

EccDNA atlas in male mice reveals features protecting genes against transcription-induced eccDNA formation

Received: 21 January 2024

Accepted: 29 January 2025

Published online: 22 February 2025

 Check for updates

Xue Liang^{1,2,7}, Gerard Arrey^{1,7}, Yating Qin^{1,2}, Lucía Álvarez-González^{3,4}, Judith Mary Hariprakash¹, Jie Ma², Sylvester Holt¹, Peng Han^{1,2}, Yonglun Luo⁵, Hanbo Li², Aurora Ruiz-Herrera^{3,4}, Henriette Pilegaard⁶ & Birgitte Regenberg¹ ✉

eccDNA is a driver of many cancers and a potential intermediate in other age-related disorders. However, little is known about the mechanisms underlying eccDNA formation in healthy tissue and how aging affects these processes. Here, we present an atlas of eccDNA across seven tissues of male mice spanning four ages. EccDNA correlates with open chromatin characterized by signatures of H3K27ac and H3K4me1. Additionally, the mutational load of eccDNA on genes correlates with tissue-specific transcription and increases logarithmically as a function of transcript level. Still, a population of intron-dense genes with many splice forms remains sheltered from eccDNA formation. We also find that the total number of eccDNA molecules does not increase as mice age, unlike other types of mutations. Our data reveal a link between eccDNA formation and transcript level that may drive gene architecture in mammals.

EccDNA (extrachromosomal circular DNA) of chromosomal origin is a latent source of unstable genetic variation in eukaryotic cells. eccDNA has been found to originate from any region of the eukaryotic genomes, including intergenic regions, fragments of genes, and intact genes^{1–4}. The latter is especially important in the evolution of cancer because oncogenes can rapidly increase their copy number and expression levels if they are found within eccDNA (often referred to as ecDNA)^{5–9}. However, little is known about the features that determine the formation of eccDNA in healthy animal tissues.

eccDNA can be a result of DNA damage¹⁰. Genomic instability is regarded as a hallmark of aging¹¹, whereby DNA mutations such as single nucleotide variants (SNVs) and indels accumulate with age¹² and even lead to mammalian tissue-specific signatures¹³. It is still unknown whether the genomic burden of eccDNA increases in mammals during

aging. Furthermore, studies in humans, mice, and birds have revealed a high frequency of eccDNA from genic regions^{2,14,15}, suggesting an effect of transcription on eccDNA formation, which is in line with a similar relationship previously shown for selected genes in yeast¹⁶. Therefore, this study aims to investigate whether eccDNA is associated with transcription and how aging impacts the level and origin of eccDNA in healthy mammalian tissues. To achieve this, we generated a comprehensive atlas of eccDNA across different tissues and ages in the house mouse (*Mus musculus*).

In this work, we assemble the atlas of eccDNA in mice and demonstrate that the number of eccDNA increases logarithmically with transcript levels, providing strong evidence that eccDNA formation is linked to transcriptional activity. Our study further reveals that some tissue-specific genes, which have numerous splice forms and

¹Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. ²Lars Bolund Institute of Regenerative Medicine, Qingdao-Europe Advanced Institute for Life Sciences, BGI Research, Qingdao 266555, China. ³Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain. ⁴Genome Integrity and Instability Group, Institut de Biociències i Biomèdica, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain. ⁵Department of Biomedicine, Aarhus University, Aarhus, Denmark. ⁶Section for Cell Biology and Physiology, Department of Biology, University of Copenhagen, Copenhagen, Denmark. ⁷These authors contributed equally: Xue Liang, Gerard Arrey. ✉ e-mail: bregenberg@bio.ku.dk

high intron density, are partially protected from transcription-induced circularization. Additionally, we uncover that eccDNA does not accumulate in tissues as animals age. In summary, we provide insights into the mechanisms driving gene evolution and suggest important avenues for research in eccDNA.

Results

An atlas of eccDNA in healthy mice

We adapted the Circle-Seq protocol²¹⁷ to purify eccDNA from a total of 205 samples from wild-type C57BL/6NRj male mice. These samples included tissues isolated from the brain (cortex, $n = 26$; and hippocampus, $n = 28$), inguinal subcutaneous adipose tissue (SAT, $n = 27$), also referred here as Sfat, epididymal visceral adipose tissue (VAT, $n = 27$), also referred here as Vfat, liver ($n = 27$), pancreas ($n = 24$) and skeletal muscle ($n = 25$). Samples were collected from animals of three different ages: 3-month-old (young), 12-month-old (adult), and 22-month-old (old) mice (Fig. 1a and Supplementary Data 1). Additionally, we also purified eccDNA from mouse embryos stage E17.5 (neuronal tissue: $n = 11$, non-neuronal tissue: $n = 10$) from three different litters that showed no differences in biometric measurements (Supplementary Fig. 1a–d).

The 22-month-old group showed a distinctive aging phenotype compared to the 3-month-old and 12-month-old groups. The body weight of 22-month-old mice (44.4 ± 7.0 g) was significantly higher than 12-month-old mice (28.6 ± 1.8 g) and 3-month-old mice (25.9 ± 1.5 g) (Supplementary Fig. 1e). The aged animals also showed a decrease in white blood cells in circulation ($3.35 \times 10^6 \pm 1.1 \times 10^6$ cells/mL) compared to 12-month-old mice ($5.46 \times 10^6 \pm 0.92$ cells/mL) and 3-month-old mice ($5.1 \times 10^6 \pm 1.09$ cells/mL) (Supplementary Fig. 1f).

We included at least seven biological replicates from each of the 7 tissues and 3 ages. We extracted approximately 10^6 cells from each sample (Supplementary Fig. 1g). Nuclear eccDNA was enriched from the samples by linearization of mitochondrial DNA in vitro using a CRISPR-Cas9 system with two sgRNAs specific for mitochondrial DNA, and all linear chromosomal and mitochondrial DNA were subsequently removed using exonuclease V¹⁷. The remaining eccDNA was amplified by rolling circle amplification with the $\phi 29$ polymerase and sequenced with 150 bp-paired-end sequencing with at least 70 million reads per sample. The reads were subsequently aligned to the mouse reference genome (GRCm38/mm10) using the BWA-MEM algorithm¹⁸. eccDNA was detected by the Circle-Map pipeline¹⁹ and supported by at least two soft-clipped reads mapping around the circle junction (or one soft-clipped read and at least one discordant read) and at least 95% mean base coverage within the eccDNA detection coordinates. Circle-Map was chosen for the identification of eccDNA because it is the most commonly used eccDNA pipeline, and it out-performs other pipelines¹⁹. We tested whether the majority of the eccDNA in the samples had been detected by randomly sampling and mapping fractions of sequence reads (20%, 40%, 60%, 80%, 99%) from the liver, cortex, and hippocampus of three different ages. Saturation plots with the eccDNA count at each sequencing depth showed that the current sequencing depth was sufficient to record most of the eccDNA in the samples (Supplementary Fig. 1h–j).

Chromosomes produce eccDNA numbers in proportion to their size

We detected 567,963 high-confidence eccDNAs distributed across different tissues and ages (Supplementary Data 2). Of these, 16.8% were below 2000 bp in size, which has previously been designated microDNA¹⁴. The average size of eccDNA detected in the study was 4786.5 bp (Fig. 1b), and the size range of eccDNA varied from <100 bp to 258,322 bp. This is similar to the main peak of eccDNA found in human somatic tissue and substantially higher than previously reported sizes of eccDNA in mice, which were found to be 60–2000 bp¹⁴.

The eccDNA size distribution showed a pronounced size periodicity at around 378 bp. We also observed a smaller but visible periodicity of roughly half (~189 bp), which has previously been associated with nucleosome fragments during apoptosis²⁰ (Fig. 1c). All tissues had this size distribution of eccDNA independent of age except for eccDNA from the seven VAT samples of 22-month-old mice that harbored significantly larger eccDNA than other tissues (Supplementary Fig. 2a–c). The eccDNA size distribution was likely a reflection of the underlying natural distribution. Still, a fraction of the eccDNAs was large enough to carry whole genes (Supplementary Data 3). Additionally, we explored the presence of repetitive elements in eccDNA, as eccDNA may be generated as an intermediate of retrotransposition. We observed various classes of repetitive elements overlapping with eccDNA in all the tissue types (Supplementary Fig. 3).

We compared the number of uniquely mapped eccDNA to assess whether steady-state levels increased as mice aged. Although we did not observe any significant differences in the number of eccDNA between ages, there was a non-significant trend suggesting that brain tissues contained more eccDNA in all ages, particularly in embryo, where this difference approached statistical significance (Fig. 1d, Kruskal–Wallis test, $p = 0.057$). We did not find any differences in other age stages (3, 12, and 22-month-old) (Fig. 1d, Kruskal–Wallis test, $p = 0.98, 0.32$, and 0.99 , respectively). Also, the number of eccDNA was similar between non-embryonic tissues and ages (Fig. 1e, Kruskal–Wallis test between tissues and age groups, $p = 0.09$ and $p = 0.33$, respectively). Taken together, we did not detect any evidence that the number of unique eccDNA accumulates with aging.

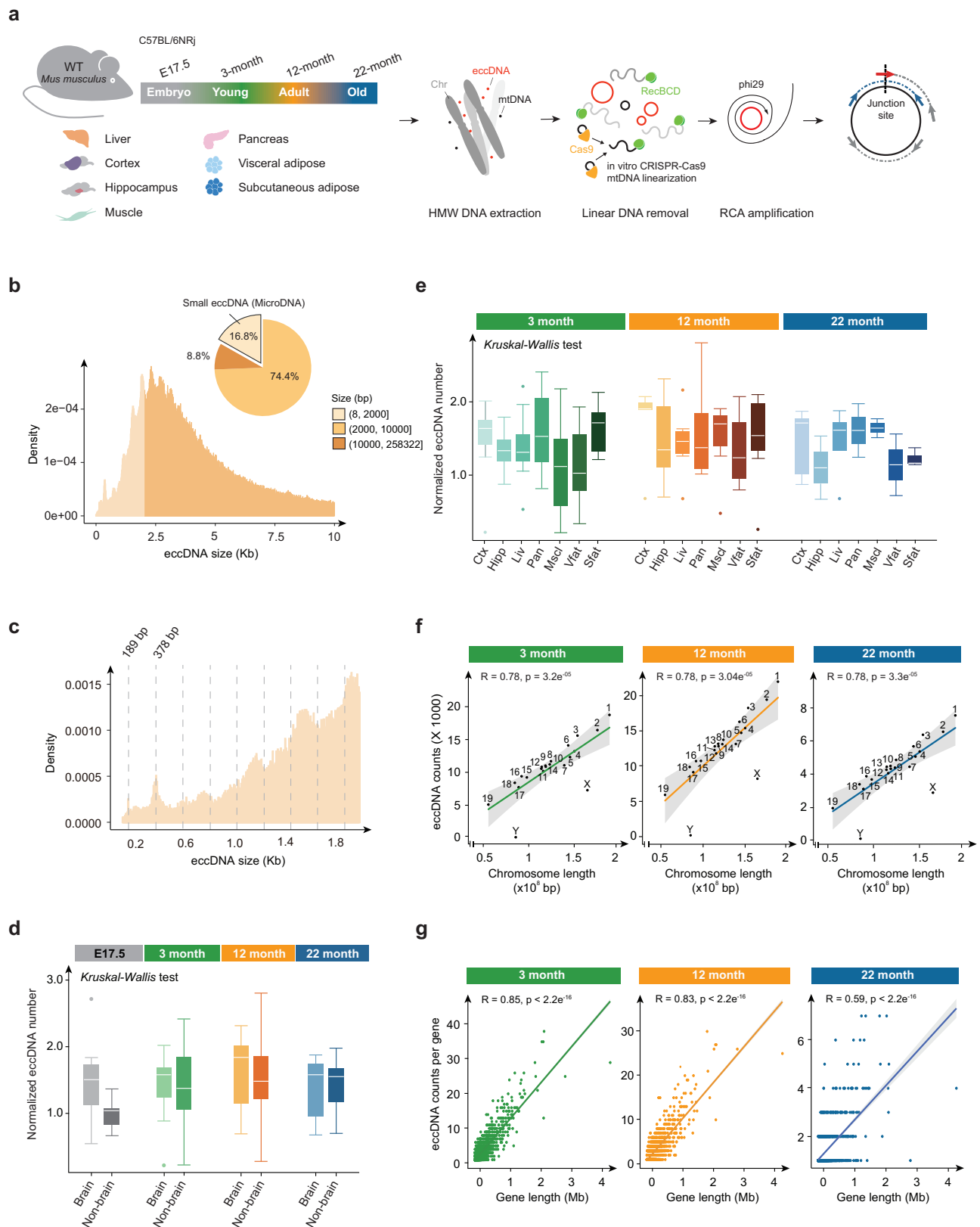
Next, we estimated the number of eccDNA per cell using internal plasmid controls added to samples before eccDNA purification. Plasmids were recaptured in all samples. Based on these plasmid controls, we detected an average of 45 unique eccDNAs per nucleus (Supplementary Data 4).

To determine if genes could be compared with respect to their eccDNA counts, we next mapped all gene fragments and full genes contained within eccDNA. First, the number of eccDNA showed a very clear correlation with chromosome length for all age groups (Fig. 1f), where chromosomes produce eccDNA linearly proportional to its length. Similarly, the number of eccDNA a gene produces is also linearly proportional to its length (Fig. 1g). This correlation between the number of eccDNA and gene length was observed in each tissue across 3, 12, and 22-month-old age groups (Supplementary Fig. 4a), which meant that genes with the highest contribution to eccDNA were consistently the same between samples and corresponded to the longest genes in the mouse genome (Supplementary Fig. 4b). Gene Ontology terms associated with long genes such as synapse, cell junction/adhesion, and neurodevelopment were overrepresented among genes that formed the highest count of eccDNA, regardless of tissue source (Supplementary Fig. 4c). Thus, we reasoned that genes could not solely be compared based on the eccDNA count due to the confounding effect of gene length (Supplementary Fig. 4d), and therefore we subsequently normalized eccDNA counts to gene length in the next analysis.

A link between eccDNA formation and transcription

Next, we sought to investigate if transcription could influence eccDNA formation by introducing a higher mutational burden, as previously shown for the *CUP1* gene in yeast *Saccharomyces cerevisiae*¹² and suggested for the human titin gene *TTN*².

We purified RNA from matching samples of the liver, cortex, hippocampus, SAT, VAT, pancreas, and skeletal muscle to evaluate the potential association between transcription and eccDNA levels in mitotic and post-mitotic tissue (Supplementary Data 5). In liver, transcriptomes clustered together according to age in a Principal Component Analysis (PCA) plot (Fig. 2a), which corresponded to specific transcriptomic profiles for each age group. Interestingly, the



transcriptomic distributions became increasingly broader with aging (Fig. 2b), suggesting unique liver transcript profiles associated with aging in biological replicates.

We also observed that the eccDNA profiles became increasingly different as livers aged. We assessed eccDNA similarities between biological replicates by calculating the Sørensen-Dice Similarity Index (DSI) between each pair of samples. Briefly, this index was used to

calculate the overlap between eccDNA origin coordinates (bp) between samples (Supplementary Fig. 5). We used technical replicates (same sample sequenced multiple times) to estimate the maximal similarity level of our technique and used that to normalize biological replicate comparisons (Supplementary Fig. 5). We found that eccDNA samples became increasingly different with age, as evidenced by the larger standard deviation of DSI indexes in the oldest group (Fig. 2c).

Fig. 1 | eccDNA atlas overview. **a** High Molecular Weight DNA was extracted from seven different tissues across 3, 12, and 22-month-old wildtype mice. Mitochondrial DNA was linearized using CRISPR-Cas9, and linear DNA was removed with exonuclease RecBCD enzyme. Remaining exonuclease resistant fraction was amplified using phi29 enzyme and sequenced using Circle-Seq. **b** Density size distribution of all detected eccDNA across samples, with a pie chart representing percentage in each size group. **c** Density size distribution of segment from 0 to 2 kilobase, with grey lines indicating periodicities for each peak. **d** Number of unique eccDNA detected in brain (cortex and hippocampus) and non-brain (muscle, liver, visceral adipose, subcutaneous adipose and pancreas) across embryo ($n = 10, 11$), 3 ($n = 12, 49$), 12 ($n = 12, 40$), and 22-month-old ($n = 7, 13$) age groups. The eccDNA number is normalized to million reads per million genomes and the number of genomes was estimated using averages between Qubit and qPCR estimates (see methods). Significance was assessed using two-side Kruskal–Wallis test with pairwise comparisons and Bonferroni correction (embryo: $p = 0.057$, 3-month-old: $p = 0.98$, 12-

month-old: $p = 0.32$, 22-month-old: $p = 0.99$). The median is marked with a white line, with margins showing the interquartile range (IQR), whiskers extending up to 1.5 IQR, and outliers showing as individual data points. Only samples with $\geq 70\%$ of reads mapped to the genome were included. **e** Number of unique eccDNA detected per tissue across 3 ($n = 7, 5, 10, 11, 8, 9, 11$), 12 ($n = 5, 7, 6, 9, 7, 8, 10$) and 22-month-old ($n = 5, 2, 4, 2, 2, 2, 3$) age groups. eccDNA number is normalized similarly. The eccDNA number is normalized to million reads per million genomes and the number of genomes was estimated using averages between Qubit and qPCR estimates. Significance was assessed using two-side Kruskal–Wallis test with pairwise comparisons and Bonferroni correction (tissue group: $p = 0.09$, age group: $p = 0.33$). **f, g** Number of eccDNA per each chromosome as a function of chromosome length (**f**) and number of eccDNA per gene as a function of gene length (**g**) with two-sided t test with Pearson regression. Pearson correlation (R), p -value, and gray shadow representing 95% confidence interval of the regression are shown.

Thus, both eccDNA and RNA profiles diverged increasingly with age (Fig. 2d).

We detected RNA from 80–90% of the genes and gene fragments mapped on the eccDNA profiles in the liver, suggesting tissue-specific formation and supporting a link between transcription and eccDNA formation (Fig. 2e). To test if genes with high transcript levels had increased eccDNA formation rates, we performed correlation analysis based on the Kernell-density gradients and “Locally Weighted Scatterplot Smoothing–LOESS” between gene transcript level (RNA count per gene per Mb) and eccDNA formation rate (number of eccDNA per gene per Mb) for all age groups. The regression analysis results indicated a clear positive relationship between the two processes, suggesting that genes with higher expression levels also gave rise to more eccDNAs derived from them (Fig. 2f–h and Supplementary Figs. 6a–c, 7a–c, 15a–c, 16a–c, and 17a–c). Interestingly, there was a positive, yet segmented, relationship between transcript levels and eccDNA load, with two different effects on eccDNA formation for the liver (Supplementary Fig. 8a–c), cortex (Supplementary Fig. 8d–f), hippocampus (Supplementary Fig. 8g–i), muscle (Supplementary Fig. 8j–l), SAT (Supplementary Fig. 8m–o) and VAT (Supplementary Fig. 8p–r). In the first phase, the eccDNA number did not correlate with increasing transcript level, while in the second phase, the eccDNA number correlated logarithmically with the transcript level. We then confirmed this dual effect by fitting a non-linear segmented hinge function and identifying the transcript level corresponding to the phase change, called “hinge” or “threshold”. For the liver, the threshold levels decreased with age (liver: 3-month hinge = 11.49, 12-month hinge = 10.77, 22-month hinge = 8.24; Supplementary Fig. 8a–c), suggesting that the transcript level at which eccDNA formation increases declines with age. In agreement with this, we observed that the overall eccDNA count per transcript level (eccDNA counts per Mb/RNA count per Mb) was elevated with age, suggesting that at the same gene expression level, eccDNA formation is accentuated with age (Fig. 2i, 3-month median = -7.99 , 12-month median = -7.79 , 22-month median = -7.24).

We also found a two-phased correlation between transcript level and eccDNA formation for the hippocampus, cortex, muscle, SAT and VAT, and also, in these tissues, the threshold at which transcription conferred to log increase of eccDNA changed with age (Supplementary Fig. 8d–r). However, the overall eccDNA count per transcript level did not increase with age for the hippocampus (Supplementary Fig. 6d, 3-month median = -9.83 , 12-month median = -9.98 , 22-month median = -10.6), cortex (Supplementary Fig. 7d, 3-month median = -9.84 , 12-month median = -9.45 , 22-month median = -10.4), muscle (Supplementary Fig. 15d, 3-month median = -9.74 , 12-month median = -7.99 , 22-month median = -8.92), SAT (Supplementary Fig. 16d, 3-month median = -9.97 , 12-month median = -9.3 , 22-month median = -11.64), VAT (Supplementary Fig. 17d, 3-month median = -9.55 , 12-month median = -8.92 , 22-month median = -10.42).

To understand if the association between eccDNA and transcription applied for all sizes of eccDNA, we tested this association for eccDNA in sizes smaller than 2000 bp and larger than 2000 bp for liver, cortex, pancreas, SAT, VAT, skeletal muscle and hippocampus. Transcription correlated with eccDNA formation for both small and large eccDNA. We confirmed this effect for both small eccDNAs (<2000 bp) and large eccDNAs (≥ 2000 bp) (Supplementary Fig. 9).

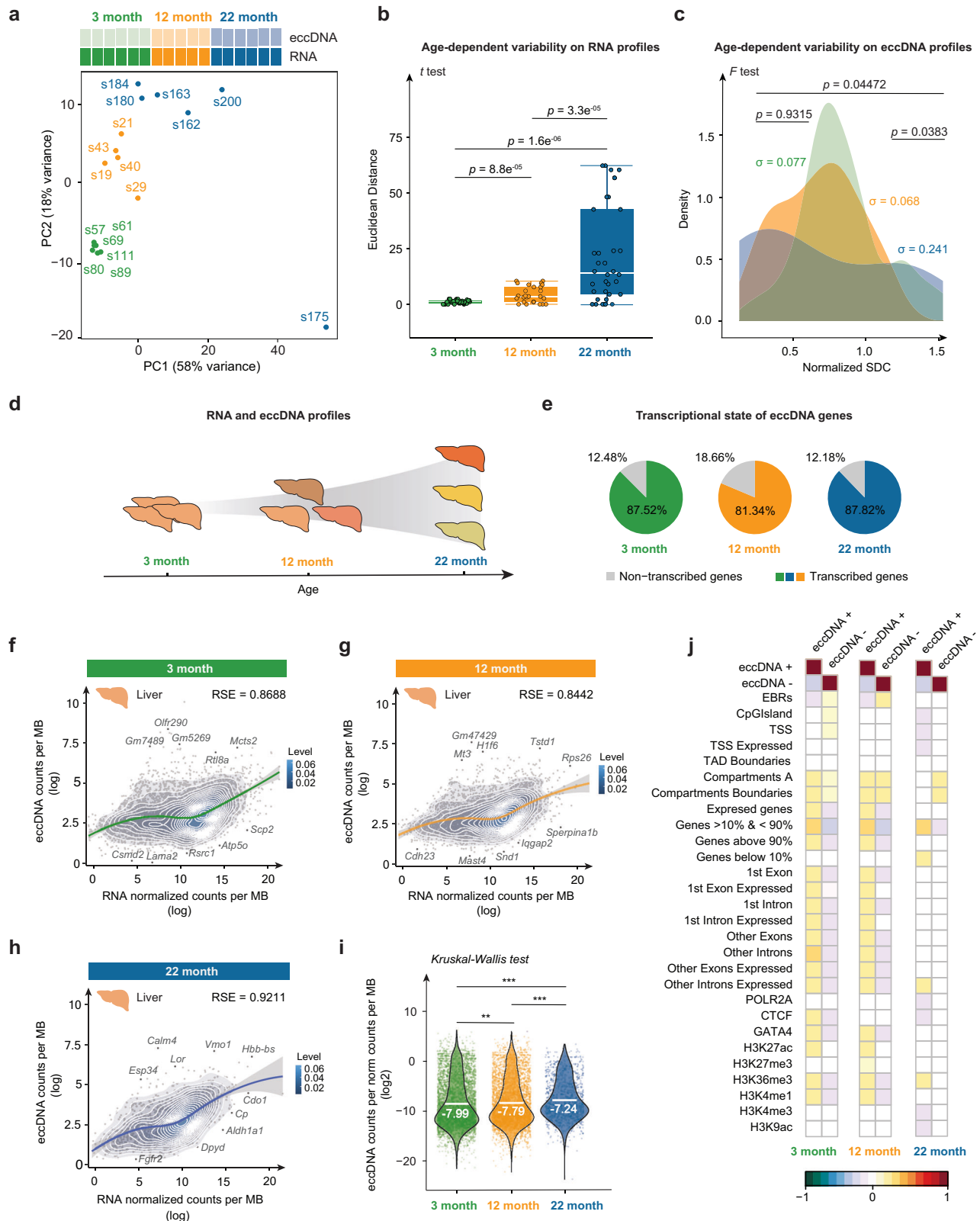
To test if eccDNA primarily came from transcribed genes, we employed permutation analysis^{21,22} based on 10,000 permutations to a number of publicly available genome features matching tissue and age and using 50 kb genome windows containing eccDNA (eccDNA+) or not containing eccDNA (eccDNA–) (See methods). Overall, eccDNA+ shows positive correlations with open chromatin (Compartment A), compartment boundaries, expressed genes, exons, and introns in the liver from 3 and 12-month-old groups. Additionally, we observed a positive correlation with open chromatin histone marks such as H3K27ac and H3K4me1 (Fig. 2j and Supplementary Fig. 10).

Interestingly, most of these positive associations, including histone marks with eccDNA formation, were lost in the 22-month-old group (Fig. 2j and Supplementary Fig. 10). This is consistent with the epigenetic “erosion” aging model whereby genome and folding marks are progressively lost with aging²³. Remarkably, only the H3K36me3 histone mark was consistently associated with eccDNA in all ages, including the 22-month-old group.

Collectively, these findings provide evidence that the number of eccDNA per gene is affected by the level of transcription, and the eccDNA profiles are thereby tissue-specific. Specifically, it appears that there is a logarithmic increase in the eccDNA level per gene as a function of transcription at a higher transcriptional level, revealing that the levels of eccDNA are especially affected in genes with high transcript levels.

Intron-rich genes with many splice forms are protected against eccDNA formation

Next, we sought to identify transcribed genes that are more vulnerable to eccDNA formation or protected from eccDNA formation. We used the 10% and 90% quantiles as cutoffs to detect genes that either formed more eccDNA than expected (eccDNA exposed) or less than expected (eccDNA protected) than the rest of the genes (bulk) (Fig. 3a and Supplementary Data 6). We consistently identified the same genes being protected from circularization in the liver when transcribed across all ages (Fig. 3b). Specifically, we identified 125 genes (between 3 and 12-month-old mice), 32 genes (between 3 and 22-month-old mice) and 35 genes (between 12 and 22-month-old mice). From those, 11 genes were found in all age groups (Fig. 3b, hypergeometric test, $p = 0.04$, Supplementary Data 7). On the contrary, we did not find any significantly recurrent genes exposed to eccDNA circularization in either the liver or the other analyzed tissues (Fig. 3c and



Supplementary Data 7). This indicated that some genic features might aid the DNA repair machinery in preventing circularization events during the transcription of these genes.

To identify the genic characteristics that the protected genes had in common, we first tested for differences in exonic and intronic densities. We found that the protected genes had more exons than in randomly chosen gene set (Fig. 3e). Interestingly, we found that intron

density (total intron length/gene length) was also significantly higher in both the protected genes and recurrently protected genes than bulk genes across ages (Fig. 3f). The number of transcripts isoforms was also significantly higher in both the protected genes and recurrently protected genes than bulk genes across ages (Fig. 3g). To rule out the confounding effects of gene length and exon number, we also used the metric isoforms per exon number per Mb, in which both the protected

Fig. 2 | A link between eccDNA formation and transcription. **a** Schematic overview of RNA samples sequenced in liver. PCA plot of RNA profiles between age groups, with the percentage of variance explained by PC1 and PC2 components shown in brackets. **b** Euclidean distance between all samples across 3 age groups (3-month: $n = 6$; 12-month: $n = 5$; 22-month: $n = 6$) in the PCA space, assessed using two-side t test. Significance (***) is shown if $p < 0.001$ (3-month vs. 12-month: $p = 8.8e^{-05}$, 3-month vs. 22-month: $p = 1.6e^{-06}$, 12-month vs. 22-month: $p = 3.3e^{-05}$). The median is marked with a white line, with margins showing the interquartile range (IQR), whiskers extending up to 1.5 IQR, and outliers showing as individual data points. **c** Variance in the Sørensen-Dice Similarity Coefficient (SDC) assessing similarities between eccDNA profiles in each age group (see methods and Supplementary Fig. 5). Variance equality was tested using two-side F test. Significance (*) is shown if $p < 0.05$. **d** Schematic overview of RNA and eccDNA profiles showing

variability increasingly different as liver aged. **e** Percentage of genes detected within eccDNA profiles and expressed in tissues and ages. Gray indicates genes detected in eccDNA profiles with no transcript detected. **f–h** Logarithmic eccDNA number per Mb for each gene as a function of normalized read counts per Mb of liver. LOESS regression fitted (colored line) with kernel 2-densities plotted. Residual Standard Error (RSE) is shown for each LOESS regression. **i** eccDNA level at the same gene expression level: (\log_2 (eccDNA count per Mb/normalized RNA count per Mb)) across age groups. The Kruskal–Wallis test was used to assess significant differences (3 vs. 12-month: $p = 0.0061$; 3 vs. 22-month: $p = 1.4e^{-15}$, 12 vs. 22-month: $p = 4.4e^{-10}$). The white bar represents median values (3-month = -7.99 , 12-month = -7.79 , 22-month = -7.24). **j** Heatmap shows the association of eccDNA with interested genomic features within 50 kb genome windows. eccDNA+: exists eccDNA in 50 kb windows. eccDNA-: does not exist.

genes and recurrently protected genes showed significantly higher numbers than bulk genes (Fig. 3h). This suggests that genes with higher intronic densities and a high number of transcript isoforms are protected from transcription-induced eccDNA formation events.

An even larger number of genes was found to be recurrently protected from eccDNA formation in all samples of the hippocampus (Fig. 4a, b, hypergeometric test, p -value: $3 \cap 12 \cap 22$ months = $7.11e^{-14}$, Supplementary Data 7), cortex (Supplementary Fig. 11a, b, hypergeometric test, p -value: $3 \cap 12 \cap 22$ months = $2.28e^{-06}$, Supplementary Data 7), muscle (Supplementary Fig. 12a, b, hypergeometric test, p -value: $3 \cap 12 \cap 22$ months = 0.0029 , Supplementary Data 7), SAT (Supplementary Fig. 13a, b, hypergeometric test, p -value: $3 \cap 12 \cap 22$ months = 0.0002 , Supplementary Data 7) and VAT (Supplementary Fig. 14a, b, hypergeometric test, p -value: $3 \cap 12 \cap 22$ months = $5.85e^{-05}$, Supplementary Data 7). These recurrently protected genes also had higher intron density, a larger number of transcript isoforms, and a larger number of isoforms per exon number per Mb than bulk genes (Fig. 4d–f and Supplementary Figs. 11d–f, 12d–f, 13d–f, and 14d–f).

Some of the eccDNA-protected genes in the liver are linked not only to liver-specific functions but, more importantly, to liver diseases^{24–28} (Fig. 3d). This highlights the critical need for genomic stability of these genes to maintain proper liver functionality. Therefore, their protection from eccDNA formation may play a key role in ensuring their stability by preventing up/downregulation or isoform switch. For the other tissues, we confirmed that the functions of eccDNA-protected genes are related to their tissue functionalities (Fig. 4g and Supplementary Figs. 11g, 12g, 13g, and 14g), albeit with some GO terms related to neuron and axon function in non-neuronal tissues, likely due to infiltration of nervous tissues. Altogether, this suggests that eccDNA protection might contribute to gene stability and tissue homeostasis.

Discussion

Our atlas of eccDNA across tissues and ages has revealed that the total number of eccDNA does not appear to accumulate in healthy tissue during aging. This is surprising because other types of mutations, such as SNVs and indels, have been found to accumulate with age in studies in epithelium¹² and stem cells^{13,29} with aging. A possible explanation for the lack of accumulation could be that eccDNA, which does not replicate, gets diluted out as cells divide. This has previously been shown for yeast³⁰, where the majority of eccDNA has no replication origins and disappear as yeast cells undergo replicative aging. Alternatively, animal cells could have mechanisms for clearing eccDNA, though there is yet no evidence for this in healthy cells. Another possibility is that some of the cells that accumulate eccDNA might enter senescence and disappear, causing the eccDNA to disappear.

Accumulating chromosomal DNA damage is known to impact actively transcribed genes severely. During aging, basal transcription is compromised by stalling RNA polymerases due to DNA mutations,

leading to a decrease in transcript productivity and skewing aging transcriptomes toward shorter isoform length^{31,32}. Additionally, transcription itself can introduce more stress in the form of supercoiling and torsional stress, R-loops, or clashes between transcription and replication bubbles³³. R-loop formation correlates strongly with transcription levels, and they can cause DNA double-strand breaks³¹ a strong determinant of eccDNA excision¹⁴. Our observation that highly transcribed genes correlate with an increased number of eccDNA indicates that a fraction of DNA repair events during transcriptional stress leads to circularization and eccDNA formation. Alternatively, the correlation between eccDNA and transcript levels could be explained by eccDNA formation regulating transcription. However, this explanation seems unlikely because eccDNAs are rare phenomena that cannot regulate common mechanisms such as transcription. Future experiments where highly transcribed genes are repressed through promoter deletion or deletion of transcription factor will assist the understanding of how transcription is associated with eccDNA frequency in mammalian cells.

We also show that besides the random formation of eccDNA, open chromatin, and transcribed genes form more eccDNA, and its formation is proportional to its transcript level. This means that the same genes that are particularly highly expressed in a tissue are also the genes that generate more mutations resulting in eccDNA. However, not all genes underlie this apparent result of transcriptional stress. Genes with high intron density and many splice isoforms are partially protected against eccDNA formation (Fig. 4h). One explanation for this connection could be that components of the RNA processing machinery, such as RNA splicing factors, could modulate the action of DNA repair proteins, a crosstalk previously suggested³¹. The fact that intron density affects the rate of eccDNA production is on its own suggestive that the transcripts contribute to DNA excision. The spliceosome is thought to protect RNA from invading the transcribed gene to form R-loops³⁴ and high levels of splicing may thereby protect a gene from eccDNA formation^{35,36}.

Our results indicate that high intron density might protect against eccDNA formation and concomitant deleterious mutations in response to transcriptional stress. This connection could potentially imply that high intron density might have been selected during the evolution of animal cells to reduce deleterious mutations formed during eccDNA generation.

The profiles of eccDNA in healthy tissue are expected to differ from those in tumor tissue, where there is a selection for amplification of oncogenes on eccDNA/ecDNA³⁷, which is not expected in healthy tissue. Furthermore, cancer is associated with genome instability due to mutations in DNA repair pathways and chromothripsis, which may cause high levels and altered profiles of eccDNA in tumors³⁸. Still, we expect eccDNA with intact oncogenes to arise occasionally in healthy and premalignant tissue and provide the basis for tumorigenesis. An example of this has been observed in Barrett's esophagus (BE), a condition predisposed to malignant esophageal adenocarcinoma, where BE patients occasionally carry amplified oncogenes on eccDNA⁸.

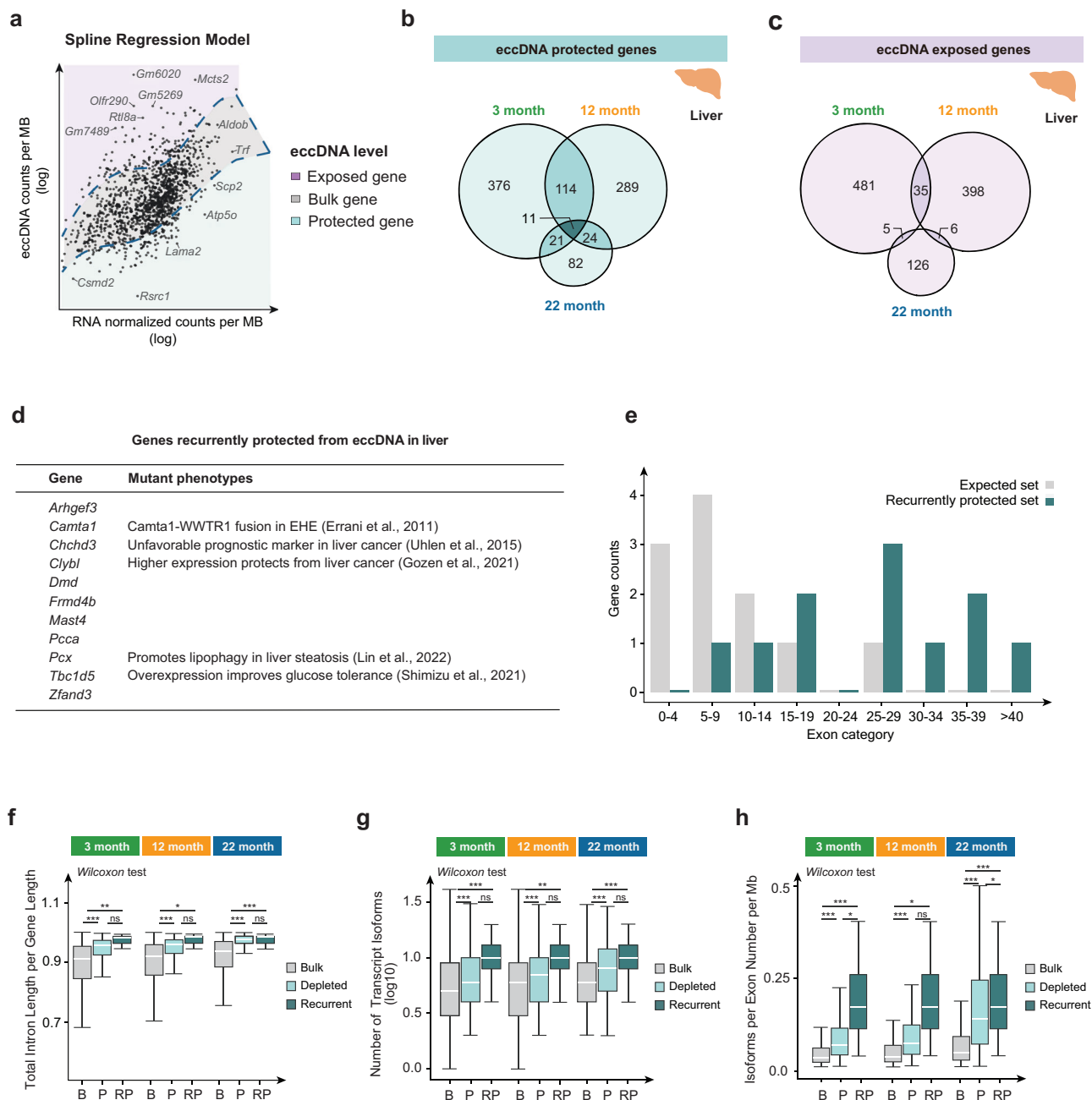
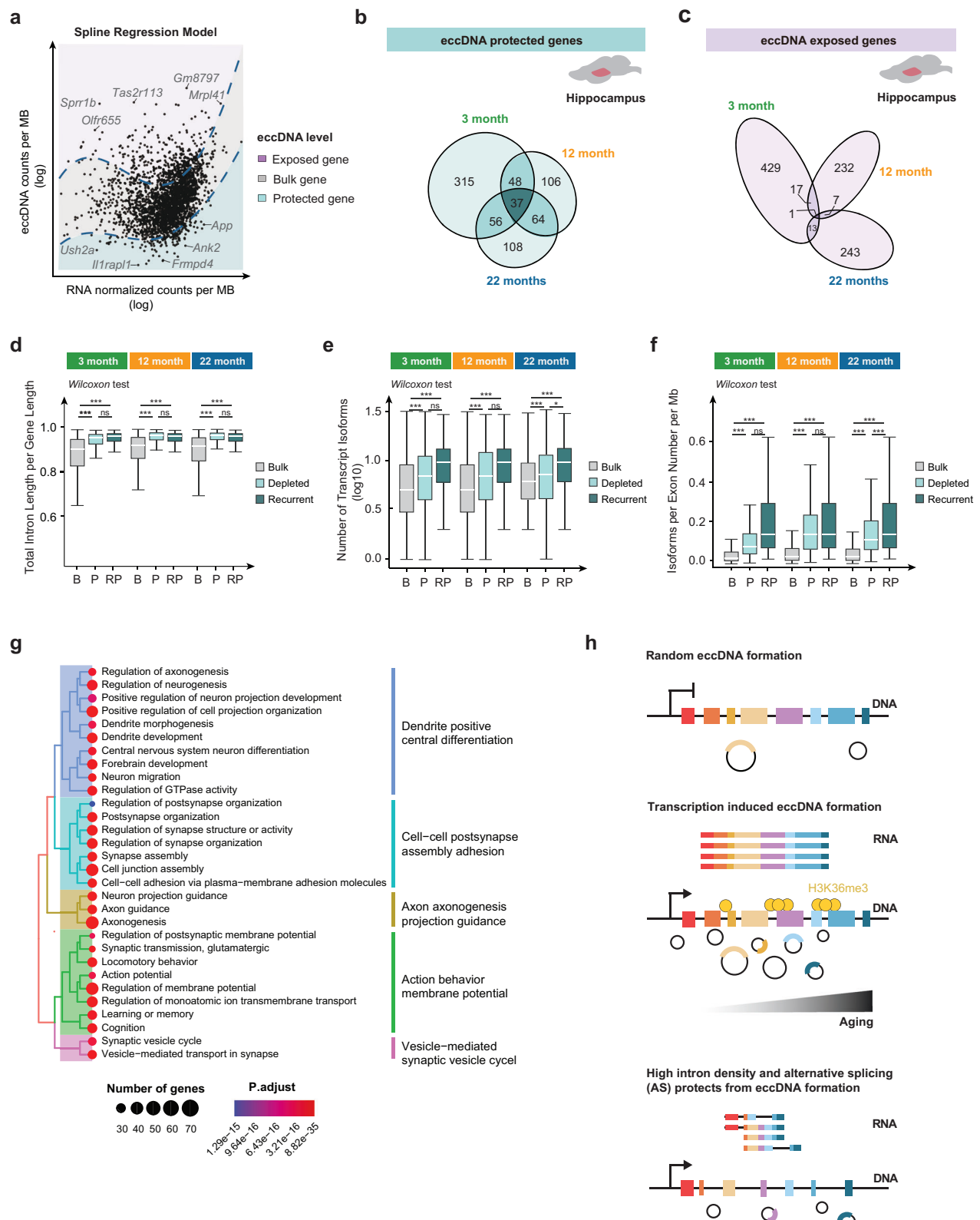


Fig. 3 | Liver-specific intron-rich genes with many splice forms are protected against eccDNA formation. **a** Schematic overview of a spline regression on eccDNA per gene as a function of normalized RNA read counts from the liver of 22-month-old mice. Genes between 10% and 90% quantiles are defined as bulk genes (grey), those above 10% quantiles are defined as eccDNA-exposed genes (light purple), while those below 90% quantiles are defined as eccDNA-protected genes (light green). **b** Venn diagram displaying the overlap and differences in eccDNA-protected genes across 3, 12, and 22-month-old mouse liver. Significance was assessed using one-side hypergeometric test, with *p*-values summarized in Supplementary Data 7. **c** Venn diagram displaying the overlap and differences in eccDNA-exposed genes across 3, 12, and 22-month-old mouse liver. Significance was assessed using one-side hypergeometric test, with *p*-values summarized in Supplementary Data 7. **d** A list of 11 recurrently protected from eccDNA formation across three ages. **e** Number of protein-coding genes within each specified exon number category (grouped). The recurrently protected set (dark green) shows the number of exons in the 11

genes. The expected set (gray) shows exon numbers from the 11 genes randomly selected from the mouse genome. **f** Relative density of intron among B (Bulk genes), P (Protected genes), and RP (Recurrently Protected genes) across 3 (*n* = 6), 12 (*n* = 5) and 22-month-old mouse liver (*n* = 2), calculated by dividing total intron length by gene length. **g** Relative level of transcripts isoforms from bulk, protected, and recurrently protected genes across 3 (*n* = 6), 12 (*n* = 5) and 22-month-old mouse liver (*n* = 2), measured by log10 (number of transcripts isoforms). **h** Relative level of isoforms from bulk, protected, and recurrently protected genes across 3 (*n* = 6), 12 (*n* = 5) and 22-month-old mouse liver (*n* = 2), measured by isoform per exon number per Mb. Significance was assessed using two-sided Kruskal–Wallis test with Wilcoxon multiple comparisons and Bonferroni correction, with *p*-values provided in Supplementary Data 9. The median is marked with a line across each box; box margins mark the interquartile range (IQR), and whiskers extend up to 1.5 IQR, with individual data points shown for outliers.



Collectively, these findings contribute to our understanding of transcription and eccDNA formation, which has so far only been demonstrated for the *CUP1* gene in the eukaryote *S. cerevisiae*. While our results are primarily correlative, we believe they are of significance and provide insights into the etiology of eccDNA, particularly in the stages leading to cancer. Besides giving fundamental insight into the underlying cause of eccDNA formation in healthy tissue, the current eccDNA atlas will also be a reference to future studies of mice mutants

that are expected to affect eccDNA. The atlas can thereby provide an important understanding of eccDNA biology and the clinical implications eccDNA might have for cancer and other diseases.

Methods

Ethics

Mice were housed at the University of Copenhagen, Denmark according to regulations. The ethical approval was granted by the

Fig. 4 | Hippocampus-specific intron-rich genes with many splice forms are protected against eccDNA formation. **a** Schematic overview of a spline regression on eccDNA per gene as a function of normalized RNA read counts from the hippocampus of 22-month-old mice. Genes between 10% and 90% quantiles are defined as bulk genes (grey), those above 10% quantiles are defined as eccDNA-exposed genes (light purple), while those below 90% quantiles are defined as eccDNA-protected genes (light green). **b** Venn diagram displaying the overlap and differences in eccDNA-protected genes across 3, 12, and 22-month-old mouse hippocampus. Significance was assessed using one-side hypergeometric test, with *p*-values summarized in Supplementary Data 7. **c** Venn diagram displaying the overlap and differences in eccDNA-exposed genes across 3, 12, and 22-month-old mouse hippocampus. Significance was assessed using one-side hypergeometric test, with *p*-values summarized in Supplementary Data 7. **d** Relative density of intron from B (Bulk genes), P (Protected genes), and RP (Recurrently Protected genes across ages) across 3 (*n* = 4), 12 (*n* = 4) and 22-month-old mouse

hippocampus (*n* = 4), calculated by dividing total intron length by gene length. **e** Relative level of transcripts isoforms from bulk, protected, and recurrently protected genes across 3 (*n* = 4), 12 (*n* = 4) and 22-month-old mouse hippocampus (*n* = 4), measured by log10 (number of transcripts isoforms). **f** Relative level of isoforms from bulk, protected, and recurrently protected genes across 3 (*n* = 4), 12 (*n* = 4) and 22-month-old mouse hippocampus (*n* = 4), measured by isoform per exon number per Mb. Significance was assessed using two-sided Kruskal–Wallis test with Wilcoxon multiple comparisons and Bonferroni correction, with *p*-values provided in Supplementary Data 9. The median is marked with a line across each box; box margins mark the interquartile range (IQR), and whiskers extend up to 1.5 IQR, with individual data points shown for outliers. **g** Gene ontology biological function terms of all protected genes across 3, 12, and 22-month-old mouse hippocampus. Significance was assessed using one-sided hypergeometric test. **h** Model of eccDNA formation in genes with low and high transcript levels, high intron density and alternative splicing.

Danish Veterinary and Food Administration (Fødevarestyrelsen) to Jørgen FP Wojtaszewski (Number 2019-15-0201-01659).

Mice

Male C57BL/6NRj mice (3 and 12-month-old) and pregnant female mice were purchased from Janvier, France, and housed under the conditions of standard rodent chow (SAFE D30, Scientific Diets, Augy, France), temperature: 22 °C (± 2 °C) and humidity: 55% (± 10%) at the University of Copenhagen. A group of 12-month-old male mice were aged to 22 months. Embryonic tissues were collected from pregnant female mice, non-embryonic tissues were collected from male mice at 3, 12, and 22-month-old age, flash-frozen in liquid nitrogen, and stored at −80 °C until further use. All mice were euthanized by cervical dislocation, embryos were decapitated, and there was no antemortem abandonment.

White blood cell quantification

Fresh blood was collected right after decapitation into EDTA-coated Eppendorf tubes. Leukocyte count was measured with a NucleoCounter NC-3000 from Chemometec. Briefly, fresh blood was mixed with Blood Lysis Buffer Solution 17 (Chemometec, Cat. No. 910-3017) and loaded into NC-Slide A8 for counting according to the manufacturer's instructions.

Purification of eccDNA

EccDNA was enriched and amplified in the following five steps:

DNA isolation and genome quantification. Small amounts of tissue (>25 mg) were cut and weighed in a −20 °C environment. Tissues were digested overnight in Proteinase K (Qiagen, Cat. No. 1913) at 56 °C and 900 rpm. Following digestion, three yeast plasmids were spiked in according to the amounts indicated (see next section). The next day, High molecular Weight DNA (HMW DNA) was purified using MagAttract HMW DNA Kit (Qiagen, Cat. No. 67563), following the manufacturer protocol. DNA concentration was assessed by Qubit HS DNA dsDNA High Sensitivity assay on Qubit 3.0 Fluorometer (Thermo Fisher). The number of genomes extracted was calculated using Qubit DNA concentration and the weight of a diploid mouse genome to assess genome number (ca. 6 picogram). Additionally, we also estimated the number of genomes via qPCR by first measuring the copy number of the *Cox5b* gene using a standard curve of known amounts of the target region for qPCR. We purchased a 120-bp ssDNA oligo corresponding to the *Cox5b* genomic region (1:36691363-36691482). We used primers *Cox5b* Fwd (5' CAGGGTGACTGACGCCAC 3') and *Cox5b* Rev (5' CGCTAAAAGTAGACGCCGCA 3') to perform the amplification of both the oligo and samples using 95 °C + (95 °C 15 s + 60 °C 15 s + 72 °C 15 s) 25 cycles + 72 °C in a Quant Studio 7 Flex (Cat. No. 4485701, Life Technologies). We used the average between the two methods to estimate the number of genomes extracted for each

sample and to normalize the number of unique eccDNA present in each sample.

Removal of linear DNA. For linear DNA removal, approximately 4.8 µg of HMW DNA from each sample was split into two independent tubes and treated in parallel. Each tube started the digestion with 30 U of exonuclease V (RecBCD) (New England Biolabs, Cat. No. M0345L), together with NEBuffer 4 and ATP, according to the RecBCD recommendations. The samples were incubated at 37 °C in a heating block for seven days. Every 24 h, fresh nuclease (25 U of RecBCD), ATP, and buffer were added. Exonuclease was heat inactivated for 30 min at 70 °C. Next, samples were cleaned using 1.8× AMPure XP beads (Agencourt AMPure XP, Beckman Cat. No. A63881). Complete removal of linear DNA was confirmed by qPCR of the *Cox5b* gene with primers Fwd 5' CAGGGTGACTGACGCCAC 3' and Rev 5' CGCTAAAAGTAGACGCCGCA 3'.

Removal of mitochondrial DNA. Mitochondrial DNA (mt-DNA) was removed in vitro using CRISPR-Cas9 and two sgRNA targeting two distinct sites in the mouse mt-DNA as previously described¹⁷. The sgRNAs (30 nM each) and 30 nM Cas9 *Streptococcus pyogenes* (New England Biolabs, Cat. No. M0386T) were first assembled and incubated for 20 min at 25 °C without shaking to allow the sgRNA to bind the Cas9. Next, the assembled reaction was mixed with the linear-free DNA and incubated at 37 °C for at least 90 min. Then, the Cas9 was heat-inactivated for 10 min at 65 °C, and the samples were allowed to cool off at room temperature. On top of RecBCD (20 U), 10 mM ATP, and 10× 3.1 NEBuffer were added and incubated overnight at 37 °C to digest the linearized mt-DNA molecules. DNA was then also purified using the AMPure XP magnetic beads (Beckman, Cat. No. A62881). Removal of mt-DNA was confirmed by qPCR.

Amplification of exonuclease-resistant DNA. After mt-DNA removal, the DNA was amplified via Multiple Displacement Amplification or MDA using a Phi29 enzyme with a random primase (TruePrime RCA kit, 4BaseBio, Cat. No. SKU 390100). Reactions were incubated at 30 °C for 48 h and enzymes heat inactivated at 65 °C for 10 min. DNA concentrations were measured using Qubit and samples were run on a 0.7% agarose gel.

Internal plasmid controls. For each sample, we spiked in three different yeast plasmids: p4339 from C. Boone Lab (p160 50000 copies/sample), YGPM25009 from Open Biosystems (p162 20000 copies/sample), and *Escherichia coli* plasmid pBR322 (p166 10000 copies/sample).

Sequencing and eccDNA identification

Phi29-amplified DNA was sequenced at BGI-Qingdao, China, using the MGI DNBSEQ-T5 platform with a 2×150 pair-end sequencing strategy.

SOAPnuke (v2.1.7) was first used to remove (1) low-quality reads containing more than 50% bases with base quality less than or equal to 12, (2) adaptor-polluted reads, and (3) high content of unknown base reads containing more than 10% 'N' nucleotides from raw sequencing reads. BWA-MEM (v0.7.17) was used to align the high-quality clean reads to the mouse (GRCm38/mm10) reference genome with the '-q' option, then samtools (v1.9) was used to sort and index the BAM files. Circle-Map (v1.1.4, <https://github.com/iprada/Circle-Map>) with both ReadExtractor and Realign modules was used to identify the coordinates of eccDNAs. Finally, high-confidence eccDNAs were preserved based on the following filtering criteria: (1) at least two split reads mapping around the circle junction (or one split read and at least one discordant read) and (2) at least 95 % mean base coverage within the eccDNA detection coordinates. There, we only selected samples with at least 70% of the reads mapping to the mouse genome or mitochondrial DNA to avoid differences in the circle count due to poor read count.

Semi-quantitative assessment of eccDNA number

Prior to the rolling-circle amplification by phi29 polymerase, we added three different types of spiked-in plasmids to all purified eccDNA samples: plasmid 160 (p160), plasmid 162 (p162), and plasmid 166 (p166). Because the amount of plasmid amplification (% mapped reads to plasmids) would indirectly indicate the number of eccDNA present, knowing the quantity of plasmid in each sample made it possible for us to compare samples (Supplementary Data 8). Firstly, we calculated the percentage of reads in each sample that mapped to each plasmid as the following formula: %reads/plasmid = (total number of mapped reads on plasmids × 100)/total number of all mapped reads. Then, we calculated the relative numbers of eccDNAs per sample as the following formula: relative number of eccDNAs/sample = (number of detected eccDNAs/samples)/(%reads/plasmid). Finally, the normalized eccDNA abundance per sample were calculated by taking the average of the three types of plasmids: normalized eccDNA abundance/sample = (the relative numbers of eccDNAs per sample using p160+ the relative numbers of eccDNAs per sample using p162+ the relative numbers of eccDNAs per sample using p166)/3.

Genomic annotation of eccDNA

Repetitive elements annotation was performed using the UCSC table browser, bedtools (v2.17.0) was used to identify the overlapping regions of eccDNA and the repetitive region annotations. The normalized ratio was calculated as follows:

$$\text{Normalized ratio of repetitive elements} = \frac{\text{percentage of DNA falling in element A}}{\text{percentage of element A over the length of whole genome}} \%$$

RNA isolation and sequencing

RNA from tissue adjacent to eccDNA tissue samples was extracted using RNeasy Mini Kit (Qiagen, Cat. No. 74104). Samples were first pulverized in high-speed agitation TissueLyser (Qiagen, Cat. No. 85600) by adding a metallic bead and disruption buffer RLT supplemented with β-mercaptoethanol, as stated in the manufacture protocol. Pancreas samples (high endogenous RNase levels) were incubated with RNAlater-ICE (ThermoFisher, Cat. No. AM7030) overnight at -20 °C prior to disruption in order to improve the integrity of the RNA. After purification, RNA quantity and integrity were checked using Qubit HS RNA assay on Qubit 3.0 Fluorometer (Thermo Fisher) and 2100 Bioanalyzer (Agilent) using RNA Nano Chips (Agilent, Cat. No. 5067-1511). Samples were then used to prepare RNA-seq libraries using MGIEasy RNA kit (MGI Tech, Cat. No. 1000006384) with mRNA polyA enrichment by

Dynabeads® mRNA Purification Kit (Thermo Fisher, Cat. No. 61006) and sequenced in a DNBSEQ-T1.

RNA-seq data processing

Quality control and low-quality read trimming were conducted using FastQC (v0.11.7) and fastp (v0.19.5) with default parameters to obtain cleaned high-quality reads. Read alignment against the mouse genome (mm10) was performed using hisat2 (v2.2.0) within all samples, then samtools (v1.9) was used to sort and index BAM files. After quantifying reads with featurecounts (v2.0.3), the original count matrix was normalized using DESeq2 (v1.36.0) R package.

Euclidean distance calculation

In the PCA two-dimensional space, the Euclidean distance between the RNA profiles of any two samples within the same age group was calculated in R based on their coordinates in each dimension. The Euclidean distance between the points (a, b) and (c, d) is calculated as: $\sqrt{(a - c)^2 + (b - d)^2}$.

Sørensen-Dice similarity

To calculate the Sørensen-Dice similarity index, we used the overlap length between two eccDNA sample profiles (.bed files) to construct a similarity matrix between our biological replicates for each age group. We next used technical replicates (the same sample processed independently three times) to normalize the similarity score, assuming that the technical replicates would provide the upper limit of similarity between two samples in that tissue.

Data regression and fitting

To identify the genes that are potentially protected from making eccDNA despite transcription, the quantile regression with cubic spline (5° of freedom) was performed using quantreg (v5.94) and splines (v4.2.1) R package. In the regression model, the transcription level (log-transformed normalized read counts per gene per MB, based e) and eccDNA level (log-transformed eccDNA number per gene per MB, based e) were taken as a dependent variable and independent variable, respectively. LOESS method was fitted by using default parameters in ggplot2 (v3.4.1) R package. Finally, non-linear segmented hinge functions were fitted using chngpt (v2023.1-30) R package.

Random dataset generation

The random dataset was generated artificially from the whole mouse

genome with custom script. It consisted of 5000 random chromosomal regions with the same size as 5000 randomly extracted eccDNAs from the current study. Ten sets were generated and used in the analyses.

Multi-association analyses

The statistical significance of the eccDNA overlap with other features was performed using a permutation test. A permutation test is a non-parametric statistical method used to determine the significance of an overlap by comparing it to the distribution of overlapping under the null hypothesis (non-overlapping). To achieve this, we repeatedly shuffled the data (in this case, 10,000 times) to create multiple permuted datasets that represent the null hypothesis scenario. For each permuted dataset, the test statistic was recalculated, producing a distribution of test statistics under the null hypothesis. The

significance of the observed test statistic was assessed by determining the proportion of permuted test statistics that were as extreme as or more extreme than the observed value (p -value). Once the null hypothesis was rejected (p -value > 0.005), we calculated the associated normalized z -score to assess how significant the overlapping between two features was in the multi-comparison context. Therefore, the p -value calculation was adjusted using the Benjamini–Hochberg procedure. The z -score value for association with an adjusted p -value greater than 0.005 was considered 0. The z -score calculated was subsequently normalized by dividing it by the square root of n , where n is the number of regions in the permuted regions. As input datasets, we considered the entire eccDNA region. We implemented a series of functions to allow the calculation of associations between multiple regions, including the genomic positions of A/B compartments, topologically associated domains, the ChIP-seq peaks of POLR2A, GATA4, CTCF, H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, H3K9ac in addition to the TSS, CpG Island, and introns/exons. To perform the overlapping calculations and normalization, we employed the statistical tool *regioneReloaded*³⁹ as previously described²². *RegioneReloaded* uses the statistical framework previously developed in *regioneR*²¹ that allows the calculation of associations of different region sets simultaneously.

Gene ontology enrichment

Gene ontology (GO) enrichment analyses of genes protected from eccDNA formation were performed using *clusterProfiler* (v4.8.1) *R* package with the *enrichGO* function. P values were adjusted with the BH method.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequence data from Circle-seq and RNA-seq generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) database. The Circle-seq and RNA-seq data are available under accession code [PRJNA5957769](https://www.ncbi.nlm.nih.gov/sra/PRJNA5957769). The relevant analysis data generated in this study are provided in Supplementary Information. Source data are provided as a Source Data file. The ChIP-seq peaks of POLR2A (*ENCSR000CBR*), GATA4 (*ENCSR194GNY*), CTCF (*ENCSR000CBU*), H3K27ac (*ENCSR000CDH*), H3K27me3 (*ENCSR000CEN*), H3K36me3 (*ENCSR000CEO*), H3K4me1 (*ENCSR000CAO*), H3K4me3 (*ENCSR000CAP*) and H3K9ac (*ENCSR000CEQ*) were obtained from the ENCODE database. The TSS, CpG Island, and introns/exons coordinates were obtained from the USCS mm10 at <http://hgdownload.soe.ucsc.edu/downloads.html#mouse>. Source data are provided with this paper.

Code availability

The custom scripts for this study are available at GitHub [https://github.com/regenbergelab/MOUSE_ATLAS] and Zenodo [<https://doi.org/10.5281/zenodo.14195725>].

References

- Paulsen, T., Kumar, P., Koseoglu, M. M. & Dutta, A. Discoveries of extrachromosomal circles of DNA in normal and tumor cells. *Trends Genet.* **34**, 270–278 (2018).
- Møller, H. D. et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat. Commun.* **9**, 1069 (2018).
- Henriksen, R. A. et al. Circular DNA in the human germline and its association with recombination. *Mol. Cell* **82**, 209–217.e7 (2022).
- Paulsen, T., Shibata, Y., Kumar, P., Dillon, L. & Dutta, A. Small extrachromosomal circular DNAs, microDNA, produce short regulatory RNAs that suppress gene expression independent of canonical promoters. *Nucleic Acids Res.* **47**, 4586–4596 (2019).
- Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
- deCarvalho, A. C. et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat. Genet.* **50**, 708–717 (2018).
- Lange, J. T. et al. The evolutionary dynamics of extrachromosomal DNA in human cancers. *Nat. Genet.* **54**, 1527–1533 (2022).
- Luebeck, J. et al. Extrachromosomal DNA in the cancerous transformation of Barrett’s oesophagus. *Nature* **616**, 798–805 (2023).
- Stöber, M. C. et al. Intercellular extrachromosomal DNA copy-number heterogeneity drives neuroblastoma cell state diversity. *Cell Rep.* **43**, 114711 (2024).
- Oobatake, Y. & Shimizu, N. Double-strand breakage in the extrachromosomal double minutes triggers their aggregation in the nucleus, micronucleation, and morphological transformation. *Genes, Chromosomes Cancer* **59**, 133–143 (2020).
- López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: an expanding universe. *Cell* **186**, 243–278 (2023).
- Huang, Z. et al. Single-cell analysis of somatic mutations in human bronchial epithelial cells in relation to aging and smoking. *Nat. Genet.* **54**, 492–498 (2022).
- Blokzijl, F. et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).
- Dillon, Laura W. et al. Production of extrachromosomal MicroDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Rep.* **11**, 1749–1759 (2015).
- Møller, H. D., Ramos-Madrugal, J., Prada-Luengo, I., Gilbert, M. T. P. & Regenberg, B. Near-random distribution of chromosome-derived circular DNA in the condensed genome of pigeons and the larger, more repeat-rich human genome. *Genome Biol. Evol.* **12**, 3762–3777 (2020).
- Hull, R. M. et al. Transcription-induced formation of extrachromosomal DNA during yeast ageing. *PLoS Biol.* **17**, e3000471 (2019).
- Feng, W. et al. Targeted removal of mitochondrial DNA from mouse and human extrachromosomal circular DNA with CRISPR-Cas9. *Comput. Struct. Biotechnol. J.* **20**, 3059–3067 (2022).
- Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv preprint arXiv:1303.3997* (2013).
- Prada-Luengo, I., Krogh, A., Maretty, L. & Regenberg, B. Sensitive detection of circular DNAs at single-nucleotide resolution using guided realignment of partially aligned reads. *BMC Bioinforma.* **20**, 663 (2019).
- Schliephacke, T., Meinl, A., Kratzmeier, M., Doenecke, D. & Albig, W. The telomeric region is excluded from nucleosomal fragmentation during apoptosis, but the bulk nuclear chromatin is randomly degraded. *Cell Death Differ.* **11**, 693–703 (2004).
- Gel, B. et al. *regioneR*: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289–291 (2016).
- Álvarez-González, L. et al. 3D chromatin remodelling in the germ line modulates genome evolutionary plasticity. *Nat. Commun.* **13**, 2608 (2022).
- Yang, J.-H. et al. Loss of epigenetic information as a cause of mammalian aging. *Cell* **186**, 305–326.e27 (2023).
- Errani, C. et al. A novel WWTR1-CAMTA1 gene fusion is a consistent abnormality in epithelioid hemangioendothelioma of different anatomic sites. *Genes, Chromosomes Cancer* **50**, 644–653 (2011).
- Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

26. Gözen, D. et al. Transcriptome profiles associated with selenium-deficiency-dependent oxidative stress identify potential diagnostic and therapeutic targets in liver cancer cells. *Turkish J. Biol.* **45**, 149–161 (2021).
27. Selen, E. S. et al. Requirement of hepatic pyruvate carboxylase during fasting, high fat, and ketogenic diet. *J. Biol. Chem.* **298**, 102648 (2022).
28. Seaman, M. N., Mukadam, A. S. & Breusegem, S. Y. Inhibition of TBC1D5 activates Rab7a and can enhance the function of the retromer cargo-selective complex. *J. Cell Sci.* **131**, jcs217398 (2018).
29. Brunner, S. F. et al. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**, 538–542 (2019).
30. Prada-Luengo, I. et al. Replicative aging is associated with loss of genetic heterogeneity from extrachromosomal circular DNA in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **48**, 7883–7898 (2020).
31. Gyenis, A. et al. Genome-wide RNA polymerase stalling shapes the transcriptome during aging. *Nat. Genet.* **55**, 268–279 (2023).
32. Stoeger, T. et al. Aging is associated with a systemic length-associated transcriptome imbalance. *Nat. Aging* **2**, 1191–1206 (2022).
33. Vitelli, V. et al. Recent advancements in DNA damage-transcription crosstalk and high-resolution mapping of DNA breaks. *Annu. Rev. Genomics Hum. Genet.* **18**, 87–113 (2017).
34. Sorrells, S. et al. Spliceosomal components protect embryonic neurons from R-loop-mediated DNA damage and apoptosis. *Dis. Models Mech.* **11**, dmm031583 (2018).
35. Zylka, M. J., Simon, J. M. & Philpot, B. D. Gene length matters in neurons. *Neuron* **86**, 353–355 (2015).
36. McCoy, M. J. & Fire, A. Z. Intron and gene size expansion during nervous system evolution. *BMC Genomics* **21**, 1–9 (2020).
37. Kim, H. et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).
38. Shoshani, O. et al. Chromothripsis drives the evolution of gene amplification in cancer. *Nature* **591**, 137–141 (2021).
39. Malinverni, R., Corujo, D., Gel, B. & Buschbeck, M. regioneReloaded: evaluating the association of multiple genomic region sets. *Bioinformatics* **39**, btad704 (2023).

Acknowledgements

The authors thank the staff at BGI Qingdao sequencing facility for library preparation and sequencing, especially Wen Zhuang and Chunhua Li. We also thank Stine Ringholm for her help during dissection and the rest of the members of the Regenber Laboratory for fruitful discussions of results and the manuscript. This work is supported by the Novo Nordisk Foundation (NNF18OC0053139 and NNF21OC0072023) to G.A., H.P., and B.R. GA also received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 801199. X.L., Y.Q., and P.H. were funded by the Qingdao-Europe Advanced Institute for Life Sciences. AR-H. was funded by the Spanish Ministry of Science and Innovation (PID2020-112557GB-I00), the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, 2021SGR00122) and the Catalan Institution for Research and

Advanced Studies (ICREA). LA-G. was supported by a FPI predoctoral fellowship from the Ministry of Economy and Competitiveness (PRE-2018-083257).

Author contributions

G.A., H.P., and B.R. designed and conceptualized the study. H.P. performed tissue dissections supported by G.A. and B.R. Experimental methodology was done by G.A., Y.Q., and J.M. X.L. performed upstream sequencing data processing and circle identification. G.A., X.L., A.R.-H., L.A.-G., J.M.H., S.H., P.H., and B.R. did downstream data analysis. G.A., X.L., and B.R. curated figures and wrote the manuscript with input from all authors. Y.L., H.L., H.P., and B.R. provided equipment and resources.

Competing interests

B.R. is co-founder of CARE-DNA. X.L., Y.Q., P.H. are funded by Qingdao-Europe Advanced Institute for Life Sciences (QEI). J.M., Y.L., H.L. are or were employed by BGI. G.A., L.A.-G., J.M.H., S.H., A.R.-H., and H.P. declare that they have no known competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-57042-y>.

Correspondence and requests for materials should be addressed to Birgitte Regenber.

Peer review information *Nature Communications* thanks Anton Henssen, and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025