

# SCIENTIFIC REPORTS



Corrected: Publisher Correction

OPEN

## Lipidome determinants of maximal lifespan in mammals

Katarzyna Bozek<sup>1,2,13</sup>, Ekaterina E. Khrameeva<sup>3,4</sup>, Jane Reznick<sup>5</sup>, Damir Omerbašić<sup>6</sup>, Nigel C. Bennett<sup>7</sup>, Gary R. Lewin<sup>8</sup>, Jorge Azpurua<sup>8</sup>, Vera Gorbunova<sup>8</sup>, Andrei Seluanov<sup>8</sup>, Pierrick Regnard<sup>9</sup>, Fanelie Wanert<sup>9</sup>, Julia Marchal<sup>10</sup>, Fabien Pifferi<sup>10</sup>, Fabienne Aujard<sup>10</sup>, Zhen Liu<sup>11</sup>, Peng Shi<sup>11</sup>, Svante Pääbo<sup>2</sup>, Florian Schroeder<sup>12</sup>, Lothar Willmitzer<sup>12</sup>, Patrick Gialvalisco<sup>12</sup> & Philipp Khaitovich<sup>1,2,3</sup>

Maximal lifespan of mammalian species, even if closely related, may differ more than 10-fold, however the nature of the mechanisms that determine this variability is unresolved. Here, we assess the relationship between maximal lifespan duration and concentrations of more than 20,000 lipid compounds, measured in 669 tissue samples from 6 tissues of 35 species representing three mammalian clades: primates, rodents and bats. We identify lipids associated with species' longevity across the three clades, uncoupled from other parameters, such as basal metabolic rate, body size, or body temperature. These lipids clustered in specific lipid classes and pathways, and enzymes linked to them display signatures of greater stabilizing selection in long-living species, and cluster in functional groups related to signaling and protein-modification processes. These findings point towards the existence of defined molecular mechanisms underlying variation in maximal lifespan among mammals.

In contrast to average life expectancy that may change depending on living conditions, maximal lifespan (MLS) is a stable characteristic of a species. For instance, human average life expectancy has increased substantially over the past centuries, but our MLS remained constant at approximately 120 years<sup>1</sup>. At the same time, MLS varies greatly among species. In mammals it ranges from 3–4 years in small rodents to as long as 150–200 years in bow-head whales<sup>2</sup>. More remarkably, MLS evolves rapidly, resulting in markedly different lifespans and, consequently, different time of aging onset, even among closely related species. For instance, humans and macaques diverged only around 30 million years ago (MYA), yet over this time their MLS has diverged as much as three-fold<sup>3</sup>.

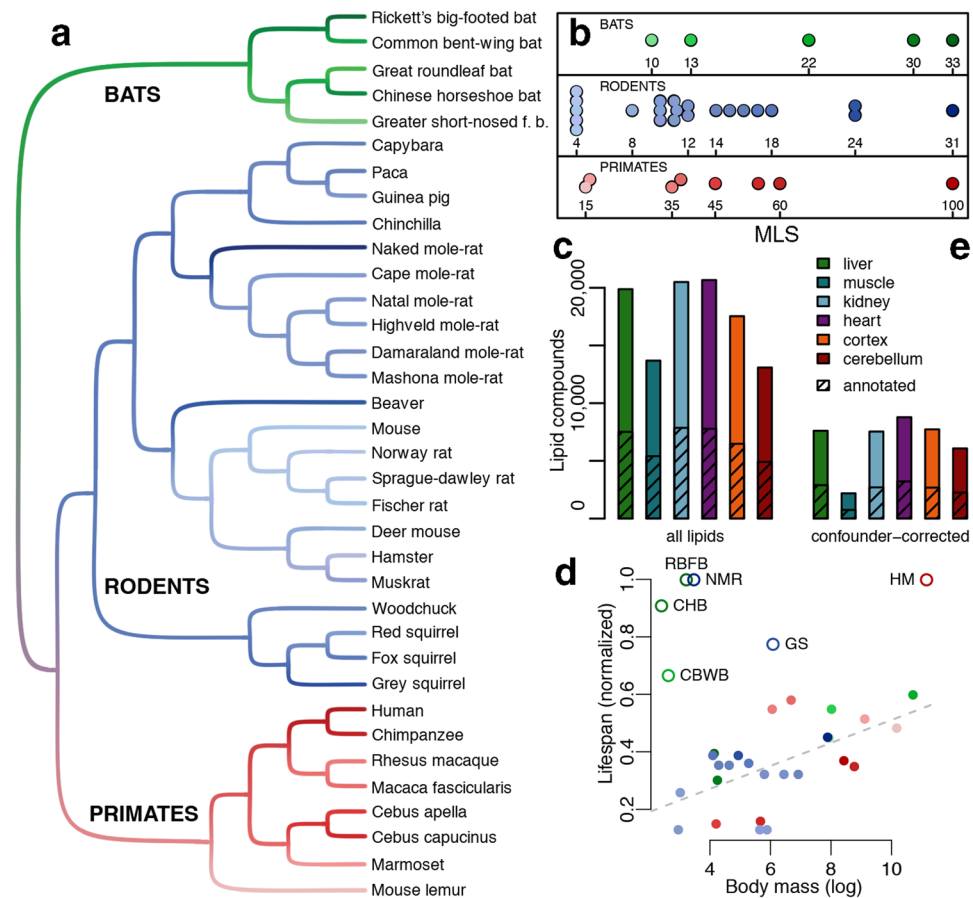
Despite remarkable evolutionary plasticity of MLS, the existence and nature of the molecular mechanisms controlling this variation remains unclear. Most studies so far have focused on mechanisms of lifespan plasticity within species, resulting in identification of longevity-related pathways, such as insulin signaling pathways and targets of rapamycin pathway, shared across a wide range of animal species: from worms to mice<sup>4–8</sup>. Genetic, dietary and pharmacological manipulation targeting these pathways resulted in more than 10-fold lifespan extension in short-living nematode worms<sup>9</sup>, but only approximately 40% (1.4-fold) lifespan extension in short-living mammals such as laboratory mice<sup>10</sup>. At the same time, natural variation in MLS among mammalian species exceeds 50-fold<sup>2</sup>.

<sup>1</sup>CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, 320 Yue Yang Road, Shanghai, 200031, China. <sup>2</sup>Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103, Leipzig, Germany. <sup>3</sup>Skoltech Center for Computational and Systems Biology, Skolkovo Institute for Science and Technology, Novaya st. 100, Skolkovo, 143025, Russia. <sup>4</sup>Institute for Information Transmission Problems, Russian Academy of Sciences, Bolshoy Karetny per. 19/1, Moscow, 127051, Russia. <sup>5</sup>Max-Delbrück Center for Molecular Medicine, Department of Zoology and Entomology, University of Pretoria, Pretoria, Republic of South Africa, South Africa. <sup>6</sup>Max-Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13092, Berlin, Germany. <sup>7</sup>Department of Zoology and Entomology, University of Pretoria, Pretoria, Republic of South Africa, South Africa. <sup>8</sup>University of Rochester, 434 Hutchison Hall, Rochester, NY, 14627-0211, USA. <sup>9</sup>SILABE, Primatology Center, University Louis Pasteur, Strasbourg, 4 Rue Blaise Pascal, 67070, France. <sup>10</sup>UMR 7179/CNRS-MNHN, 1 avenue du Petit Chateau, 91800, Brunoy, France. <sup>11</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 Jiaochang Dong Lu, Kunming, Yunnan, 650223, China. <sup>12</sup>Max Planck Institute for Molecular Plant Physiology, Am Mühlenberg 1, 14476, Potsdam, Germany. <sup>13</sup>Present address: Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Kunigami-gun, Okinawa, 904-0495, Japan. Correspondence and requests for materials should be addressed to K.B. (email: [kasia.bozek@oist.jp](mailto:kasia.bozek@oist.jp)) or P.G. (email: [gialvalisco@mpimp-golm.mpg.de](mailto:gialvalisco@mpimp-golm.mpg.de)) or P.K. (email: [khaitovich@eva.mpg.de](mailto:khaitovich@eva.mpg.de))

Received: 10 June 2016

Accepted: 20 December 2016

Published online: 31 January 2017



**Figure 1.** Dataset. **(a)** Phylogenetic tree of the 35 species used in this study. Colors show the clade identity and the MLS of species, with darker shades representing longer MLS. **(b)** Species' MLS distribution in years. Each dot represents a species. Colors are as in panel a. **(c)** Number of lipid compounds measured in each tissue. **(d)** Relationship between species' MLS and body mass. The colors are as in panel a. Open circles indicate species deviating from the linear relationship between MLS and body mass. The dashed line shows linear model fit to the remaining species (F-test,  $p < 0.05$ ). MLS was normalized to the maximal MLS value within each clade. NMR – naked mole-rat, HM – human, GS – grey squirrel, CBWB – common bent-winged bat, RBFB – rickett's big-footed bat, CHB – chinese horseshoe bat. **(e)** Number of lipid compounds of each tissue after removal of the compounds related to the confounding factors.

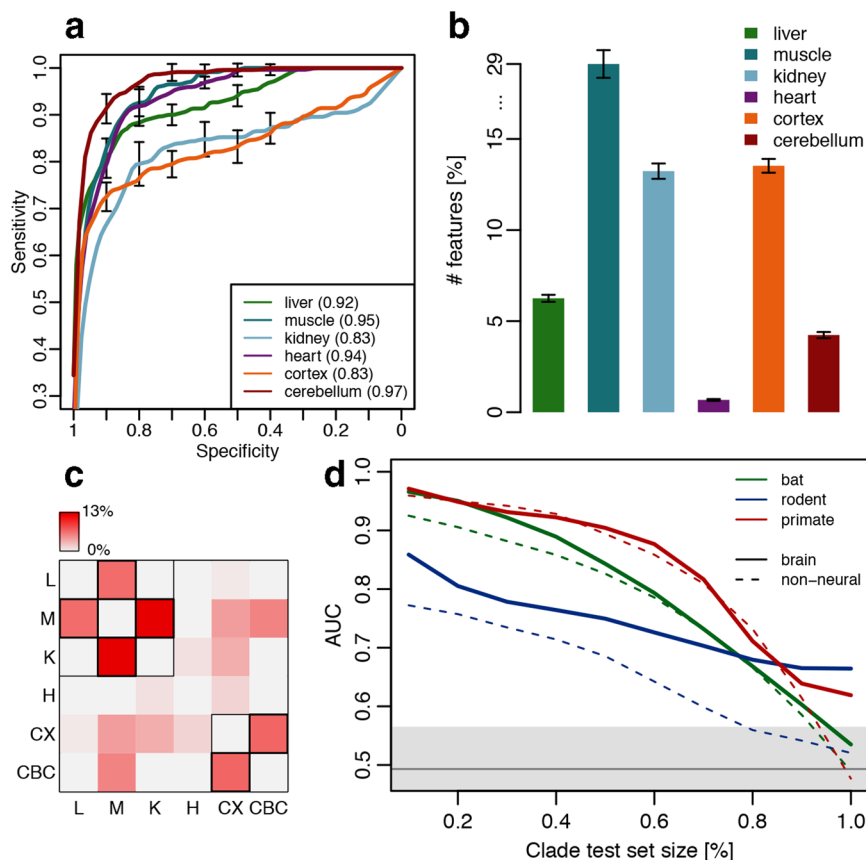
In this study we searched for a link between MLS and another marker of species' physiology – concentrations of hydrophobic metabolic compounds (henceforth referred to as “lipids” for simplicity). Recent scan of gene expression variation among 33 mammalian species with MLS differences of over 30-fold has shown that expression variation of 11–18% of analyzed 19,643 genes could be associated with MLS variation<sup>11</sup>. There is however a stronger evidence of genetic changes in genes controlling lipid metabolism to play a role in human longevity<sup>11–16</sup>, as well as in MLS differences among species<sup>17</sup>, and changes in lipid saturation levels<sup>18–25</sup>, pointing to lipids as a good potential target for investigation of molecular mechanisms underlying differences in MLS among mammalian species.

## Results

**Data samples.** To investigate the direct connection between lipid metabolic processes and MLS, we measured lipid concentrations in liver, muscle, kidney, heart, brain cortex and cerebellum samples in a total of 669 individuals of 35 mammalian species representing three phylogenetic clades: rodents, primates, and bats (Fig. 1a, Tables S1 and S2). Within clades, MLS varied between 4 and 31 years in rodents, 15 and 100 years in primates, and 10 and 33 years in bats (Fig. 1b)<sup>26</sup>. As long-living species we consider those with MLS exceeding 0.9 of the highest MLS in each clade. Based on this cutoff human, naked mole-rat and two bat species (chinese horseshoe bat and rickett's big-footed bat) are regarded as long-living throughout this study.

Using liquid chromatography coupled with high-precision mass spectrometry in positive and negative ionization modes, we quantified between 13,000 and 21,000 hydrophobic compounds with molecular weights below 3,000 Daltons, detected in at least 50% of individuals in one tissue of one species. Among them, between 5,000 and 8,000 (38%) were annotated using computational matching to the lipid compound database (Fig. 1c, Table S3)<sup>27</sup>.

Several factors, such as species' basal metabolic rate (BMR), body mass and temperature, diet, and ability to hibernate, have been shown to explain some, but not all, variability in species' lifespan. To discriminate the



**Figure 2.** Long lifespan prediction. **(a)** Accuracy of predictive models in each tissue. **(b)** Proportion of lipid compounds identified as MLS predictors in each tissue. **(c)** Proportion of MLS predictors shared between tissues. Black borders indicate significant overlap ( $p < 0.05$ ). **(d)** AUC of predictive models, constructed on training sets excluding an increasing proportion of randomly sampled individuals of one clade, with the models' accuracy, tested on the clade individuals excluded from the training set. Identity of the test clade is indicated by the curve color. The horizontal gray line and shaded region represent mean and standard deviation of the random predictor accuracy calculated by randomizing the output variable 100 times.

effect of these factors from the effect of lipid concentrations on MLS, our dataset included species where MLS was uncoupled from one or more of these factors (Figs 1d, S1). All lipids that were significantly related to these factors, as well as to individuals' age and sex (Figure S2, Table S4), were removed from further analyses (Fig. 1e).

**Predictive models of MLS.** To search for combinations of lipidome features that might be predictive of species' MLS, we constructed models distinguishing long-living species in all three clades from other species, based on the concentrations of lipid compounds in each tissue. The models were based on logistic regression with elastic net penalty – an approach which by applying a penalty measure on the number of lipids selected for the model during the training process selects only their limited number. This way we use this predictive model as a tool to both assess how accurately can we estimate long lifespan based on lipidome, as well as to investigate which lipid compounds are essential for this estimation and therefore potentially play a role in the species' exceptionally long lifespans. In each tissue, we performed  $10 \times 10$  cross-validation to determine optimal model parameters and to assess predictive accuracy. Since other classification methods, as well as linear models, resulted in comparable results (Figures S3–5, Table S5), we used models based on logistic regression throughout this study.

Remarkably, in each of the six tissues, lipid concentrations identified the species with long MLS with an average accuracy of 0.91, and a maximum accuracy of 0.97, estimated as the area under the receiver operating characteristic curve (AUC) (Fig. 2a). Notably, the lipids identified as long MLS predictors were not predictive of species' BMR, body mass and temperature, or other confounding factors when these were used as an output variable (Figure S6). The long MLS predictions were also robust with respect to uncertainties in the reported MLS of some of the species. We additionally assessed the robustness of the MLS predictors against age-related changes in lipid concentrations using 104 additional human samples from one brain tissue (cortex) and one non-neural tissue (heart) spanning the entire human postnatal lifespan (Table S6). Of the 14,104 and 26,105 compounds detected in cortex and heart, 75% and 25%, respectively, showed significant concentration changes with age (Table S7). Excluding these age-dependent lipids from our data did not decrease MLS prediction accuracy (Figure S7). Similarly, lipid concentration changes induced by postmortem delay<sup>28</sup> showed no association with long MLS predictions (Table S8).

The proportion of lipids selected by the models as the long MLS predictors varied substantially among tissues, comprising 1% of the lipid compounds in heart, 4% in cerebellum, 7% in liver, 14% in kidney, 14% in brain cortex and 30% in skeletal muscle (Fig. 2b, Table S9). Thus, long lifespan markers are distributed over a larger portion of the lipidome in skeletal muscle, brain cortex, and kidney compared to heart, cerebellum and liver. Despite variation in the number of MLS predictive features among tissues, they overlapped significantly between brain cortex and cerebellum, as well as among three non-neural tissues: liver, muscle and kidney (Fig. 2c, Table S10). Thus, our MLS prediction models might rely on common lipid composition and pathways in these two tissue types.

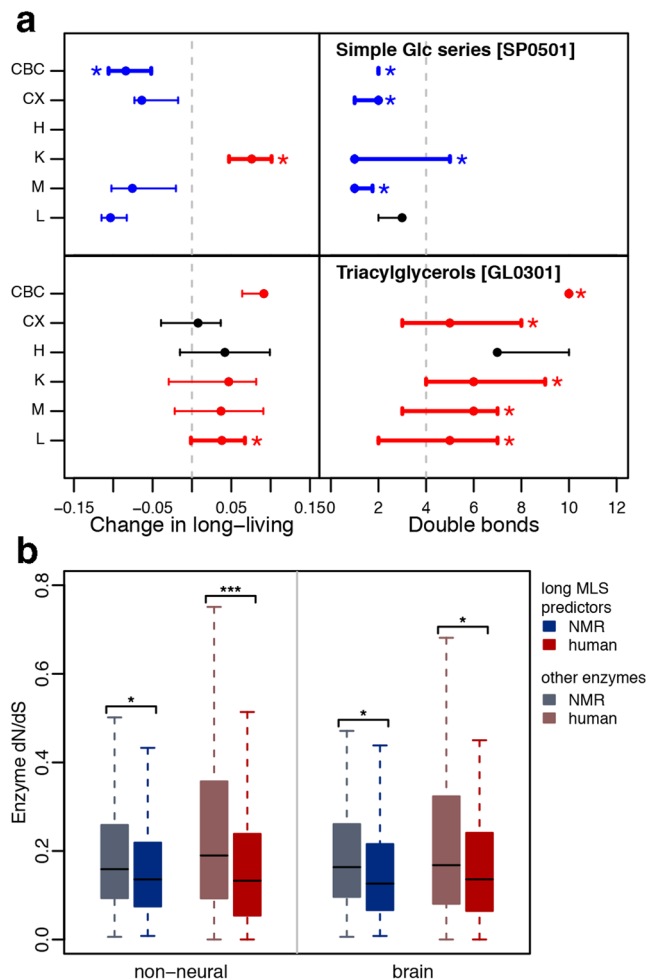
The high performance of MLS prediction models based on lipid concentrations in the three mammalian clades suggests the presence of common markers of long lifespan shared across the clades. To inspect this further, we tested the models' ability to predict MLS in a species from one clade based on the lipid information from the remaining two clades. Removal of an increasing number of individuals representing one clade from the training set resulted in a gradual reduction in prediction performance for this clade both in brain and in non-neural tissues (Figures S8 and 9). Still, the models based on brain data showed an above random performance even when all individuals of a given clade were removed from the training set (Fig. 2d). In other words, a model based solely on brain lipidome information from two clades, e.g. rodent and bats, is sufficient to make valid predictions about MLS for a species in the primate clade.

**Lipid predictors of long MLS.** Existence of common predictors of long lifespan suggests coordinated lipid concentration changes in long-living primates, rodents and bats. Indeed, changes in concentration of long lifespan predictors were positively and significantly correlated between clades, unlike changes in concentration of other lipids (permutation,  $p < 0.01$ ) (Figure S10). Lipid predictors showing coordinated concentration change in long-living species of all three clades were enriched in 12 specific lipid classes, 24 sub-classes and 12 metabolic pathways in all tissues except heart (hypergeometric test, permutation,  $p < 0.05$ ) (Table S11). What's more, concentrations of long MLS predictors were significantly shifted towards either higher or lower values in the long-living species in at least one tissue in 11 of 12 enriched lipid classes, 17 of 24 sub-classes and 10 of 12 metabolic pathways (permutation,  $p < 0.05$ ) (Table S11). Among them, triacylglycerols enriched in liver showed an increase in concentration levels in long-living species in almost all tissues; glycerophospholipids enriched in brain cortex were showing a decrease in concentration levels in most of the tissues; and sphingolipid metabolism pathway and several enriched classes of sphingolipids showed predominantly a decrease in concentration levels in all tissues except kidney (Fig. 3a, Figure S11). We inspected the relationship of the number of double bonds in lipid chains of the MLS predictors with their concentration levels in the long-living species and found several significant associations e.g. in sphingolipids, triacylglycerols, glycerolipids (Figure S11). However, no consistent pattern across all lipid classes and tissues could be observed. In order to inspect if the removal of lipids related to other confounding factors such as basal metabolic rate, body size, or body temperature could have affected these results, we performed the same analysis on the initial dataset before removal of the confounder-related lipids. A similarly inconsistent link between lipid concentration levels and double bond number was found among the enriched (Figures S12 and S13) as well as all lipid classes (Table S12) nevertheless the significant associations in sphingolipids, triacylglycerols, glycerolipids were confirmed. Overall, saturated lipids show higher correlation with species lifespan on average in all tissues except heart (Figure S14).

Several fatty acids were reported to show significant associations to longevity<sup>18–25</sup>. To verify their role in our data we manually searched and annotated fatty acids in our mass spectrometry measurements. We tested their concentration level difference between the long-living and other species and compared this difference to the distribution of the same difference across all other lipids in our dataset. The majority of fatty acids showed elevated concentration levels in the long-living species with several of them being significant in particular in the non-neural tissues. Notably, none of the saturated fatty acids showed significantly higher concentration levels in the long-living species (Table S13).

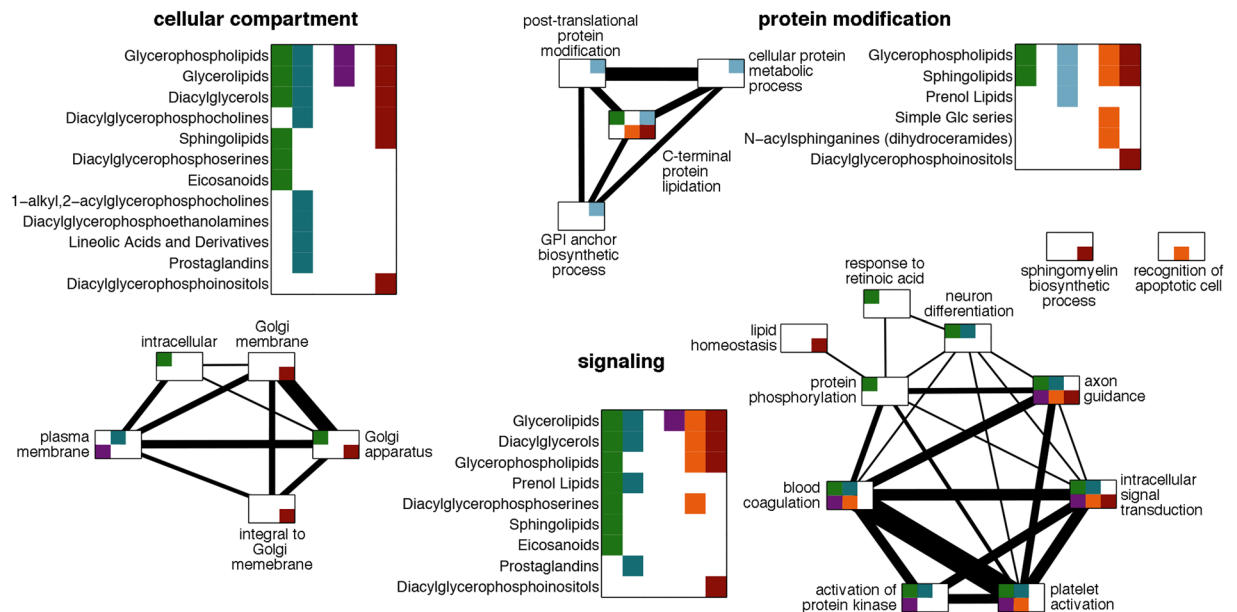
**Evolutionary processes leading to long MLS.** The presence of common lipidome features associated with long lifespan in the three clades suggests the existence of a shared molecular mechanism associated with MLS in rodents, primates, and bats. To corroborate this finding, we searched for a possible evolutionary signature of such a mechanism. We examined amino acid substitution levels, measured as dN/dS ratios, for enzymes linked to the long MLS predictors using available genomic data: genomes of long-living rodent and primate – naked mole-rat and human, as well as their short-living counterparts – mouse and macaque. We found that, in long-living species, these enzymes showed consistently and significantly lower amino acid substitution levels compared to enzymes linked to other lipid compounds detected in our data (Figs 3b, S15,16, Tables S14 and S15). By contrast, short-living species did not show a significant reduction of amino acid substitution levels in all primate tissues and all but two rodent tissues (Figure S17). This shows that metabolic processes identified as related to lifespan by our MLS prediction models indeed experienced increased evolutionary pressure in long-living species compared to short-living species, in both rodent and primate clades.

We next inspected the biological processes that involve enzymes linked to long MLS predictors, which additionally show a signature of increased conservation pressure in long-living species compared to their short-living clade counterparts. Analysis of Gene Ontology (GO) functional terms<sup>29</sup> revealed a significant enrichment of the enzymes with evidence of higher negative selection in all six tissues, with several GO terms overrepresented in more than one tissue (Table S16). The resulting network based on enzymes shared between GO terms was composed of three distinct modules related to: (i) signaling processes, (ii) protein modification processes and (iii) cellular components (Fig. 4). The signaling module contained five GO terms universally present across tissues, including “intracellular signal transduction”, “activation of protein kinase”, “platelet activation”, and “axon guidance”, and four tissue-specific terms, including “neuron differentiation” and “protein phosphorylation”. The protein modification module contained four GO terms including “post-translational protein modification”



**Figure 3.** Long-lifespan related lipids. **(a)** Concentration level change and double bond distribution of long MLS predictors belonging to two lipid sub-classes in six tissues. Change in the concentration level is calculated as the difference between mean of the concentration levels in the long-living species and all other species in each of the clades. Left panel shows median concentration change of long MLS predictors belonging to simple Glc series and triacylglycerol lipid sub-classes in long-living species. The error bars show the 0.25 to 0.75 inter-quintile intervals. Significant concentration shift towards lower and higher values in long-living species is indicated by blue and red colors, respectively. Significant enrichment of MLS predictors among detected compounds in a given lipid sub-class is shown by thick lines and asterisks. The vertical dashed line shows the median concentration change for all long MLS predictors. Right panel shows the median double bond number and its 0.25 to 0.75 inter-quintile intervals for the same lipid sub-classes as in the left panel. Significantly higher or lower number of double bonds in a given lipid group is shown by red and blue color respectively, as well as thick lines and asterisks. The vertical dashed line shows the median number of double bonds in all long MLS predictors. **(b)** Enzyme DNA sequence variation. Distributions of the dN/dS values of the enzymes linked to the lipid predictors of long MLS (darker colors), and the enzymes linked to other lipids detected in our dataset (lighter colors), in brain and non-neural tissues (liver, muscle, and kidney) in the long-living species. Asterisks indicate p-value range \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

and “C-terminal protein lipidation”. The cellular component module contained five GO terms including “Golgi apparatus components” and “plasma membrane” (Fig. 4). Long MLS predictors associated with these modules, in turn, belonged to specific lipid sub-classes. They included: glycerolipids, diacylglycerols, and glycerophospholipids, which were associated with the signaling and cellular compartment modules in multiple tissues; sphingolipids, associated with the protein modification module; diacylglycerophosphoinositols, simple Glc series, and N-acylsphingamines, associated with biological processes specific to brain; and prenol lipids and prostaglandins associated with processes specific to non-neural tissues (Tables S14 and S16). These lipid sub-classes intersected significantly with the ones significantly enriched in the long MLS predictors (permutation,  $p < 0.01$ ) and showed common concentration levels shifts in long-living species (Table S11). Among them, lipids in the cellular compartment term showed lower concentration levels in the long-living species, while signaling and protein modification terms contained both lipid classes with elevated and reduced concentration levels in the long-living species.



**Figure 4.** Functional modules. The network nodes represent GO terms significantly enriched in enzymes linked to the lipid predictors of long MLS, and having lower dN/dS values in the long-living species compared to the short-living species. The edges represent enzymes shared between terms with line width proportional to the number of shared enzymes: from 1 to 12. Colored squares within each GO term node indicate the tissue in which these enzymes are linked to the lipid predictors of long MLS: green – liver, dark blue – muscle, light blue – kidney, magenta – heart, orange – brain cortex, brown – cerebellum. Sub-classes of the lipid predictors of long MLS linked to the enzymes in each functional group are listed next to the corresponding network module. The colors indicate the tissues where they are used as predictors of long MLS, according to the same color code as the GO terms.

## Discussion

Our results show that long lifespan is associated with distinct lipidome features shared across three mammalian clades. Lipid predictors of long MLS differ across tissues, but overlap within brain and non-neural tissue types. The long MLS predictors identified in brain are especially conserved among clades, allowing long-living species identification within a clade solely based on lipid predictors from the other two clades.

The existence of lipid predictors shared among the three clades further manifested itself in form of specific lipid classes, sub-classes and pathways up- or down-regulated within long-living species of all three clades. While little can be said about functional significance of these changes one potentially important feature stands out: the genomes of long-living primate and rodent species show increased evolutionary selection acting upon the amino acid sequences of enzymes linked to lipid predictors of long MLS. These enzymes cluster in specific functional categories associated with signaling and protein modification processes, as well as the corresponding cellular compartments: Golgi apparatus and plasma membrane.

Another interesting observation is the role of double bonds in lipid chain in relationship to long lifespan. In contrast to energy lipids, structural lipids, such as sphingolipids, the major constituent of cell membrane show higher saturation degree in long-living species presumably as means of protection from continuous oxidative stress (Figs 3a, S11). Lower number of double bonds in these lipids potentially represents a mechanism for maintaining their stability over the extended lifespans of respective species. The relationship between the number of double bonds and species lifespan is more complex for energy-related lipids. The observed higher concentration of unsaturated lipids such as triacylglycerols in long-living species compared to other species could be related to the efficiency of energy release of these lipids. As saturated lipids do not require creation of a double bond within the beta-oxidation process it allows them for an accelerated energy production. Therefore, comparatively slower energy production of the unsaturated energy lipids could be related to the slower metabolism characteristic of long-living animals.

The relationship between the number of lipid chain double bond and its concentration level in long-living species could not be established across tissues and lipid classes, however several significant observations were found (Figure S11). As lipid saturation was associated with lifespan by a number of studies<sup>18–25</sup> to fully elucidate this mechanism of action specific tissue and lipid class examples should be investigated in detail and further validated experimentally.

The complex collection procedure of the large set of tissue samples used in our study did not allow us to make a comprehensive study of postmortem delay effect on our results. For this reason, we relied on a previously published data in which lipid degradation with postmortem time was extensively assessed<sup>28</sup>. We found a small and insignificant overlap of the long-lifespan predictors in our study with the postmortem delay-related lipids reported previously (Table S8). Therefore, in spite of the limited power of this method of assessing the effect of postmortem delay in our data, we conclude that it should not have a dramatic effect on the major results of this study.

In our approach, we were not able to discriminate between the causal lipidome determinants of a species' MLS and its byproducts. Nonetheless, the existence of a defined signature for long MLS, shared among three mammalian clades, which was reflected not only by lipid concentrations, but also enzyme conservation and functional clustering, strongly suggests the presence of a distinct molecular mechanism controlling MLS variation among mammalian species. Identification of this mechanism may give us a unique opportunity to understand the determinants of lifespan duration and the nature of the ageing process.

## Materials and Methods

**Samples.** Human samples were obtained from the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland, the Netherlands Brain Bank, and the Chinese Brain Bank Center (CBBC, <http://cbbc.scuec.edu.cn>, Wuhan, China). According to the protocol of these institutions, specific permission for brain autopsy and use of the brain tissue for research purpose was given by the donors or their relatives. Use of human autopsy tissue is considered non-human subject research and is IRB exempt under NIH guidelines. All subjects were defined as healthy with respect to the sampled tissue, by forensic pathologists at the corresponding tissue bank. All subjects suffered sudden death with no prolonged agony state.

Primate samples were obtained from Simian Laboratory Europe (SILABE) in Strasburg, research unit CNRS-MNHM 717 in Brunoy, France, German Primate Center (DPZ) in Goettingen, Max Planck Institute for Anthropology in Leipzig, Germany, Suzhou Experimental Animal Center in China, the Anthropological Institute and Museum of the University of Zürich-Irchel, Switzerland, and the Biomedical Primate Research Centre, the Netherlands. All non-human primates used in this study suffered sudden deaths for reasons other than their participation in this study and without any relation to the tissue used.

Rodent samples were obtained from the University of Rochester Biology Department, MDC Berlin, Department of Zoology and Entomology, University of Pretoria, and the animal center at the Shanghai Institute for Biological Sciences. All mice were from the C57/BL6 strain with no genetic modifications. Bat samples were obtained from Kunming Institute of Zoology of the Chinese Academy of Sciences. The use and care of the animals in this research was reviewed and approved by Biological Research Ethics Committee, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. All animals were lawfully acquired and their retention and use were in every case in compliance with national and local laws and regulations, and in accordance with the Institute for Laboratory Animal Research (ILAR) Guide for Care and Use of Laboratory Animals.

Whenever possible, cortex samples were dissected from the anterior part of the superior frontal gyrus, a cortical region corresponding to Brodmann area 10 in humans. For smaller animals, a frontal part of the frontal cortex was taken. The CBC samples were dissected from the lateral part of the cerebellar hemispheres. For all brain samples we took special care to dissect gray matter only. The kidney samples were dissected from the renal cortex. Muscle tissue was dissected from thigh skeletal muscle, mainly from *M. vastus lateralis*. The heart samples were dissected from heart muscle tissue avoiding fat and heart valves. Liver tissue was dissected from the peripheral part of the liver lobes.

Each tissue sample was of approximately 20 mg of weight and was dissected from the frozen postmortem tissue on dry ice without thawing. The tissue samples were first powdered using a mortar and pestle cooled by liquid nitrogen and then prepared for the lipidome extraction procedure.

**MS sample preparation and measurements.** Lipids were extracted from frozen tissue powder by solution of methanol, methyl-*tert*-butyl-ether 1:3 (v/v) according to Giavalisco *et al.*<sup>30</sup>. Briefly, 20 mg of frozen powdered tissue material was resuspended in 1 ml extraction solution containing 0.5 µg of corticosterone, 1.5 µg of 1,2-diheptadecanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids, 850360P, City State), 0.5 µg of 13C sorbitol and 0.25 µg of ampicillin. The samples were incubated for 10 min at 4 °C on an orbital shaker. This step was followed by ultrasonication in an ice-cooled bath-type sonicator for 10 min. To separate the organic from the aqueous phase 500 µl of a H<sub>2</sub>O:methanol mix (3:1(v/v)) was added to the homogenate, vortexed and centrifuged (5 min; 14,000 g). Finally, 500 µl of the supernatant (MTBE phase containing lipid compounds) was collected to a fresh 1.5 ml Eppendorf tube for lipid analysis. This aliquot was concentrated to complete dryness in a speed vacuum at room temperature.

Prior to analysis the dried pellets were re-suspended in 500 µL acetonitrile:isopropanol (7:3) (v:v), ultrasonicated, and centrifuged (5 min; 14,000 g). 200 µL of the cleared supernatant was transferred to fresh glass vials, of which 2 µL were loaded onto a UPLC system (Acquity, Waters) equipped with a C<sub>8</sub> reverse phase column (100 mm × 2.1 mm × 1.7 µm particles, Waters). In addition to the individual tissue samples we measured mixtures of tissue samples (pooled samples). These pools were prepared by mixing 10 µL of each individual sample after resuspension of the dried lipid pellet. Pooled samples were measured after every 40<sup>th</sup> sample, providing us information on system performance (sensitivity and retention time consistency), sample reproducibility and compound stability over the time of the MS-based analysis.

The mobile phase contained water (UPLC MS grade, BioSolve) with 10 mM NH<sub>4</sub>ac, 0.1% acetic acid (Buffer A), and acetonitrile/isopropanol (7:3, UPLC grade Biosolve) containing 10 mM NH<sub>4</sub>ac, 0.1% acetic acid (Buffer B). The flow rate of the system was set to 400 µL/min, and the gradient to: 1 min 45% A, 3 min linear gradient from 45% A to 35% A, 8 min linear gradient from 25% A to 11% A, 3 min linear gradient from 11% A to 1% A. After washing the column for 3 min with 1% A, the buffer is set back to 45% A, and the column is re-equilibrated for 4 min (22 min total run time).

The mass spectra were acquired using an Exactive mass spectrometer (Thermo-Fisher, Bremen, Germany). The spectra of the pooled samples were recorded using altering full scan and all ion fragmentation scan mode, covering a mass range from 100–1500 m/z. The individual samples were measured using two time segments where the measurements were performed either in negative (0–13 min) or in positive (13–20 min) ionization mode. This strategy allowed us to reduce the number of measurements by 50%. The resolution for all measurements was set

to 25,000 providing 5 scans per second, restricting the loading time of the orbitrap cell to a maximum of 100 ms. The capillary voltage was 3 kV with a sheath gas flow value of 60, an auxiliary gas flow of 35, and a capillary temperature of 150 °C. The drying gas in the heated electro spray source was 350 °C, while the skimmer voltage and the tube lens were 25 V and 130 V, respectively. The spectra were recorded from 0 min to 20 min of the UPLC gradients.

**Peak extraction and alignment.** Peaks were extracted using the Progenesis QI metabolomics software (Version 1.0.5165, Nonlinear Dynamics, New Castle, UK) allowing only the detection of a single isotope  $(M + H)^+$  in positive- and  $(M - H)^-$  in negative ionization mode. The software automatically detects the different isotope peaks and clusters them together while reporting the summed intensity under the monoisotopic mass retention time feature. Accordingly we define as a peak, throughout this manuscript, as an individual mass trace of the first isotope. In a later step the different, possible adducts of each compound class were matched, as described in below to remove the redundancy from the peak annotation.

**Normalization.** All peak concentration values were normalized by the internal standard (IS) PC 34:0 concentration levels and log transformed. Samples with IS concentration levels outside of a 2.5 standard deviation range around the mean of that IS in the tissue were removed from the dataset.

**Peak annotation.** We used mass searches with a mass tolerance of 10 ppm, allowing adducts listed in Table S18 in the positive and negative ionization modes<sup>31</sup>. We tested a lower cutoff for database search of 5 ppm and obtained qualitatively the same results. The obtained  $m/z$  values were next searched against the LIPID MAPS<sup>27</sup> database for the initial annotation of the positive and negative ionization mode lipid datasets. For each lipid class we defined a list of possible adducts, Table S18, used in this step of data base search. In the second step, all the remaining, less common, adducts were used in an additional database search to test the support of the database matches found in the first step of the annotation procedure (Table S19). Next, peaks showing a correlation  $>0.7$ , a difference in RT of  $<0.05$  and matched to the same lipid compound in the database search were merged together. The average concentration level of the merged peaks in all samples was used in further analysis. This procedure resulted in merging 427–880 peaks in different tissues into composite peaks of size up to 6. As a result, the datasets contained between 5,313 and 13,067 compounds in different tissues and ionization modes and between 13,089 (cerebellum) and 20,669 (heart) compounds in both ionization modes together.

We used LIPID MAPS and HMDB database annotation to assign each identified compound to its lipid class and subclass. We tested the overrepresentation of lipid classes and subclasses in a given compound group using a hypergeometric test corrected for multiple testing by random sampling of the same number of lipid compounds from our dataset 1000 times.

To determine the number of double bonds of the compounds assigned to each peak we used their InChI identifiers<sup>32</sup>.

**Confounder correlation search.** We searched the literature<sup>33,34</sup> for the reported species basal metabolic rate (BMR), body temperature, diet, body mass, and hibernation information (Table S20). In addition, we collected individual sex and age information (Table S1). The relation of lipid concentration levels to these factors was tested based on (a) t-test for sex and hibernation factors, (b) analysis of variance (ANOVA) for diet factors, (c) linear regression analysis for BMR, body temperature, body mass, and age factors. The resulting p-values were further corrected for multiple testing using permutations. Lipid compounds showing significant ( $p < 0.01$ ) relationship with any of the factors were removed from the analysis (Table S3).

Previous work<sup>28</sup> has shown that concentrations of less than 8% of lipids detected in various primate tissues are affected by postmortem delay (PMD). We found no significant association between lipids showing these PMD-associated changes and the long MLS predictors (Table S8). Furthermore, samples collected from long-living species in our dataset are characterized by different PMD: no PMD for naked mole-rats,  $<5$  h for bats, and up to 50 h PMD for humans. Thus, long MLS predictors shared among the three clades cannot be due to the effect of PMD.

In the tests requiring matching lipid peaks between datasets, including the time-series and PMD measurements, cutoffs of 10 ppm of the mass to charge ratio and 0.1 sec RT were used. Age-related lipid change was assessed using a polynomial function of up to 3<sup>rd</sup> degree fit to each lipid compound measurement (F-test,  $p < 0.01$ ).

**Classifier construction.** We used logistic regression with elastic net penalty to construct binary classifiers distinguishing the long-living species from other species in the dataset. The  $\alpha$  parameter of the classifier was chosen based on the classification accuracy in  $10 \times 10$  cross-validation runs. Classifier accuracy was assessed through cross-validation, for the functional analysis the model was retrained on all sample data using the  $\alpha$  parameter value selected in the cross-validation procedure.

**Other models.** For comparison, we implemented a feature selection procedure for a support vector machine (SVM) classifier according to<sup>35</sup>. In every cross-validation run, iteratively, 1% of the initial number of features with the lowest coefficient in the linear SVM was removed and the model was retrained on the remaining features until there were no more than 10 features left. Out of all numbers of features tested in the iterative procedure, the best performing one was chosen and the model accuracy was assessed as the average across the cross-validation runs for the chosen number of features. For model comparison, the SVM was retrained using all sample data in the iterative procedure until the chosen number of features was reached.



In the same way as the logistic regression model we constructed a linear regression model with elastic net penalty. For the response variable in this model we used lifespan of a species normalized to the maximum lifespan of a species within each clade.

**Individual clade prediction test.** In order to inspect the robustness of the long lifespan classifier based on incomplete clade information, we performed a test of eliminating an increasing proportion of a clade individuals from the training set and using them in the test set. In this procedure we randomly sampled an increasing proportion of clade individuals from 10% to 90% in 10%-steps, 100 times in every step. In each randomization step, after calculating prediction models for all individual tissues, tissue datasets were merged into: a non-neural dataset including liver, muscle and kidney, and a brain dataset including cortex and cerebellum. The previously sampled training and test sets in the individual tissues were merged to test the prediction accuracy of models based on the merged data. Prediction accuracy of the models based on limited clade data was estimated as the average of all 100 sampling trials done for each elimination step.

Lipid datasets in the randomization procedure were merged based on the coefficients of lipid peaks in the individual tissue models. Individual tissue lipid data were first sorted according to the coefficients in the individual-tissue models, and according to the difference in the concentration levels between long- and short-living individuals in the training set. Then, peaks with zero coefficients and lowest difference between the long- and short-living individuals were removed in order to create datasets of the same sizes. Finally, the datasets were merged according to the predefined order.

**Genome analysis.** We used genomes of primate species (human, rhesus macaque and marmoset) and rodent species (naked mole-rat and mouse). Orthologous gene information for 908 lipid enzymes identified based on their link to the lipids in our dataset<sup>27,36</sup> was downloaded from NCBI Gene database ([ftp://ftp.ncbi.nih.gov/gene/ DATA/gene\\_group.gz](ftp://ftp.ncbi.nih.gov/gene/ DATA/gene_group.gz)) and concatenated into Table S17. We used *muscle*<sup>37</sup> to align exonic sequences (consensus CDS only, according to NCBI Nucleotide database) of corresponding RNAs annotated in NCBI Gene database (<ftp://ftp.ncbi.nih.gov/gene/ DATA/gene2refseq.gz>, 'reviewed' status only), in two sets of species: (1) human, rhesus macaque and marmoset as an outgroup; (2) naked mole-rat, mouse and human as an outgroup. We used *codeml* (with all zero parameters except CodonFreq = 2, model = 2, kappa = 2, omega = 2, fix\_alpha = 1, ncatG = 4)<sup>38</sup> to calculate dN/dS ratios for all tree branches separately (Table S17).

We next compared the distribution of the dN/dS ratios of enzymes linked to the long MLS predictors and enzymes linked to other lipids in our dataset. For each enzyme, we calculated the proportion of the number of linked long MLS predictors to the number of all lipids this enzyme is linked to. We defined as lifespan-related the enzymes with top 30% (top 25% in heart and top 35% in non-neural tissues) proportions of the long MLS predictors in a given tissue. The dN/dS ratios of these lifespan-related enzymes were compared to the dN/dS of other lipid enzymes.

## References

- Oeppen, J. & Vaupel, J. W. Demography. Broken limits to life expectancy. *Science* **296**, 1029–1031, doi:[10.1126/science.1069675](https://doi.org/10.1126/science.1069675) (2002).
- Foot, A. D. *et al.* Ancient DNA reveals that bowhead whale lineages survived Late Pleistocene climate change and habitat shifts. *Nature communications* **4**, 1677, doi:[10.1038/ncomms2714](https://doi.org/10.1038/ncomms2714) (2013).
- Colman, R. J. *et al.* Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. *Nature communications* **5**, 3557, doi:[10.1038/ncomms4557](https://doi.org/10.1038/ncomms4557) (2014).
- Bonafe, M. *et al.* Polymorphic variants of insulin-like growth factor I (IGF-I) receptor and phosphoinositide 3-kinase genes affect IGF-I plasma levels and human longevity: cues for an evolutionarily conserved mechanism of life span control. *The Journal of clinical endocrinology and metabolism* **88**, 3299–3304, doi:[10.1210/jc.2002-021810](https://doi.org/10.1210/jc.2002-021810) (2003).
- Holzenberger, M. *et al.* IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature* **421**, 182–187, doi:[10.1038/nature01298](https://doi.org/10.1038/nature01298) (2003).
- Jia, K., Chen, D. & Riddle, D. L. The TOR pathway interacts with the insulin signaling pathway to regulate *C. elegans* larval development, metabolism and life span. *Development* **131**, 3897–3906, doi:[10.1242/dev.01255](https://doi.org/10.1242/dev.01255) (2004).
- Kapahi, P. *et al.* Regulation of lifespan in *Drosophila* by modulation of genes in the TOR signaling pathway. *Current biology: CB* **14**, 885–890, doi:[10.1016/j.cub.2004.03.059](https://doi.org/10.1016/j.cub.2004.03.059) (2004).
- Harrison, D. E. *et al.* Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* **460**, 392–395, doi:[10.1038/nature08221](https://doi.org/10.1038/nature08221) (2009).
- Ayyadevara, S., Alla, R., Thaden, J. J. & Shmookler Reis, R. J. Remarkable longevity and stress resistance of nematode PI3K-null mutants. *Aging cell* **7**, 13–22, doi:[10.1111/j.1474-9726.2007.00348.x](https://doi.org/10.1111/j.1474-9726.2007.00348.x) (2008).
- Flurkey, K., Papaconstantinou, J., Miller, R. A. & Harrison, D. E. Lifespan extension and delayed immune and collagen aging in mutant mice with defects in growth hormone production. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6736–6741, doi:[10.1073/pnas.111158898](https://doi.org/10.1073/pnas.111158898) (2001).
- Fushan, A. A. *et al.* Gene expression defines natural changes in mammalian lifespan. *Aging cell* **14**, 352–365, doi:[10.1111/accel.12283](https://doi.org/10.1111/accel.12283) (2015).
- Barzilai, N. *et al.* Unique lipoprotein phenotype and genotype associated with exceptional longevity. *Jama* **290**, 2030–2040, doi:[10.1001/jama.290.15.2030](https://doi.org/10.1001/jama.290.15.2030) (2003).
- Barzilai, N., Gabrieli, I., Gabrieli, M., Iankowitz, N. & Sorkin, J. D. Offspring of centenarians have a favorable lipid profile. *Journal of the American Geriatrics Society* **49**, 76–79 (2001).
- Heijmans, B. T. *et al.* Lipoprotein particle profiles mark familial and sporadic human longevity. *PLoS medicine* **3**, e495, doi:[10.1371/journal.pmed.0030495](https://doi.org/10.1371/journal.pmed.0030495) (2006).
- Mooijaart, S. P. *et al.* ApoE plasma levels and risk of cardiovascular mortality in old age. *PLoS medicine* **3**, e176, doi:[10.1371/journal.pmed.0030176](https://doi.org/10.1371/journal.pmed.0030176) (2006).
- Vaarhorst, A. A. *et al.* Lipid metabolism in long-lived families: the Leiden Longevity Study. *Age* **33**, 219–227, doi:[10.1007/s11357-010-9172-6](https://doi.org/10.1007/s11357-010-9172-6) (2011).
- Jobson, R. W., Nabholz, B. & Galtier, N. An evolutionary genome scan for longevity-related natural selection in mammals. *Molecular biology and evolution* **27**, 840–847, doi:[10.1093/molbev/msp293](https://doi.org/10.1093/molbev/msp293) (2010).
- Hulbert, A. J., Pamplona, R., Buffenstein, R. & Buttemer, W. A. Life and death: metabolic rate, membrane composition, and life span of animals. *Physiol Rev* **87**, 1175–1213, doi:[10.1152/physrev.00047.2006](https://doi.org/10.1152/physrev.00047.2006) (2007).

19. Jove, M. *et al.* Plasma long-chain free fatty acids predict mammalian longevity. *Sci Rep* **3**, 3346, doi:[10.1038/srep03346](https://doi.org/10.1038/srep03346) (2013).
20. Jove, M. *et al.* A Stress-Resistant Lipidomic Signature Confers Extreme Longevity to Humans. *J Gerontol A Biol Sci Med Sci*, **10.1093/gerona/glw048** (2016).
21. Naudi, A. *et al.* Membrane lipid unsaturation as physiological adaptation to animal longevity. *Front Physiol* **4**, 372, doi:[10.3389/fphys.2013.00372](https://doi.org/10.3389/fphys.2013.00372) (2013).
22. Pamplona, R. Membrane phospholipids, lipoxidative damage and molecular integrity: a causal role in aging and longevity. *Biochim Biophys Acta* **1777**, 1249–1262, doi:[10.1016/j.bbabi.2008.07.003](https://doi.org/10.1016/j.bbabi.2008.07.003) (2008).
23. Pamplona, R. & Barja, G. Highly resistant macromolecular components and low rate of generation of endogenous damage: two key traits of longevity. *Ageing Res Rev* **6**, 189–210, doi:[10.1016/j.arr.2007.06.002](https://doi.org/10.1016/j.arr.2007.06.002) (2007).
24. Pamplona, R., Barja, G. & Portero-Otin, M. Membrane fatty acid unsaturation, protection against oxidative stress, and maximum life span: a homeoviscous-longevity adaptation? *Ann N Y Acad Sci* **959**, 475–490 (2002).
25. Pamplona, R. *et al.* Low fatty acid unsaturation protects against lipid peroxidation in liver mitochondria from long-lived species: the pigeon and human case. *Mech Ageing Dev* **86**, 53–66 (1996).
26. Tacutu, R. *et al.* Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic acids research* **41**, D1027–1033, doi:[10.1093/nar/gks1155](https://doi.org/10.1093/nar/gks1155) (2013).
27. Fahy, E. *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *Journal of lipid research* **50**(Suppl), S9–14, doi:[10.1194/jlr.R800095-JLR200](https://doi.org/10.1194/jlr.R800095-JLR200) (2009).
28. Bozek, K. *et al.* Organization and evolution of brain lipidome revealed by large-scale analysis of human, chimpanzee, macaque, and mouse tissues. *Neuron* **85**, 695–702, doi:[10.1016/j.neuron.2015.01.003](https://doi.org/10.1016/j.neuron.2015.01.003) (2015).
29. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**, 25–29, doi:[10.1038/75556](https://doi.org/10.1038/75556) (2000).
30. Giavalisco, P. *et al.* Elemental formula annotation of polar and lipophilic metabolites using (13) C, (15) N and (34) S isotope labelling, in combination with high-resolution mass spectrometry. *The Plant journal: for cell and molecular biology* **68**, 364–376, doi:[10.1111/j.1365-313X.2011.04682.x](https://doi.org/10.1111/j.1365-313X.2011.04682.x) (2011).
31. Hummel, J. *et al.* Ultra performance liquid chromatography and high resolution mass spectrometry for the analysis of plant lipids. *Frontiers in plant science* **2**, 54, doi:[10.3389/fpls.2011.00054](https://doi.org/10.3389/fpls.2011.00054) (2011).
32. Stein, S. E., Heller, S. R. & Tchekhovski, D. In *Nimes International Chemical Information Conference*. 131–143 (2003).
33. Jones, K. *et al.* PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90** (2009).
34. White, C. R. & Seymour, R. S. Mammalian basal metabolic rate is proportional to body mass<sup>2/3</sup>. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4046–4049, doi:[10.1073/pnas.0436428100](https://doi.org/10.1073/pnas.0436428100) (2003).
35. Guyon, I., Weston, J., Barnhill, J. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Machine Learning* **46**, 389–422 (2002).
36. Wishart, D. S. *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic acids research* **37**, D603–610, doi:[10.1093/nar/gkn810](https://doi.org/10.1093/nar/gkn810) (2009).
37. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797, doi:[10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340) (2004).
38. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**, 1586–1591, doi:[10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088) (2007).

## Acknowledgements

We thank Mehmet Somel and Jerome Boyd-Kirkup for help in sample procurement and manuscript preparation. This study was supported by the National Natural Science Foundation of China (grant 31420103920), Strategic Priority Research Program of the Chinese Academy of Sciences (grant XDB13010200), the National Natural Science Foundation of China (grant 91331203), the National One Thousand Foreign Experts Plan (grant WQ20123100078), the Bureau of International Cooperation, Chinese Academy of Sciences (grant GJHZ201313) and the Federal Targeted Program for Research and Development in Priority Areas of Advancement of the Russian Scientific and Technological Complex for 2014–2020 (the Ministry of Education and Science of the Russian Federation), grant № 14.615.21.0002, the Unique identifier of the agreement: RFMEFI61515×0002. Additional support was obtained from the European Research Council (advanced grant 294678 to GRL).

## Author Contributions

J.R., D.O., N.B., G.L., J.A., V.G., A.S., P.R., F.W., J.M., F.P., F.A., Z.L., P.S., S.P. provided tissue samples, P.G., L.W., F.S. conducted the experiments, P.K., K.B., designed the study, K.B., E.E.K. performed the analysis, P.K., K.B. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-00037-7](https://doi.org/10.1038/s41598-017-00037-7)

**Competing financial interests:** The authors declare no competing financial interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017