scientific data

Check for updates

OPEN

DATA DESCRIPTOR

The extrachromosomal circular DNA atlas of aged and young mouse brains

Xiaoning Hong[1,4], Jing Li[2,4], Peng Han[3,4], Shaofu Li[2], Jiaying Yu[1], Haoran Zhang[2], Jiang Li[1] ✉, Yonghui Dang[2] ✉ & Xi Xiang[1] ✉

Extrachromosomal circular DNA (eccDNA) refers to a distinct class of circular DNA molecules that exist independently from linear chromosomal DNA. Extensive evidence has firmly established the significant involvement of eccDNA in cancer initiation, progression, and evolutionary processes. However, the relationship between eccDNA and brain aging remains elusive. Here, we employed extrachromosomal circular DNA sequencing (Circle-seq) to generate a comprehensive dataset of eccDNA from six brain structures of both young and naturally-aged mice, including the olfactory bulb, medial prefrontal cortex, nucleus accumbens, caudate putamen, hippocampus, and cerebellum. Furthermore, through database annotation, we characterized the properties of mouse brain eccDNA, thereby gaining insights into the potential functions of eccDNA in the mouse brain. In conclusion, our study addresses a previously unexplored area by providing a comprehensive molecular characterization of eccDNA in brain tissues. The data presented in the study can be used as a fundamental resource to associate the molecular phenotypes of eccDNA with brain aging and gain deep insights into the biological role of eccDNA in mammalian brain aging.

Background & Summary

Aging is a complex process that leads to a decline in organ function and life quality as time passes[1]. During the aging process, changes occur in the brain, including shrinkage of brain size, remodelling of vascular structure and poor cognition[2]. At the histological level, this is manifested by decreased brain weight and atrophy, white matter atrophy occurring later than gray matter, ventricular enlargement, cerebral vascular changes, and reduced neuronal repair capacity[3]. In daily life, this is reflected in a decline in memory, learning ability, sensory perception and motor coordination[4]. This process is characterized by its irreversibility, progressive nature and cumulative effects[5]. These changes can serve as markers of brain aging and contribute to the development of neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, temporal lobe dementia and stroke[4].

Extrachromosomal circular DNA (eccDNA) refers to circular DNA molecules that originate from chromosomal DNA but exist independently in the cell nucleus and range in size from hundreds to millions of base pairs. eccDNA was first observed in the boar sperm by Alix Bassel and Yasuo Hoota in 1965 through electron microscopy[6]. There is increasing evidence indicating that eccDNAs are widely present in eukaryotic organisms, including plants, animals, and fungi, and they play important roles in various biological processes[7–9], including gene amplification[10], regulating RNA expression[11] and senescence[12]. Some studies also suggest that eccDNA may be involved in sequestering transcription factors, releasing molecules for intercellular communication and stimulating the innate immune pathways[13]. Recently, scientists have paid more attention to the role of eccDNA in tumor initiation and malignant progression, particularly the genetic heterogeneity among tumor cells caused by the amplification of oncogenes and drug resistance genes[14]. Despite its vital role in biology, relatively little is known about its distribution, function and clinical impact on brain aging. Ain Q *et al*. proposed that eccDNA plays a key role in aging and neurodegeneration of the central nervous system, but there is currently a lack of

[1]Scientific Research Center, The Seventh Affiliated Hospital of Sun Yat-Sen University, Shenzhen, Guangdong, China. [2]College of Medicine and Forensics, Xi'an Jiaotong University Health Science Center, Xi'an, Shanxi, China. [3]Department of Biology, University of Copenhagen, Copenhagen, Denmark. [4]These authors contributed equally: Xiaoning Hong, Jing Li, Peng Han. ✉e-mail: lijiang29@mail.sysu.edu.cn; psydyh@xjtu.edu.cn; xiangx25@mail.sysu.edu.cn

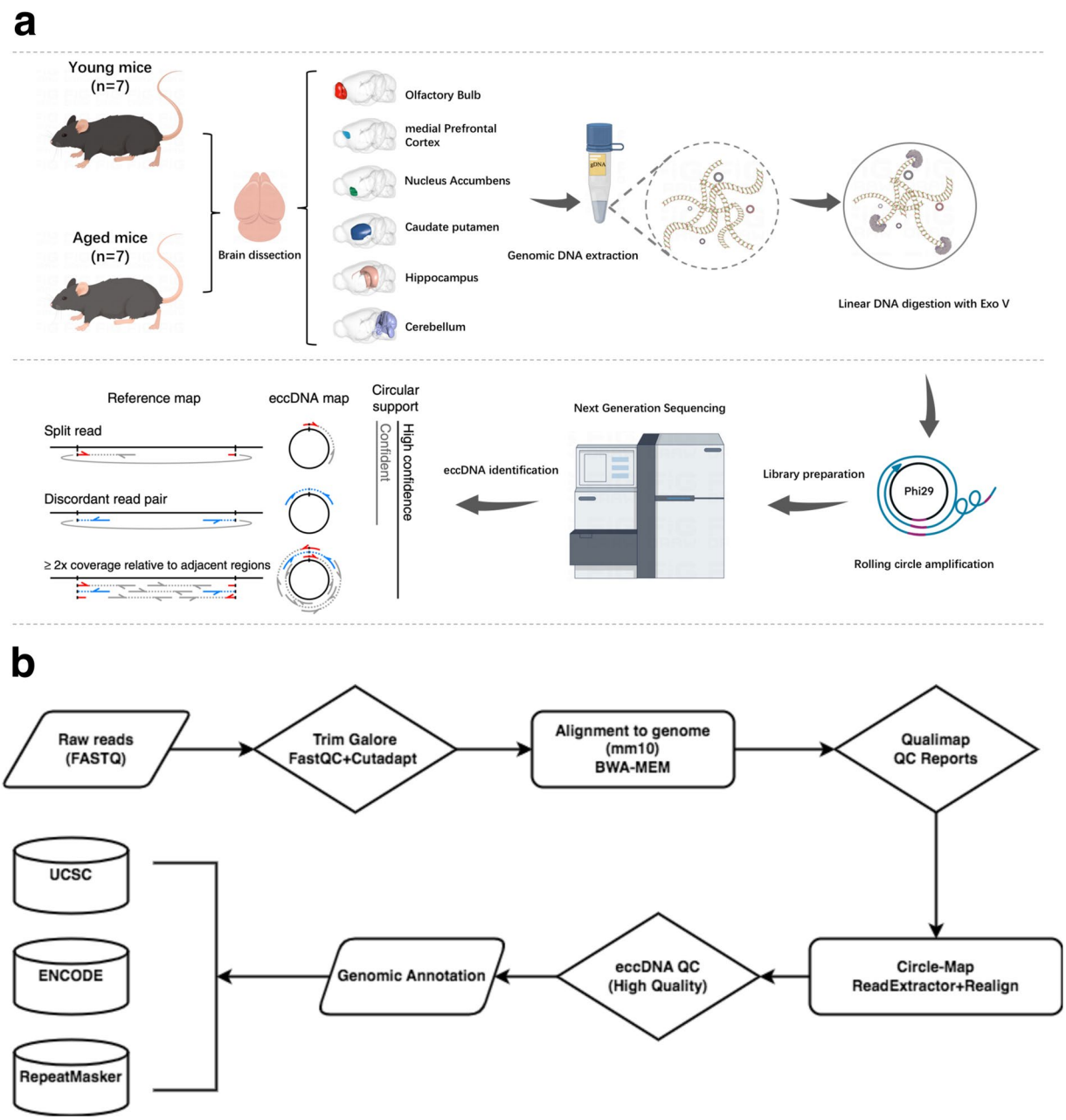

Fig. 1 Schematic overview of the experimental design and data analysis. (**a**) Workflow illustrating the process of mouse brain eccDNA purification and identification. (**b**) Bioinformatics pipeline for analysing mouse brain eccDNA.

systematic studies in this area[15]. In the present study, we described the genome atlas of eccDNA from various brain structures in mice with aging brains for the first time.

In this Data Descriptor, we conducted rigorous quality control (QC) measures to ensure the high quality of the Circle-seq data. The visualization summary of the study design and workflow is presented in Fig. 1a. The data analyses were performed using a standard pipeline (Fig. 1b). We established seven biological replicates for Circle-seq in each mouse brain structure, where seven young and aged mice were involved and eccDNA data were collected to confirm the authenticity of the datasets. The results showed that in total 876,918 and 1,168,079 eccDNAs were identified in the aged and young group, respectively (Fig. 2a). On average, the number of eccDNAs per brain structure in the aged group was 20,879 (ranging from 5,662 to 75,197), while in the young group it was 27,811 (ranging from 9,819 to 51,781) (Supplementary 1). After normalization, the average eccDNA count per million mapped reads (EPM) in the aged group was 203 (ranging from 89 to 338) per brain structure, compared to 258 (ranging from 158 to 364) in the young group (Fig. 2b,c and Supplementary Table 1). In addition, mitochondrial DNA plays a vital role as a positive control in the detection of eccDNA. The ratio of mitochondrial reads in the young and the aged group was 11.62% and 11.08% (mtDNA read counts divided by all the raw read counts), respectively. The abundance of mitochondrial DNA was quantified using the split-read counts per million reads

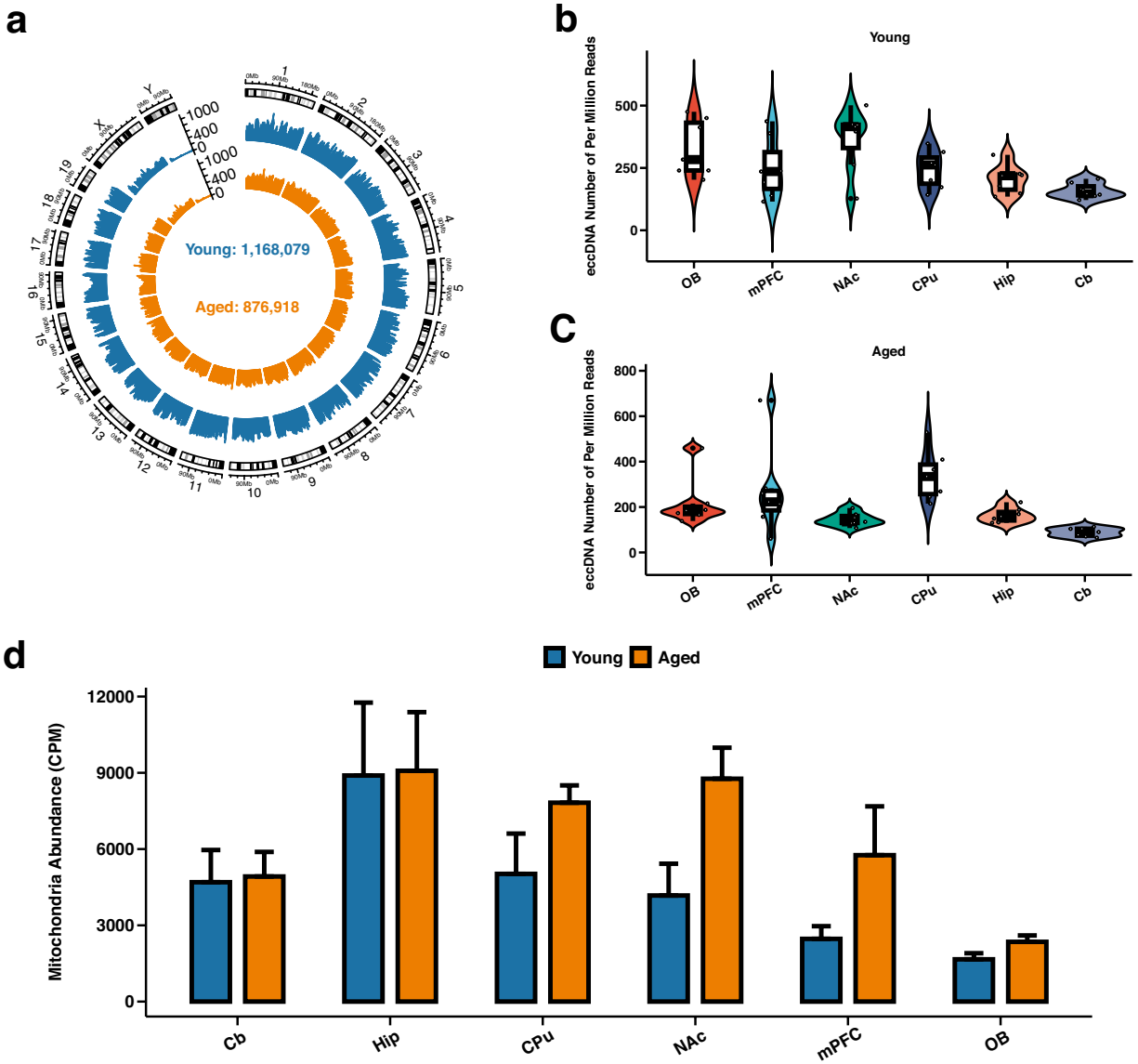

Fig. 2 EccDNA characteristics in the mouse brain. (**a**) Circos plot demonstrating the genome-wide distribution of eccDNA in both the Aged and Young groups. (**b,c**) Violin plots demonstrating the distribution of eccDNA per million mapped reads (EPM) across six brain structures for both groups. (**d**) Bar plot demonstrating the mitochondrial abundance in both the Aged and Young groups. The quantification was performed using the split-reads CPM (counts per million) normalization method.

(CPM) normalization method. The mean CPM values of mitochondrial DNA in the young and aged groups were 4486 and 6450, respectively (Fig. 2d). It indicated that the mitochondrial content of the samples was 0.45% and 0.65% (mtDNA split-read counts divided by the eccDNA split-read counts in each group), respectively.

The functional annotation results indicate that the most commonly eccDNA-carried elements are transposable elements, coding genes and cis-regulatory elements, accounting for 58.70%, 52.23% and 15% respectively in the aged group. In the young group, these elements account for 60%, 40%, and 20% respectively (Fig. 3c,d). In addition to that, the detected eccDNAs also contain considerable amounts of repetitive elements, including the SINE, LTR and LINE repeats (Fig. 3c,d).

We provided annotated bed files for each sample, which contain the location information of eccDNAs for each brain structure. These data not only can be a valuable resource for investigating the molecular phenotype association between eccDNA and brain aging, but also contribute to the research in pathogenesis and therapeutics of neurodegenerative diseases.

Methods

Experimental animal and samplings. The animals used in this study were approved by the Biomedical Ethics Committee of the Medical School of Xi'an Jiaotong University (No:2017-648). In this study, a total of seven male mice at 3 months of age (equivalent to ~20 years in humans) were collected for the young group, and seven male mice at 19 months of age (equivalent to ~60 years in humans) were collected for the aged group[16]. The mice

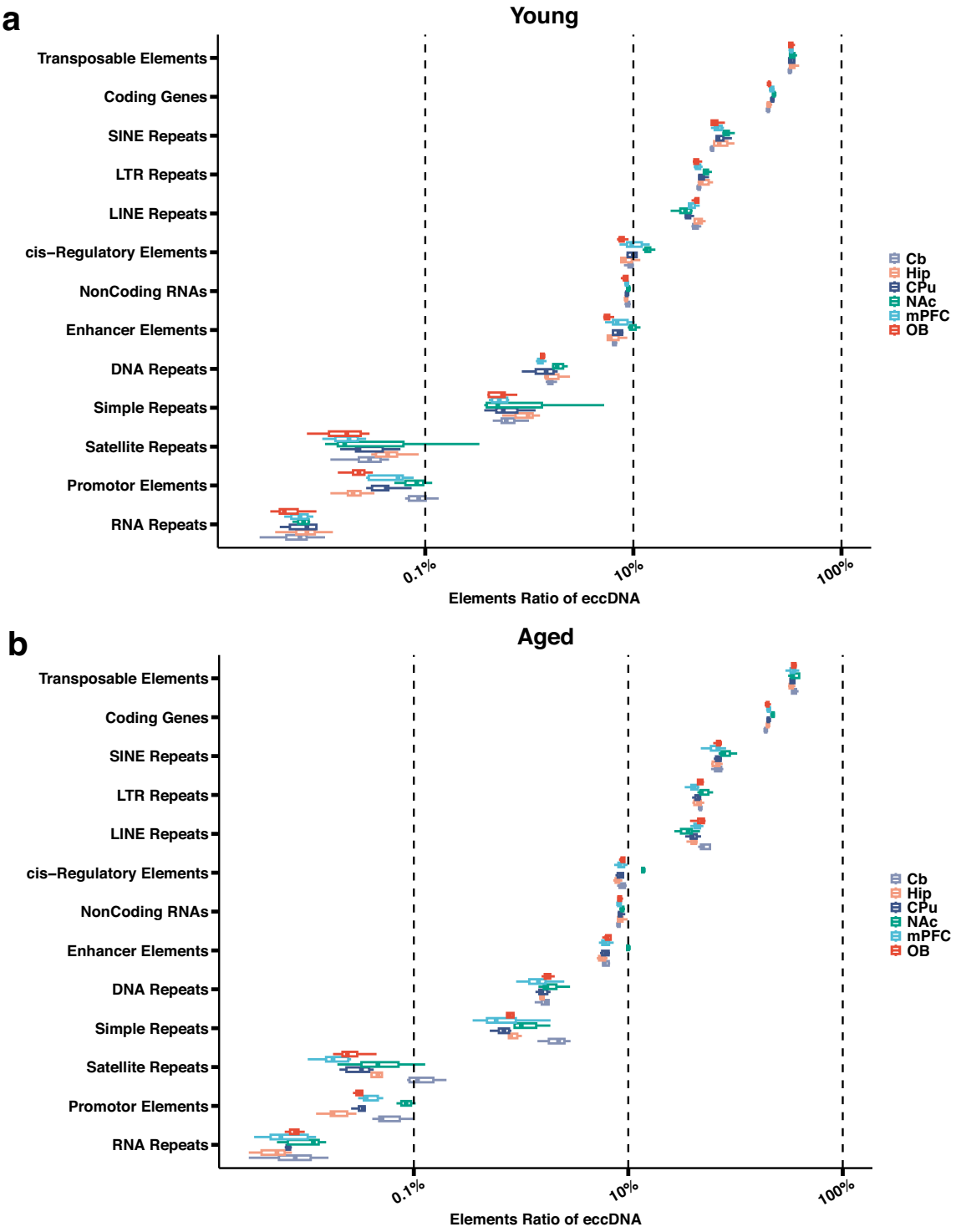

Fig. 3 Genomic annotation of mouse brain eccDNA. (**a,b**) The ratio of genomic element from eccDNA annotation across six brain structures for both the Aged and Young groups.

were euthanized using the cervical dislocation method, and six brain structures, including the olfactory bulb (OB), medial prefrontal cortex (mPFC), nucleus accumbens (NAc), caudate putamen (CPu), hippocampus (Hip) and cerebellum (Cb), were isolated.

eccDNA isolation and purification from brain tissue. Following the protocol of our previous study, we isolated and amplified eccDNA from mouse brain tissue samples[17]. Briefly, total genomic DNA was extracted from each tissue sample (~ 10 mg) by the MagAttract HMW DNA Kit (QIAGEN, Germany) according to the manufacturer's instruction. 100 ~ 1000 ng genomic DNA of each sample (depending on the tissue size and

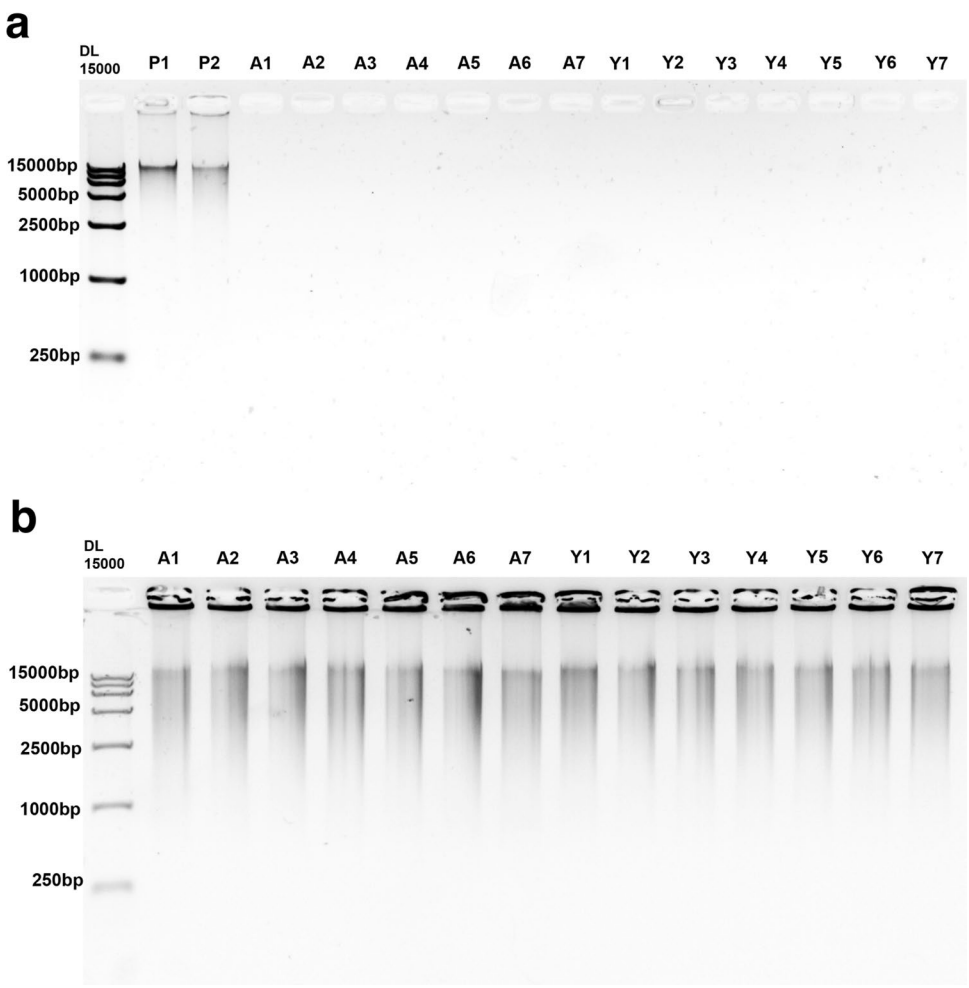

Fig. 4 Validation of the purification of mouse brain eccDNA. (**a**) The results of agarose gel electrophoresis demonstrated that no bands indicative of linear DNA was detected in the samples obtained from mouse brain tissue, except for the positive control (P1 and P2). (**b**) The results of agarose gel electrophoresis consistently exhibited bands corresponding to the amplified products of eccDNA after rolling circle amplification.

the resulting DNA amount. For example, the mouse mPFC tissue was very small and the input DNA amount was ranging from 129 ng to 1000 ng in this study) was treated with plasmid-safe ATP-dependent DNase (PSD) (Lucigen) to remove linear DNA. A 50 μL reaction mixture consisting of 400 ~ 1000 ng genomic DNA, 2 μL PSD, 5 μL 10 X PSD Buffer, 5 μL ATP solution (25 mM), and ddH2O was carried out at 37°C continuously for 7–12 days. Additional ATP solution and PSD (10 μL supplementary system containing 2 μL 10 X PSD Buffer, 2 μL ATP solution (25 mM), 2 μL PSD and 4 μL ddH2O) was added into the reaction mixture every 24 hours. Agarose gel electrophoresis was conducted on DNA samples both before and after PSD digestion to confirm the thorough elimination of linear DNA in each sample. Only the PSD-digested products lacking visible DNA bands were used for the subsequent rolling circle amplification (RCA) reaction. After gel test and before RCA, the PSD-digested mixture was inactivated at 70°C for 30 minutes and purified using VAHTS DNA Clean Beads (Vazyme). The resulting eccDNA product was eluted in 30 μL nuclease-free water and stored in −80°C refrigerator or subjected to the subsequent RCA reaction.

Rolling circle amplification of eccDNA. The purified eccDNA was subjected to RCA using the Phi29 Polymerase (Thermo Scientific) as previously described[17]. Firstly, a 20 μL reaction system consisting of 14 μL eccDNA (PSD-digested product), 4 μL 10 × phi29 reaction buffer and 2 μL Exo-resistant Random Primer (Invitrogen) was denatured at 95 °C for 5 min and ramped down to 4 °C at a ramping rate of – 5 °C/min. Then an additional mixture containing 4 μL dNTP (10 mM each), 0.8 μL DTT (100 mM), 1 μL phi29 polymerase (Thermo Scientific) and 14.2 μL ddH2O was added into the reaction system and incubated at 30°C continuously for 72 h. 1 μL RCA product of each sample was tested by agarose gel electrophoresis to confirm the amplification effect. Subsequently, the RCA product was purified using the VAHTS DNA Clean Beads (Vazyme) and eluted in 80 μL nuclease-free water. The concentration of the resulting RCA product was quantified by Qubit 4.0.

Library preparation and sequencing. For the library construction for next-generation sequencing, 1 μg RCA product from each sample was fragmented into 300–400 bp fragments using sonication with Covaris

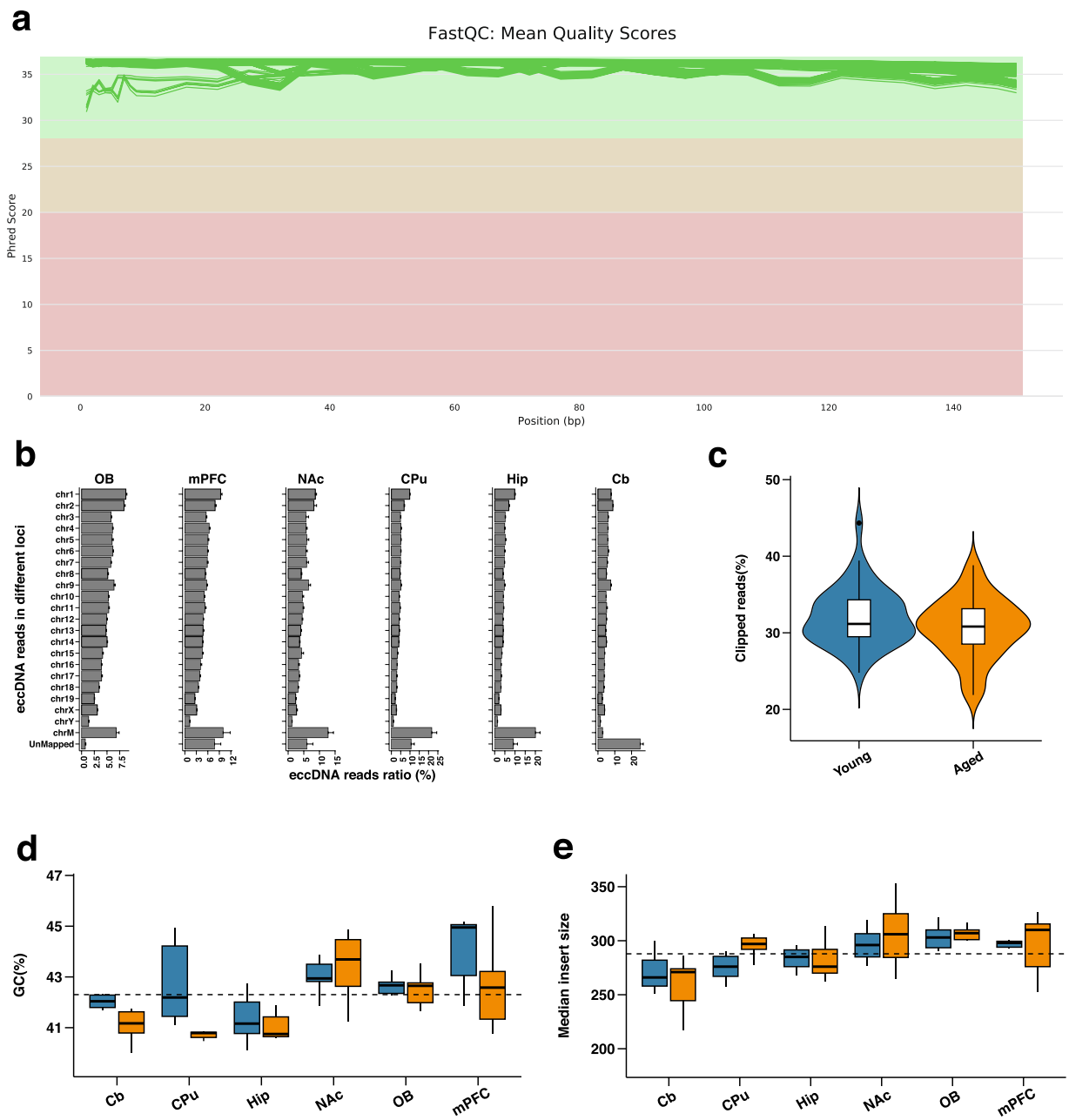

Fig. 5 Quality control of Circle-seq data. (**a**) The line plot from FastQC demonstrates consistent high Phred Scores across the Circle-seq dataset in all samples. (**b**) Mapped ratios of eccDNA data across different chromosomes in six brain structure sequencing reads of mice. (**c**) Violin plot demonstrates the percentage of soft-clipped reads for both groups. (**d,e**) Box plot demonstrates the GC content and median insert size of the two groups separately. (Blue: young group and Orange: aged group).

LE220. The fragmented DNA was subsequently recovered and used to construct the sequencing library with the MGIEasy DNA Library Preparation Kit from MGI-BGI in China. The quality and length distribution of the libraries were evaluated using the Bioanalyzer 2100 from Agilent. Subsequently, the samples' libraries were sequenced on the MGISeq-2000 platform (BGI) in paired-end mode, generating 150-base pair reads (PE150).

Data pre-processing and alignment. Adapters or low-quality raw reads were trimmed using Trim Galore (v0.5.0) software (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore) and Cutadapt (v1.18)[18] with the following parameters: "--paired -q 28--fastqc--gzip". The quality of processed clean data was evaluated using MultiQC (V1.15)[19] and Qualimap (v2.3)[20] with their default parameters. BWA-MEM (v 0.7.17)[21] with default parameters was used to align the high-quality clean reads to the mouse reference genome (GRCm38/mm10), and the BAM files were sorted by Samtools[22]. Additionally, the aligned BAM files were sorted by sequence name using Samtools for further analysis.

Identification of eccDNA by Cricle-Map. The Circle-Map (v1.1.4) software (https://github.com/iprada/Circle-Map) was used to extract circular reads from the sorted BAM files and identify eccDNAs based on the read coordinates in BAM files using default parameters. To improve the accuracy of eccDNA detection, multiple filtering steps were applied with the following specific criteria: (1) Split reads ≥ 2, (2) Circle score ≥ 200, (3) Coverage continuity ≤ 0.9, and (4) The standard deviation of coverage is smaller than the mean coverage across the entire eccDNA region.

Annotation and analysis of eccDNA. The GFF3 annotation file of genome features was obtained from GENCODE version M23 (https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M23/gencode.vM23.annotation.gff3.gz). The annotation data of candidate cis-Regulatory Elements (cCREs) was download from the ENCODE Project (https://www.encodeproject.org/). The data for repetitive DNA annotation was obtained from RepeatMasker open-4.0.5 (http://repeatmasker.org)[23]. To calculate the count of eccDNA mapped to the specific elements, we employed the multicov and groupby functions provided by BedTools[24] with default parameters.

Data Records

All the FASTQ files generated in this study by Circle-seq have been submitted and deposited in the NCBI Sequence Read Archive (SRA) with the BioProject accession number (PRJNA1012841)[25]. Additionally, the bed files containing information of eccDNA and the corresponding annotation data, as well as the detected results of eccDNA based on the mm39 genome, have also been uploaded to Figshare (https://doi.org/10.6084/m9.figshare.24086121.v4)[26].

Technical Validation

Quality check of the purified eccDNA samples. To verify if linear DNA from chromosomes was completely digested, we examined the status of DNA on a 0.8% agarose gel. The results indicated that, except for two positive controls, no DNA content was detected in the remaining samples. (Fig. 4a). Subsequently, Rolling Circle Amplification was performed on the eccDNA and DNA detection was carried out again using a 0.8% agarose gel. The outcomes showed the existence of amplified products of eccDNA in all samples (Fig. 4b).

Quality control of raw reads and sample statistics. We performed quality control on the Circle-seq data (Fig. 1b) and generated a series of QC metrics for Circle-seq (Supplementary Table 2). The overall quality of the Circle-seq dataset was deemed satisfactory at the levels of raw and mapped data in several features: (1) the Phred quality scores consistently exhibited a high level of quality in all samples (Fig. 5a); (2) the average ratio of mapping for all samples was 90.1% (Fig. 5b); (3) the average soft-clipped rate of 30.64% was determined for the aged group, while the young group had an average soft-clipped rate of 32.07% (Fig. 5c and Supplementary Table 3). This indicates the presence of a significant number of "split-reads" in the samples, which is also one of the characteristics of eccDNA; (4) the average GC content was approximately 41.88% for the aged group and 42.72% for the young group (Fig. 5d); (5) the average insert size (median length of the DNA fragment in the library) of each group ranged from 259 to 306 bp (Fig. 5e).

Usage Notes

The Circle-seq data processing pipeline, which includes raw data filtering, reads alignment, eccDNA identification and genomic annotation, was executed on a Linux operating system. All the source codes in R and Python, along with the optimized parameters used for downstream data analyses and visualization, are available online for access.

Code availability

The codes used to analyze the data in this study are available in the GitHub repository at the following URL: (https://github.com/XiaoningHong/MouseBrain_ScientificData).

Received: 12 October 2023; Accepted: 14 March 2024;
Published online: 27 March 2024

References

1. Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243–278, https://doi.org/10.1016/j.cell.2022.11.001 (2023).
2. Peters, R. Ageing and the brain. *Postgrad Med J* **82**, 84–88, https://doi.org/10.1136/pgmj.2005.036665 (2006).
3. Drayer, B. P. Imaging of the aging brain. Part I. Normal findings. *Radiology* **166**, 785–796, https://doi.org/10.1148/radiology.166.3.3277247 (1988).
4. Mattson, M. P. & Arumugam, T. V. Hallmarks of Brain Aging: Adaptive and Pathological Modification by Metabolic States. *Cell Metab* **27**, 1176–1199, https://doi.org/10.1016/j.cmet.2018.05.011 (2018).
5. Guo, J. *et al*. Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. *Signal Transduct Target Ther* **7**, 391, https://doi.org/10.1038/s41392-022-01251-0 (2022).
6. Hotta, Y. & Bassel, A. Molecular Size and Circularity of DNA in Cells of Mammals and Higher Plants. *Proc Natl Acad Sci USA* **53**, 356–362, https://doi.org/10.1073/pnas.53.2.356 (1965).
7. Moller, H. D., Parsons, L., Jorgensen, T. S., Botstein, D. & Regenberg, B. Extrachromosomal circular DNA is common in yeast. *Proc Natl Acad Sci USA* **112**, E3114–3122, https://doi.org/10.1073/pnas.1508825112 (2015).
8. Cohen, S., Houben, A. & Segal, D. Extrachromosomal circular DNA derived from tandemly repeated genomic sequences in plants. *Plant J* **53**, 1027–1034, https://doi.org/10.1111/j.1365-313X.2007.03394.x (2008).
9. Cohen, S., Yacobi, K. & Segal, D. Extrachromosomal circular DNA of tandemly repeated genomic sequences in Drosophila. *Genome Res* **13**, 1133–1145, https://doi.org/10.1101/gr.907603 (2003).

10. Mansisidor, A. *et al*. Genomic Copy-Number Loss Is Rescued by Self-Limiting Production of DNA Circles. *Mol Cell* **72**, 583–593 e584, https://doi.org/10.1016/j.molcel.2018.08.036 (2018).
11. 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, https://doi.org/10.1093/nar/gkz155 (2019).
12. Qiu, G. H., Zheng, X., Fu, M., Huang, C. & Yang, X. The decreased exclusion of nuclear eccDNA: From molecular and subcellular levels to human aging and age-related diseases. *Ageing Res Rev* **67**, 101306, https://doi.org/10.1016/j.arr.2021.101306 (2021).
13. 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, https://doi.org/10.1016/j.tig.2017.12.010 (2018).
14. Yang, L. *et al*. Extrachromosomal circular DNA: biogenesis, structure, functions and diseases. *Signal Transduct Target Ther* **7**, 342, https://doi.org/10.1038/s41392-022-01176-8 (2022).
15. Ain, Q., Schmeer, C., Wengerodt, D., Witte, O. W. & Kretz, A. Extrachromosomal Circular DNA: Current Knowledge and Implications for CNS Aging and Neurodegeneration. *Int J Mol Sci* **21** https://doi.org/10.3390/ijms21072477 (2020).
16. Flurkey, K., Currer, J. M. & Harrison, D. E. Mouse Models in Aging Research. *The Mouse in Biomedical Research (Second Edition)* **III**, 637–672 (2007).
17. Jiang, X. *et al*. Genome-wide characterization of extrachromosomal circular DNA in gastric cancer and its potential role in carcinogenesis and cancer progression. *Cell Mol Life Sci* **80**, 191, https://doi.org/10.1007/s00018-023-04838-0 (2023).
18. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
19. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048, https://doi.org/10.1093/bioinformatics/btw354 (2016).
20. Okonechnikov, K., Conesa, A. & Garcia-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294, https://doi.org/10.1093/bioinformatics/btv566 (2016).
21. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).
22. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, https://doi.org/10.1093/bioinformatics/btp352 (2009).
23. Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* **16**, 418–420, https://doi.org/10.1016/s0168-9525(00)02093-x (2000).
24. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**, 11 12 11–34, https://doi.org/10.1002/0471250953.bi1112s47 (2014).
25. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP458406 (2023).
26. Hong, X. *et al*. Annotation files for eccDNA in mouse brain tissue. *figshare* https://doi.org/10.6084/m9.figshare.24086121.v4 (2023).

Acknowledgements

This study was funded by the Seventh Affiliated Hospital of Sun Yan-Sen University, grant number ZSQYBRJH0024; the Research Start-up Fund of The Seventh Affiliated Hospital, Sun Yat-sen University, grant number 592026; the Shenzhen Science and Technology Innovation Commission, grant number JCYJ20220530145014033 and JCYJ20230807110302006; Guangdong Provincial Key Laboratory of Digestive Cancer Research, grant number 2021B1212040006 and the Open Fund of Guangdong Provincial Key Laboratory of Digestive Cancer Research, grant number GPKLDCR202206M.

Author contributions

Conceptualization, X.X., Y.D. and J.L.; Methodology, X.H., J.L., S.L., H.Z. and J.Y.; software, X.H.; validation, X.H. and J.L.; formal analysis, X.H., P.H. and J.Y.; resources, Y.D. and X.X.; data curation, J.L. and P.H.; writing—original draft preparation, X.H., J.Y. and J.L.; writing—review and editing, J.L., X.H., J.L., J.Y., P.H, S.L., H.Z., Y.D. and X.X.; visualization, X.H. and P.H.; supervision, X.X., Y.D. and J.L.; project administration, X.X., Y.D. and J.L.; funding acquisition, X.X., Y.D. and J.L. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-024-03146-x.

Correspondence and requests for materials should be addressed to Jiang Li, Y.D. or X.X.