nature portfolio

Corresponding author(s): Seth Blackshaw
David Hyde
Jiang Qian

Last updated by author(s): 2023-9-4

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Single-cell RNA and ATAC libraries were prepared according to the 10xMultiome ATAC + Gene Expression protocol and subjected for Illumina NovaSeq sequencing at ~500 million reads per library. The multi-omic sequencing files were processed for demultiplexing and analyzed using Cell Ranger ARC (version 2.0.2) The genes were mapped and referenced using the zebrafish reference genome DanRer11. To address potential ambient RNA contamination in the cell-by-gene expression matrix, we employed scAR module from scvi-tools (version 0.19.0) and SoupX (version 1.6.2) for further ambient RNA cleaning for each sample.
Data analysis	All test datasets and custom scripts are available in: Zenodo DOI: 10.5281/zenodo.8317611 and Github: https://github.com/Pinlyu3/Zebrafish-retina-GRNs We analysis scRNA-seq, snRNA-seq and scATAC-seq with existing softwares: For scRNA-seq and snRNA-seq datasets, we utilized the Scrublet package (version 0.2.3) and the solo module from scvi-tools (version 0.19.0) to identify doublet cells. The Seurat software (version 4.0.6) was employed for clustering, visualization, cell type identification, sample integration, and differential gene analysis. Additionally, we utilized the Slingshot package (version 2.2.1) for constructing trajectories. For scATAC-seq datasets, the ArchR package (version 1.0.2) was used to perform dimensional reduction, clustering, and differential peaks analysis. The chromVAR package (version 1.16.0) was used to calculate motif activities. The differential motif activities were calculated by the Signac package (version 1.9.0). To identify significantly enriched Gene Ontology (GO) terms and KEGG terms among the Differentially Expressed Genes (DEGs) between biological conditions, the gene set enrichment analysis was performed using the "clusterProfiler" package (version 4.9.0). We construct gene regulatory networks with existing software and custom codes:

1 Inferring activators and repressors by expression and motif activity

For each TF-motif pair in each datasets (LD,NMDA,Injury,Development), custom scripts were used to calculate the correlation between TF expression and motif activities (chomVAR score) at single cell level in R. Activator and repressor TF-motif pairs were identified if their correlation are large than 0.05 or less than -0.05.

2 Identifying Cis-regulatory elements

Firstly, the categorization of all peaks into three groups based on their genomic location relative to gene loci using the custom script: 1) Promoter (within 500bp of TSS), 2) Gene Body, and 3) Intergenic. Subsequently, peak-target pairs are generated using the following methods: 1) The target genes for Promoter and Gene Body peaks are determined by the genes they overlap with. 2) The target genes for Intergenic peaks encompass all genes located within 200kb of the peak's location. Next, PtoG correlations for each peak with its surrounding genes (200kb) are calculated using the "addPeak2GeneLinks" function in the ArchR (version 1.0.2) package. Finally, the retention of the peak-target pairs is based on meeting the following criteria for their PtoG correlations: abs(correlation) > 0.25 and FDR < 0.01.

3 Predicting TF Binding Sites

The TF-peak pairs were constructed by predicting TF binding sites inferred based on motif information and scATACseq footprint signals within the identified cis-regulatory elements using custom scripts.

Initially, Position Weight Matrices (PWMs) were extracted from the TRANSFAC2018 and CIS-BP databases. The binding regions were then identified by matching these motifs to the DNA sequences of the cis-regulatory elements using the motifmatchr package (version 1.16.0). Subsequently, the scATACseq corrected footprint signals were separately calculated for the Light-damage Injury, NMDA injury, combined injury, and Development datasets. These merged fragments were converted to BAM format and processed through the TOBIAS (version 0.12.0) to obtain bias-corrected Tn5 signals. For each binding region of the motif, footprint scores were computed. Finally, the TF binding sites were retained based on the following criteria: NR+NL-2*NC > 0.1. Additionally, the binding regions for motifs whose corresponding TFs were not expressed in the MG cells were removed.

4 TF-target correlation

The TF-target relationship was calculated using the arboreto (version 0.1.6) in Python. The TF-gene pairs were filtered based on their important scores, with pairs that had scores lower than the 95th quantile being removed. Additionally, the Pearson correlation between each TF-gene pair was computed according to the cell-by-gene expression matrix. If the correlation exceeded 0.03, the TF-gene pair was annotated as "positive" regulation; if the correlation was below -0.03, it was annotated as "negative" regulation. Any other TF-gene pairs were filtered out.

5.Construction of TF-peak-target links

Custom scripts were used to construct GRNs by integrating data from the previous steps. The following procedure was employed: The TF-peak pairs from step 3 and the peak-target pairs from step 2 were merged to form TF-peak-target triples. Subsequently, these TF-peak-target triples were filtered using the following criteria: 1) The triples were retained only if TF activity are in the same direction with TF-gene correlation. (Activtor with positive TF-gene correlation, Repressor with negative TF-gene correlation) 2) The triples were retained only if the TF's expression levels are enriched in MG cell groups (MG, ActMG, MGPCs). 3) Any duplicate triples were eliminated, and we retained the highest footprint score for each TF-peak-target pair.

6. Identification of enriched gene regulatory sub-networks

The enriched sub-GRNs were extracted from the total GRNs generated in step 5, based on the logFC change of TFs, peaks, and target genes (as shown in Figure 4E and Figure 6E) using custom scripts.

7. Identication of key activator TFs

Custom scripts were used to identify the key activators (TFs) in the GRNs. we initially reduce the triple pairs (TFpeak-target) into double pairs (TF-target) for each GRNs. For each TF in the network, we calculate coverage score and p-value using hypergeometric test with "phyper" function in R for each TF and DEG cluster. all the TFs with p-value < 0.001 and coverage > 0.01 were identified as key activators.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

scRNA-Seq, snRNA-Seq, and scATAC-Seq raw and processed data have been deposited in GEO (GSE239410).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable.
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable.
Population characteristics	Not applicable.
Recruitment	Not applicable.
Ethics oversight	Not applicable.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Data from 5 to 30 retinas were used for all immunohistochemical samples analyzed. Single cell RNA-Seq and multiome libraries were constructed from at least 4-6 retinas from 4-6 individuals.
Data exclusions	No data was excluded from analysis.
Replication	Data from 5 to 30 retinas were used for all immunohistochemical samples analyzed. These were generated over the course of several months, and no statistically significant differences were observed among different batches analyzed.
Randomization	Zebrafish were assigned randomly into treatment and control groups.
Blinding	Immunohistochemical samples were blinded prior to quantification.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The following primary antibodies were used: mouse anti-PCNA monoclonal antibody (Sigma P8825, 1:500 dilution), rabbit anti-GFAP (Dako Z0334, 1:300 dilution), mouse anti-HuC/D monoclonal antibody (Invitrogen A21271, 1:300 dilution), mouse anti-4c12 monoclonal antibody (gift from Dr. Fadool, Florida State University, 1:200 dilution), rabbit anti-PKCa (Sigma Life Science P4334, 1:300 dilution), rabbit anti-UV cone opsin, 1:1000 dilution), rabbit anti-green cone opsin, 1:500 dilution), mouse anti-Zpr1 monoclonal antibody (ZIRC, 1:200) rabbit anti-Lcp1 (GeneTex GTX134697, 1:500). Secondary antibodies (diluted 1:500) included: goat anti-mouse 488 (Life Technologies A11029), goat anti-mouse 594 (Life Technologies A11032), goat anti-mouse 647 (Life Technologies A21236), goat anti-rabbit 488 (Life Technologies A11034), goat anti-rabbit 594 (Life Technologies A11037), goat anti-chicken 488 (Life Technologies A11039).
Validation	Each antibody used selectively labeled the expected cell type in regenerating zebrafish retina, as detailed in Figures 1, 3, 7 and Supplemental Figures S1, S2, S5, and S7.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	6 month-old albinob4/b4 , mmp9mi5003 mutant, and transgenic albino Tg(gfap:EGFP)nt11 19 zebrafish were used in this study standard operating policies and procedures

Wild animals	Not applicable.
Reporting on sex	A random mixture of male and female zebrafish were used for this study. No sex-dependent efffects on retina development or regeneration have previously been reported in zebrafish.
Field-collected samples	Not applicable.
Ethics oversight	Animals were bred and maintained in the Freimann Life Science Center Zebrafish Facility in accordance with procedures approved by the IACUC committee of the University of Notre Dame.

Note that full information on the approval of the study protocol must also be provided in the manuscript.