.1093/icc/ibc/062)
7. Kelleher FC, O'Sullivan H. 2014 Oxford and the savannah: can the Hippo provide an explanation for Peto's paradox? *Clin. Cancer Res.* **20**, 557–564. (doi:10.1158/1078-0432.CCR-13-2010)
8. Abegglen LM *et al.* 2015 Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *JAMA* **314**, 1850. (doi:10.1001/jama.2015.13134)
9. Dang CV. 2015 A metabolic perspective of Peto's paradox and cancer. *Phil. Trans. R. Soc. B* **370**, 20140223. (doi:10.1098/rstb.2014.0223)
10. Dang CV. 2012 Links between metabolism and cancer. *Genes Dev.* **26**, 877–890. (doi:10.1101/gad.189365.112)
11. Adelman R, Saul RL, Ames BN. 1988 Oxidative damage to DNA: relation to species metabolic rate and life span. *Proc. Natl Acad. Sci. USA* **85**, 2706–2708. (doi:10.1073/pnas.85.8.2706)
12. Gillooly JF, McCoy MW, Allen AP. 2007 Effects of metabolic rate on protein evolution. *Biol. Lett.* **3**, 655–660. (doi:10.1098/rsbl.2007.0403)
13. Gillooly JF, Allen AP, West GB, Brown JH. 2005 The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proc. Natl Acad. Sci. USA* **102**, 140–145. (doi:10.1073/pnas.0407735101)
14. Savage VM, Gillooly JF, Woodruff WH, West GB, Allen AP, Enquist BJ, Brown JH. 2004 The predominance of quarter-power scaling in biology. *Funct. Ecol.* **18**, 257–282. (doi:10.1111/j.0269-8463.2004.00856.x)
15. West GB, Woodruff WH, Brown JH. 2002 Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proc. Natl Acad. Sci. USA* **99**, 2473–2478. (doi:10.1073/pnas.012579799)
16. Speakman JR. 2005 Body size, energy metabolism and lifespan. *J. Exp. Biol.* **208**, 1717–1730. (doi:10.1242/jeb.01556)
17. Banavar JR, Moses ME, Brown JH, Damuth J, Rinaldo A, Sibly RM, Maritan A. 2010 A general basis for quarter-power scaling in animals. *Proc. Natl Acad. Sci. USA* **107**, 15 816–15 820. (doi:10.1073/pnas.1009974107)
18. Kozłowski J, Konarzewski M, Gawelczyk AT. 2003 Cell size as a link between noncoding DNA and metabolic rate scaling. *Proc. Natl Acad. Sci. USA* **100**, 14 080–14 085. (doi:10.1073/pnas.2334605100)
19. Price CA, Enquist BJ, Savage VM. 2007 A general model for allometric covariation in botanical form and function. *Proc. Natl Acad. Sci. USA* **104**, 13 204–13 209. (doi:10.1073/pnas.0702242104)
20. Demetrius L, Tuszynski JA. 2010 Quantum metabolism explains the allometric scaling of metabolic rates. *J. R. Soc. Interface* **7**, 507–514. (doi:10.1098/rsif.2009.0310)
21. Takemoto K. 2015 Heterogeneity of cells may explain allometric scaling of metabolic rate. *Biosystems* **130**, 11–16. (doi:10.1016/j.biosystems.2015.02.003)
22. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. 2013 Cancer genome landscapes. *Science* **339**, 1546–1558. (doi:10.1126/science.1235122)
23. Walker EJ *et al.* 2012 Monoallelic expression determines oncogenic progression and outcome in benign and malignant brain tumors. *Cancer Res.* **72**, 636–644. (doi:10.1158/0008-5472.CAN-11-2266)
24. Roche B, Sprouffske K, Hbid H, Missé D, Thomas F. 2013 Peto's paradox revisited: theoretical evolutionary dynamics of cancer in wild populations. *Evol. Appl.* **6**, 109–116. (doi:10.1111/eva.12025)
25. Roche B, Hochberg ME, Caulin AF, Maley CC, Gatenby RA, Misse D, Thomas F. 2012 Natural resistance to cancers: a Darwinian hypothesis to explain Peto's paradox. *BMC Cancer* **12**, 387. (doi:10.1186/1471-2407-12-387)
26. Keane M *et al.* 2015 Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep.* **10**, 112–122. (doi:10.1016/j.celrep.2014.12.008)
27. 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)
28. Loeb LA, Loeb KR, Anderson JP. 2003 Multiple mutations and cancer. *Proc. Natl Acad. Sci. USA* **100**, 776–781. (doi:10.1073/pnas.0334858100)
29. Park JY, An Y-R, An C-M, Kang J-H, Kim EM, Kim H, Cho S, Kim J. 2016 Evolutionary constraints over microsatellite abundance in larger mammals as a potential mechanism against carcinogenic burden. *Sci. Rep.* **6**, 25246. (doi:10.1038/srep25246)
30. Takemoto K, Kawakami Y. 2015 The proportion of genes in a functional category is linked to mass-specific metabolic rate and lifespan. *Sci. Rep.* **5**, 10008. (doi:10.1038/srep10008)
31. Jones KE *et al.* 2009 PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90**, 2648–2648. (doi:10.1890/08-1494.1)
32. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014 Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42**, D199–D205. (doi:10.1093/nar/gkt1076)
33. Ashburner M *et al.* 2000 Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29. (doi:10.1038/75556)
34. Huang DW, Sherman BT, Lempicki RA. 2009 Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13. (doi:10.1093/nar/gkn923)
35. Felsenstein J. 1985 Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15. (doi:10.1086/284325)
36. Garland T, Harvey PH, Ives AR. 1992 Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* **41**, 18–32. (doi:10.1093/sysbio/41.1.18)
37. Garland T, Bennett AF, Rezende EL. 2005 Phylogenetic approaches in comparative physiology. *J. Exp. Biol.* **208**, 3015–3035. (doi:10.1242/jeb.01745)
38. Takemoto K. 2016 Habitat variability does not generally promote metabolic network modularity in flies and mammals. *Biosystems* **139**, 46–54. (doi:10.1016/j.biosystems.2015.12.004)
39. Läubli H, Balmelli C, Bossard M, Pfister O, Glatz K, Zippeli A. 2015 Acute heart failure due to autoimmune myocarditis under pembrolizumab treatment for metastatic melanoma. *J. Immunother. Cancer* **3**, 11. (doi:10.1186/s40425-015-0057-1)
40. Elad S, Zadik Y, Zeevi I, Miyazaki A, de Figueiredo MAZ, Or R. 2010 Oral cancer in patients after hematopoietic stem-cell transplantation: long-term follow-up suggests an increased risk for recurrence. *Transplantation* **90**, 1243–1244. (doi:10.1097/TP.0b013e3181f9caaa)
41. Zendejdel K, Nyrén O, Ostenson C-G, Adami H-O, Ekborn A, Ye W. 2003 Cancer incidence in patients with type 1 diabetes mellitus: a population-based cohort study in Sweden. *J. Natl Cancer Inst.* **95**, 1797–1800. (doi:10.1093/jnci/djg105)

42. Sarrouilhe D, Clahaut J, Defamie N, Mesnil M. 2015 Serotonin and cancer: what is the link? *Curr. Mol. Med.* **15**, 62–77. (doi:10.25601469)
43. Santos CR, Schulze A. 2012 Lipid metabolism in cancer. *FEBS J.* **279**, 2610–2623. (doi:10.1111/j.1742-4658.2012.08644.x)
44. Lane AN, Fan TW-M. 2015 Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Res.* **43**, 2466–2485. (doi:10.1093/nar/gkv047)
45. Rodriguez E, Weber J-M, Page B, Roubik DW, Suarez RK, Darveau C-A. 2015 Setting the pace of life: membrane composition of flight muscle varies with metabolic rate of hovering orchid bees. *Proc. R. Soc. B* **282**, 20142232. (doi:10.1098/rspb.2014.2232)
46. Hulbert AJ, Else PL. 1999 Membranes as possible pacemakers of metabolism. *J. Theor. Biol.* **199**, 257–274. (doi:10.1006/jtbi.1999.0955)
47. Hulbert AJ, Else PL. 2005 Membranes and the setting of energy demand. *J. Exp. Biol.* **208**, 1593–1599. (doi:10.1242/jeb.01482)
48. Rakoff-Nahoum S, Medzhitov R. 2009 Toll-like receptors and cancer. *Nat. Rev. Cancer* **9**, 57–63. (doi:10.1038/nrc2541)
49. Jensen-Jarolim E *et al.* 2008 AllergoOncology: the role of IgE-mediated allergy in cancer. *Allergy* **63**, 1255–1266. (doi:10.1111/j.1398-9995.2008.01768.x)
50. Ranoa DRE *et al.* 2016 Cancer therapies activate RIG-I-like receptor pathway through endogenous non-coding RNAs. *Oncotarget*. **7**, 26 496–26 515. (doi:10.18632/oncotarget.8420)
51. Mathew R, Karantza-Wadsworth V, White E. 2007 Role of autophagy in cancer. *Nat. Rev. Cancer* **7**, 961–967. (doi:10.1038/nrc2254)
52. Wu J, Lanier LL. 2003 Natural killer cells and cancer. *Adv. Cancer Res.* **90**, 127–156. (doi:10.1016/B0000-0000(03)00049-9)
53. Dranoff G. 2004 Cytokines in cancer pathogenesis and cancer therapy. *Nat. Rev. Cancer* **4**, 11–22. (doi:10.1038/nrc1252)
54. Griffith OL, Moodie GEE, Civetta A. 2003 Genome size and longevity in fish. *Exp. Gerontol.* **38**, 333–337. (doi:10.1016/S0531-5565(02)00204-8)
55. McAloose D, Newton AL. 2009 Wildlife cancer: a conservation perspective. *Nat. Rev. Cancer* **9**, 605. (doi:10.1038/nrc2698)
56. Martin AP, Palumbi SR. 1993 Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl Acad. Sci. USA* **90**, 4087–4091. (doi:10.1073/pnas.90.9.4087)
57. Ducasse H *et al.* 2015 Can Peto's paradox be used as the null hypothesis to identify the role of evolution in natural resistance to cancer? A critical review. *BMC Cancer* **15**, 792. (doi:10.1186/s12885-015-1782-z)
58. Tomasetti C, Vogelstein B. 2015 Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81. (doi:10.1126/science.1260825)
59. Breitkreutz D, Hlatky L, Rietman E, Tuszynski JA. 2012 Molecular signaling network complexity is correlated with cancer patient survivability. *Proc. Natl Acad. Sci. USA* **109**, 9209–9212. (doi:10.1073/pnas.1201416109)
60. Takemoto K, Kihara K. 2013 Modular organization of cancer signaling networks is associated with patient survivability. *Biosystems* **113**, 149–154. (doi:10.1016/j.biosystems.2013.06.003)
61. Thuesen EV, Childress JJ. 1993 Metabolic rates, enzyme activities and chemical compositions of some deep-sea pelagic worms, particularly *Nectonemertes mirabilis* (Nemertea; Hoplonemertea) and *Poebius meseres* (Annelida; Polychaeta). *Deep Sea Res. Part I* **40**, 937–951. (doi:10.1016/0967-0637(93)90082-E)
62. Thuesen EV, Childress JJ. 1993 Enzymatic activities and metabolic rates of pelagic chaetognaths: lack of depth-related declines. *Limnol. Oceanogr.* **38**, 935–948. (doi:10.4319/lo.1993.38.5.0935)
63. Brawand D *et al.* 2011 The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348. (doi:10.1038/nature10532)
64. Stefano GB. 2014 Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Med. Sci. Monit. Basic Res.* **20**, 138–142. (doi:10.12659/MSMBR.892101)
65. Radivojac P *et al.* 2013 A large-scale evaluation of computational protein function prediction. *Nat. Methods* **10**, 221–227. (doi:10.1038/nmeth.2340)
66. Fortunato S. 2010 Community detection in graphs. *Phys. Rep.* **486**, 75–174. (doi:10.1016/j.physrep.2009.11.002)
67. Handorf T, Christian N, Ebenhöf O, Kahn D. 2008 An environmental perspective on metabolism. *J. Theor. Biol.* **252**, 530–537. (doi:10.1016/j.jtbi.2007.10.036)
68. Kanehisa M. 2013 Chemical and genomic evolution of enzyme-catalyzed reaction networks. *FEBS Lett.* **587**, 2731–2737. (doi:10.1016/j.febslet.2013.06.026)
69. Muto A, Kotera M, Tokimatsu T, Nakagawa Z, Goto S, Kanehisa M. 2013 Modular architecture of metabolic pathways revealed by conserved sequences of reactions. *J. Chem. Inf. Model.* **53**, 613–622. (doi:10.1021/ci3005379)