

DNA repair in species with extreme lifespan differences

Sheila L. MacRae¹, Matthew McKnight Croken¹, R.B. Calder¹, Alexander Aliper², Brandon Milholland¹, Ryan R. White¹, Alexander Zhavoronkov², Vadim N. Gladyshev⁴, Andrei Seluanov³, Vera Gorbunova³, Zhengdong D. Zhang¹, and Jan Vijg¹

¹Department of Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA;

²InSilico Medicine, Inc., Johns Hopkins University, Baltimore, MD 21218, USA;

³Department of Biology, University of Rochester, Rochester, NY 14627, USA;

⁴Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.

Key words: genome maintenance, aging, longevity, RNA-seq, transcriptome, naked mole rat, DNA repair

Received: 11/12/15; **Accepted:** 11/29/15; **Published:** 12/30/15

Correspondence to: Jan Vijg, PhD; Sheila L. MacRae, PhD; **E-mail:** jan.vijg@einstein.yu.edu, lola.macrae@phd.einstein.yu.edu

Copyright: MacRae et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Abstract: Differences in DNA repair capacity have been hypothesized to underlie the great range of maximum lifespans among mammals. However, measurements of individual DNA repair activities in cells and animals have not substantiated such a relationship because utilization of repair pathways among animals—depending on habitats, anatomical characteristics, and life styles—varies greatly between mammalian species. Recent advances in high-throughput genomics, in combination with increased knowledge of the genetic pathways involved in genome maintenance, now enable a comprehensive comparison of DNA repair transcriptomes in animal species with extreme lifespan differences. Here we compare transcriptomes of liver, an organ with high oxidative metabolism and abundant spontaneous DNA damage, from humans, naked mole rats, and mice, with maximum lifespans of ~120, 30, and 3 years, respectively, with a focus on genes involved in DNA repair. The results show that the longer-lived species, human and naked mole rat, share higher expression of DNA repair genes, including core genes in several DNA repair pathways. A more systematic approach of signaling pathway analysis indicates statistically significant upregulation of several DNA repair signaling pathways in human and naked mole rat compared with mouse. The results of this present work indicate, for the first time, that DNA repair is upregulated in a major metabolic organ in long-lived humans and naked mole rats compared with short-lived mice. These results strongly suggest that DNA repair can be considered a genuine longevity assurance system.

INTRODUCTION

Maintaining the integrity of the genome is among the most critical functions of a cell. The accumulation of DNA damage and mutations in multicellular organisms increases the risk of cancer and is linked to aging [1-3]. Essential genome maintenance functions include cell cycle control, regulation of cellular death and senescence, and DNA damage signaling and repair. DNA repair pathways are highly conserved, and because of their critical importance for cell survival and

genome integrity they have long been hypothesized to represent primary longevity assurance systems. Indeed, ample research has been devoted to measuring DNA repair activities in cells and animals from species with differences in maximum lifespan. Most notably, nucleotide excision repair activity in cultured fibroblasts has been reported as strongly correlated with maximum lifespan [4]. However, later, this correlation was mainly attributed to differences between humans and rodents in the utilization of excision repair for removal of UV-induced lesions, which rodents, due to their fur and

nocturnal habits, experience at a much lower rate than humans [5]. While many other attempts have been made to demonstrate correlations between DNA repair activities and lifespan, thus far, there has been no conclusive evidence that long-lived humans have DNA repair activities that are superior to those of short-lived mice.

It is now well documented that even among phylogenetically related rodents, extreme differences in lifespan are common. Most notably, the naked mole rat (NMR), a rodent species with an extreme resistance to cancer, has a maximum lifespan of approximately 30 years, i.e., about 10 times the lifespan of mice [6]. In previous work we showed that the gene mutation rate is significantly higher in the mouse than in NMR or human [7]. Because of the extreme longevity and cancer resistance of the naked mole rat, we speculated that these characteristics were based, at least in part, on more effective genome maintenance, which could be mediated by higher expression levels of genes involved in DNA repair. Most DNA repair genes are constitutively expressed [8] at a level that, in human cancer cell lines, was found to be inversely correlated with chromosomal instability [9]. Hence, we considered it of interest to compare the expression of DNA repair genes *in vivo* between humans and mice as well as between NMR and mice, reasoning that if longevity is controlled mainly by genome maintenance we should

find this reflected in the level of DNA repair-related transcripts *in vivo*.

In this study, we looked for longevity-associated enrichment of genes involved in DNA repair in RNA-seq data sets obtained from liver in the three different species: human, NMR, and mouse. Liver is the target organ of a large number of DNA damaging agents, and chronic liver diseases are characterized by increased oxidative stress [10]. It therefore requires proficient DNA repair processes for its long-term integrity. Our results show that the expression of genes encoding core DNA repair enzymes is indeed significantly higher in human and NMR than in the mouse, and that most DNA repair signaling pathways are upregulated in the long-lived species. This work provides further evidence that DNA repair functions as a longevity assurance mechanism in mammals.

RESULTS

To compare DNA repair gene expression in liver from the three species—human, NMR, and mouse—we produced RNA-seq data for three individuals per species. Figure 1 shows a multidimensional scaling (MDS) plot of the nine sequencing datasets, in which the human, mouse, and NMR samples cluster by species. Table S1 provides details on library sequencing read counts and alignments.

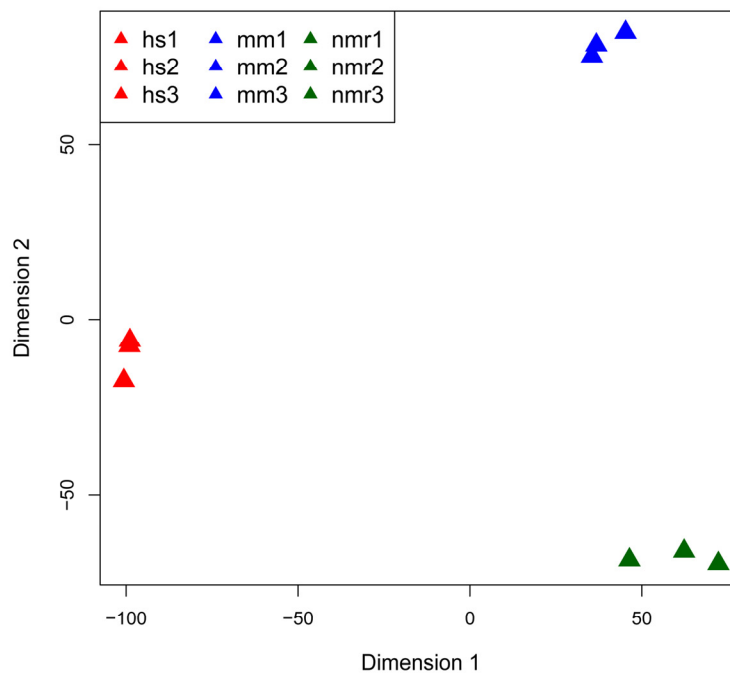


Figure 1. Multi-dimensional scaling (MDS) plot of RNA-seq datasets. MDS plot of the 500 highest expressed genes shows that each of the three species clusters together in first two dimensions.

Comparing RNA-seq data sets between species poses many challenges. It is particularly difficult to identify orthologs, especially for species with less well-annotated genomes. For our analysis, we used mouse protein-coding genes to identify orthologs in the NMR genome. To identify regions of the NMR genome (hetGla2 [11]) homologous to mouse genes, we isolated mouse entries from UCSC cross-species BLAT results. This provided the genomic coordinates for putative NMR orthologous genes. Initially, there were 16,708 protein-coding mouse genes that mapped to the NMR genome. Of these, 13,993 mouse genes aligned to the NMR genome without overlapping any other alignment. From this list, 13,209 genes exist as putative orthologous gene pairs between mouse and human. We further narrowed this list down to 13,181 genes with unambiguous annotation (Table 1).

In addition, there are caveats to directly comparing expression of orthologs between species, as biologically the systems are not the same. For example, orthologs may have different gene lengths in the different species, confounding the normalization of read counts. There are also many biological characteristics, such as transcription factors and other gene transcription regulatory mechanisms, that may differ among species and therefore one must be careful in drawing conclusions from direct comparisons of ortholog expression between different species. To mitigate technical biases we used quantile normalization of our read counts, which also factored in gene length [12]. To overcome biological biases that could possibly affect our analysis, we took several approaches to compare expression of DNA repair genes from these three species, including both differential expression analysis

Table 1. Orthologs compared and their expression in each species.

Species	Genes compared	Genes expressed	% expressed
Human	13181	5510	41.80%
Mouse	13181	5459	41.42%
NMR	13181	5496	41.70%

The total number of protein-coding orthologs genes analyzed and the number expressed in each species.

Table 2. Essential DNA repair genes having higher expression in human and NMR compared with mouse.

Gene symbol	Full gene name
MBD4	methyl-CpG binding domain protein 4
MSH3	mutS homolog 3 (E. coli)
MUTYH	mutY homolog (E. coli)
NEIL1	nei endonuclease VIII-like 1 (E. coli)
NEIL2	nei like 2 (E. coli)
NHEJ1	nonhomologous end-joining factor 1
POLK	polymerase (DNA directed) kappa
POLL	polymerase (DNA directed), lambda
TDG	thymine-DNA glycosylase
TP53	tumor protein p53
UBE2N	ubiquitin-conjugating enzyme E2N (UBC13 homolog, yeast)
XRCC6	X-ray repair complementing defective repair in Chinese hamster cells 6; 70 kDa subunit, Ku70

Includes genes with probability of differential expression greater than 0.95 and fold change greater than two.

and pathway analysis—the latter of which considers all of the genes in a pathway and is less sensitive to biases introduced when directly comparing expression of individual orthologs between species. First, we looked at differential expression of 130 genes encoding proteins characterized as directly involved in DNA repair in mouse and human. Next, we performed a signaling pathway activation analysis with a focus on signaling pathways involved in DNA repair. Finally, to take into account the overall picture of gene expression differences between the three species in liver, we performed an unbiased differential expression analysis of all 13,181 identified protein-coding homologs.

DNA repair genes are more highly expressed in human and NMR than mouse

To determine whether genes involved in DNA repair are more highly expressed in the two long-lived species, we first compiled a set of 130 genes involved in DNA repair activities in mouse and human (Table S2), [13, 14]. In comparing the expression of all of these DNA repair genes among the three species, we found that, as a whole, they are significantly more highly expressed in the longer-lived species, human and NMR, than in the mouse (Fig. 2A). By filtering these genes for those with more than two-fold difference in expression, we found that 33 of these genes had significantly higher expression in human than mouse, and 34 had significantly higher expression in NMR than mouse (Tables S4 and S5). Table 2 lists the 12 DNA repair genes having higher expression in both human and NMR than mouse. These include essential enzymes for mismatch repair, base excision repair, non-homologous end-joining, as well as the tumor suppressor gene TP53, which plays an essential role in regulating excision repair pathways [15]. This list also includes five of the eleven human DNA glycosylases, which recognize and remove damaged or incorrect bases [16]. In this analysis mouse had higher expression of only six genes involved in DNA repair (Table S3).

We then compared expression of the genes encoding proteins involved in each of the five major DNA repair pathways: base excision repair (BER), mismatch repair (MMR), nucleotide excision repair (NER), homologous recombination repair (HRR), and non-homologous end-joining (NHEJ), as well as DNA damage response signaling (DDRS), which includes genes involved in DNA damage sensing and signaling. Figure 2B shows expression levels of these DNA repair pathways in each species. Our results show that human liver had higher overall expression of genes encoding proteins involved in the BER, HRR, NHEJ, and MMR pathways; however, only the BER pathway expression was significantly higher ($p = 0.0010$). In fact, human liver had significantly higher expression of genes involved in BER than both mouse and NMR. We also found higher expression of HRR and NHEJ genes in NMR than mouse, and genes in the DDRS and MMR pathways show significantly higher expression levels in the NMR than in the mouse ($p=0.0496$ and 0.0011). This is not surprising, given the extreme cancer resistance of the NMR, and the high rates of cancer in mice. None of the DNA repair pathways had significantly higher expression in mouse than either human or NMR.

Pathway activation analysis reveals upregulated DNA repair pathways in longer lived species

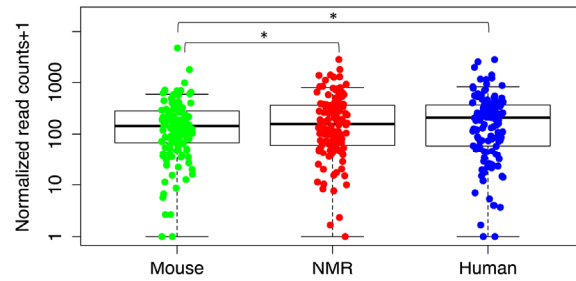
To test whether signaling pathways involved in DNA repair are upregulated in the long-lived species, we performed a signaling pathway analysis (SPA). This analysis enabled us to compare the differential activation of signaling pathways using a recently published algorithm, OncoFinder, for calculating pathway activation strength (PAS) values [17]. This approach is more biologically informative than comparing overall expression of genes or functional enrichment of differentially expressed genes, as it deals with the functional annotation of each gene product in a pathway and considers its role as an activator or repressor of the signal transduction within the pathway.

Table 3. Differentially expressed orthologs.

Comparison	Up in human	Up in mouse	Up in NMR	Differentially expressed
Human vs. Mouse	2056	2181		4237
Human vs. NMR	2294		2426	4720
Mouse vs. NMR		2184	2332	4516

For differential expression orthologs were limited to a read count of 200 or greater and a probability of differential expression of 0.95 or higher.

A



B

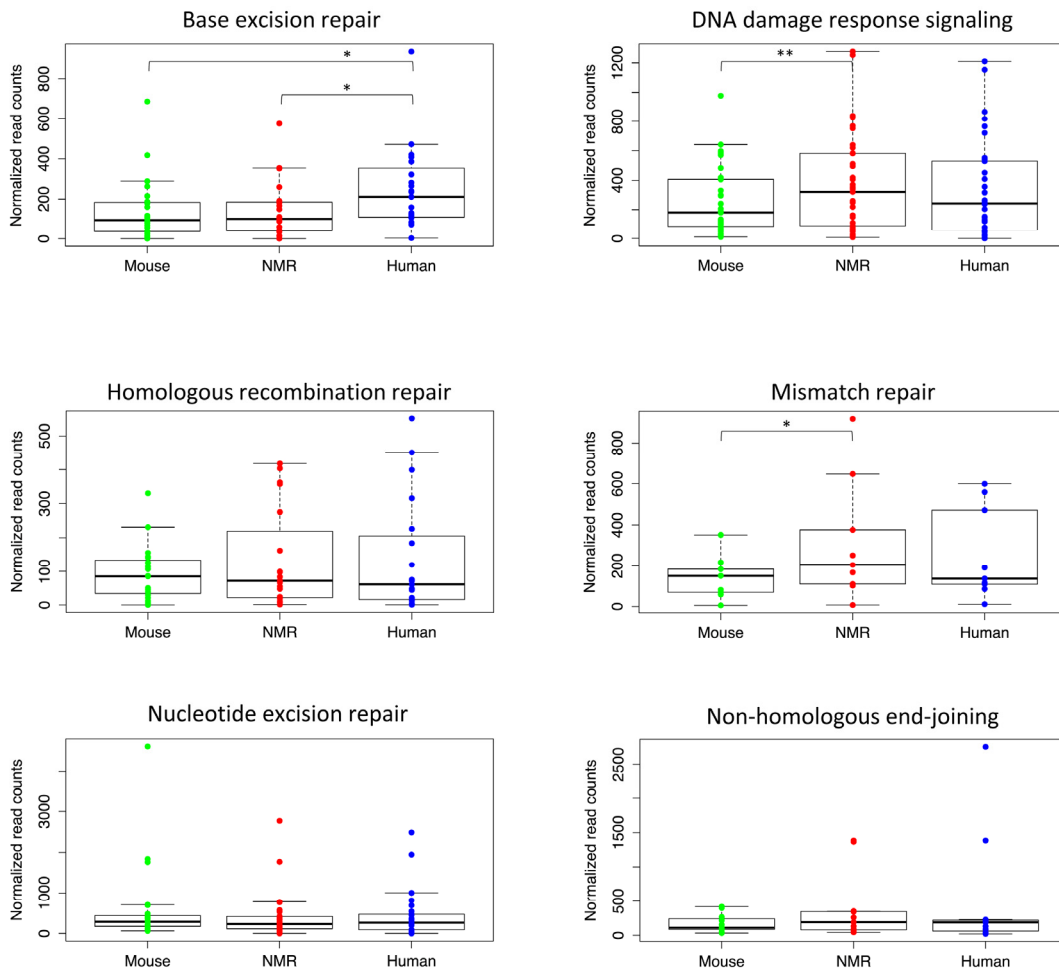


Figure 2. Higher expression of essential DNA repair genes in long-lived species. (A) Expression of DNA repair genes is significantly higher in livers of naked mole rat ($p=0.0167$) and human ($p=0.0429$) than mouse (Wilcoxon matched-pairs signed rank test). (B) For individual DNA repair pathways, human and NMR had higher expression of many essential repair proteins in liver. Human had significantly higher expression than mouse ($p=0.0010$) and NMR ($p=0.0039$) of genes in the BER pathway (Wilcoxon matched-pairs signed rank test), and NMR had significantly higher expression than mouse ($p=0.0496$) of genes involved in MMR (paired t test) and DDRS ($p=0.0011$, Wilcoxon matched-pairs signed rank test).

The absolute value of the PAS indicates the degree of functional changes in the regulation of a signaling pathway, and the sign of PAS indicates whether it is up- or down-regulated, enabling the quantification of signaling changes within a tissue.

Using the OncoFinder algorithm we analyzed 375 signaling pathways, annotated from KEGG and SABiosciences, and obtained PAS profiles for human and NMR using mouse as a reference. Of these pathways, over 200 were significantly upregulated in human and NMR compared with mouse (Tables S6 and S7), and from this list we focused on the 18 pathways directly involved in DNA repair. Figure 3 shows the 18 DNA repair PAS profiles for NMR versus mouse, and human versus mouse. Positive PAS values reflect upregulated signaling pathways, and negative values represent down-regulated pathways. PAS scores of zero indicate similarly acting pathways in mouse compared with human or NMR. Sub-pathways for p53, ATM, and BRAC1 are shown in Fig. S1.

With this SPA approach, we found that all of the DNA repair pathways were upregulated in NMR compared with mouse, except for the BER pathway. Most of the DNA repair pathways were also upregulated in human compared to mouse; however, homologous recombination repair (HRR), non-homologous end-joining (NHEJ), transcription-coupled repair, and the ATM (DNA repair) and p53 pathways were down-regulated in human compared with mouse. The PTEN pathways for DNA repair and genomic stability, and the chromatin main pathway were also more upregulated in human than both NMR and mouse.

Differential expression analysis and functional annotation

Apart from DNA repair genes and pathways, human, mouse, and NMR are likely to differ greatly in a variety of liver-specific functions. Some of these differences may also relate to the extreme lifespan differences between these species, but most are probably a con-

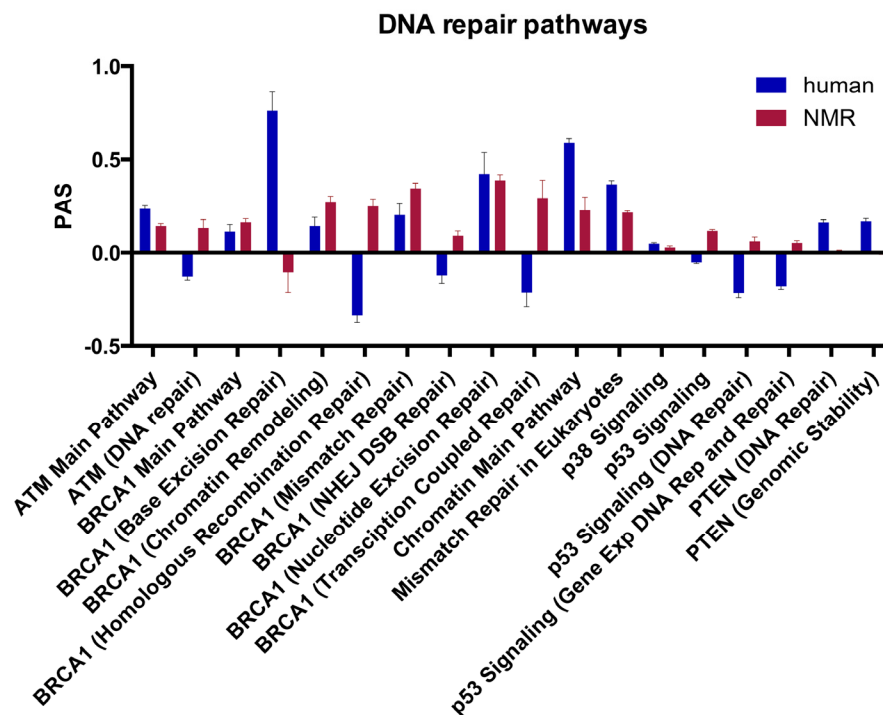
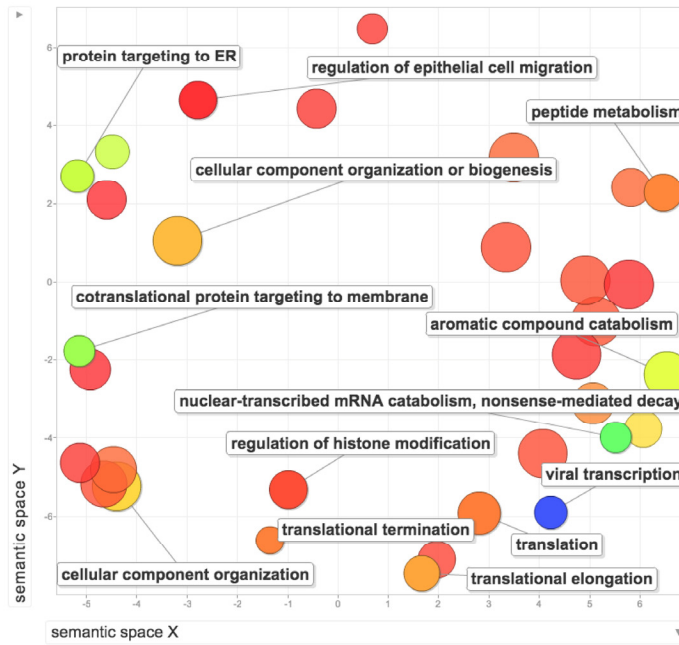


Figure 3. Pathway activation strength (PAS) for DNA repair signaling pathways. PAS values for DNA repair pathways in human vs. mouse and NMR vs. mouse ($p < 0.05$). Positive PAS values reflect upregulated signaling pathways in human or NMR compared with mouse, and negative values represent down-regulated pathways. PAS scores of zero indicate similarly acting pathways in mouse compared with human or NMR.

sequence of differences in lifestyle, habitat, and metabolism. In order to conduct an unbiased differential expression analysis we applied a nonparametric test using the R package NOIseq to compare expression of 13,181 orthologs between the three species [18]. We limited this analysis to genes with normalized

expression to counts of 200 or more and a probability of differential expression of 0.95 or higher, and we found that in each of the three pairwise comparisons (mouse vs. human, mouse vs. NMR, and NMR vs. human), around 4,500 genes were differentially expressed (Table 3).

A



B

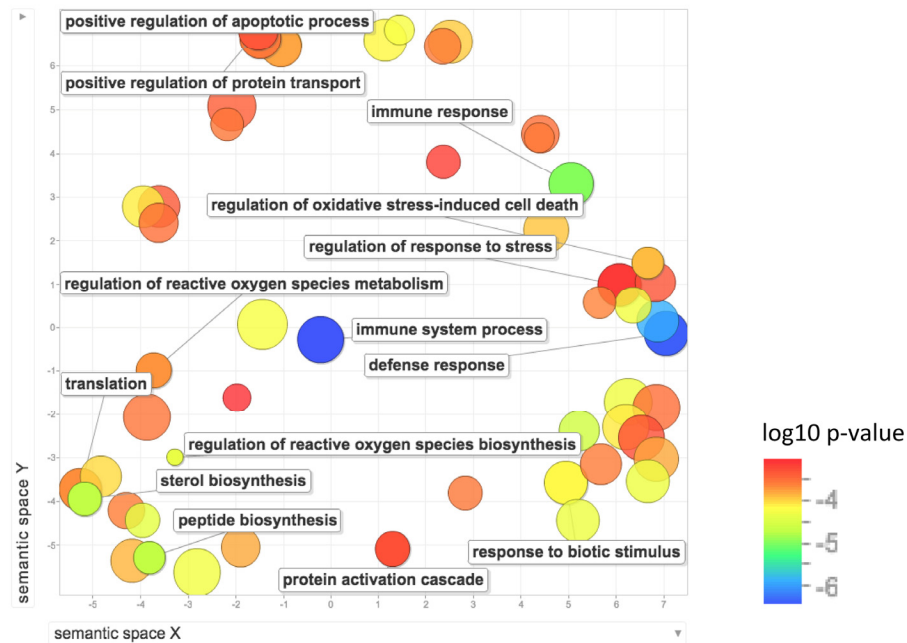
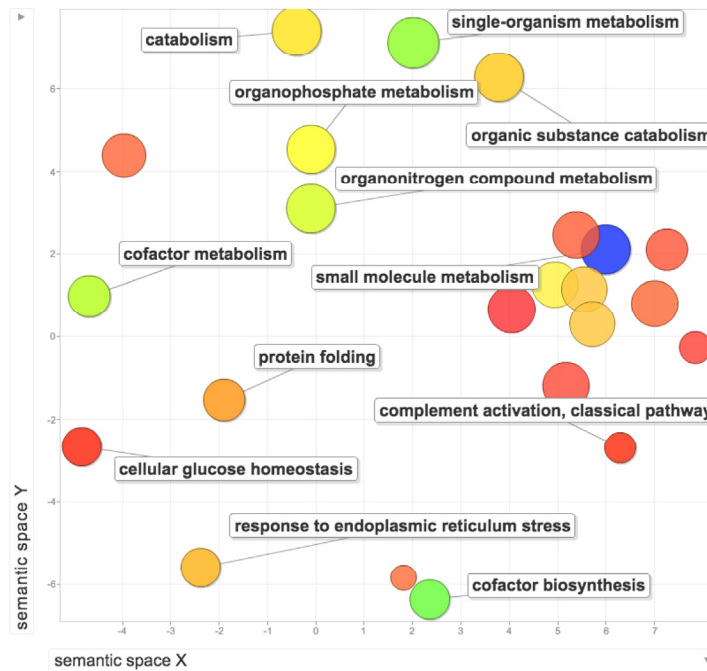


Figure 4. Functional annotation of protein-coding genes more highly expressed in long-lived species. Enrichment analysis visualized as an MDS plot for GO biological processes that are (A) higher in NMR than mouse (B) higher in human than mouse. Plots are generated based on a matrix of semantic similarities in space (x, y). Clusters of circles closer together represent terms that are more closely related. Circle color and size indicates log₁₀ p-value.

A



B

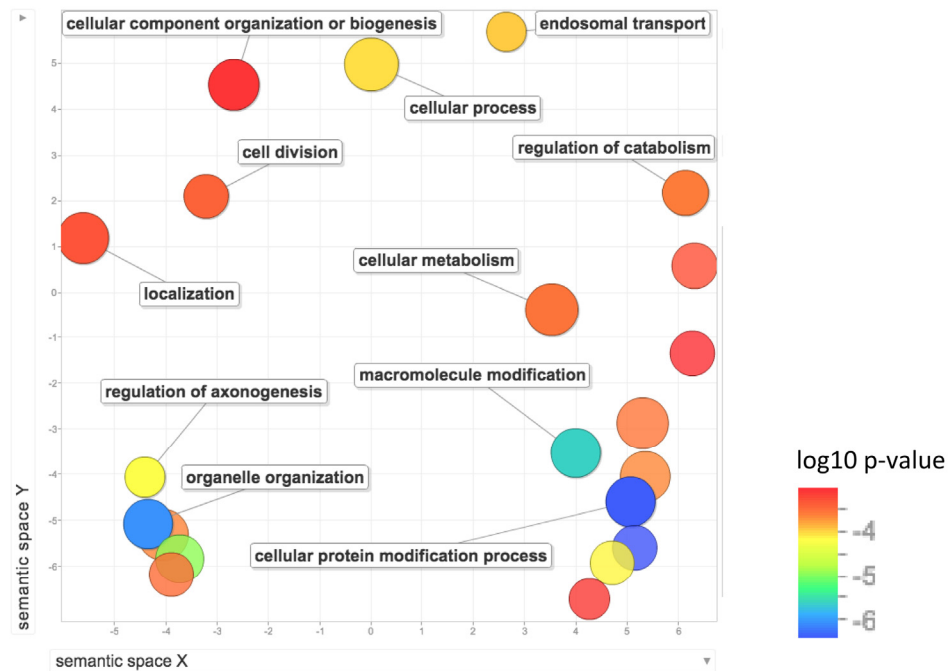


Figure 5. Functional annotation of protein-coding genes more highly expressed in short-lived species. Enrichment analysis visualized as an MDS plot for GO biological processes that are (A) higher in mouse than NMR (B) higher in mouse than human. Plots are generated based on a matrix of semantic similarities in space (x, y). Clusters of circles closer together represent terms that are more closely related. Circle color and size indicates log₁₀ p-value.

To look at biological differences in expression of protein-coding genes among human, mouse, and NMR, with a focus on expression in the long-lived species compared to mouse, we performed functional annotation of these differentially expressed genes using GOrilla [19], comparing human with mouse and NMR with mouse. We found 57 gene ontology (GO) biological process terms were significantly upregulated in NMR compared with mouse, and 92 significantly upregulated in human compared with mouse (Tables S8 and S9). We found 35 GO biological process terms that were significantly higher in mouse than NMR, and 30 that were significantly higher in mouse than human (Tables S10 and S11). As expected from gene expression in liver, the majority of the enriched terms were related to metabolism.

We then used REVIGO, a tool that parses GO annotations using semantic similarity, in order to remove redundant terms and to visualize clusters of related terms [20]. Figure 4 shows the GO biological process terms that were significantly upregulated in the long-lived species. Terms that were enriched for higher expression in NMR than mouse include: protein localization and targeting to the endoplasmic reticulum; mRNA and RNA catabolism; translation, including initiation, elongation, and termination; gene expression; and regulation of histone modification and chromatin organization. Terms that were enriched for higher expression in human than mouse include: regulation of reactive oxygen species metabolism, regulation of oxidative stress induced cell death, positive regulation of cell death and apoptotic processes, response to stress, regulation of response to oxidative stress, response to oxygen-containing compound, translation, protein transport, and several biological processes related to immune system response. These results suggest higher regulation of processes involved in gene expression, including chromatin modification, translation, and protein targeting in liver in the long-lived species, and higher response to oxidative stress and regulation of cell death in human liver.

Figure 5 shows the GO biological process terms that are significantly upregulated in mouse compared with human and NMR. GO terms enriched for higher expression in mouse than NMR include lipid metabolism, response to endoplasmic reticulum stress, protein folding, oxidation-reduction, cellular glucose homeostasis, and nucleoside metabolism. Compared with human, mouse also showed significant enrichment of GO terms including protein modification, including ubiquitination and de-ubiquitination, chromatin modification, and cell division.

DISCUSSION

The results of our analysis of the overall expression of DNA repair genes among the three species show that compared with mouse, human and NMR have higher expression levels of DNA repair genes in liver, including significantly higher expression levels of core DNA repair enzymes that are essential for DNA damage sensing and signaling, and the MMR, NHEJ, and BER pathways (Table 2). This supports the hypothesis that the two long-lived species have superior genome maintenance than mouse. The DNA repair genes more highly expressed in the two long-lived species include genes encoding tumor suppressor *TP53*, DNA mismatch repair protein *MSH3*, and NHEJ repair proteins *NHEJ1* and *Ku70* (*XRCC6*). We found it interesting that *Ku70* was more highly expressed in human and NMR liver, as *Ku70* mRNA levels have been shown to decline with aging in human cells [21, 22]. This list also includes five of the eleven human DNA glycosylases, *MBD4*, *MUTYH*, *NEIL1*, *NEIL2*, and *TDG*, which are essential for removing and replacing damaged DNA bases. Also more highly expressed in human and NMR than mouse were polymerases λ (*POLL*) and κ (*POLK*). *POLL* is involved in NHEJ and BER, and is required for cell cycle progression [23], and *POLK* is required for translesion synthesis [24]. We also looked at genes implicated in both longevity and genome maintenance and found that *SIRT6* and *PARP* are more highly expressed in human liver [25].

Most DNA repair genes are not transcriptionally regulated during the DNA damage response, but are constitutively expressed and regulated by post-transcriptional modification. Therefore, having high constitutive levels of mRNA transcripts for these proteins available in the cell is critical for maintaining genome stability. However, some DNA repair genes are transcriptionally induced upon genotoxic stress, including many of the key components of the NER pathway (*DDB1*, *DDB2*, *ERCC1*, *XPC*, *ERCC4*, and *ERCC5*) [26]. This could perhaps explain why we did not find that genes in the NER pathway were more highly expressed in human and NMR compared with mouse. We were also unable to confidently identify two major components of the NER pathway in the NMR genome, *DDB1* and *ERCC1*, so they are unfortunately not included in this analysis. It is important to note that previous studies have shown that rodent fibroblasts have less efficient NER, and specifically global genomic NER (GG-NER), than human cells. However this may not be the case for liver tissue, where NER may be less essential than in UV-exposed skin cells [4, 5].

In our analysis of overall expression of genes in the individual DNA repair pathways, we found that the human and NMR liver samples had higher expression of most of the DNA repair pathways, and that human liver had significantly higher expression of genes involved in BER than both mouse and NMR, which supports our hypothesis that longer-lived species have better DNA repair.

Our findings were confirmed by a more rigorous statistical analysis in which we used the OncoFinder signaling pathway activation algorithm to test whether signaling pathways involved in DNA repair are upregulated in long-lived species compared to the short lived mouse. The results confirm that the long-lived species, human and NMR, have higher activation of DNA repair signaling pathways in liver. The upregulation of nearly all DNA repair pathways in NMR compared with mouse could explain not only its significantly longer lifespan, but also its superior cancer resistance.

We were surprised to find, however, that in our human samples, compared with mouse, the pathways for double-strand DNA break repair, HRR and NHEJ, were down-regulated. In our functional enrichment analysis we found that humans had higher expression of genes involved in GO biological processes regulating cell death and apoptosis, so it is possible that in human liver, DNA double-strand breaks (DSBs) are more likely to lead to cell death than repair. This may be because hepatocytes are predominately quiescent and it is to their advantage to undergo cell death upon suffering DSBs rather than risking cancerous mutations by undergoing error-prone NHEJ repair, the more common DSB repair pathway in mammals used in non-dividing cells. There is evidence that defects in DSB repair contribute to aging in mice as persistent double-strand breaks and chromosomal rearrangements increase in livers of aged mice [27, 28].

In comparing overall gene expression differences of 13,181 orthologous genes through functional annotation, we found that human and NMR have higher expression in liver of genes regulating gene expression, including chromatin modification, RNA decay, translation, and protein localization. It has been shown that NMR has superior translational fidelity than mouse, so it is not surprising to find enrichment for genes regulating translation in the NMR as well [29]. The human liver samples also had higher expression of genes responding to oxidative stress and regulating cell death. As many of these enriched GO terms involve cellular functions related to control of tissue homeostasis and gene expression, and these functions

have been shown to deteriorate with age in mice, they possibly reflect the longevity of these species compared with mouse [30]. While the unbiased functional annotation of differentially expressed genes did not show enrichment for DNA repair-related GO biological processes, we speculate that this because in most organs, and especially in liver, the majority of genes expressed are related to cellular metabolism. As these genes show the greatest differences in expression between human and mouse [31], it is difficult to detect other, subtle differences in gene expression for functions such as DNA repair, which involves fewer than 200 genes that are not the most highly expressed. Therefore, functional annotation of differentially expressed genes is perhaps not the best approach for distinguishing expression differences in small sets of genes that are not very highly expressed in either sample.

The results of this study support the hypothesis that the longer-lived species, human and NMR, have superior DNA repair than the short-lived mouse, and confirm that DNA repair is indeed a longevity assurance system.

METHODS

Animals and tissue collection. All procedures involving animals were approved by the Institutional Animal Care and Use Committee (IACUC) of Albert Einstein College of Medicine. Three four-month-old Balb/C mice were procured from the National Institutes on Aging. Mice were sacrificed and harvested liver samples were immediately flash-frozen. Three one-year-old naked mole rats were provided by the laboratory of Dr. Vera Gorbunova. They were sacrificed and liver tissues were harvested and immediately flash frozen. Three adult human non-tumor liver samples were obtained from the liver tissue biorepository and Albert Einstein College of Medicine. These samples were preserved in RNA later immediately upon harvesting.

RNA-sequencing and library construction. Directional RNA sequencing libraries were construction as described previously [28]. Libraries were multiplexed at six samples per lane and sequenced on the Illumina HiSeq.

Analysis and statistics. To identify regions of the naked mole genome (hetGla2) [11] homologous to mouse genes, we isolated mouse entries from the UCSC cross species BLAT results (<http://hgdownload.cse.ucsc.edu/goldenPath/hetGla2/database/xenoRefFlat.txt.gz>) [32]. This provided the genomic coordinates for putative NMR homologous genes. We then used UCSC blat

search results as NMR "genes". Preliminarily, 16,708 mouse genes (protein coding) mapped to the NMR genome, and 13,993 of these mouse genes aligned to the NMR genome without overlapping any other alignment. From this list, 13,209 genes exist as homologous gene pairs between mouse and human. We narrowed this down to 13,181 genes due to annotation.

Reads were aligned to the NMR (hetGla2), mouse (mm10), and human (hg19) genomes using STAR [33]. Counts were calculated using HTSeq [34]. Read counts were quantile normalized using EDASeq [12], also accounting for gene length within each species.

NOIseq was used to quantify differential gene expression across species [18]. For comparisons of overall expression of genes in specific pathways, the Wilcoxon matched-pairs signed rank test was used for data with non-normal distribution and paired t tests were used when the data were normally distributed.

Gene ontology (GO) analysis was performed using Gorilla, where four genes were required to be considered in each annotation and a q-value cutoff <0.01 [19]. Gene ontology enrichment was visualized using REVIGO with a similarity cutoff set at 0.7 [20].

Pathway activation strength scores were calculated using the OncoFinder algorithm [17], which uses SABiosciences and KEGG pathway annotations. For this analysis mouse expression data was used as the reference data set.

Figures were generated using R [35] and Prism by GraphPad.

ACKNOWLEDGMENTS

We would like to thank the Marion Bessin Liver Research Center at the Albert Einstein College of Medicine for providing human liver samples.

Author contributions

S.L.M and J.V conceived and designed the experiments. S.L.M. prepared the sequencing libraries, performed functional annotation/GO analysis as well as downstream analysis of differential expression and pathway activation, and wrote the manuscript. M.M.C. performed alignment, data normalization, and differential expression analysis. A.A. performed normalization and signaling pathway analysis. V.G. and A.S. provided NMR liver tissues. Z.D., A.Z., A.S, and

V.G. provided critical input for the analysis and, along with V.N.G, helped edit the manuscript.

Funding

This work was supported by the NIH grants AG017242, AG038072 and AG047200, and the Glenn Foundation for Medical Research.

Data in this paper are from a thesis to be submitted by S.L.M in partial fulfillment of the requirements for the Degree of Ph.D. in the Graduate Division of Medical Sciences, Albert Einstein College of Medicine.

Conflict of interest statement

The authors declare no conflict of interest.

REFERENCES

1. Hoeijmakers JH. DNA damage, aging, and cancer. *The New England journal of medicine*. 2009; 361:1475-1485.
2. Vijg J and Suh Y. Genome instability and aging. *Annual review of physiology*. 2013; 75:645-668.
3. Freitas AA and de Magalhaes JP. A review and appraisal of the DNA damage theory of ageing. *Mutat Res*. 2011; 728:12-22.
4. Hart RW and Setlow RB. Correlation between deoxyribonucleic acid excision-repair and life-span in a number of mammalian species. *Proc Natl Acad Sci U S A*. 1974; 71:2169-2173.
5. Hanawalt PC. Revisiting the rodent repairadox. *Environ Mol Mutagen*. 2001; 38:89-96.
6. Gorbunova V, Seluanov A, Zhang ZD, Gladyshev VN and Vijg J. Comparative genetics of longevity and cancer: insights from long-lived rodents. *Nat Rev Genet*. 2014; 15:531-540.
7. MacRae SL, Zhang Q, Lemetre C, Seim I, Calder RB, Hoeijmakers J, Suh Y, Gladyshev VN, Seluanov A, Gorbunova V, Vijg J and Zhang ZD. Comparative analysis of genome maintenance genes in naked mole rat, mouse, and human. *Aging Cell*. 2015; 14:288-291.
8. Christmann M and Kaina B. Transcriptional regulation of human DNA repair genes following genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation. *Nucleic Acids Res*. 2013; 41:8403-8420.
9. Roschke AV, Glebov OK, Lababidi S, Gehlhaus KS, Weinstein JN and Kirsch IR. Chromosomal instability is associated with higher expression of genes implicated in epithelial-mesenchymal transition, cancer invasiveness, and metastasis and with lower expression of genes involved in cell cycle checkpoints, DNA repair, and chromatin maintenance. *Neoplasia*. 2008; 10:1222-1230.
10. Cichoż-Lach H and Michalak A. Oxidative stress as a crucial factor in liver diseases. *World J Gastroentero*. 2014; 20:8082-8091.
11. Keane M, Craig T, Alfoldi J, Berlin AM, Johnson J, Seluanov A, Gorbunova V, Di Palma F, Lindblad-Toh K, Church GM and de Magalhaes JP. The Naked Mole Rat Genome Resource: facilitating analyses of cancer and longevity-related adaptations. *Bioinformatics*. 2014; 30:3558-3560.

- 12.** Bullard JH, Purdom E, Hansen KD and Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*. 2010; 11:94.
- 13.** Wood RD, Mitchell M and Lindahl T. Human DNA repair genes, 2005. *Mutat Res*. 2005; 577:275-283.
- 14.** Milanowska K, Krwawicz J, Papaj G, Kosinski J, Poleszak K, Lesiak J, Osinska E, Rother K and Bujnicki JM. REPAIRtoire--a database of DNA repair pathways. *Nucleic Acids Res*. 2011; 39:D788-792.
- 15.** Smith ML and Seo YR. p53 regulation of DNA excision repair pathways. *Mutagenesis*. 2002; 17:149-156.
- 16.** Christmann M, Tomicic MT, Roos WP and Kaina B. Mechanisms of human DNA repair: an update. *Toxicology*. 2003; 193:3-34.
- 17.** Buzdin AA, Zhavoronkov AA, Korzinkin MB, Venkova LS, Zenin AA, Smirnov PY and Borisov NM. Oncofinder, a new method for the analysis of intracellular signaling pathway activation using transcriptomic data. *Front Genet*. 2014; 5:55.
- 18.** Tarazona S, Furio-Tari P, Turra D, Pietro AD, Nueda MJ, Ferrer A and Conesa A. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res*. 2015; 21:e140.
- 19.** Eden E, Navon R, Steinfeld I, Lipson D and Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009; 10:48.
- 20.** Supek F, Bosnjak M, Skunca N and Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. 2011; 6:e21800.
- 21.** Prall WC, Czibere A, Jager M, Spentzos D, Libermann TA, Gattermann N, Haas R and Aivado M. Age-related transcription levels of KU70, MGST1 and BIK in CD34+ hematopoietic stem and progenitor cells. *Mech Ageing Dev*. 2007; 128:503-510.
- 22.** Ju YJ, Lee KH, Park JE, Yi YS, Yun MY, Ham YH, Kim TJ, Choi HM, Han GJ, Lee JH, Lee J, Han JS, Lee KM, et al. Decreased expression of DNA repair proteins Ku70 and Mre11 is associated with aging and may contribute to the cellular senescence. *Exp Mol Med*. 2006; 38:686-693.
- 23.** Zucca E, Bertolotti F, Wimmer U, Ferrari E, Mazzini G, Khoronenkova S, Grosse N, van Loon B, Dianov G, Hubscher U and Maga G. Silencing of human DNA polymerase lambda causes replication stress and is synthetically lethal with an impaired S phase checkpoint. *Nucleic Acids Res*. 2013; 41:229-241.
- 24.** Haracska L, Prakash L and Prakash S. Role of human DNA polymerase kappa as an extender in translesion synthesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99:16000-16005.
- 25.** Mao Z, Hine C, Tian X, Van Meter M, Au M, Vaidya A, Seluanov A and Gorbunova V. SIRT6 promotes DNA repair under stress by activating PARP1. *Science*. 2011; 332:1443-1446.
- 26.** Christmann M and Kaina B. Transcriptional regulation of human DNA repair genes following genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation. *Nucleic Acids Research*. 2013; 41:8403-8420.
- 27.** Dollé ME, Giese H, Hopkins CL, Martus HJ, Hausdorff JM and Vijg J. Rapid accumulation of genome rearrangements in liver but not in brain of old mice. *Nature Genetics*. 1997; 17:431-434.
- 28.** White RR, Milholland B, de Bruin A, Curran S, Laberge RM, van Steeg H, Campisi J, Maslov AY and Vijg J. Controlled induction of DNA double-strand breaks in the mouse liver induces features of tissue ageing. *Nat Commun*. 2015; 6:6790.
- 29.** Azpurua J, Ke Z, Chen IX, Zhang Q, Ermolenko DN, Zhang ZD, Gorbunova V and Seluanov A. Naked mole-rat has increased translational fidelity compared with the mouse, as well as a unique 28S ribosomal RNA cleavage. *Proc Natl Acad Sci U S A*. 2013; 110:17350-17355.
- 30.** White RR, Milholland B, MacRae SL, Lin M, Zheng D and Vijg J. Comprehensive transcriptional landscape of aging mouse liver. *BMC Genomics*. 2015; 16:899.
- 31.** Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, Dobin A, Zaleski C, Beer MA, Chapman WC, Gingeras TR, Ecker JR and Snyder MP. Comparison of the transcriptional landscapes between human and mouse tissues. *Proc Natl Acad Sci U S A*. 2014; 111:17224-17229.
- 32.** Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM and Haussler D. The human genome browser at UCSC. *Genome Res*. 2002; 12:996-1006.
- 33.** Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29:15-21.
- 34.** Anders S, Pyl PT and Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015; 31:166-169.
- 35.** R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.