Article

Somatic mutation rates scale with lifespan across mammals

https://doi.org/10.1038/s41586-022-04618-z

Received: 17 August 2021

Accepted: 7 March 2022

Published online: 13 April 2022

Open access

Check for updates

Alex Cagan[1,15]✉, Adrian Baez-Ortega[1,15], Natalia Brzozowska[1], Federico Abascal[1], Tim H. H. Coorens[1], Mathijs A. Sanders[1,2], Andrew R. J. Lawson[1], Luke M. R. Harvey[1], Shriram Bhosle[1], David Jones[1], Raul E. Alcantara[1], Timothy M. Butler[1], Yvette Hooks[1], Kirsty Roberts[1], Elizabeth Anderson[1], Sharna Lunn[1], Edmund Flach[3], Simon Spiro[3], Inez Januszczak[3,4], Ethan Wrigglesworth[3], Hannah Jenkins[3], Tilly Dallas[3], Nic Masters[3], Matthew W. Perkins[5], Robert Deaville[5], Megan Druce[6,7], Ruzhica Bogeska[6,7], Michael D. Milsom[6,7], Björn Neumann[8,9], Frank Gorman[10], Fernando Constantino-Casas[10], Laura Peachey[10,11], Diana Bochynska[10,12], Ewan St. John Smith[13], Moritz Gerstung[14], Peter J. Campbell[1], Elizabeth P. Murchison[10], Michael R. Stratton[1] & Iñigo Martincorena[1]✉

The rates and patterns of somatic mutation in normal tissues are largely unknown outside of humans[1–7]. Comparative analyses can shed light on the diversity of mutagenesis across species, and on long-standing hypotheses about the evolution of somatic mutation rates and their role in cancer and ageing. Here we performed whole-genome sequencing of 208 intestinal crypts from 56 individuals to study the landscape of somatic mutation across 16 mammalian species. We found that somatic mutagenesis was dominated by seemingly endogenous mutational processes in all species, including 5-methylcytosine deamination and oxidative damage. With some differences, mutational signatures in other species resembled those described in humans[8], although the relative contribution of each signature varied across species. Notably, the somatic mutation rate per year varied greatly across species and exhibited a strong inverse relationship with species lifespan, with no other life-history trait studied showing a comparable association. Despite widely different life histories among the species we examined—including variation of around 30-fold in lifespan and around 40,000-fold in body mass—the somatic mutation burden at the end of lifespan varied only by a factor of around 3. These data unveil common mutational processes across mammals, and suggest that somatic mutation rates are evolutionarily constrained and may be a contributing factor in ageing.

Somatic mutations accumulate in healthy cells throughout life. They underpin the development of cancer[9] and, for decades, have been speculated to contribute to ageing[10–12]. Directly studying somatic mutations in normal tissues has been challenging owing to the difficulty of detecting mutations present in single cells or small clones in a tissue. Only recent technological developments, such as in vitro expansion of single cells into colonies[13,14], microdissection of histological units[8,15], single-cell sequencing[16,17] or single-molecule sequencing[18], are beginning to enable the study of somatic mutation in normal tissues.

Over the last few years, studies in humans have started to provide a detailed understanding of somatic mutation rates and the contribution of endogenous and exogenous mutational processes across normal tissues[8,13,14,19,20]. These studies are also revealing how, as we age, some human tissues are colonized by mutant cells that contain cancer-driving mutations, and how this clonal composition changes with age and disease. With the exception of some initial studies, far less is known about somatic mutation in other species[1–7]. Yet, comparative analyses of somatic mutagenesis would shed light on the diversity of mutagenic processes across species, and on long-standing questions regarding the evolution of somatic mutation rates and their role in cancer and ageing.

A decades-long hypothesis on the evolution of somatic mutation rates pertains to the relationship between body mass and cancer risk. Some models predict that the risk of cancer should increase proportionally to the number of cells at risk of transformation. However, there appears to be no correlation between body mass and cancer risk across

[1]Cancer, Ageing and Somatic Mutation (CASM), Wellcome Sanger Institute, Hinxton, UK. [2]Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands. [3]Wildlife Health Services, Zoological Society of London, London, UK. [4]The Natural History Museum, London, UK. [5]Institute of Zoology, Zoological Society of London, London, UK. [6]Division of Experimental Hematology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [7]Heidelberg Institute for Stem Cell Technology and Experimental Medicine GmbH (HI-STEM), Heidelberg, Germany. [8]Wellcome Trust–Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. [9]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. [10]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK. [11]Bristol Veterinary School, Faculty of Health Sciences, University of Bristol, Langford, UK. [12]Department of Pathology, Faculty of Veterinary Medicine, Universitatea de Stiinte Agricole si Medicina Veterinara, Cluj-Napoca, Romania. [13]Department of Pharmacology, University of Cambridge, Cambridge, UK. [14]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK. [15]These authors contributed equally: Alex Cagan, Adrian Baez-Ortega. ✉e-mail: ac36@sanger.ac.uk; im3@sanger.ac.uk

species[21,22]. This observation, known as Peto's paradox, suggests that the evolution of larger body sizes is likely to require the evolution of stronger cancer suppression mechanisms[23,24]. Whether evolutionary reduction of cancer risk across species is partly achieved by a reduction of somatic mutation rates remains unknown.

A second long-standing hypothesis on the evolution of somatic mutation rates relates to the proposed role of somatic mutations in ageing. Multiple forms of molecular damage, including somatic mutations, telomere attrition, epigenetic drift and loss of proteostasis, have been proposed to contribute to ageing, but their causal roles and relative contributions remain debated[25,26]. Evolutionary theory predicts that species will evolve protection or repair mechanisms against life-threatening damage to minimize death from intrinsic causes, but that selection is too weak to delay ageing far beyond the typical life expectancy of an organism in the wild (Supplementary Note 1). If somatic mutations contribute to ageing, theory predicts that somatic mutation rates may inversely correlate with lifespan across species[27,28]. This prediction has remained largely untested owing to the difficulty of measuring somatic mutation rates across species.

Detection of somatic mutations across species

The study of somatic mutations with standard whole-genome sequencing requires isolating clonal groups of cells recently derived from a single cell[8,13,14]. To study somatic mutations across a diverse set of mammals, we isolated 208 individual intestinal crypts from 56 individuals across 16 species with a wide range of lifespans and body sizes: black-and-white colobus monkey, cat, cow, dog, ferret, giraffe, harbour porpoise, horse, human, lion, mouse, naked mole-rat, rabbit, rat, ring-tailed lemur and tiger (Supplementary Table 1). We chose intestinal crypts for several reasons. First, they are histologically identifiable units that line the epithelium of the colon and small intestine and are amenable to laser microdissection. Second, human studies have confirmed that individual crypts become clonally derived from a single stem cell and show a linear accumulation of mutations with age, which enables the estimation of somatic mutation rates through genome sequencing of single crypts[8]. Third, in most human crypts, most somatic mutations are caused by endogenous mutational processes common to other tissues, rather than by environmental mutagens[8,18].

A colon sample was collected from each individual, with the exception of a ferret from which only a small intestine sample was available. This sample was included because results in humans have shown that the mutation rates of colorectal and small intestine epithelial stem cells are similar[14,20] (Extended Data Fig. 1). We then used laser microdissection on histological sections to isolate individual crypts for whole-genome sequencing with a low-input library preparation method[29] (Fig. 1a, Extended Data Fig. 2, Supplementary Table 2), with the exception of human crypts, for which sequencing data were obtained from a previous study[8]. A bioinformatic pipeline was developed to call somatic mutations robustly in all these species despite the variable quality of their genome assemblies (Methods). The distribution of variant allele fractions of the mutations detected in each crypt confirmed that crypts are clonal units in all species, enabling the study of somatic mutation rates and signatures (Extended Data Fig. 3).

We found substantial variation in the number of somatic single-base substitutions across species and across individuals within each species (Fig. 1b). For five species with samples from multiple individuals (dog, human, mouse, naked mole-rat and rat), linear regression confirmed a clear accumulation of somatic mutations with age (Fig. 1c, Extended Data Fig. 4, Supplementary Table 3). All linear regressions were also consistent with a non-significant intercept. This resembles observations in humans[20] and suggests that the time required for a single stem cell to drift to fixation within a crypt is a small fraction of the lifespan of a species. This facilitates the estimation of somatic mutation rates across species by dividing the number of mutations in a crypt by the age of the individual (Supplementary Table 4). The number of somatic insertions and deletions (indels) was consistently lower than that of substitutions in all crypts (Fig. 1b), in agreement with previous findings in humans[8].

Mutational signatures across mammals

Somatic mutations can be caused by multiple mutational processes, involving different forms of DNA damage and repair. Different processes cause characteristic frequencies of base substitution types and indels at different sequence contexts, often referred to as mutational signatures, which can be inferred from mutation data[30]. Across species, the mutational spectra showed clear similarities, with a dominance of cytosine-to-thymine (C>T) substitutions at CpG sites, as observed in human colon, but with considerable variation in the frequency of other substitution types (Fig. 2a). To quantify the contribution of different mutational processes to the observed spectra, we applied mutational signature decomposition[8,30]. We used a Bayesian model to infer mutational signatures de novo, while accounting for differences in genome sequence composition across species, and using the COSMIC human signature SBS1 (C>T substitutions at CpG sites) as a fixed prior to ensure its complete deconvolution[31] (Methods). This approach identified two signatures beyond SBS1, labelled SBSB and SBSC, which resemble COSMIC human signatures SBS5 and SBS18, respectively (cosine similarities 0.93 and 0.91) (Fig. 2b).

This analysis suggests that the same three signatures that dominate somatic mutagenesis in the human colon are dominant in other mammals: SBS1, which is believed to result from the spontaneous deamination of 5-methylcytosine[8,32]; SBSB (SBS5), a common signature across human tissues that may result from endogenous damage and repair[18,33]; and SBSC (SBS18), which is dominated by C>A substitutions and attributed to oxidative DNA damage[30]. Signature SBSC contains a minor component of T>A substitutions (resembling COSMIC SBS34), which appear to be the result of DNA polymerase slippage at the boundaries between adjacent adenine and thymine homopolymer tracts, but could also reflect assembly errors at those sites[33]. Although all of the species that we examined shared the three mutational signatures, their contributions varied substantially across species (Fig. 2c). SBSC was particularly prominent in mouse and ferret, and the ratio of SBS1 to SBSB/5 varied from approximately 1.2 in rat or rabbit to 6.4 in tiger. In several species with data from multiple individuals, separate linear regressions for each signature confirmed that mutations from all three signatures accumulate with age (Fig. 2d, Extended Data Fig. 5).

Although signature deconvolution identified three signatures that are active across species, we noticed some differences in the mutational profile of signature SBSB among species. To investigate this further, we inferred independent versions of SBSB from each species, while accounting for differences in genome sequence composition (Methods). This revealed inter-species variability in the mutational profile of this signature, particularly in the C>T component (Extended Data Fig. 6). Species-specific versions of SBSB showed different similarities to the related human signatures SBS5 and SBS40. For example, SBSB inferred from the human data showed a stronger similarity with the reference human signature SBS5 (cosine similarities with SBS5 and SBS40: 0.93 and 0.84), whereas SBSB from rabbit more closely resembled the reference human signature SBS40 (0.87 and 0.91). These observations are consistent with the hypothesis that SBS5 and SBS40 result from a combination of correlated mutational processes, with some variation across human tissues[18,33] and across species.

Analysis of the indel mutational spectra revealed a dominance of the human indel signatures ID1 and ID2, which are characterized by single-nucleotide indels at A/T homopolymers, and probably caused by strand slippage during DNA replication[30] (Extended Data Fig. 7a). The ratio of insertions (ID1) to deletions (ID2) appears to vary across species, possibly reflecting a differential propensity for slippage of the template and nascent DNA strands[30]. In addition, the indel spectra suggest a potential contribution of signature ID9 (the aetiology of which

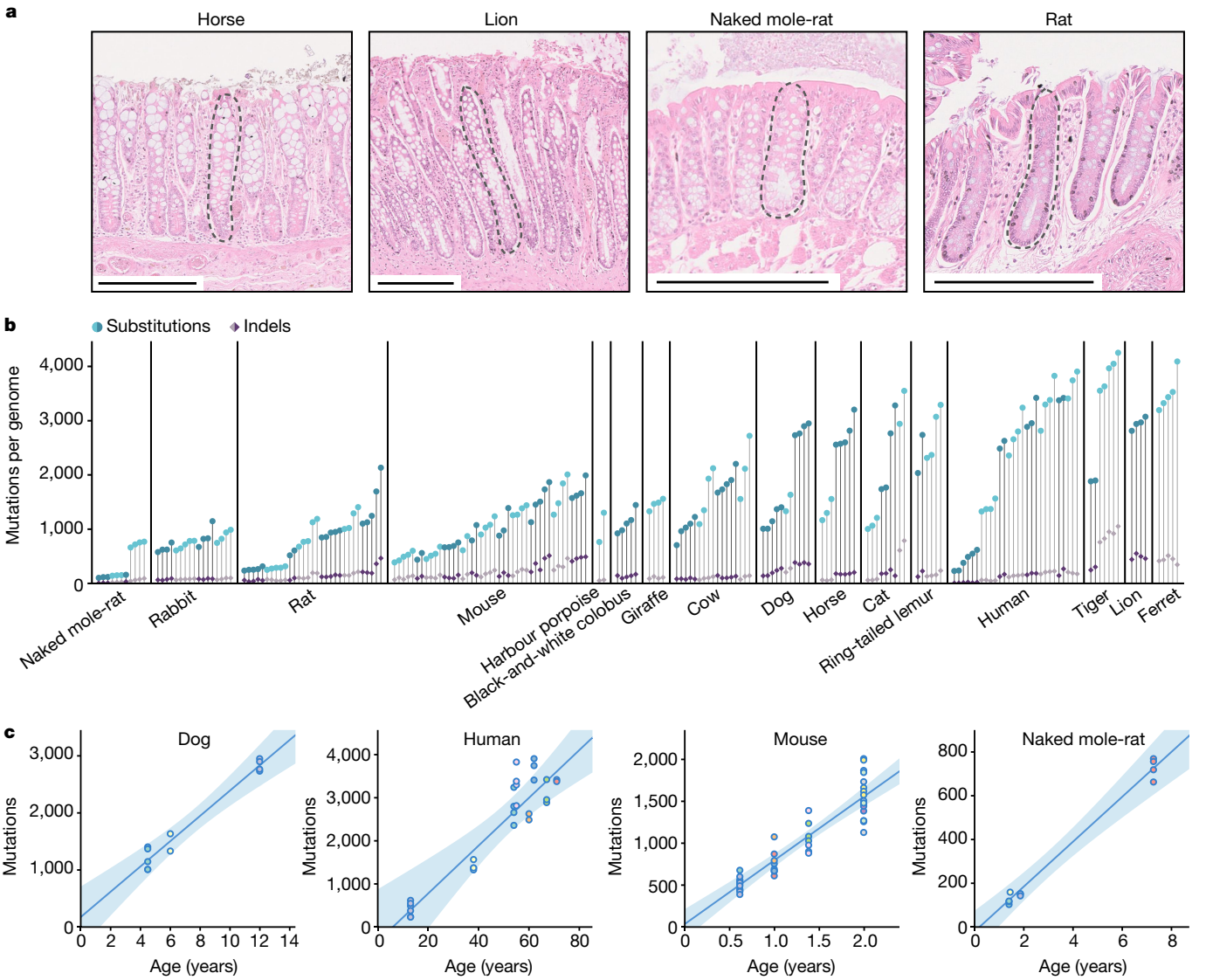


Fig. 1 | Somatic mutation burden in mammalian colorectal crypts. a, Histology images of colon samples from horse, lion, naked mole-rat and rat, with one colorectal crypt marked in each. Scale bars, 250 µm. **b**, Burden of somatic substitutions and indels per diploid genome in each colorectal crypt sample (corrected for the size of the analysable genome). Samples are grouped by individual, with samples from the same individual coloured in the same shade. Species, and individuals within each species, are sorted by mean mutation burden. **c**, Linear regression of somatic substitution burden (corrected for analysable genome size) on individual age for dog, human, mouse and naked mole-rat samples. Samples from the same individual are shown in the same colour. Regression was performed using mean mutation burdens per individual. Shaded areas indicate 95% confidence intervals of the regression line.

remains unknown) to human, colobus, cow, giraffe and rabbit. Analysis of indels longer than one base pair also suggested the presence of a signature of four-base-pair insertions at tetrameric repeats, which was particularly prevalent in mouse and tiger; a pattern of insertions of five or more base pairs at repeats in colobus; and a pattern of deletions of five or more base pairs at repeats, which was prominent in rabbit and resembles ID8 (a signature possibly caused by double-strand break repair through non-homologous end joining[30]) (Extended Data Fig. 7a).

Other mutational processes and selection

The apparent lack of additional mutational signatures is noteworthy. A previous study of 445 colorectal crypts from 42 human donors found that many crypts were affected by a signature that was later attributed to colibactin, a genotoxin produced by *pks*$^{+}$ strains of *Escherichia coli*[8,34,35]. Analysing the original human data and our non-human data with the same methodology, we found evidence of colibactin mutagenesis in 21% of human crypts, but only uncertain evidence of colibactin in one non-human crypt (0.6%) (Extended Data Fig. 7b, Methods). This revealed a significant depletion of colibactin mutagenesis in the non-human crypts studied (Fisher's exact test, $P = 7 \times 10^{-14}$). The apparent difference in colibactin mutagenesis observed between species, or between the cohorts studied, might result from a different prevalence of *pks*$^{+}$ *E. coli* strains[36] or a different expression of colibactin by *pks*$^{+}$ *E. coli* across species[37]. Finally, we also searched for evidence of APOBEC signatures (SBS2 and SBS13), which have been reported in a small number of human crypts and are believed to be caused by APOBEC DNA-editing cytidine deaminases. We detected APOBEC signatures in 2% ($n = 9$) of human crypts and found only uncertain evidence in one non-human crypt ($P = 0.30$).

Beyond substitutions and indels, crypts from the eight species with chromosome-level genome assemblies were inspected for large-scale copy number changes (at least 1 Mb) (Methods). Studies in humans have found that large-scale copy number changes are relatively rare in

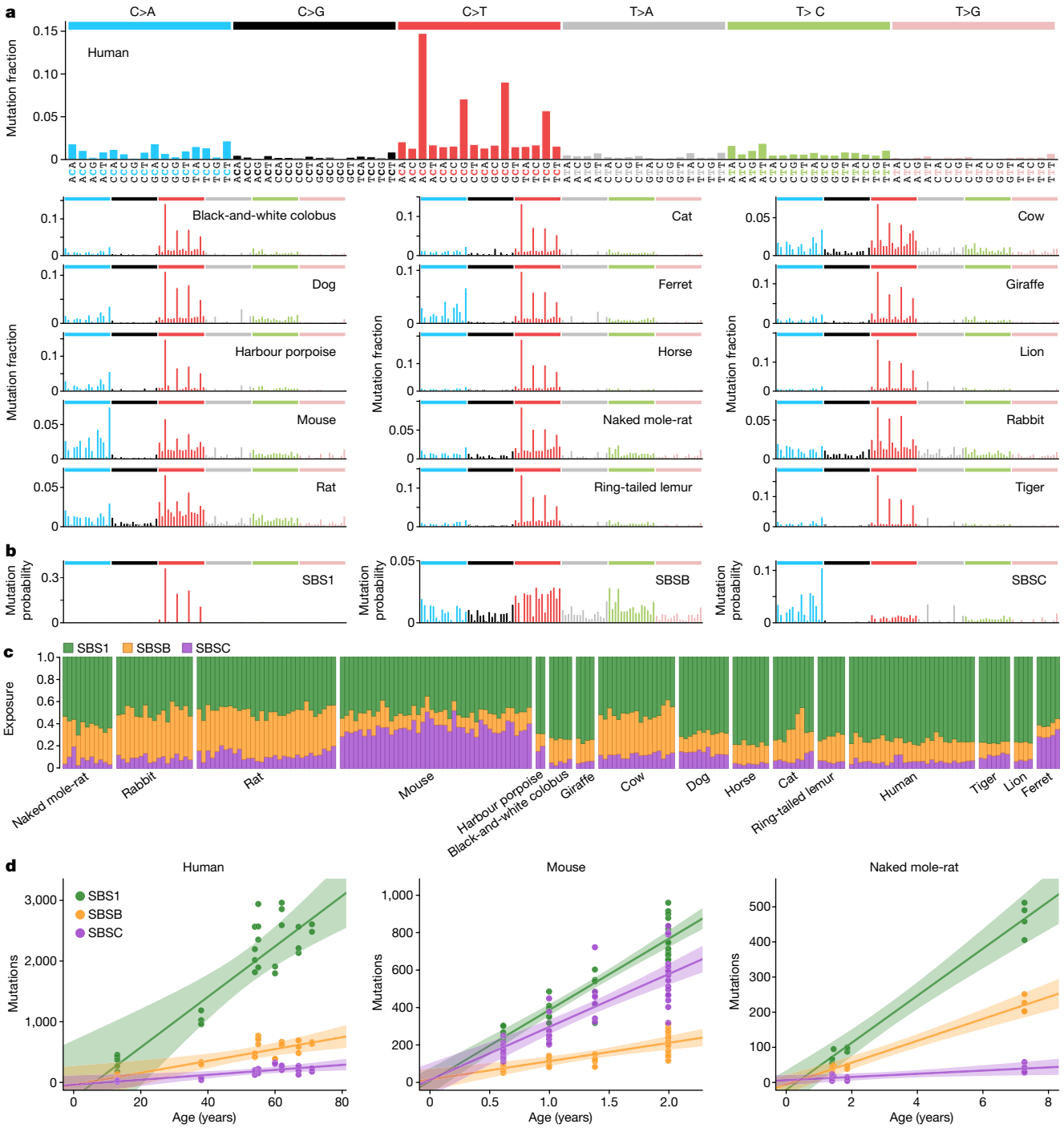


Fig. 2 | Mutational processes in the mammalian colon. a, Mutational spectra of somatic substitutions in each species. The *x* axis shows 96 mutation types on a trinucleotide context, coloured by base substitution type. **b**, Mutational signatures inferred from (SBSB, SBSC) or fitted to (SBS1) the species mutational spectra shown in **a**, and normalized to the human genome trinucleotide frequencies. The *y* axis shows mutation probability. **c**, Estimated contribution of each signature to each sample. Samples are arranged horizontally as in Fig. 1b. **d**, Linear regression of signature-specific mutation burdens (corrected for analysable genome size) on individual age for human, mouse and naked mole-rat samples. Regression was performed using mean mutation burdens per individual. Shaded areas indicate 95% confidence intervals of the regression line.

normal tissues, including colorectal epithelium[8]. Consistent with these results, we only identified 4 large copy number changes across the 162 crypts included in this analysis: 2 megabase-scale deletions in 2 crypts from the same cow; the loss of an X chromosome in a female mouse crypt; and a 52-Mb segment with copy-neutral loss of heterozygosity in a human crypt (Extended Data Fig. 8, Methods). These results suggest that large-scale somatic copy number changes in normal tissues are also rare in other mammalian species.

Previous analyses in humans have shown that most somatic mutations in colorectal crypts accumulate neutrally, without clear evidence of negative selection against non-synonymous mutations and with a low frequency of positively selected cancer-driver mutations[8]. To study somatic selection in our data, we calculated the exome-wide ratio of non-synonymous to synonymous substitution rates (dN/dS) in each of the 12 species with available genome annotation. To do so and to detect genes under positive selection, while accounting for the effects of trinucleotide

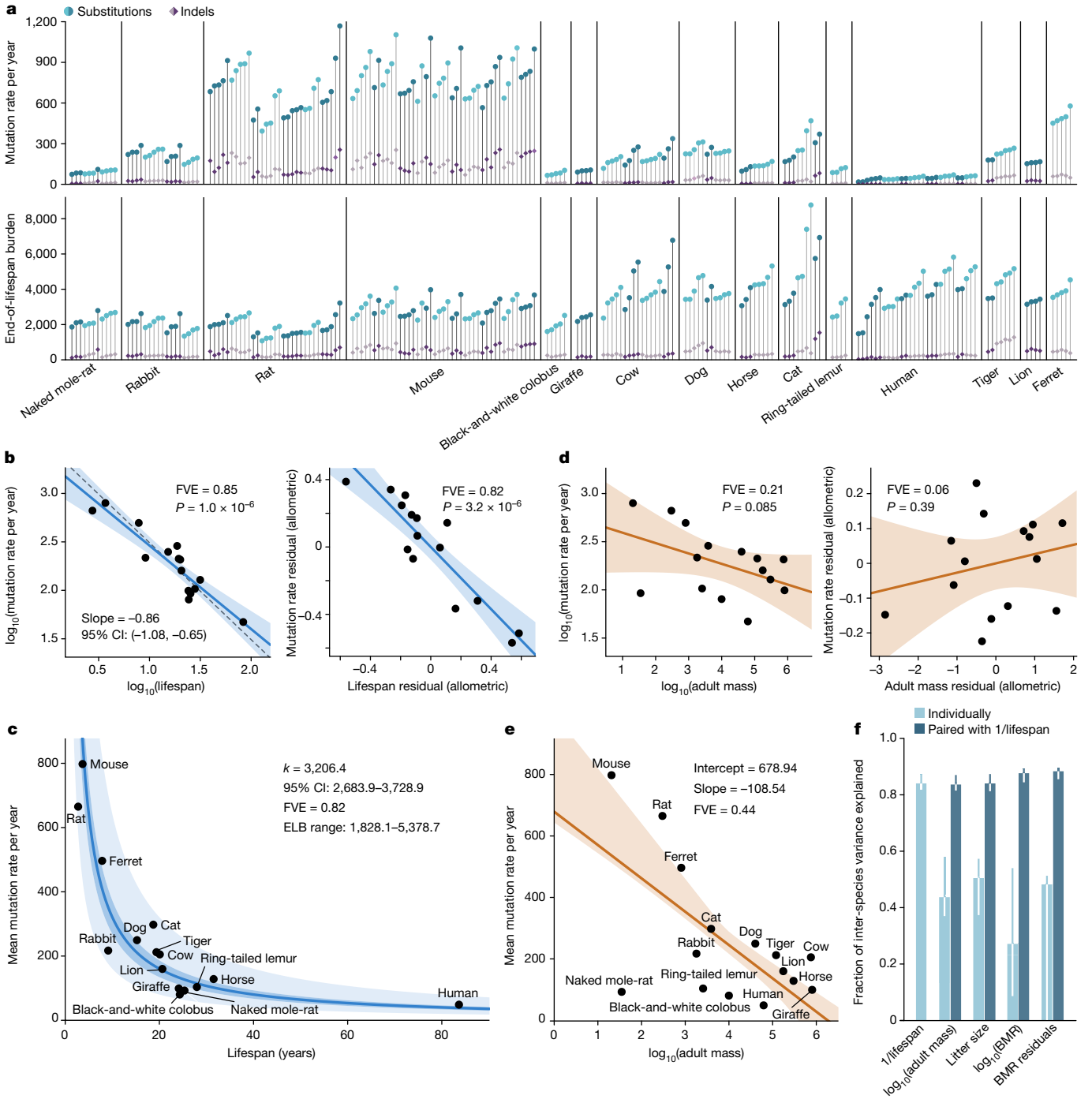


Fig. 3 | Associations between somatic mutation rates and life-history traits. **a**, Somatic mutation rate per year and expected end-of-lifespan mutation burden (ELB) per crypt. Samples are arranged horizontally as in Fig. 1b; harbour porpoise samples were excluded owing to the age of the sampled individual being unknown. **b**, Left, allometric regression of somatic mutation rate on lifespan. Right, regression of body-mass-adjusted residuals for somatic mutation rate and lifespan (partial correlation; Methods). Regression was performed using mean mutation rates per species. Shaded areas represent 95% confidence intervals (CI) of regression lines. FVE and *P* values (by *F*-test) are indicated (note that, for simple linear regression, FVE = R^2). The dashed line denotes a reference slope of −1. **c**, Zero-intercept LME regression of somatic mutation rate on inverse lifespan (1/lifespan), presented on the scale of untransformed lifespan (*x* axis). For simplicity, the *y* axis shows mean mutation rates per species, although rates per crypt were used in the regression. The darker shaded area indicates 95% CI of the regression line, and the lighter shaded area marks a twofold deviation from the line. Point estimate and 95% CI of the regression slope (*k*), FVE and range of end-of-lifespan burden are indicated. **d**, Allometric regression and linear regression of lifespan-adjusted residuals, for somatic mutation rate and body mass (elements as described in **b**). **e**, Free-intercept LME regression of somatic mutation rate on log-transformed body mass. The *y* axis shows mean mutation rates per species, although rates per crypt were used in the regression. Shaded area indicates 95% bootstrap interval of the regression line ($n = 10{,}000$ replicates). Point estimates of the regression intercept and slope, and FVE, are indicated. **f**, FVE values for free-intercept LME models using 1/lifespan or other life-history variables (alone or combined with 1/lifespan) as explanatory variables. Error bars indicate 95% bootstrap intervals ($n = 10{,}000$).

Table 1 | Variation in adult body mass, lifespan, somatic mutation rate and end-of-lifespan mutation burden across the 16 mammalian species surveyed

Variable	Minimum	Maximum	Fold variation
Adult mass (g)	20.50	800,000.00	39,024.39
Lifespan (years)	2.75	83.67	30.44
Mutation rate per year (substitutions per genome)	47.12	796.42	16.90
End-of-lifespan burden (substitutions per genome)	1,828.08	5,378.73	2.94

Species-level estimates are provided in Supplementary Tables 3 and 6.

sequence context and mutation rate variation across genes, we used the dNdScv model[38] (Methods). Although the limited number of coding somatic mutations observed in most species precluded an in-depth analysis of selection, exome-wide dN/dS ratios for somatic substitutions were not significantly different from unity in any species, in line with previous findings in humans[8] (Extended Data Fig. 9). Gene-level analysis did not find genes under significant positive selection in any species, although larger studies are likely to identify rare cancer-driver mutations[8].

Correlation with life-history traits

Whereas similar mutational processes operate across the species surveyed, the mutation rate per genome per year varied widely. Across the 15 species with age information, we found that substitution rates per genome ranged from 47 substitutions per year in humans to 796 substitutions per year in mice, and indel rates from 2.5 to 158 indels per year, respectively (Fig. 3a, Supplementary Table 4, Methods).

To investigate the relationship between somatic mutation rates, lifespan and other life-history traits, we first estimated the lifespan of each species using survival curves. We used a large collection of mortality data from animals in zoos to minimize the effect of extrinsic mortality (Extended Data Fig. 10). We defined lifespan as the age at which 80% of individuals reaching adulthood have died, to reduce the effects of outliers and variable cohort sizes that affect maximum lifespan estimates[39] (Methods). Notably, we found a tight anticorrelation between somatic mutation rates per year and lifespan across species (Fig. 3b). A log-log allometric regression yielded a strong linear anticorrelation between mutation rate per year and lifespan (fraction of inter-species variance explained (FVE) = 0.85, $P = 1 \times 10^{-6}$), with a slope close to and not significantly different from −1. This supports a simple model in which somatic mutation rates per year are inversely proportional to the lifespan of a species (rate ∝ 1/lifespan), such that the number of somatic mutations per cell at the end of the lifespan (the end-of-lifespan burden; ELB) is similar in all species.

To further study the relationship between somatic mutation rates and life-history variables, we used linear mixed-effects (LME) regression models. These models account for the hierarchical structure of the data (with multiple crypts per individual and multiple individuals per species), as well as the heteroscedasticity of somatic mutation rate estimates across species (Methods). Using these models, we estimated that the inverse of lifespan explained 82% of the inter-species variance in somatic substitution rates (rate = k/lifespan) ($P = 2.9 \times 10^{-9}$; Fig. 3c), with the slope of this regression (k) representing the mean estimated ELB across species (3,206.4 substitutions per genome per crypt, 95% confidence interval 2,683.9–3,728.9). Of note, despite uncertainty in the estimates of both somatic mutation rates and lifespans, and despite the diverse life histories of the species surveyed—including around 30-fold variation in lifespan and around 40,000-fold variation in body mass—the estimated mutation load per cell at the end of lifespan varied by only around threefold across species (Table 1). Analogous results were obtained when repeating the analysis with estimates of the protein-coding mutation rate, which may be a better proxy for the functional effect of somatic mutations (85% of variance explained; ELB: 31 coding substitutions per crypt) (Extended Data Fig. 11, Methods).

We next examined the association between somatic mutation rates and adult body mass, which is known to be a common confounder in correlations that involve lifespan[40,41]. An anticorrelation between somatic mutation rates and body mass may be expected if the modulation of cancer risk across species of vastly different sizes has been a major factor in the evolution of somatic mutation rates. We observed that log-transformed adult body mass was less strongly associated with somatic substitution rates than the inverse of lifespan (allometric regression FVE = 0.21, Fig. 3d; LME regression FVE = 0.44, Fig. 3e). Given that lifespan is correlated with body mass, we performed two tests to assess whether body mass explained any variation in somatic mutation rates that was not explained by lifespan. First, including both the inverse of lifespan and log-transformed adult body mass in the regression model suggested that body mass does not explain a significant amount of variance in somatic mutation rates across species after accounting for the effect of lifespan (likelihood ratio tests: $P = 0.16$ for body mass on a model with lifespan; $P < 10^{-4}$ for lifespan on a model with body mass; Fig. 3f, Methods). Second, partial correlation analyses using allometric regressions further confirmed that the association between somatic mutation rates and lifespan is unlikely to be mediated by the effect of body mass on both variables (lifespan residuals: $P = 3.2 \times 10^{-6}$, FVE = 0.82, Fig. 3b; body mass residuals: $P = 0.39$, FVE = 0.06, Fig. 3d; Methods).

The fact that the variation in somatic mutation rates across species appears to be dominated by lifespan rather than body size is also apparent when looking at particularly informative species. Giraffe and naked mole-rat, for instance, have similar somatic mutation rates (99 and 93 substitutions per year, respectively), in line with their similar lifespans (80th percentiles: 24 and 25 years, respectively), despite a difference of around 23,000-fold in adult body mass (Fig. 3c, e). Similarly, cows, giraffes and horses weigh much more than an average human, and yet have somatic mutation rates that are several fold higher, in line with expectation from their lifespan but not their body mass. Altogether, the weak correlation between body mass and somatic mutation rates after correction for lifespan suggests that the evolution of larger body sizes may have relied on alternative or additional strategies to limit cancer risk, as has been speculated[24,42] (Supplementary Note 2). Of note, the low somatic mutation rate of naked mole-rats, which is unusual for their body mass but in line with their long lifespan (Fig. 3c, e), might contribute to the exceptionally low incidence rates of cancer in this species[43].

We found similar results for other life-history variables that have been proposed to correlate with lifespan, namely basal metabolic rate (BMR) and litter size[44] (Fig. 3f). With the caveat that estimates for these variables vary in quality, they showed weaker correlations with the somatic mutation rate as single predictors, and small non-significant increases in explanatory power when considered together with lifespan (likelihood ratio tests: $P = 0.92$ for litter size; P = 0.083 for log-BMR; $P = 0.79$ for allometric BMR residuals; Fig. 3f, Methods). We note that the results above are robust to the use of alternative measures of the somatic mutation rate, including the rate per exome or mutations per Mb (Extended Data Fig. 11, Methods); alternative estimates of lifespan, including maximum lifespan (Extended Data Fig. 12, Methods); alternative regression models, including a Bayesian hierarchical model and a phylogenetic generalised least-squares regression, which accounts for the effect of phylogenetic relationships (Extended Data Fig. 13a, b, Methods); and bootstrapping analyses at the level of individuals or species (Extended Data Fig. 13c, Methods).

Mutational processes and lifespan

To investigate whether a single biological process could drive the association between somatic mutation rates and lifespan, we analysed each mutational signature separately. SBS1, SBSB/5 and SBSC/18 are believed to result from different forms of DNA damage and are expected to be

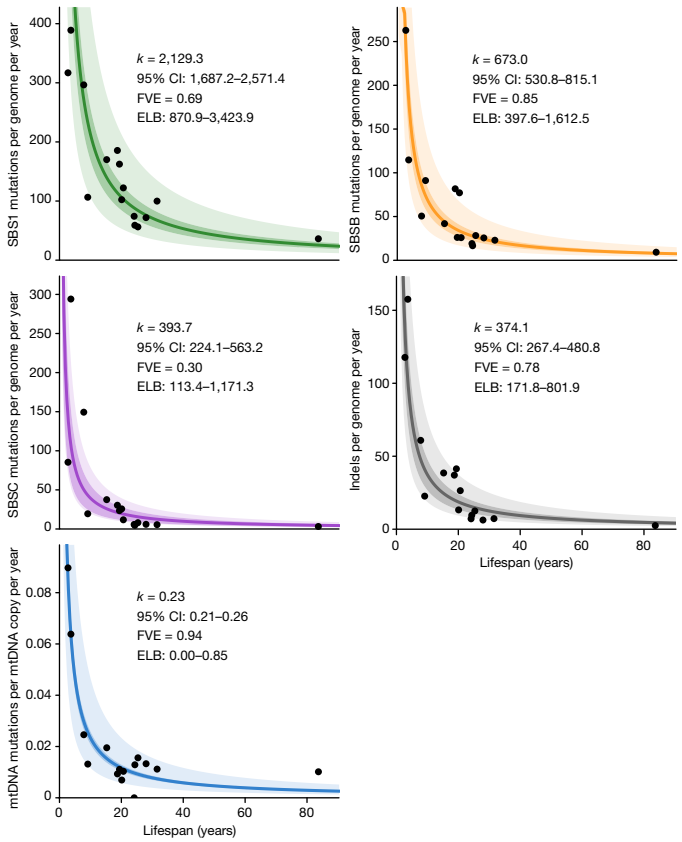


Fig. 4 | Association between mutation rate subtypes and species lifespan. Zero-intercept LME regression of somatic rates of signature-specific substitutions, indels and mtDNA mutations on inverse lifespan (1/lifespan), presented on the scale of untransformed lifespan (*x* axis). For simplicity, *y* axes present mean mutation rates per species, although mutation rates per crypt were used in the regressions. The darker shaded areas indicate 95% confidence intervals (CI) of the regression lines, and the lighter shaded areas mark a twofold deviation from the regression lines. Point estimates and 95% CI of the regression slope (*k*), fraction of inter-species variance explained by for each model (FVE) and ranges of end-of-lifespan burden (ELB) are indicated.

subject to different DNA repair pathways[18,33]. They also appear to differ in their association with the rate of cell division in humans, with SBS1 being more common in fast-proliferating tissues, such as colon and embryonic or foetal tissues, and SBS5 dominating in post-mitotic cells in the absence of cell division[14,18,20]. Overall, we found clear anticorrelations between mutation rates per year and lifespan for the three substitution signatures and for indels, suggesting that a single biological process or DNA repair pathway is unlikely to be responsible for this association (Fig. 4). The total mutation burden also appears to show a closer fit with lifespan than individual mutational processes, as measured by the range of end-of-lifespan burden for each process across species (Fig. 4). This might be expected if the observed anticorrelation were the result of evolutionary pressure on somatic mutation rates.

DNA damage and somatic mutations in the mitochondrial genome have also attracted considerable interest in the ageing field[45]. Our whole-genome sequencing of individual crypts provided high coverage of the mitochondrial genome, ranging from 2,188- to 29,691-fold. Normalized against the nuclear coverage, these data suggest that colorectal crypts contain on the order of around 100–2,000 mitochondrial genomes per cell (Extended Data Fig. 14a). Using a mutation-calling algorithm that is sensitive to low-frequency variants, we found a total of 261 mitochondrial mutations across 199 crypts (Extended Data Fig. 14a, Methods). The mutational spectra across species appeared broadly consistent with that observed in humans, with a dominance of C>T and A>G substitutions that are believed to result from mitochondrial DNA (mtDNA) replication errors rather than DNA damage[46] (Extended Data Fig. 14b). Although the low number of mitochondrial mutations detected per species precludes a detailed analysis, the estimated number of somatic mutations per copy of mtDNA also appears to show an anticorrelation with lifespan. Across species, we obtained an average of 0.23 detectable mutations per copy of the mitochondrial genome by the end of lifespan (Fig. 4, Methods)—a considerable burden given the coding-sequence density and the functional relevance of the mitochondrial genome.

Discussion

Using whole-genome sequencing of 208 colorectal crypts from 56 individuals, we provide insights into the somatic mutational landscape of 16 mammalian species. Despite their different diets and life histories, we found considerable similarities in their mutational spectra. Three main mutational signatures explain the spectra across species, albeit with varying contributions and subtle variations in the profile of signature SBSB. These results suggest that, at least in the colorectal epithelium, a conserved set of mutational processes dominate somatic mutagenesis across mammals.

The most notable finding of this study is the inverse scaling of somatic mutation rates with lifespan—a long-standing prediction of the somatic mutation theory of ageing[11,27]. Considering evolutionary and mechanistic models of ageing together provides a framework for discussing the possible implications of these results for ageing (see Supplementary Note 1). Jointly, these models predict ageing to be multifactorial, with multiple forms of molecular and cellular damage contributing to organismal ageing owing to evolutionary limits to selection acting on the rates of these processes. The inverse scaling of somatic mutation rates and lifespan is consistent with somatic mutations contributing to ageing and with somatic mutation rates being evolutionarily constrained, although we discuss alternative explanations below. This interpretation is also supported by studies reporting more efficient DNA repair in longer-lived species[47,48]. Somatic mutations could contribute to ageing in different ways. Traditionally, they have been proposed to contribute to ageing through deleterious effects on cellular fitness[11,49], but recent findings question this assumption (Supplementary Note 1). Instead, the discovery of widespread clonal expansions in ageing human tissues[19,50–52] raises the possibility that some somatic mutations contribute to ageing by driving clonal expansions of functionally altered cells at a cost to the organism[49,53,54]. Examples include the possible links between clonal haematopoiesis and cardiovascular disease[54]; between mutations in liver disease and insulin resistance[55]; and between driver mutations in cavernomas and brain haemorrhages[49,53,56]. Detailed studies on the extent and effect of somatic mutations and clonal expansions on age-related diseases and ageing phenotypes may help to clarify the precise role—if any—of somatic mutations in ageing. Even if clear causal links between somatic mutations and ageing are established, ageing is likely to be multifactorial. Other forms of molecular damage involved in ageing could be expected to show similar anticorrelations with lifespan and, indeed, such anticorrelations have been reported for telomere shortening and protein turnover[57,58].

Alternative non-causal explanations for the observed anticorrelation between somatic mutation rates and lifespan need to be considered. One alternative explanation is that cell division rates could scale with lifespan and explain the observed somatic mutation rates. Available estimates of cell division rates, although imperfect and limited to a few species, do not readily support this argument (Methods). More importantly, studies in humans have shown that cell division rates are not a major determinant of somatic mutation rates across human tissues[14,18]. Another alternative explanation for the observed anticorrelation might be that selection acts to reduce germline mutation rates in species with longer reproductive spans, which in turn causes an anticorrelation of somatic mutation rates and lifespan. Although selective pressure on

germline mutation rates could influence somatic mutation rates, it is unlikely that germline mutation rates tightly determine somatic mutation rates: somatic mutation rates in humans are 10–20 times higher than germline mutation rates, show variability across cell types and are influenced by additional mutational processes[18,20]. Overall, the strong scaling of somatic mutation rates with lifespan across mammals, despite the different rates between germline and soma and the variable contributions of different mutational processes across species, suggests that somatic mutation rates themselves have been evolutionarily constrained, possibly through selection on multiple DNA repair pathways. Alternative explanations need to be able to explain the strength of the scaling despite these differences.

Altogether, this study provides a detailed description of somatic mutation across mammals, identifying common and variable features and shedding light on long-standing hypotheses. Scaled across the tree of life and across tissues, in species with markedly different physiologies, life histories, genome compositions and mutagenic exposures, similar studies promise to transform our understanding of somatic mutation and its effects on evolution, ageing and disease.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04618-z.

1. Garcia, A. M. et al. Age- and temperature-dependent somatic mutation accumulation in *Drosophila melanogaster*. *PLoS Genet.* **6**, e1000950 (2010).
2. Behjati, S. et al. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422–425 (2014).
3. Milholland, B. et al. Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* **8**, 15183 (2017).
4. Schmid-Siegert, E. et al. Low number of fixed somatic mutations in a long-lived oak tree. *Nat. Plants* **3**, 926–929 (2017).
5. Jager, M. et al. Deficiency of nucleotide excision repair is associated with mutational signature observed in cancer. *Genome Res.* **29**, 1067–1077 (2019).
6. Orr, A. J. et al. A phylogenomic approach reveals a low somatic mutation rate in a long-lived plant. *Proc. R. Soc. B* **287**, 20192364 (2020).
7. López, E. H. & Palumbi, S. R. Somatic mutations and genome stability maintenance in clonal coral colonies. *Mol. Biol. Evol.* **37**, 828–838 (2020).
8. Lee-Six, H. et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532–537 (2019).
9. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
10. Szilard, L. On the nature of the aging process. *Proc. Natl Acad. Sci. USA* **45**, 30–45 (1959).
11. Morley, A. A. The somatic mutation theory of ageing. *Mutat. Res.* **338**, 19–23 (1995).
12. Vijg, J. & Dong, X. Pathogenic mechanisms of somatic mutation and genome mosaicism in aging. *Cell* **182**, 12–23 (2020).
13. Welch, J. S. et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
14. Blokzijl, F. et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).
15. Moore, L. et al. The mutational landscape of normal human endometrial epithelium. *Nature* **580**, 640–646 (2020).
16. Lodato, M. A. et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).
17. Zhang, L. et al. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc. Natl Acad. Sci. USA* **116**, 9014–9019 (2019).
18. Abascal, F. et al. Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405–410 (2021).
19. Martincorena, I. et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
20. Moore, L. et al. The mutational landscape of human somatic and germline cells. *Nature* **597**, 381–386 (2021).
21. Peto, R., Roe, F. J., Lee, P. N., Levy, L. & Clack, J. Cancer and ageing in mice and men. *Br. J. Cancer* **32**, 411–426 (1975).
22. Vincze, O. et al. Cancer risk across mammals. *Nature* **601**, 263–267 (2022).
23. Peto, R. Epidemiology, multistage models, and short-term mutagenicity tests. *Int. J. Epidemiol.* **45**, 621–637 (2016).
24. Tollis, M., Boddy, A. M. & Maley, C. C. Peto's paradox: how has evolution solved the problem of cancer prevention? *BMC Biol.* **15**, 60 (2017).
25. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
26. Schumacher, B., Pothof, J., Vijg, J. & Hoeijmakers, J. H. J. The central role of DNA damage in the ageing process. *Nature* **592**, 695–703 (2021).
27. Burnet, M. *Intrinsic Mutagenesis: a Genetic Approach to Ageing* (Springer, 1974).
28. Kirkwood, T. B. & Holliday, R. The evolution of ageing and longevity. *Proc. R. Soc. B* **205**, 531–546 (1979).
29. Ellis, P. et al. Reliable detection of somatic mutations in solid tissues by laser-capture microdissection and low-input DNA sequencing. *Nat. Protoc.* **16**, 841–871 (2021).
30. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
31. Gori, K. & Baez-Ortega, A. sigfit: flexible Bayesian inference of mutational signatures. Preprint at *bioRxiv* https://doi.org/10.1101/372896 (2020).
32. Lindahl, T. & Nyberg, B. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* **13**, 3405–3410 (1974).
33. Zou, X. et al. A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat. Cancer* **2**, 643–657 (2021).
34. Wilson, M. R. et al. The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**, eaar7785 (2019).
35. Pleguezuelos-Manzano, C. et al. Mutational signature in colorectal cancer caused by genotoxic *pks*$^+$ *E. coli*. *Nature* **580**, 269–273 (2020).
36. Smati, M. et al. Quantitative analysis of commensal *Escherichia coli* populations reveals host-specific enterotypes at the intra-species level. *MicrobiologyOpen* **4**, 604–615 (2015).
37. Oliero, M. et al. Oligosaccharides increase the genotoxic effect of colibactin produced by *pks*$^+$ *Escherichia coli* strains. *BMC Cancer* **21**, 172 (2021).
38. Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041 (2017).
39. Moorad, J. A., Promislow, D. E. L., Flesness, N. & Miller, R. A. A comparative assessment of univariate longevity measures using zoological animal records. *Aging Cell* **11**, 940–948 (2012).
40. Speakman, J. R. Correlations between physiology and lifespan—two widely ignored problems with comparative studies. *Aging Cell* **4**, 167–175 (2005).
41. de Magalhães, J. P., Costa, J. & Church, G. M. An analysis of the relationship between metabolism, developmental schedules, and longevity using phylogenetic independent contrasts. *J. Gerontol. A* **62**, 149–160 (2007).
42. Vazquez, J. M. & Lynch, V. J. Pervasive duplication of tumor suppressors in Afrotherians during the evolution of large bodies and reduced cancer risk. *eLife* **10**, e65041 (2021).
43. Smith, E. S. J., Schuhmacher, L.-N. & Husson, Z. The naked mole-rat as an animal model in biomedical research: current perspectives. *Open Access Anim. Physiol.* **7**, 137–148 (2015).
44. Millar, J. S. & Zammuto, R. M. Life histories of mammals: an analysis of life tables. *Ecology* **64**, 631–635 (1983).
45. Kauppila, T. E. S., Kauppila, J. H. K. & Larsson, N.-G. Mammalian mitochondria and aging: an update. *Cell Metab.* **25**, 57–71 (2017).
46. Ju, Y. S. et al. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *eLife* **3**, e02935 (2014).
47. Hall, K. Y., Hart, R. W., Benirschke, A. K. & Walford, R. L. Correlation between ultraviolet-induced DNA repair in primate lymphocytes and fibroblasts and species maximum achievable life span. *Mech. Ageing Dev.* **24**, 163–173 (1984).
48. Zhang, L. et al. Maintenance of genome sequence integrity in long- and short-lived rodent species. *Sci. Adv.* **7**, eabj3284 (2021).
49. Smith, J. M. Review lectures on senescence—I. The causes of ageing. *Proc. R. Soc. B* **157**, 115–127 (1962).
50. Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
51. Martincorena, I. et al. Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018).
52. Yokoyama, A. et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312–317 (2019).
53. Ren, A. A. et al. *PIK3CA* and CCM mutations fuel cavernomas through a cancer-like mechanism. *Nature* **594**, 271–276 (2021).
54. Jaiswal, S. & Libby, P. Clonal haematopoiesis: connecting ageing and inflammation in cardiovascular disease. *Nat. Rev. Cardiol.* **17**, 137–144 (2020).
55. Ng, S. W. K. et al. Convergent somatic mutations in metabolism genes in chronic liver disease. *Nature* **598**, 473–478 (2021).
56. Jaiswal, S. & Ebert, B. L. Clonal hematopoiesis in human aging and disease. *Science* **366**, eaan4673 (2019).
57. Swovick, K. et al. Interspecies differences in proteome turnover kinetics are correlated with life spans and energetic demands. *Mol. Cell. Proteomics* **20**, 100041 (2021).
58. Whittemore, K., Vera, E., Martínez-Nevado, E., Sanpera, C. & Blasco, M. A. Telomere shortening rate predicts species life span. *Proc. Natl Acad. Sci. USA* **116**, 15122–15127 (2019).

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022