



OPEN

DATA DESCRIPTOR

Integrated Transcriptome Analysis of Long Noncoding RNA and mRNA in Developing and Aging Mouse Retina

Kangjie Kong^{1,5}, Peiyuan Wang^{1,5}, Zihong Xie^{2,5}, Lu Wang^{3,5}, Jiaxuan Jiang¹, Yaoming Liu¹, Shaolin Du⁴, Jingwen Jiang¹, Yunhe Song¹, Fengbin Lin¹, Wei Wang¹, Xiuli Fang¹, Zhuoxing Shi¹✉, Xiulan Zhang¹✉ & Shida Chen¹✉

Mice have emerged as a widely employed model for investigating various retinal diseases. However, the availability of comprehensive datasets capturing the entire developmental and aging stages of the mouse retina, particularly during the elderly period, encompassing integrated lncRNA and mRNA expression profiles, is limited. In this study, we assembled a total of 18 retina samples from mice across 6 distinct stages of development and aging (5 days, 3 weeks, 6 weeks, 10 weeks, 6 months, and 15 months) to conduct integrated lncRNA and mRNA sequencing analysis. This invaluable dataset offers a comprehensive transcriptomic resource of mRNA and lncRNA expression profiles during the natural progression of retinal development and aging. The discoveries stemming from this investigation will significantly contribute to the elucidation of the underlying molecular mechanisms associated with various retinal diseases, such as congenital retinal dysplasia and retinal degenerative diseases.

Background & Summary

Mice serve as an invaluable model for investigating retinal diseases due to their resemblance to the human retina in terms of structure and developmental patterns^{1–3}. Upon reaching the neonatal stage, mice undergo the pivotal event of eye opening and light perception, during which retinal progenitor cells differentiate into distinct cell types, leading to the formation of retinal layers⁴. The maturation of synapses, dendrites, and cell junctions contributes to the progressive development of the retina, ultimately culminating in the specification and differentiation of retinal ganglion cells (RGC), Müller cells, bipolar cells, horizontal cells, amacrine cells, and rod/cone photoreceptor cells. The precise orchestration of this developmental process relies on the intricate regulation of transcription factors (TFs)^{5–7}. Exploring the transcriptome networks involved in the differentiation of retinal progenitor cells offers valuable insights into the developmental patterns of the retina and the broader central nervous system. Furthermore, identifying retinal aging can be challenging through current clinical imaging technology in human and pathological examination of animal tissue, especially in the early old stage. Previous studies have highlighted various hallmarks of cellular aging, including nuclear abnormalities, telomere attrition, mitochondrial dysfunction, genetic instability, and epigenetic alterations⁸. Transcriptome analysis presents a powerful approach to comprehensively investigate the genetic and epigenetic changes associated with the aging process and age-related retinal disorders.

Long noncoding RNAs (lncRNAs) are a class of RNA characterized by their length longer than 200nt and lack of protein coding function⁹. In comparison to messenger RNA (mRNA), most lncRNAs are typically less annotated and their functions are largely unexplored. However, lncRNAs may bind to microRNAs or TF, which

¹State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Guangdong Provincial Clinical Research Center for Ocular Diseases, Guangzhou, 510060, China. ²Joint School of Life Sciences, Guangzhou Medical University and Guangzhou Institutes of Biomedicine and Health (Chinese Academy of Sciences), Guangzhou, 510060, China. ³The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, 510060, China. ⁴Dongguan Tung Wah Hospital, Dongguan, 523000, China. ⁵These authors contributed equally: Kangjie Kong, Peiyuan Wang, Zihong Xie, Lu Wang. ✉e-mail: zhuoxing1988@163.com; zhangxl2@mail.sysu.edu.cn; chenshd3@mail.sysu.edu.cn

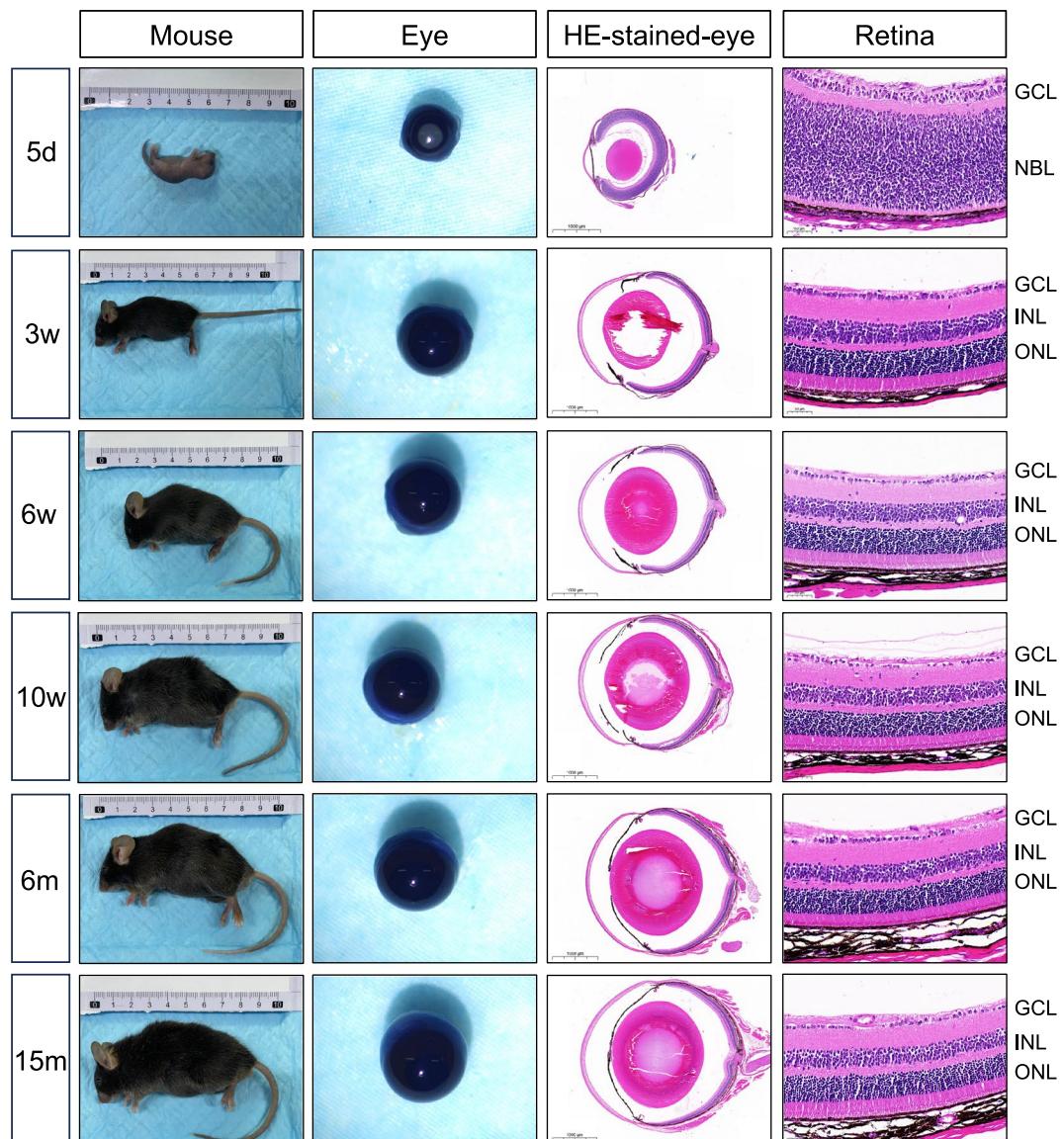


Fig. 1 Overview of six distinct stages of developing and aging in mice eyes and retinas. The body length of the mice and eye size gradually increased from CE to OP. In HE stained retinal sections, after retinal progenitor differentiation to multiple retinal cells, retina with different layers were gradually formed from CE to SP. CE: close eye period; SP: suckling period; PP: puberty period; AP: adult period; MP: middle period; OP: old period; d: days; w- weeks; HE: hematoxylin and eosin; GCL: ganglion cell layer; NBL: neuroblast layer; INL: inner nuclear layer; ONL: outer nuclear layer.

can regulate the expression of downstream genes¹⁰. In this way, lncRNAs may play crucial roles in multiple biological and pathological processes such as retinal cell differentiation, maintain retinal structure and homeostasis, retinal aging and degeneration in mouse models^{11–14}. Chen *et al.*¹⁵ revealed 2600 lncRNAs in developmental mice retina from embryonic period to adult period using whole transcriptome sequencing. Wan *et al.*^{16,17} characterized 5404 lncRNA genes and 940 intergenic lncRNAs in retinas from the embryonic day of 12.5 to the neonatal day of P28. However, our understanding of the integrated expression profiles of lncRNAs and mRNAs throughout the various stages of retinal development, particularly during the aging process, remains limited.

In this study, we utilized a total of 18 mouse retina samples spanning six distinct developmental periods, encompassing the transition from neonatal to aging. These samples were subjected to integrated lncRNA and mRNA sequencing using Illumina paired-end 150 bp (PE150) sequencing. Critical steps in the experimental pipeline, such as the evaluation of raw and clean reads, correlation and clustering of sequencing libraries by gene and transcript quantification, assembly and validation of lncRNA/mRNA transcripts, and other bioinformatic analyses, received high scores and demonstrated exceptional reliability. This transcriptome dataset represents a highly valuable resource that will significantly contribute to diverse fields of retinal research, including explorations into retinal biomarker discovery, gene therapy investigations, and the advancement of retinal organoid models.

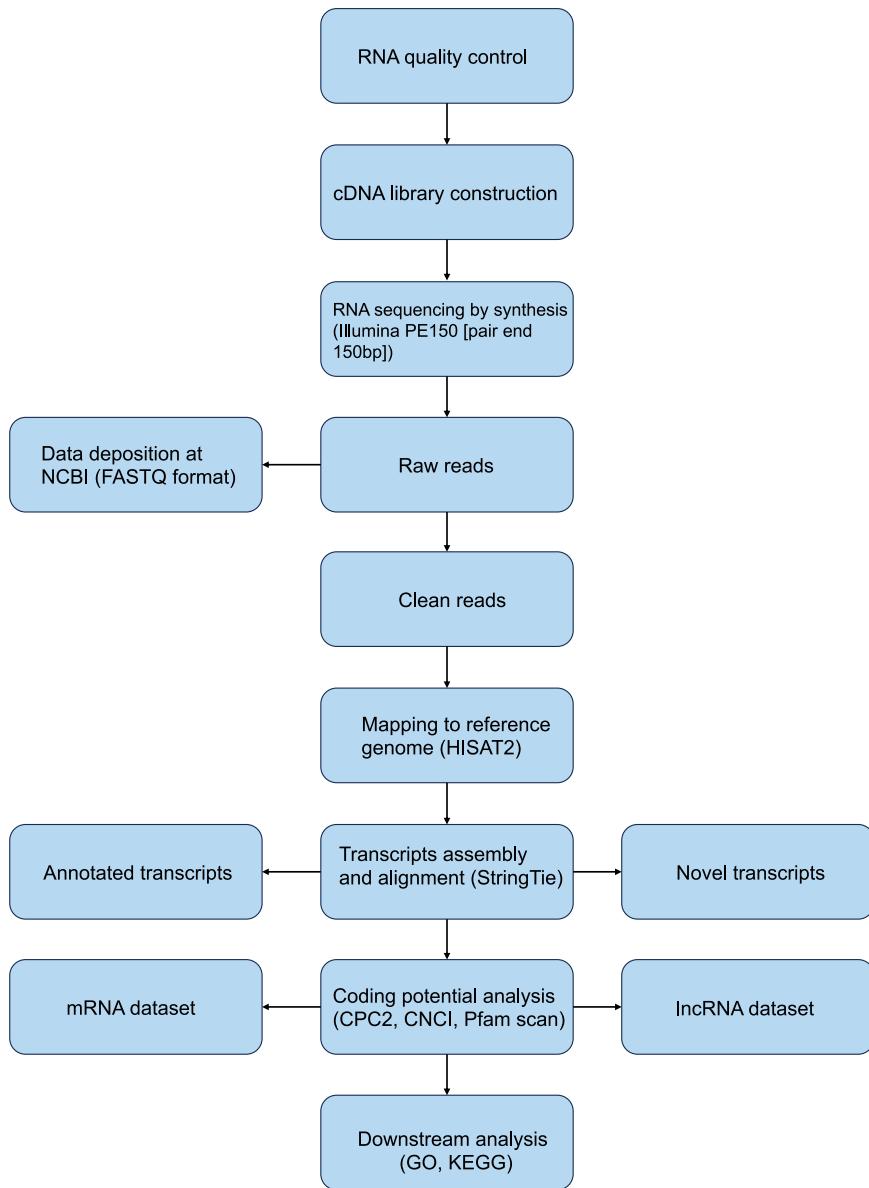


Fig. 2 Workflow of RNA sequencing and analysis.

Methods

Ethical approval. This study received approval from the ethics committee of the institutional review board at Zhongshan Ophthalmic Center. All procedures were conducted according to the ethical standards of the research committee. Male C57BL/6 mice were purchased from Jiesijie Laboratory, Shanghai, China.

The postnatal time points selected for the study, including 5 days, 3 weeks, 6 weeks, 10 weeks, 6 months, and 15 months, corresponded to distinct developmental stages, including the neonatal stage or close eye (CE) period, suckling period (SP), puberty period (PP), adult period (AP), middle period (MP), and old period (OP), respectively (Fig. 1).

Neuro-retina collection and histology. All mice were anesthetized through the intraperitoneal injection of 0.3% pentobarbital sodium, and then sacrificed for analysis. The eyes were enucleated and washed with PBS. Cornea was dissected along with limbus before removing lens and iris. Then neuro-retina was separated from RPE-choroid-sclera and washed with PBS to remove residual pigment (Supplementary Fig. 1). Neuro-retina was placed in 1.5 ml EP tubes and stored at -80°C .

The eyes were enucleated and fixed in 4% paraformaldehyde for 24 hours. After immersion in 70% alcohol, the eyes were embedded in paraffin and sectioned at 5- μm intervals in the sagittal plane. Sections through the optic papilla were collected and stained with hematoxylin and eosin (HE). HE stained sections were recorded and using CaseViewer software (v2.2). The area 250–300 μm from the optic nerve was intercepted and magnified as a typical picture of retina.

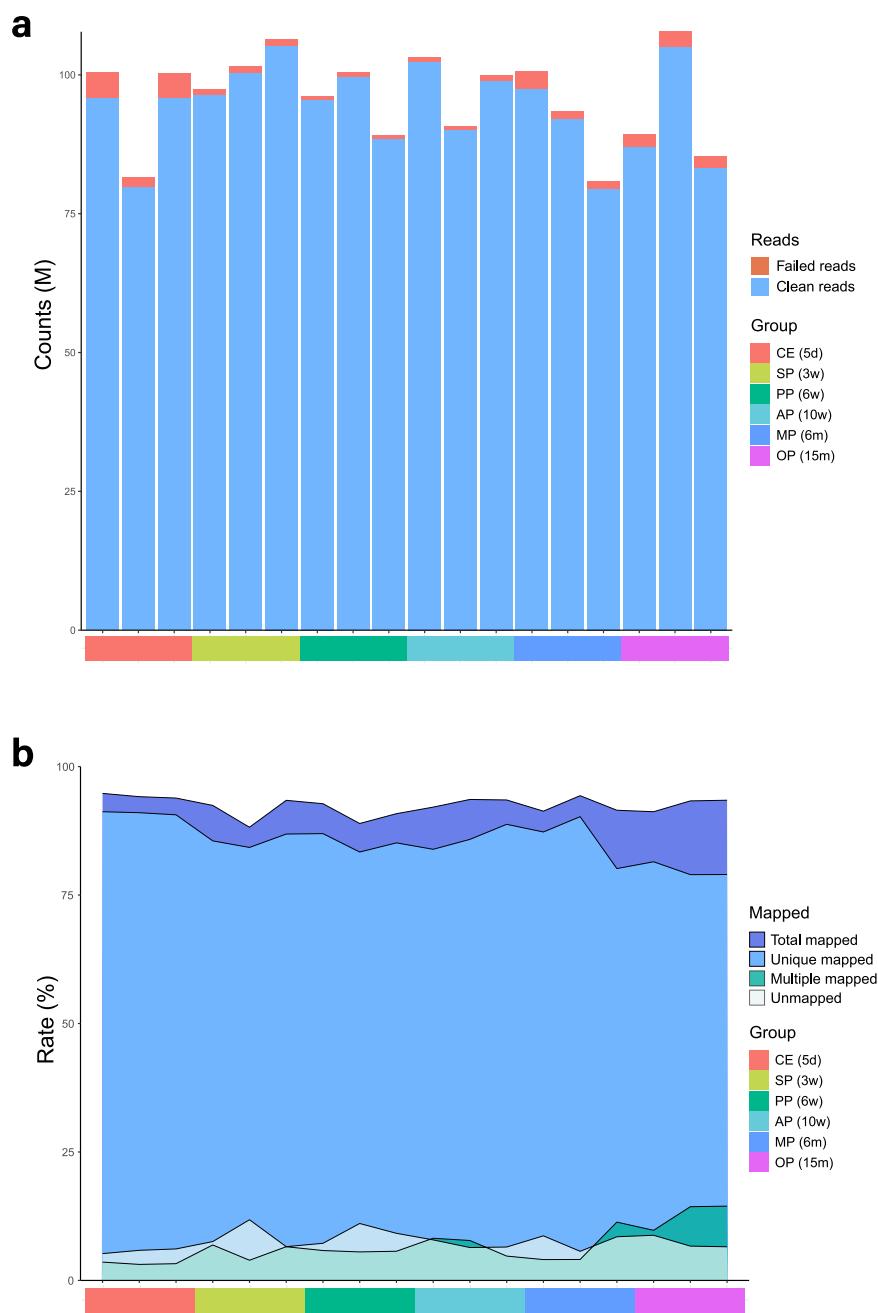


Fig. 3 Raw reads filtering and clean reads mapping. **(a)** Counts of clean reads and failed reads in raw reads of 18 samples. **(b)** Mapping rate to mouse reference of clean reads of 18 samples.

Total RNA isolation and qualification. Total RNA was extracted from retina samples by using TRIzol reagent (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions, including treatment with DNase. Bilateral retinas from one mouse were pooled as a sample in all groups. RNA degradation and contamination was monitored on 1% agarose gels with 180 V for 16 min. RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA), and samples with an optical density 260/280 and 260/230 ratio between 1.8 and 2.1 were used. RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA) to calculate an RNA Integrity Number (RIN). Only RNA samples with a RIN > 8 were used for sequencing analysis. A total amount of 3 μ g RNA per sample was used as input material for the RNA sequencing library preparations.

Strand-specific library construction and qualification. The workflow of RNA sequencing was shown (Fig. 2). After removing ribosomal RNA (rRNA), total RNA was broken into short fragments of 250–300 bp. The fragmented RNA was used as a template and random oligonucleotides were used as primers to synthesize the first strand of cDNA, then RNA templates were degraded with RNase H. The second strand of cDNA was synthesized in DNA polymerase I system with the first strand of cDNA as templates and dNTPs (N = A, C, G, U)

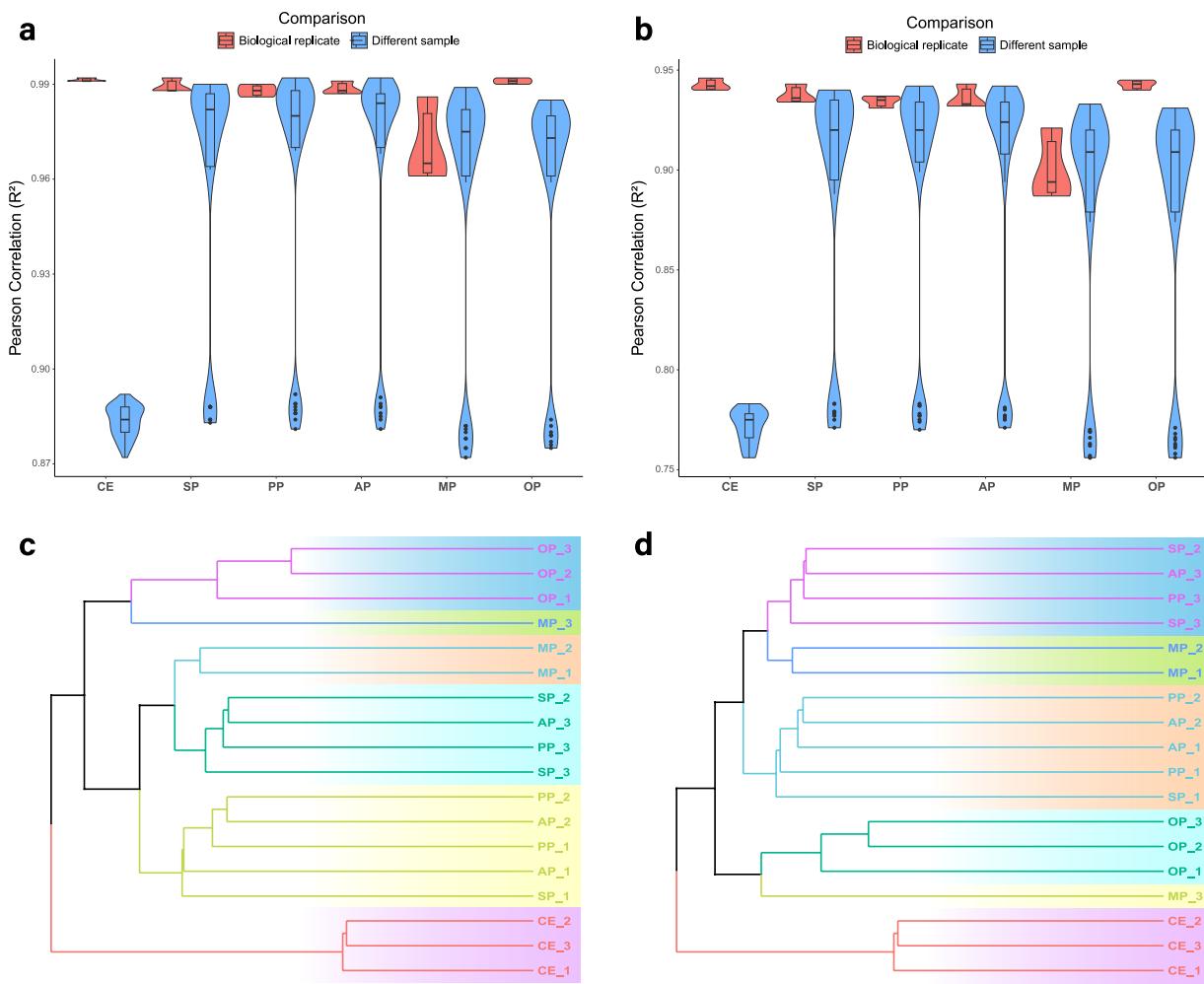


Fig. 4 Pearson correlation analysis and cluster analysis between samples in gene and transcript level. Pearson correlation analysis between biological replicates both got high scores (a) in gene level and (b) in transcript level, CE group showed great difference from other samples. Cluster analysis showed that CE and OP groups were independent from other groups (c) in gene level and (d) in transcript level. CE: close eye period; SP: suckling period; PP: puberty period; AP: adult period; MP: middle period; OP: old period.

as raw materials. The double-stranded cDNA got through end repair, adding A-tail and adaptors, and the cDNAs of about 200 bp were selected using AMPure XP beads (Beckman Coulter, Beverly, USA). The second strand of the cDNAs containing dUTPs were digested using 3 μ l USER enzyme (NEB, USA) at 37 °C for 15 min followed by 5 min at 95 °C, and finally PCR was performed to amplify and construct the library.

After library construction, initial quantification was performed using Qubit 2.0 and dilute the library to 1.5 ng/ μ l, followed by detection of the insert size of the library using Agilent 2100. Then the effective concentration of the library (>3 nM) was accurately quantified using qRT-PCR.

Illumina sequencing. After passing the library qualification, the library was pooled and sequenced by Illumina PE150 (paired-end 150 bp), which referred to high-throughput paired-end sequencing of 150 bp at each end, and the insert cDNA was the unit directly sequenced in the constructed library. The basic principle of sequencing was sequencing by synthesis. Four fluorescently labeled dNTPs ($N = A, C, G, T$), DNA polymerase and splice primers were added to the sequencing flow cell for amplification. The sequence information of the fragments was obtained by capturing the fluorescence signal and converting into sequencing peaks.

Reads filtering and mapping. The raw image files are converted into sequenced reads by CASAVA (v1.8) base identification and stored in FASTQ format. Raw data (raw reads) of FASTQ format were firstly processed through in-house perl scripts. During this process, clean data (clean reads) were obtained after removing low-quality reads ($>50\%$ of bases with $sQ < 5$) and reads containing adapter or ploy-N ($>10\%$ of uncertain base) from raw data. Reference genome and gene model annotation files were downloaded from genome website directly. Index of the reference genome was built and clean reads were aligned to the reference genome using HISAT2 (v2.0.5)¹⁸, which might generate a precise mapping result for junction reads.

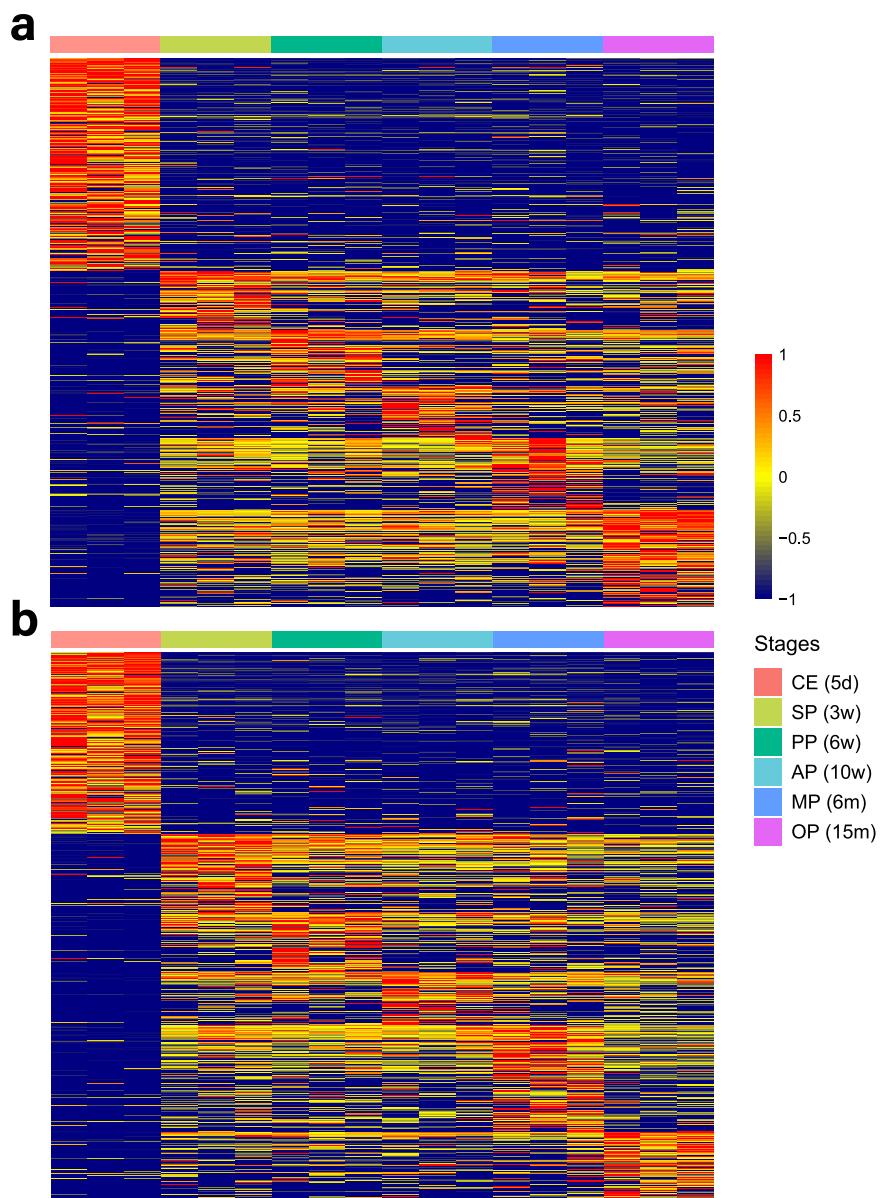


Fig. 5 Heatmaps of differentially expressed genes in 18 samples. Heatmaps showed the overview of quantification of (a) differentially expressed mRNA genes and (b) differentially expressed lncRNA genes in 18 samples.

Transcripts assembly and classification. Transcripts were assembled using StringTie (v1.3.1)¹⁹. After assembly, transcripts with length >200 bp and exon number ≥ 2 were selected. Then, these transcripts were aligned to annotated database using Cuffcompare (v2.2.1)²⁰ and classified into annotated transcripts and novel transcripts. The mainstream coding potential analysis methods (CPC2, v3.2.0; Pfam scan, v1.3; CNCI) were integrated to performed the screening of coding potential. The next selection and naming of candidate novel lncRNA was referred to the HGNC (The HUGO Gene Nomenclature Committee)²¹ guideline criteria. In order to further validate the accuracy of our transcript classification approach, we compared the transcript length, exon number, and open reading frame length of the novel lncRNAs with those of mRNAs. Following the aforementioned procedures, the transcripts were categorized into annotated mRNAs, annotated lncRNAs, novel mRNAs, and novel lncRNAs.

LncRNA/mRNA transcript quantification. We utilized StringTie (v1.3.1) for quantifying transcript-level expression. StringTie employs advanced algorithms to effectively reconstruct transcript structures and estimate their abundance using RNA-seq reads aligned to a reference genome. We inputted spliced alignments in coordinate-sorted BAM file and obtained a GTF output containing assembled transcript structures, along with their corresponding expression levels reported as FPKM (Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced). These FPKM values were utilized for downstream analyses.

Article	Time points	Methods	Platform	No. of lncRNAs	Notes
Wan <i>et al.</i> BMC Genomics. 2019 Jul 8;20(1):559.	E12.5, E14.5, E16.5, E18.5, P0, P3, P5, P7, P14, P21 and P28	Full-length transcript sequencing	PacBio Sequel System	5404 lncRNAs (940 intergenic lncRNAs)	/
Chen <i>et al.</i> BMC Genomics. 2021 Oct 30;22(1):779.	E14.5, P1, P7, P12, P17, P56	Whole-transcriptome (lncRNAs, mRNAs, miRNAs and mRNAs) sequencing	Illumina HiSeq PE150	2600 lncRNAs, 13647 mRNAs, 27319 circRNAs, and 704 miRNAs	/
Yu <i>et al.</i> BMC Genomics. 2023 May 10;24(1):252.	E12.5, E13.5, E14.5, E15.5, E16.5, E17.5, E18.5, P0, P1, P3, P5, P7, P14, P21 and P28	Short-read sequencing	Illumina HiSeq PE150	4523 lncRNAs	Combined with secondary analysis of Wan <i>et al.</i> BMC Genomics. 2019 Jul 8;20(1):559.
Chen <i>et al.</i> Invest Ophthalmol Vis Sci. 2017 Dec 1;58(14):6308–6317.	P0, P56	LncRNA microarray	Affymetrix Mouse Transcriptome Array (MTA 1.0)	910 lncRNAs (P0) and 616 lncRNAs (P56)	Dataset encompassing six ocular tissues (cornea, lens, retina, RPE, choroid, and sclera)
This study	P5, P21, P42, P70, P180, P450	Integrated lncRNA and mRNA sequencing	Illumina HiSeq PE150	11222 lncRNAs and 18809 mRNAs	/

Table 1. Comparison of lncRNA datasets of developing C57 mice retinas.

Differential expression analysis. Differential expression analysis between any two groups was performed using the EdgeR R package (v3.12.1)²². EdgeR was used to adjust read counts prior through one scaling normalized factor for each sequenced library. This package utilizes a statistical framework based on the negative binomial distribution to determine differential expression. The resulting P-values were adjusted using the Benjamini and Hochberg's method to control the false discovery rate. Genes and transcripts with an adjusted P-value < 0.05, as determined by EdgeR, were considered as differentially expressed.

GO and KEGG enrichment analysis. We utilized the clusterProfiler R package (v4.6.2)²³ to conduct statistical enrichment analysis of differentially expressed genes. This analysis included Gene Ontology (GO) enrichment analysis²⁴ and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis²⁵. GO enrichment analysis provided insights into the biological processes, cellular components, and molecular functions associated with the differentially expressed genes. KEGG pathway analysis was employed to identify relevant signaling pathways. GO terms with a corrected P-value below 0.05 were considered significantly enriched by the differentially expressed genes.

Quantitative real-time PCR (qRT-PCR). After isolation of total RNA, complementary DNA was synthesized with Evo M-MLV RT Kit (AG11707, Accurate Biotechnology, Hunan, China) and amplified with SYBR® Green Premix Pro Taq HS qPCR Kit (AG11701, Accurate Biotechnology) on CFX96 real-time PCR system (Bio-Rad), using GAPDH as an internal reference. Then, relative expression was calculated using the $2^{-\Delta\Delta CT}$ formula. Primers of the selected lncRNAs and mRNAs were synthesized by Sangon Biotech, Shanghai, China, and the sequences were as follows:

lncRNA-otx2os1-F: 5'-GCAACTCTGTCGCTTGTG-3';
 lncRNA-otx2os1-R: 5'-CCCTAGACGTCTGCAAAGCA-3';
 GAPDH-F: 5'-TGTGTCGTCGTGGATCTGA-3';
 GAPDH-R: 5'-TTGCTGTTGAAGTCGCAGGAG-3'.

Fluorescence *in situ* hybridization (FISH). FISH staining was completed according to the instruction of RNA FISH Detection Reagent kit (Servicebio, Wuhan, China) as previously described²⁶, and finally photographed using a confocal microscopy (Leica, Wetzlar, Germany). Sequencing of lncRNA-otx2os1 probe mix were as follows:

5'-GTATCACGAGCAAAGACAAGCCCTG-3';
 5'-AAGAGCAATTTGCAACTTTCCAG-3';
 5'-CGGACAGAGTTGCTTATTCTCAGGG-3';
 5'-ACACATCCTGAGCCCCTAGACGTCT-3';
 5'-GAGTGTCTTTGCAGGGCACATAA-3'

Data Records

Raw reads of all samples were deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) as FASTQ files with SRP accession number, SRP445437²⁷. The files of reference mapping, assembly and lncRNADfilter, quantification, differential expression analysis, GO and KEGG enrichment analysis were also deposited in figshare²⁸ (<https://doi.org/10.6084/m9.figshare.23614980.v1>).

Technical Validation

RNA-Seq raw data quality and filtering. Quality control and reads statistics were shown (Supplementary Table 1). Clean reads/ raw reads (mean \pm SD) were $98.17\% \pm 1.17\%$, error rate was $0.013\% \pm 0.005\%$ (Fig. 3a). Raw_bases(G) and Clean_bases(G) were 14.37 ± 1.17 and 14.11 ± 1.16 . Q20, Q30 and GC content of the clean data were $97.21 \pm 0.53\%$, $93.04 \pm 1.15\%$ and $47.66 \pm 1.50\%$. The normally distributed GC content indicated the sequencing data were not contaminated.

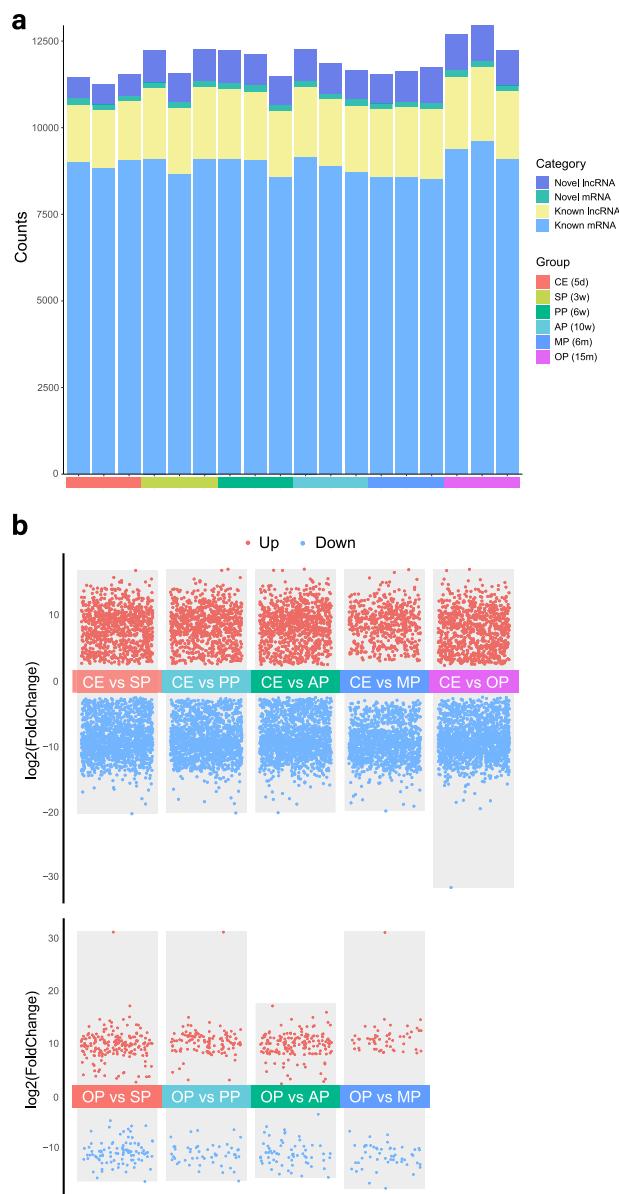
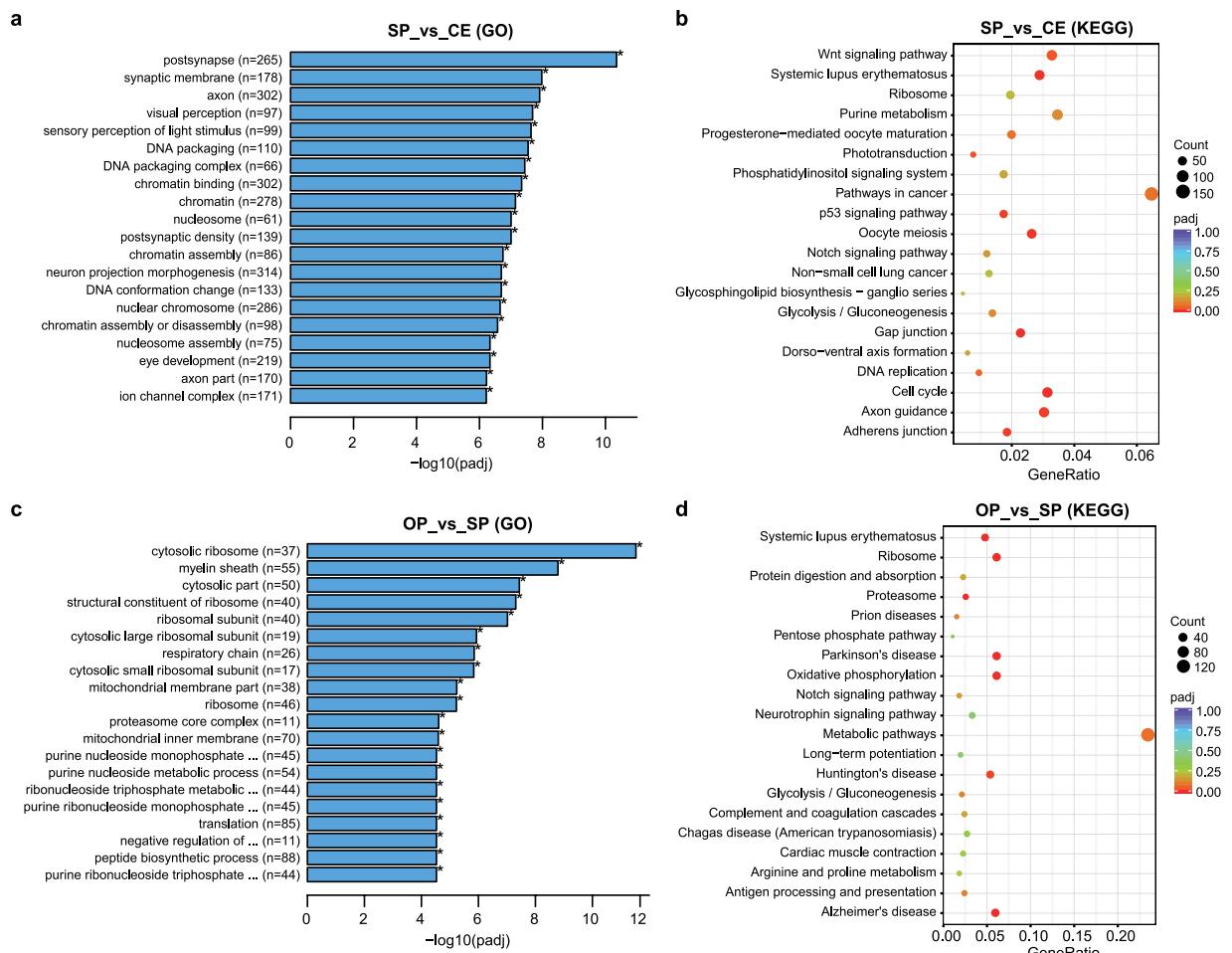


Fig. 6 Differentially expressed transcripts between two groups. **(a)** Differentially expressed transcripts in each sample were classified as novel lncRNA, novel mRNA, known lncRNA and known mRNA. **(b)** Vocalno plots showed that numbers of differentially expressed lncRNA transcripts in CE vs other groups were larger than that in OP vs other groups.

Clean reads mapping. The clean reads were aligned to the mouse genome (GRCm38) provided by the Genome Reference Consortium. The total mapped reads for all samples exceeded 90%, except for PP2 and SP2, which were 88.92% and 88.19% respectively. Uniquely mapped reads for all samples were over 80%, except for OP2 and OP3, which were 78.97% and 78.99% respectively. The percentage of reads mapped to exonic or intronic regions of the genome exceeded 90% for all samples, except for OP1 and MP3, which were 88.82% and 83.02% respectively (Fig. 3b, Supplementary Table 2).

Reproducibility of biological replicates. To evaluate the reproducibility of biological replicates, we computed Pearson's correlation coefficients for FPKM expression at both the gene and transcript levels among all pairwise combinations of the 18 samples. At the gene level, all correlation scores between samples within the same developmental period group exceeded 0.96 (Fig. 4a). At the transcript level, all correlation scores between samples within the same developmental period group were greater than 0.93, except for MP3 vs MP1 and MP3 vs MP2, which were 0.894 and 0.887, respectively (Fig. 4b). Cluster analysis of FPKM expression at the gene and transcript levels revealed that the CE and OP groups formed distinct clusters independent from other groups at the gene level (Fig. 4c) and at the transcript level (Fig. 4d). The SP, PP, AP, and MP groups clustered together due to their relatively high similarity in sample profiles.



LncRNA and mRNA transcriptome. In gene-level expression analysis, we detected a total of 11,222 lncRNA genes (Fig. 5a) and 18,809 mRNA genes (Fig. 5b) across the 18 samples, which showed better sequencing coverage and larger number of candidate lncRNAs than previous studies (Table 1). At the transcript level, we identified 4,589 lncRNA transcripts and 14,471 mRNA transcripts. Upon aligning these transcripts to the annotation database, we discovered 1,527 novel lncRNAs and 288 novel mRNAs (Fig. 6a). In terms of gene-level analysis, we identified 8,124 (SP vs CE), 80 (PP vs SP), 52 (AP vs PP), 49 (MP vs AP), and 137 (OP vs MP) differentially expressed lncRNAs between adjacent age periods. The numbers of differentially expressed lncRNAs between CE and each of the other groups were 7,470 (PP vs CE), 8,015 (AP vs CE), 5,226 (MP vs CE), and 8,352 (OP vs CE). Additionally, there were 2,152 (OP vs SP), 859 (OP vs PP), and 1,354 (OP vs AP) differentially expressed lncRNAs when comparing OP with the other groups (Fig. 6b). LncRNAs that showed differential expression profiles in one period compared to the other five periods were defined as period-specific lncRNAs. We identified 4,278 period-specific lncRNAs in CE, 8 in SP, 3 in PP, 1 in AP, 5 in MP, and 49 in OP, respectively. Collectively, there were more differentially expressed lncRNAs in the CE and OP groups compared to the other groups, suggesting a crucial role of lncRNAs in retinal development and aging.

In gene-level analysis, we found 13,418 (SP vs CE), 129 (PP vs SP), 71 (AP vs PP), 67 (MP vs AP), and 235 (OP vs MP) differentially expressed mRNAs between adjacent age periods. The numbers of differentially expressed mRNAs between CE and each of the other groups were 12,245 (PP vs CE), 13,272 (AP vs CE), 8,525 (MP vs CE), and 13,988 (OP vs CE). Additionally, there were 3,603 (OP vs SP), 1,403 (OP vs PP), and 2,281 (OP vs AP) differentially expressed mRNAs when comparing OP with the other groups. We identified 6,935 period-specific mRNAs in CE, 10 in SP, 3 in PP, 3 in AP, 4 in MP, and 94 in OP, respectively.

Functional annotation database enrichment analysis. GO analysis revealed that the top 20 enriched GO terms in the SP vs CE comparison could be classified into three main categories. The first category included

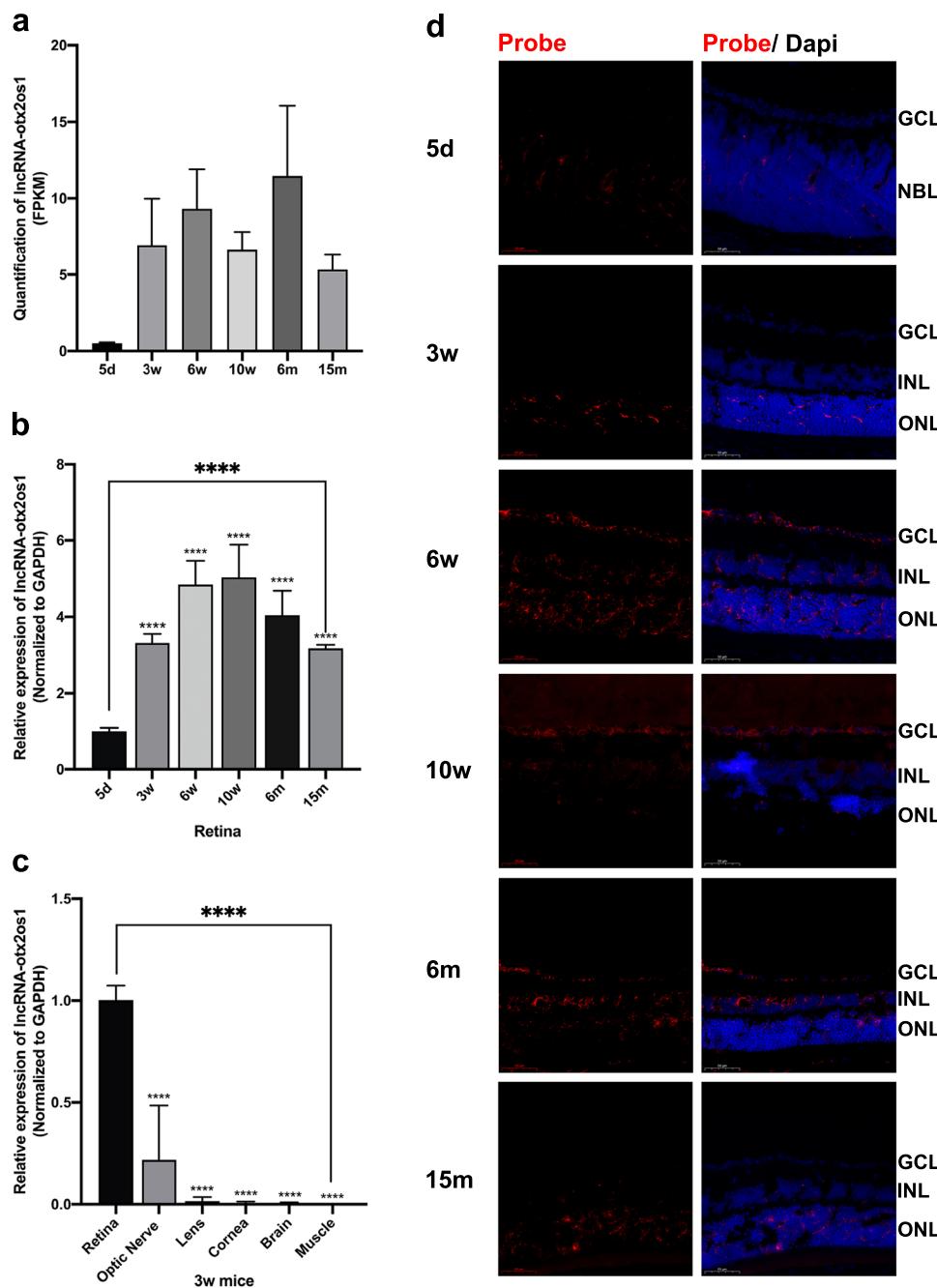


Fig. 8 Laboratory validation of lncRNA-otx2os1 with potential of regulating retinal differentiation. (a) RNA sequencing and (b) qRT-PCR (ANOVA, $n=6$, *** $P < 0.001$) showed similar expression patterns of lncRNA-otx2os1 in retina in six periods. (c) Results of qRT-PCR showed lncRNA-otx2os1 highly expressed in retina compared to other five tissues. (d) Fluorescence *in situ* hybridization showed that fluorescence intensity of lncRNA-otx2os1 corresponded with quantification in RNA sequencing. LncRNA-otx2os1 mainly located in neuroblast layer or outer nuclear layer, and increased in inner nuclear layer and retinal ganglion cell layer after differentiation and development. CE: close eye period; SP: suckling period; PP: puberty period; AP: adult period; MP: middle period; OP: old period; d: days; w- weeks; GCL: ganglion cell layer; NBL: neuroblast layer; INL: inner nuclear layer; ONL: outer nuclear layer.

terms related to light perception, such as visual perception, sensory perception of light stimulus, and eye development. The second category involved retinal neuron structural development, including terms like postsynapse, synaptic membrane, axon, postsynaptic density, neuron projection morphogenesis, and axon part. The third category encompassed DNA and chromatin events, including terms such as DNA packaging, DNA packaging complex, chromatin binding, chromatin, and nucleosome (Fig. 7a). KEGG pathway analysis demonstrated that the phototransduction pathway and axon guidance pathway were significantly enriched in the comparison between the SP and CE groups (Fig. 7b). These findings indicated that the epigenetic differences observed in lncRNAs and

mRNAs between SP and CE were associated with phenotypic changes, including eye development, light perception, and retinal differentiation. The GO terms enriched in the comparison between the OP and SP groups were mainly related to ribosome, mitochondrial activities, and purine metabolism (Fig. 7c). KEGG pathway analysis revealed that the host genes in this comparison were primarily associated with Parkinson's disease, Alzheimer's disease, metabolic pathways, and oxidative phosphorylation (Fig. 7d). Additionally, KEGG analysis of OP vs PP, OP vs AP, and OP vs MP comparisons showed that the most significantly enriched pathways included oxidative phosphorylation as well as Parkinson's and Alzheimer's diseases. These results indicate that differentially expressed lncRNAs and mRNAs play a crucial role in the aging process of the retina and central nervous system. Furthermore, the CE and OP groups exhibited a more abundant transcriptomic information of lncRNAs and mRNAs compared to the other groups.

Experimental validation of lncRNA-otx2os1. After conducting a literature review and performing functional prediction, we selected lncRNA-otx2os1 (ENSMUST00000183522.8), one of the differentially expressed lncRNAs in the SP vs CE comparison, for experimental validation. Previous studies have shown that the host coding gene of lncRNA-otx2os1, Otx2, is involved in eye development and retinal differentiation^{29–31}. The RNA sequencing results revealed that the expression of lncRNA-otx2os1 increased in the retina after eye opening, reached its peak during the adult/middle period, and decreased during the aging period (Fig. 8a). The qRT-PCR results further confirmed the developmental expression pattern of lncRNA-otx2os1 (Fig. 8b). Moreover, lncRNA-otx2os1 exhibited high tissue specificity in the retina compared to the optic nerve, lens, cornea, brain and muscle (Fig. 8c). The FISH results demonstrated that the fluorescence intensity of lncRNA-otx2os1 corresponded to its quantification in RNA sequencing. LncRNA-otx2os1 was predominantly located in the outer nuclear layer (ONL) and showed increased expression in the inner nuclear layer (INL) and RGC layer during differentiation and development (Fig. 8d). These experimental validation results of lncRNA-otx2os1 confirmed the reliability of the RNA sequencing data and served as an illustrative example of experimental validation of key lncRNAs.

Usage Notes

This transcriptome dataset comprises comprehensive developmental and aging periods of mouse retinas, obtained from 18 samples with 3 biological replicates. The data generation and analysis processes were meticulously conducted, ensuring the high quality and reliability of the sequencing data. Notably, the CE and OP periods exhibited a greater abundance of lncRNA and mRNA transcriptomic information compared to other periods, reflecting retinal differentiation and aging, respectively. The utilization of this transcriptome dataset can unveil extensive gene and transcript expression information associated with retinal development, differentiation, homeostasis, and aging, thereby enhancing our understanding of physiological and pathological processes in the retina. Additionally, this dataset of normal mouse retinas can serve as a control group for mouse models of retinal diseases, presenting new targets for various retinal diseases, particularly age-related macular degeneration and retinitis pigmentosa. Furthermore, this bulk RNA sequencing dataset can serve as a reference for integrated analysis of transcriptional profiles in retinal single-cell sequencing, enabling the identification of cell-specific lncRNAs and mRNAs. Compared with previous studies of mouse retinal lncRNA sequencing with the oldest time point of 2 months, 15 months were the older time point which could capture aging efficiently (Table 1), but this dataset might be more valuable to add sequencing data at 21 months or older in future studies.

All raw RNA sequencing data are stored in FASTQ files, named as sample-1 or 2-fq.gz. The raw data for GO analysis, KEGG analysis, and differential expression analysis are stored in XLS files, which can be efficiently explored and analyzed using Microsoft Excel software.

Code availability

The following open access software were used for quality control and data analysis as described in the main text: CASAVA, version 1.8.2, was used to convert raw image files into sequenced reads.

HISAT2, version 2.0.5, was used to map reads to the reference genome. (<http://daehwankimlab.github.io/hisat2/>).

StringTie, version 1.3.1, was used to assemble transcripts and calculate the numbers of transcripts mapped to the transcripts of database. (<https://www.cnblogs.com/raisok/p/11046403.html>).

Cuffcompare, version 2.2.1, was used to align transcripts to annotated database.

CNCI; CPC2, version 3.2.0; Pfam scan, version 1.3, were used to predict coding potential.

EdgeR R package, version 3.12.1, was used to perform differential expression analysis of two conditions/groups.

ClusterProfiler, version 4.6.2, was used to perform GO and KEGG pathway analysis. (<http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html>).

Received: 4 July 2023; Accepted: 13 September 2023;

Published online: 23 September 2023

References

1. Ramkumar, H. L., Zhang, J. & Chan, C. C. Retinal ultrastructure of murine models of dry age-related macular degeneration (AMD). *Prog Retin Eye Res* **29**, 169–190 (2010).
2. Sharma, R. K., O'Leary, T. E., Fields, C. M. & Johnson, D. A. Development of the outer retina in the mouse. *Brain Res Dev Brain Res* **145**, 93–105 (2003).
3. Dyer, M. A., Livesey, F. J., Cepko, C. L. & Oliver, G. Prox1 function controls progenitor cell proliferation and horizontal cell genesis in the mammalian retina. *Nat Genet* **34**, 53–58 (2003).
4. McKeown, A. S. & Kraft, T. W. Adaptive potentiation in rod photoreceptors after light exposure. *J Gen Physiol* **143**, 733–743 (2014).
5. Al-Khindi, T. *et al.* The transcription factor Tbx5 regulates direction-selective retinal ganglion cell development and image stabilization. *Curr Biol* **32**, 4286–4298.e4285 (2022).

6. de Melo, J., Peng, G. H., Chen, S. & Blackshaw, S. The Spalt family transcription factor Sall3 regulates the development of cone photoreceptors and retinal horizontal interneurons. *Development* **138**, 2325–2336 (2011).
7. Groman-Lupa, S., Adewumi, J., Park, K. U. & Brzezinski, J. I. The Transcription Factor Prdm16 Marks a Single Retinal Ganglion Cell Subtype in the Mouse Retina. *Invest Ophthalmol Vis Sci* **58**, 5421–5433 (2017).
8. Bao, H. *et al.* Biomarkers of aging. *Sci China Life Sci* **66**, 893–1066 (2023).
9. Kopp, F. & Mendell, J. T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **172**, 393–407 (2018).
10. Yang, Z. *et al.* LncRNA: Shedding light on mechanisms and opportunities in fibrosis and aging. *Ageing Res Rev* **52**, 17–31 (2019).
11. Hu, X. *et al.* LncRNA NEAT1 Recruits SFPQ to Regulate MITF Splicing and Control RPE Cell Proliferation. *Invest Ophthalmol Vis Sci* **62**, 18 (2021).
12. Yan, B. *et al.* lncRNA-MIAT regulates microvascular dysfunction by functioning as a competing endogenous RNA. *Circ Res* **116**, 1143–1156 (2015).
13. Zhu, W. *et al.* Identification of lncRNAs involved in biological regulation in early age-related macular degeneration. *Int J Nanomedicine* **12**, 7589–7602 (2017).
14. Ghanam, A. R. *et al.* Alternative transcribed 3' isoform of long non-coding RNA Malat1 inhibits mouse retinal oxidative stress. *iScience* **26**, 105740 (2023).
15. Chen, G. *et al.* Whole transcriptome sequencing identifies key circRNAs, lncRNAs, and miRNAs regulating neurogenesis in developing mouse retina. *BMC Genomics* **22**, 779 (2021).
16. Yu, D. *et al.* The landscape of the long non-coding RNAs in developing mouse retinas. *BMC Genomics* **24**, 252 (2023).
17. Wan, Y. *et al.* Systematic identification of intergenic long-noncoding RNAs in mouse retinas using full-length isoform sequencing. *BMC Genomics* **20**, 559 (2019).
18. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357–360 (2015).
19. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**, 1650–1667 (2016).
20. Ghosh, S. & Chan, C. K. Analysis of RNA-Seq Data Using TopHat and Cufflinks. *Methods Mol Biol* **1374**, 339–361 (2016).
21. Wright, M. W. A short guide to long non-coding RNA gene nomenclature. *Hum Genomics* **8**, 7 (2014).
22. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
23. Yu, G., Wang, L. G., Han, Y. & He, Q. clusterProfiler: an R package for comparing biological themes among gene clusters. *Oomics* **16**, 284–287 (2012).
24. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**, R14 (2010).
25. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
26. Jiang, R. *et al.* The sodium new houttuynonate suppresses NSCLC via activating pyroptosis through TCONS-14036/miR-1228-5p/PRKCDP pathway. *Cell Prolif*, e13402 (2023).
27. NCBI Sequence Read Archive. <https://identifiers.org/ncbi/insdc.sra:SRP445437> (2023).
28. Kong, K. *et al.* Integrated Transcriptome Analysis of Long Noncoding RNA and mRNA in Developing and Aging Mouse Retina. *figshare*. <https://doi.org/10.6084/m9.figshare.23614980.v1> (2023).
29. An, M. J. *et al.* c-Jun N-terminal kinase 1 (JNK1) phosphorylates OTX2 transcription factor that regulates early retinal development. *Genes Genomics* **45**, 429–435 (2023).
30. Diacou, R. *et al.* Cell fate decisions, transcription factors and signaling during early retinal development. *Prog Retin Eye Res* **91**, 101093 (2022).
31. Chew, S. H., Martinez, C., Chirco, K. R., Kandoi, S. & Lamba, D. A. Timed Notch Inhibition Drives Photoreceptor Fate Specification in Human Retinal Organoids. *Invest Ophthalmol Vis Sci* **63**, 12 (2022).

Acknowledgements

This article was supported by the National Key Research and Development Program of China (2022YFC2502800); the National Natural Science Foundation of China (82070955); the Natural Science Foundation of Guangdong province of China (2020A1515011282); the Science and Technology Program of Guangzhou, China (2024A03J00515); the High-level Hospital Construction Project, Zhongshan Ophthalmic Center, Sun Yat-sen University (303020104). The funding organizations had no role in the design or conduct of this article.

Author contributions

S.C., X.Z. and Z.S. conceived the research. RNA-seq libraries were constructed by K.K., P.W., Z.X. and L.W., Sequencing data acquisition, data management and scientific support was performed and overseen by K.K., P.W., Z.X., L.W., J.J., Y.L., S.D., J.J., Y.S., F.L. and X.F., Data analysis and interpretation were performed by K.K., P.W., J.J., X.F., J.J., Y.S., W.W. and F.L., The manuscript was written and revised by K.K., Z.S., S.C. and X.Z., All authors reviewed and approved the submitted 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/s41597-023-02562-9>.

Correspondence and requests for materials should be addressed to Z.S., X.Z. or S.C.

Reprints and permissions information is available at 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 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 licence, and indicate if changes were made. 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/4.0/>.

© The Author(s) 2023