Title: Cross-species transcriptomic and epigenomic analysis reveals key regulators of injury response and neuronal regeneration in vertebrate retinas.

Authors: Thanh Hoang[1]*, Jie Wang[2]*, Patrick Boyd[7-9], Fang Wang[2], Clayton Santiago[1], Lizhi Jiang[1], Manuela Lahne[7-9], Levi J. Todd[10], Cristian Saez[1], Sooyeon Yoo[1], Casey Keuthan[11], Isabella Palazzo[10], Natalie Squires[11], Warren A. Campbell[11], Meng Jia[7-9], Fatemeh Rajaii[2], Trisha Payail[2], Guohua Wang[2], John Ash[11], Andy J. Fischer[10], David R. Hyde[7-9], Jiang Qian[2], and Seth Blackshaw[1-6]

Affiliations: [1]Solomon H. Snyder Department of Neuroscience, [2]Department of Ophthalmology, [3]Department of Neurology, [4]Center for Human Systems Biology, [5]Institute for Cell Engineering, [6]Kavli Neuroscience Discovery Institute, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA; [7]Department of Biological Sciences, [8]Center for Stem Cells and Regenerative Medicine, [9]Center for Zebrafish Research, University of Notre Dame, Notre Dame, IN, 46556 USA; [10]Department of Neuroscience, Ohio State University Wexner Medical Center, Columbus, OH, 43210 USA; [11]Department of Ophthalmology, University of Florida School of Medicine, Gainesville, FL, 32610 USA

*These authors contributed equally.
Address all correspondence to dhyde@nd.edu (D.R.H.), jiang.qian@jhmi.edu (J.Q), sblack@jhmi.edu (S.B.)

Abstract

Injury induces retinal Müller glia of cold-blooded, but not mammalian, vertebrates to generate neurons. To identify gene regulatory networks that control neuronal reprogramming in retinal glia, we comprehensively profiled injury-dependent changes in gene expression and chromatin conformation in Müller glia from zebrafish, chick and mice using bulk RNA and ATAC-Seq, as well as single-cell RNA-Seq. Cross-species integrative analysis of these data, together with functional validation of candidate genes, identified evolutionarily conserved and species-specific gene networks controlling glial quiescence, gliosis and neurogenesis. In zebrafish and chick, transition from quiescence to gliosis is a critical stage in acquisition of neurogenic competence, while in mice a dedicated network suppresses this transition and rapidly restores quiescence. Selective disruption of NFI family transcription factors, which maintain and restore quiescence, enables Müller glia to proliferate and robustly generate neurons in adult mice after retinal injury. These comprehensive resources and findings will facilitate the design of cell-based therapies aimed at restoring retinal neurons lost to degenerative disease.

Introduction

The degeneration of retinal neurons is the key pathological feature of blinding diseases such as macular degeneration, retinitis pigmentosa and glaucoma. While there are therapies that can slow the progression of vision loss, there is currently no effective way to restore retinal neurons that are lost due to diseases. Although mammals lack the ability to spontaneously regenerate retinal neurons, retinal glia in

many other species retain some degree of injury-induced neurogenic competence. In zebrafish, for example, loss of retinal neurons robustly stimulates regeneration of all neural cell subtypes even into adulthood [1,2]. This regeneration is due to the Müller glia (MG) reprogramming and reentering the cell cycle, dividing asymmetrically to produce a multipotent Müller glia-derived progenitor cell (MGPC), which proliferates and generates retinal neurons. In contrast, although MG of post-hatch chicks proliferate robustly in response to injury, they retain limited neurogenic competence, and give rise exclusively to small numbers of inner retinal neurons [3,4]. Müller glia of mice and humans do not even proliferate robustly in response to neuronal injury, and altogether lack neurogenic competence.

Recent studies in zebrafish have identified a number of genes that are rapidly induced in MG following injury and regulate neurogenic competence. Multiple extrinsic signaling pathways, including Wnt, BMP, TGFβ, FGF, Hippo, TNFα, and cytokines, have all been found to regulate the formation of MGPCs [5–14]. These ultimately induce the neurogenic factors, such as *Ascl1a* [15] and the RNA binding protein *Lin28a* [16], which are necessary for reprogramming. Most of these signaling pathways are activated in response to injury in zebrafish, chick and mouse, but they fail to induce *Ascl1* or *Lin28a* in mammals. Interestingly, forced expression of *Ascl1* in mouse MG, in combination with a histone deacetylase inhibitor and neuronal injury, can induce direct conversion of MG to inner retinal cells that show neuronal morphology and light responsiveness, but do not correspond to any native retinal neuronal type [17]. Conversely, inhibition of the Hippo pathway robustly stimulates MG proliferation in mice, but does not induce neurogenic competence [5,18]. This suggests we have not yet identified key molecular components of MG reprogramming. A more complete understanding of the genetic and epigenetic barriers to induction of proliferative and neurogenic competence in mammalian MG can potentially inform and improve approaches for cell-based therapies aimed at treating retinal dystrophies.

To comprehensively identify transcriptional and epigenetic regulators of neurogenic competence in MG, we profiled RNA levels and chromatin accessibility in zebrafish, chick and mouse in response to multiple neuronal injury models, as well as growth factor treatment. By performing both bulk and single-cell RNA-Seq (scRNA-Seq) analysis of MG, in combination with ATAC-Seq profiling, we were able to reconstruct both evolutionarily conserved and species-specific gene regulatory networks controlling glial quiescence, gliosis, proliferation and neurogenic competence. Our findings highlight key similarities and differences in the temporal response of MG to injury in species with dramatically different regenerative ability, and point towards potential mechanisms that actively suppress the formation of MGPCs in mammals.

Results

Comprehensive profiling of MG from zebrafish, chick and mouse retinas

To comprehensively profile changes in mRNA levels and chromatin conformation associated with injury response and proliferative and neurogenic competence in MG, we performed an integrated comparative analysis using bulk and scRNA-Seq data in combination with bulk ATAC-Seq data in zebrafish, chick and mouse (Fig. 1A-B). In zebrafish, for bulk RNA-Seq and ATAC-Seq analysis, we used the reporter line Tg[*gfap:EGFP*]nt11 [19] and cell sorting to enrich MG cells, purifying 9.8% of total input

retinal cells (Fig. S1A). In mouse, we performed intraperitoneal injection of tamoxifen at postnatal day (P) 21 to induce MG-specific Sun1-GFP expression in *GlastCreERT2;CAG-lsl-Sun1-GFP* mice [20], purifying 3.5% of total input cells (Fig. S1A). Based on RT-qPCR, we had an overall enrichment of 30-fold of canonical MG marker genes such as *rlbp1a/Rlbp1* and *Glul* in the GFP-positive (GFP+) fraction in both species (Fig. S1B). A minimum of two biological replicate bulk RNA-Seq samples were generated from both the GFP+ and GFP- fractions in zebrafish and mouse. Duplicate ATAC-Seq samples were also generated from the GFP+ cells. In both species, we analyzed retinas from both undamaged retinas and a time course following inner retinal damage induced by NMDA excitotoxicity or outer retinal damage induced by bright light. In total, we generated 100 bulk RNA-Seq libraries and 40 bulk ATAC-Seq libraries (Fig. 1A; Table S1; Data S1).

In parallel, we conducted scRNA-Seq based analysis of whole retinas for the same time points as used for the bulk RNA-Seq following either NMDA or light damage in zebrafish and mouse. In addition, we profiled multiple timepoints during zebrafish retinal development, P10 chick retina both before and at multiple time points after NMDA injury, and in zebrafish at multiple time points following treatment with TNFα and the gamma secretase inhibitor RO4929097 (T+R), which induces MG reprogramming independent of neuronal injury [21]. In total, we profiled 287,790 single cells across all these conditions (Fig.1C). We observed high concordance between experimental replicates for bulk RNA-Seq, scRNA-Seq and bulk ATAC-Seq (Fig.1D, S1C-E).

RNA-Seq analysis reveals species-specific MG response to retinal injury

We first compared bulk RNA-Seq data from both zebrafish and mouse GFP+ and GFP- samples to identify a core set of MG-enriched genes (Fig. 2S1A). Reads from both zebrafish and mouse showed high mappability (Fig. S2B), and hierarchical cluster analysis revealed that libraries obtained from GFP+ and GFP- cells are similar to one another (Fig. S2C). Comparative analysis of GFP+ and GFP- samples confirmed MG enrichment of virtually all MG-specific genes previously described in the literature (Fig. S2D-E; Data S2).

Principal component analysis (PCA) of bulk RNA-Seq data from zebrafish and mouse revealed common species-specific patterns of transcriptional changes in MG in response to both NMDA and light damage (Fig. 2A-B). Both forms of injury induced rapid and similar changes in each species. In mouse, the most dramatic transcriptional changes occur rapidly at early time points, within 3 and through 12 hrs, following injury. At later time points following both NMDA and light damage, MG progressively revert to a transcriptional state that closely resembles resting, uninjured MG (Fig. 2A). However, in zebrafish, both NMDA and light damage induced MG to adopt a transcriptional state that progressively diverges from a resting state (Fig. 2B). Extensive overlap in expression was observed between genes that are differentially regulated by light damage and NMDA treatment in MG in both zebrafish and mouse at all time points examined (Fig. S3A-B; Data S3-S4).

We hypothesized that some genes specifically expressed in zebrafish might promote MGPC formation, while other genes specifically expressed in mouse might repress MGPC formation (Fig. S2A). We predicted that these genes would fall into four broad categories. The first of these would be injury-regulated genes that lacked an

ortholog in the other species. We identified 355 mouse-specific injury regulated genes, and 192 zebrafish-specific genes in these categories (Fig. 2C-D, S3C; Data S5). The second category consists of genes for which orthologs are present in both species, but which show injury-regulated expression in only one of the species (Fig. 2C-2D, S3C). These included 1476 genes that are either induced or repressed in response to injury in the mouse but that are either not expressed or do not show injury-regulated changes in zebrafish MG. Similarly, 1007 differentially regulated zebrafish genes are either not expressed or differentially expressed in mouse MG. The third category consists of 354 genes for which orthologs are present and have injury-induced expression changes in both species (Fig. 2C-2D, S3C). The final category consists of genes that do not show injury-regulated changes in expression, but which show large differences in baseline expression in either species, and which may thus serve to promote or inhibit competence to form MGPCs. We observed 165 mouse genes and 203 zebrafish genes in this final category (Fig. 2C, S3C; Data S5).

Differentially expressed genes in the first three categories show three broad patterns of injury-induced regulation in both zebrafish and mouse (Fig. 2E-F). A large subset of genes showed high and selective expression in resting MG, which is then rapidly repressed following injury. The two remaining categories corresponded to differentially expressed genes that are induced rapidly (i.e. within 3-4 hours) following injury, and genes that show a slower time course of injury-induced expression. Gene ontology (GO) analysis revealed major species-specific differences in the function of these differentially expressed genes. For instance, genes that regulate the TNF, MAPK, Nfkb and Hippo signaling pathways are enriched and rapidly induced in expression in the injured mouse MG (Fig. 2G). In contrast, MG genes that are induced rapidly following zebrafish retinal injury are enriched for involvement in RNA processing and translation (Fig. 2H). These functional categories are enriched among genes that show a delayed induction in the injured mouse MG, while in zebrafish, this delayed induction category is enriched for cell cycle-related genes. This progression of MG gene expression in the damaged zebrafish retina is consistent with the MG beginning to express *PCNA* at 25 hours into the light damage treatment and the first cell divisions occurring at 36 hours following injury [22–24].

ScRNA-Seq analysis identifies resting and reprogrammed MG in species

We next performed scRNA-Seq analysis of damaged zebrafish, chick and mouse retinas. We first analyzed the unique molecular identifiers (UMI) and genes detected in each sample. On average, we observed ~1000-1500 genes and ~2500-3250 UMI per cell. Some consistent differences were seen across species and treatment conditions. For instance, injured retinas expressed higher numbers of genes and UMI (Fig. S4A). T-stochastic nearest neighbor embedding (tSNE) analysis readily separated retinal cell types into distinct clusters (Fig. 3A-F; Data S6), and no clear batch effects were observed among different samples (Fig. S4B, S5A, S5C, S5G and S6A-B).

Using well-established marker genes (Data S7), along with MG-enriched genes that we identified in our bulk RNA-Seq analysis (Fig. S2D-E; Data S2), we successfully annotated all major retinal cell types in each of the three species analyzed (Fig. 3A-F, S4C, S5B, S5D, S5E, S5G and S6A-B). In mouse, we readily distinguished rod bipolars from cone ON and OFF bipolar cells, and also distinguished GABAergic and glycinergic

amacrine cells (Fig. 3A-B, S5G). Small subpopulations of non-neuroretinal cells, including microglia, vascular/endothelial cells, astrocytes, pericytes, and RPE cells were also detected. In chick, we detected all major neuronal subtypes along with two glial cell types -- oligodendrocytes and non-astrocytic inner retinal glia (NIRG) -- in which NIRG is found specifically in chick [25,26] (Fig. 3C, S5B). In zebrafish, we likewise observed all the major retinal cell types, along with a molecularly distinct subset of progenitor cells in the uninjured retina, which strongly expresses *nr2e3* (Fig. 3E-F, S5D) This appears to correspond to multipotent progenitors of the retinal ciliary margin and rod-biased progenitors derived from MG in the central retina [27], which serve as progenitors for the persistent neurogenesis observed in the adult zebrafish retina. All major cell types were detected in each of the different injury models, as well as following T+R treatment (Fig. S6A-B). However, the efficiency of capture of individual cell types varied depending on the injury model used and the time following injury. For instance, more activated MG and microglia were captured following all forms of injury, while fewer photoreceptor cells were captured following light damage (Fig. S6C-D, S7A).

In all three species, we readily distinguished resting and activated MG cells. Many MG-enriched genes showed evolutionarily conserved patterns of expression. For instance, *Rlbp1/RLBP1/rlbp1a* all showed a high and selective expression in resting MG, and reduced expression in response to injury (Fig. 3B, 3D, 3F). Other genes such as *Gfap/GFAP/gfap* showed species-specific injury-dependent expression patterns. In mouse, *Gfap* was selectively induced following injury, and essentially absent from resting MG, while it is expressed in both resting and activated MG in zebrafish, and barely detected in chick MG (Fig. S5F). In all three species, MG were over-represented relative to their actual abundance in the intact retina, representing 25.8%, 38.2%, and 16.7% of total retinal cells in zebrafish, chick and mouse, respectively (Fig. S6C-D, S7A). This likely reflects bias in cell survival and capture efficiency in the microfluidic packaging of individual cells and reagents into nanodroplets [28,29].

While cell type-specific marker genes are generally well-established for major retinal cell types in mammals, they are not as well annotated for zebrafish and chick. Comparison of cell type-specific transcriptomes, moreover, allowed us to identify new evolutionarily conserved and species-specific marker genes for each of the common cell types among all three species (Fig. 3G; Data S8-S9). For instance, we identified a new marker gene *Espn*/*ESPN*/*espn* that is selectively expressed in MG of all three species examined (Fig. 3G).

Finally, we performed a global analysis of transcriptomic similarity among all individual cell types across all species. This allowed us to determine whether transcriptional similarity between individual cell types was more strongly influenced by cell type or species of origin. We firstly compared mouse cell types with each cell type from chick and zebrafish. A majority of cell types from chick and zebrafish are close to the same cell types from mouse (Fig. S7B). Then we constructed an evolutionary tree of cell types from all three species examined (Fig. 3H). Interestingly, we observed that functionally related but distinct neuronal cell types often showed higher overall similarity within individual species. Rod and cone photoreceptors were more closely related within each species than they were related to rods and cones between species. Retinal ganglion cells and GABAergic amacrine cells from zebrafish, chick and mouse likewise all showed greater similarity within species than to orthologous cell types across

species. In contrast, individual non-neuronal cell types, such as zebrafish and mouse microglia, showed more similar patterns of gene expression between species. This was also the case for zebrafish and chick MG, which both retain proliferative and neurogenic competence. Resting MG from zebrafish and chick clustered together, as did activated MG. In contrast, mouse resting and activated MG clustered together, along with mouse astrocytes, rather than with the zebrafish and chick MG (Fig. 3H). This suggests that the differences in the transcriptomes of resting and activated MG clearly distinguish the species that retain neurogenic competence from mouse, which lacks neurogenic competence and whose resting and activated MG more closely resemble one another.

MG trajectories reveal dramatic differences in gliotic and neurogenic competence across species

To profile dynamic injury-induced changes in gene expression in MG, we conducted trajectory analysis on MG from our scRNA-Seq dataset. Consistent with bulk RNA-Seq data (Fig. 2A), mouse MG showed rapid injury-induced changes in gene expression following NMDA treatment, followed by a gradual reversion to a transcriptional state like that of resting MG (Fig. 4A-B; Data S10). In zebrafish, MG showed a progressive and consistent change in gene expression following NMDA injury (Fig.4C), again closely resembling what was seen using bulk RNA-Seq (Fig. 2B). Trajectories were similar for both NMDA and light damage in both mouse and zebrafish (Fig. S8, S9), again consistent with bulk RNA-Seq data (Fig. 2A-B). In zebrafish, T+R treatment likewise produced a similar change in gene expression (Fig. S9C-F). Interestingly, in zebrafish a small side-branch was observed in both damage models and T+R treated retinas, which may represent a transient reactive state present at 10 hours after injury (Fig. 4C, S9). This side branch is enriched for genes associated with gliosis, such as *manf*, while the major branch shows increased expression of neurogenesis and proliferation-related genes (*ascl1a* and *cdk1*) (Fig. S9A).

Chick, in contrast, exhibits a prominent branched trajectory (Fig. 4D, S10). One branch was enriched for cells profiled at 24 hours following NMDA treatment, while the second was enriched for cells from later time points following injury (48 and 72 hours) (Fig. S10A-B). We observed that a large portion of MG at later time points (48 and 72hrs) after damage start reverting back to a resting state in chick (Fig. S10A-B). This reflects the limited neurogenic competence in chick. Analysis of gene expression indicated that, while both branches show reduced expression of genes specific to resting MG such as *RLBP1,* the branch that predominantly contains cells from 24 hours is enriched in genes associated with reactive gliosis, such as *WNT6* and *PMP2* (Fig. S10C-D; Data S11). The other branch, in contrast, is enriched for genes associated with proliferation (*CDK1*) and neurogenesis (*STMN1* and *ASCL1*) [30] (Fig. S10C-D). MG from undamaged retinas treated with growth factor (insulin+FGF2) occupied trajectory branches that were enriched for genes associated with reactive gliosis and proliferation/neuro-competent progenitors, and none of the MG treated with insulin and FGF2 had a transcriptional profile that resembled that of resting MG (Fig. 4D, S10A-B). Collectively, these findings indicate that chick and zebrafish MG undergo reprogramming in injured and uninjured retinas by passing through a gliotic state prior to conversion to MGPCs.

As with the bulk RNA-Seq analysis, we observed high overlap between genes at single-cell level that show dynamic regulation in response to light damage and NMDA in both zebrafish and mouse (Fig. S11A; Data S12). T+R treatment in zebrafish also regulated similar sets of genes, as did NMDA and insulin+FGF2 treatment in chick (Fig. S11A). Seven broad categories of injury-regulated genes were identified based on the specificity of dynamic regulation across three species (Fig. 4E-H, S11B-C; Data S13-S14). We observed species-conserved changes in gene expression, such as dynamic regulation of a gliosis-related gene *Manf*/*MANF*/*manf* following retinal injury in mouse, chick and zebrafish MG (Fig. 4F-H). Here we also observed similar functional categories of genes with bulk RNA-Seq data analysis (Fig. 4I). Specifically, DNA replication and cell cycle are enriched in genes regulated in both chick and zebrafish. There are ~31% in mouse and ~23% in zebrafish of dynamically expressed genes detected in both bulk and single-cell RNA-Seq datasets (Fig. S11D; Data S15), indicating that integration of these two parallel datasets is an effective means of comprehensively identifying injury-regulated genes. Finally, we calculated the distance of cell states based on MG transcriptomes along the trajectory. We observed that resting MG from mouse and zebrafish are similar to each other, as are gliotic MG (Fig. 4J). In mouse trajectory, we found a transit and similar proliferation in comparison with zebrafish trajectory. While zebrafish MG in the rest, gliosis and proliferation branches are close to chick MG from corresponding branches (Fig. S11E).

Injury-induced MG reprogramming recapitulates a partial reversal of MG development in zebrafish

To determine whether the formation of MGPCs from resting MG represented a simple reversal of gene expression changes observed during the differentiation of glia from multipotent RPCs, we next obtained scRNA-Seq data from developing zebrafish retinas, and integrated this with our analysis of injury-induced changes in gene expression in MG. Recent scRNA-Seq studies of developing mouse retina from our group have identified two transcriptionally and developmentally distinct RPC subtypes, which have been designated primary and neurogenic RPCs [28]. Using marker genes identified from this and other studies (Data S7), we were able to identify primary and neurogenic RPCs in embryonic zebrafish retina (Fig. 5A, S12; Data S16). We next aggregated all RPC and MG cells and performed trajectory analysis (Fig. 5B; Data S17). While NMDA injury clearly induced many MG to shift to a transcriptomic state which closely resembled that of multipotent RPCs, we also observe a separate branch that is specific to the injured MG (Fig. 5C-D, S13A-B). This injury-associated branch is most clearly observed following NMDA treatment, but is also seen following light damage (Fig. S14) and T+R treatment (Fig. S15), and was primarily populated by cells profiled at early time points following injury or T+R treatment (Fig. S13A, S14E and S15E). The injury-associated branch was selectively enriched for cytokines such as *manf* and *lepb*, as well as cytokine receptors such as *il4r.1*, while the main branch was enriched for genes that regulate proliferation and neurogenic competence such as *ascl1a*, *hmgn2*, and *cdk1* (Fig S13C-G, S14F, S15F; Data S18).

Extensive overlap was observed between genes that showed dynamic expression during development and following NMDA treatment (Fig. 5E), light damage (Fig. S14A), and T+R treatment (Fig. S15A). For overlapping genes, we found that a

great majority showed strongly negative correlation, with genes highly expressed early in development showing low expression in resting MG and vice versa, as expected (Fig. 5F). Specifically, 1227 show $r^2 < -0.5$, while only 14 showed $r^2 > 0.5$ (Fig. 5G). This suggests that reprogrammed MG following injury or T+R treatment reverse the process of RPC development in zebrafish.

Gene regulatory networks controlling MG quiescence, gliosis, proliferation and neurogenesis

We next developed a computational method to construct an integrated model of the gene regulatory networks that control response to injury in both zebrafish and mouse (Fig. S16A). We term this algorithm Integrated Regulatory Network Analysis (**IReNA**). To do this, we integrated both bulk and single-cell RNA-Seq data from all the damage/treatment models in both zebrafish and mouse to obtain a comprehensive list of differentially expressed genes at each time point, and to identify transcription factors expressed in MG. In parallel, we analyzed chromatin accessibility obtained by bulk ATAC-Seq profiling in the FACS-isolated MG samples obtained from zebrafish and mouse from matched samples (Table S1; Data S1). We next used footprinting analysis [31] to identify differential footprints for MG-expressed transcription factors and their target motifs associated with these differentially expressed genes. We then inferred regulatory relationships between these transcription factors and target genes, and used K-means clustering to separate target genes into different regulatory modules (see Methods). Based on these inferred regulatory relationships, we then constructed an integrated gene regulatory network (GRN) for each module, and compared these GRNs between zebrafish and mouse (Fig. S16A).

We observed an expected size distribution of ATAC-Seq inserts in NMDA-treated and light-damaged samples in both zebrafish and mouse (Fig. S16B). A comparison of differentially accessible open chromatin regions (DARs) between NMDA and light-damaged samples revealed extensive overlap between these different injury conditions in mouse and zebrafish (Fig. S16C; Data S19-S21). Chromatin accessibility changes are overall well-correlated between the two injury models (Fig. S16D; Data S22). As expected, a majority of DARs are from candidate regulatory regions (Fig. S16E).

Overall correlation of chromatin accessibility and gene expression was consistently high (Fig. S17A-B). We observed many species-specific and conserved dynamic injury-induced changes in chromatin accessibility. For instance, while *ascl1a* and *lin28a* are both devoid of open chromatin regions in unstimulated zebrafish and mouse MG, a rapid change in chromatin accessibility in response to injury is observed at the promoters of both genes in zebrafish but not the mouse (Fig. 6A, S17C). Conversely, we observed a rapid and transient change in the promoter accessibility of genes that are selectively expressed at early time points of retinal injury in mouse MG but not zebrafish MG, such as *Nfkb1, Nfkb2, Rel* and *Stat3* (Fig. 6A, S17D).

We hypothesized that genes with shared expression patterns are more likely to be controlled by the same network of transcription factors. Integrating bulk and single-cell RNA-Seq data, we clustered 4190 and 3942 injury-regulated genes in zebrafish and mouse, respectively, each into 10 gene modules (Fig. 6B-C, S17E-F; Data S23). Based on differential footprinting analysis, we identified 26,083 and 30177 regulatory relationships among 192 and 212 MG-expressed transcription factors in zebrafish and

mouse, respectively (Data S24). We then inferred whether individual TFs activate or repress expression of their target genes by calculating the correlation coefficient between expression of the TFs and their target genes. In zebrafish and mouse, we separately identified 97 and 156 enriched TFs that significantly activate or repress any one of 10 gene modules (Fig. 6B-C; Data S25). We further constructed two regulatory networks consisting of enriched TFs in zebrafish and mouse, respectively (Fig. S18, S19; Data S26). We observed a type of regulations that TFs activate themselves or TFs in the same module, such as the regulations of *nfia* in zebrafish and *Sox9* in mouse (Fig. 6B-C, S18, S19 and S20A). There is another type of regulations in which source TFs and target TFs are from different modules. For instance, *hmga1a* activates the module in which TFs are expressed at the subsequent time in the zebrafish (Fig. S18, S20A). In the mouse, *Nfib* represses another module related to gliosis (Fig. S19, S20A).

To obtain insight into regulatory relationships among modules of co-regulated and enriched TFs, we identified regulatory relationships between individual modules by separately calculating the number of inferred positive and negative regulatory relationships for each pair of modules, and determining whether these were statistically overrepresented relative to the whole network. An intramodular regulatory network was then generated for both zebrafish and mouse (Fig. 6D-E).

In the zebrafish, we observed multiple TF modules that are most active in resting MG (Fig. 6B, 6D, S18). These modules are predicted to both positively regulate their own expression, and to also activate expression of other TFs in networks active in resting MG. These quiescence-associated networks in turn are predicted to repress expression of TFs found in a second set of networks, which are enriched for TFs known to regulate neurogenic competence and predicted to regulate gliosis (Fig. 6D, S20B). TFs found in this second group include *ascl1a*, *tgif1* and *six3b*, all of which are necessary for generating MGPCs in zebrafish [15,32]. These TFs in the gliosis/neurogenesis network in turn positively induce their own expression, and negatively regulate TFs in the quiescence-associated network. Direct downstream of the gliosis/neurogenesis network lies a third set of TF networks which regulate genes associated with DNA replication and cell cycle progression, which includes TFs known to regulate cell cycle entry such as *myca* and *e2f3*, but also includes other TFs that control neurogenesis, such as *olig2* and *foxn4* (Fig. 6D, S20B) [33–35] In zebrafish MG, we conclude that a bistable regulatory relationship between mutually opposed quiescence and neurogenesis/gliosis networks controls that transition from a resting to an activated, neurogenic state. Once MG have transitioned to a neurogenic state, this then rapidly leads to transition to a proliferative and highly neurogenic state (Fig. 6F).

In the mouse, we likewise observed three interconnected networks of TF modules (Fig. 6F). We observed a set of cross-activating networks that activates expression of other TFs in networks active in resting MG (Fig. 6C, 6E, S19). This included *Lhx2*, which is known to negatively regulate transition to reactive gliosis in mouse MG [36], as well as Notch pathway genes such as *Hes1*. Like in zebrafish, this quiescence-associated network is predicted to repress expression of genes in a similarly interconnected gliosis-associated TF module network, which in turn is predicted to repress quiescence-associated TFs, establishing a bistable, cross-repressive regulatory relationship very similar to that seen in the zebrafish (Fig. 6D-F, S20B). TFs in the gliosis-associated network included Nfkb pathway components

(*Rela, Relb, Nfkb1, Nfkb2*), and multiple TFs that were previously shown to promote glial activation (*Stat3*, *Sox9*) [37,38], along with the Hippo pathway target *Tead1*.

Interestingly, TFs associated with cell cycle entry (*Myc*) and even neurogenesis (*Sox4*) [39], are also included in this gliosis-associated network in the mouse. Expression of these genes, however, peaks relatively early following injury and drops rapidly, as eventually do all gliosis-associated genes (Fig. 2E). An indication as to how this reversion to a quiescent state occurs in mouse comes from the presence of a third network of TF modules that is not found in zebrafish (Fig. 6F). This network is also enriched for TFs expressed in resting MG, e.g., *Nfib, Myb* and *Sox5*, which showed reduced expression immediately following injury, but elevated expression at later time points (Fig. 2E). These TF modules are separated from those that are predicted to maintain MG in a quiescent state, and appear to mediate restoration of quiescence following activation (Fig. 6E). Taken together, this implies that the transition between quiescence, gliosis, and restoration of quiescence in mouse are mediated by three separate bistable, auto-activating transcriptional regulatory networks (Fig. 6F). Furthermore, they imply that reactive gliosis appears to be necessary to induce MG reprogramming and proliferation to generate MGPCs in zebrafish and chick, and that dedicated transcriptional regulatory networks exist to maintain and restore a resting state in mouse MG.

Functional validation of candidate genes in regulating proliferative and neurogenic competence

We next tested whether individual TFs that regulate gliosis and glial quiescence regulated the formation of MGPCs. To select such TFs, we integrated the regulatory networks and gene expression datasets, and compared them across species (Fig. 6G; Data S27). We examined the degree of each TF within the regulatory networks. The degree, which is defined as the number of edges in the regulatory networks, reflects the importance of the TF in the networks (Fig. S18, S19 and S20C). Similarly, we also assessed the fold changes of gene expression during the time course. It is interesting that *ascl1a* has a high degree in zebrafish regulatory network, while it has no degree in the mouse. The injury-induced change in *ascl1a* expression is correspondingly much greater in the zebrafish than in the mouse. Based on the same rule, we selected *hmga1a* for functional analysis. We expect other genes with species-specific differences in regulation may play critical roles in promoting or restricting retinal regeneration. For example, *myb* shows a high degree in the cell cycle module in the zebrafish, but also has a high degree in the restore rest module in the mouse. In addition, *nfia/b* showed decreased expression at later stages of injury in the zebrafish, but show significantly increased expression at later stages in the mouse (Fig. 6G).

We then directly tested whether individual TFs that are predicted to regulate gliosis and glial quiescence are required for MG reprogramming and proliferation. The non-histone HMG family DNA binding protein gene *hmga1a/Hmga1* is induced in reactive glia in zebrafish and mouse (Fig. 6B-F, S21A). We used morpholinos to knockdown expression of *hmga1a* to determine whether it is necessary for MG proliferation and generating MGPCs. Retinas electroporated with the *hmga1a* morpholino exhibited a significant reduction in the number of proliferating MG following 36 hours of light damage relative to the Standard Control morphant (Fig. 7A,B,D). Even

after 72 hours of constant light, the *hmga1a* morphant retina possessed significantly fewer PCNA-positive MGPCs than the Standard Control morphant retina (Fig. S21E,F,J). We also tested the role of *smarca5*, a SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 gene, that is expressed in reactive glia in the zebrafish (Fig. 6F, S21C). The morpholino-mediated knockdown of Smarca5 protein in NMDA-damaged retinas significantly reduced the number of proliferating MG relative to the Standard Control morphant retina (Fig. 7E,G,I).

To test the potential role of proliferation-related genes, we used morpholinos to knockdown the expression of Myb, a transcription factor that is predominantly expressed in proliferating cells, and PCNA, which served as a positive control (Fig.6F, S21D). Knockdown of the Proliferating Cell Nuclear Antigen gene (*pcna*) in the NMDA-damaged retina significantly reduced the number of proliferating MG relative to the Standard Control morphant (Fig. 7E,F,I). Similarly, the NMDA-damaged *myb* morphant retinas possessed significantly fewer proliferating MG compared to the Standard Control morphant retina (Fig. 7E,H,I). Finally, we examined the potential role of *yap1* in the damaged zebrafish retina, which was recently shown to be required for MG to exit quiescence in both *Xenopus* and mice [5,18]. Morpholino-mediated knockdown of Yap1 protein in the light-damaged zebrafish retina significantly reduced the number of proliferating MG relative to the Standard Control morphant (Fig. 7A,C,D). Furthermore, injection of the drug Verteporfin, which inhibits the interaction of Yap1 and TEAD in the nucleus [40], nearly eliminated PCNA-positive MG in the light-damaged retina relative to the DMSO control in a dose-dependent manner (Fig. S21G-I, K).

Although a core set of evolutionarily conserved genes are upregulated in reactive gliosis, many genes show species-specific induction during gliosis. In chick, for instance, we identified the lipid binding proteins *FABP5*, *FABP7* and *FABP8* (also known as *PMP2*) as highly expressed in reactive gliosis (Fig. S22A). The expression of FABP8 protein was further confirmed by immunohistochemistry (Fig. S22B-C). These *FABP* family genes all show substantially lower or absent induction of expression in zebrafish and mouse. To globally inhibit the function of these *FABP* family genes, we treated chick retina with BMS309403 and measured NMDA-induced formation of MGPCs. We observed a significant decrease in the number of MGPCs compared with the DMSO-injected control retina (Fig. 7J-L).

To test the functional role of TFs predicted to maintain and/or restore MG quiescence, we selected *NFIa/b/x* transcription factors which are expressed in resting MG, down-regulated rapidly after injury but elevated in the later timepoints in mouse (Fig. S23A-C). To this end, we examined the phenotype of tamoxifen-fed *GlastCreER;CAG-lsl-Sun1-GFP;Nfia/b/x*$^{lox/lox}$ mice, in which NFI family members were selectively deleted from mature MG (Fig. S23D). ScRNA-Seq analysis indicates that *Nfia/b/x* loss of function, as expected, leads to a reduction in expression of genes selectively expressed in resting MG, such as *Glul*, *Rlbp1 Aqp4 Apoe*. NFI-deficient MG are clearly distinct from the control (Fig. S23E-F). Strikingly, however, NFI-deficient MG also upregulate cell cycle regulators such as *Ccnd1* and *Ccnd3*, along with the neurogenic bHLH factor *Ascl1* (Fig. S23G-I). We next tested whether a combination of NMDA, FGF2 and insulin could induce NFI-deficient MG to proliferate and generate neurons in response to excitotoxic injury. We observed that NFI-deficient MG proliferate at substantially higher levels than wildtype MG in response to this treatment,

with EdU+/GFP+ cells observed in both inner and outer nuclear layers (Fig. 7M-Q). Furthermore, a substantial number of the MG-derived GFP+ cells expresses either the photoreceptor marker CRX or the neural precursor and amacrine/ganglion/horizontal cell marker HuC/D and NeuN (Fig. 7R-W). This indicates that NFI-deficient MGPCs are multipotent and that NFI factors act to repress proliferative and neurogenic competence in mouse.

Discussion

This study is the first effort to comprehensively profile changes in gene expression and chromatin conformation that underlie cellular reprogramming in species with dramatically different regenerative abilities. Retinal MG of zebrafish have a high capacity to proliferate and give rise to new neurons, while those of post-hatch chick retain high proliferative but limited neurogenic competence. Mouse MG, in contrast, lack substantial proliferative competence and do not generate neurons. Using both bulk and single-cell RNA-Seq, along with ATAC-Seq analysis, we have identified both evolutionarily conserved and species-specific transcriptomic and epigenomic events that occur during MG development, following outer and inner retinal injury, as well as following growth factor treatments that induce MG reprogramming in zebrafish and chick. Integration of these datasets has allowed identification of gene regulatory networks that are strong candidates for both promoting and restricting neurogenic and proliferative competence in MG. These datasets will serve as a starting point for future studies that functionally characterize these candidate genes.

This study identifies several key events that appear to be critical for regulating MG reprogramming. First, we observed that MG in all three species pass through a state resembling reactive gliosis. Reactive gliosis occurs following all forms of retinal injury, but differs considerably in its duration. It is a transient state in zebrafish, and a much more prolonged state in chick. In contrast, mouse MG arrest in gliosis, before rapidly returning to quiescence. Most reactive zebrafish and chick MG pass through this gliotic state to become MGPCs, which express high levels of neurogenic bHLH genes *Ascl1* in zebrafish, and much lower levels in chick, in line with their more limited neurogenic competence. Reactive gliosis is a critical step in formation of MGPCs, as loss of function of genes that are highly enriched in reactive MG such as *hmga1a* in zebrafish and *FABP* family genes in chick blocks formation of MGPCs. Previous studies in zebrafish have likewise identified multiple genes that show enriched expression in reactive MG -- such as *six3b*, *sox2*, *lepb* and others [23,32,41] -- to also be essential for MGPC formation.

Second, in both mouse and zebrafish, we identified dedicated gene regulatory networks that maintain or restore expression of genes specific to quiescence of MG. Notably, although reactive mouse MG transiently express cell cycle regulatory genes such as *Ccnd1, Ccnd2* and *Myc*, as well as the neurogenic *Sox4*, they return to quiescence rapidly, and have dedicated a transcriptional regulatory network that mediates this process. NFI family genes, which are members of this network, act to maintain expression of genes specific to quiescent MG and to restrict proliferative and neurogenic competence in mouse MG. NFI family genes show evolutionarily-conserved expression patterns, with injury repressing expression in MG in all three species examined. Furthermore, selective deletion of NFI genes in mouse MG, in combination

with NMDA and growth factor treatment, resulted in a substantial MG proliferation and formation of MG-derived cells expressing photoreceptor and inner neuron markers, indicating the multipotency of these NFI-deficient MGPCs. Detailed characterization of these MG-derived neurons awaits further studies. Taken together, this suggests that NFI family genes maintain MG quiescence, and prevent expression of genes associated with transition to gliosis and/or MGPC status, and suppress neurogenesis from MGPCs. Other data support this proposed role for NFI factors. Our group has recently shown that *Nfia/b/x* act in late-stage retinal progenitors to drive gliogenesis and restrict neurogenic and proliferative competence, with *Nfia/b/x* loss of function mutants leading to persistence of proliferative progenitors and generation of excess rod photoreceptors [28]. Furthermore, the observation that *Ascl1* overexpression induces direct transdifferentiation of MG to inner retinal neurons but not photoreceptors may result from the strong persistent expression of *Nfia/b/x* that is seen in these transformed cells [17,42].

Ever since the development of the Yamanaka protocol for generation of iPSCs [43], the great majority of work aimed at induced dedifferentiation or cellular reprogramming has focused on gain of function approaches, in which TFs associated with the desired fate are overexpressed. However, with the exception of the iPSC generation, the effectiveness of these approaches has been limited, in large part because genes specific to the cell of origin are not efficiently repressed. This study highlights the importance of downregulating these gene regulatory networks in the context of endogenous cellular reprogramming, and identifies potential targets for improving regenerative therapies that rely on directed reprogramming of endogenous cells.

Acknowledgements:
We thank J. Nathans and A. Kolodkin for their comments on the manuscript. We thank Berrnard Kulemeka and the Freimann Life Science Center staff for breeding and raising the zebrafish strains used in this study. We thank X. Zhou and J. Wang for curation of the RNA and ATAC-Seq data at St. Jude.
Funding: This work was supported by NIH grants U01EY027267 (D.R.H., S.B., J.Q., A.J.F., J.A.), R01EY024519 (D.R.H.) and R01EY020560 (S.B.). P.B. was partially supported by the Hiller Family Endowment for Stem Cell Research at the University of Notre Dame, M.L. was partially supported by the Center for Zebrafish Research at the University of Notre Dame.
Author contributions: D.R.H., S.B., J.Q., A.J.F., and J.A. conceived of the project and supervised the research and experimental design. T.H. generated all RNA-Seq and ATAC-Seq data and conducted all functional studies in mouse. J.W. performed all analysis of RNA-Seq and ATAC-Seq data, developed IReNA, and identified gene regulatory networks in zebrafish and mouse. F.W. and G.W. analyzed data from developing zebrafish retina. P.B. and M.L. performed light damage and NMDA damage paradigms in zebrafish and the TNFa+RO4929097 co-injections and cell sorting to isolate MG and neuronal fractions for bulk RNA-Seq, single-cell RNA-Seq, and ATAC-Seq. P.B. and M.J. performed morpholino and verteporfin studies in zebrafish. L.Z., C.Z., C.S., S.Y., F.R., and T.P. all contributed to analysis of mouse functional data. C.S., N.S., and C.K. contributed to analysis of mouse gene expression data. L.J.T. and I.P. conducted functional analysis in chick.

Competing interests: The authors declare no competing interests.
Data and materials availability: All RNA-Seq and ATAC-Seq data are deposited in GEO. Bulk RNA-Seq, bulk ATAC-Seq and scRNA-Seq data can be queried interactively at https://proteinpaint.stjude.org/F/2019.retina.scRNA.html.

Materials and Methods:

Animals

Fish maintenance

All zebrafish (*Danio rerio*) lines, including AB, *albino*; Tg[*gfap:GFP]mi2001* [44] and *albino*; Tg[*gfap*:EGFP]*nt11* [19] were maintained in the Center for Zebrafish Research at the University of Notre Dame Freimann Life Science Center. The adult zebrafish used in these studies were 6 to 12 months old (3-5cm in length). Fish were maintained under a light and dark cycle of 14 hours light and 10 hours of dark at 28.5°C. All experimental protocols were approved by the animal use committee at the University of Notre Dame and are in compliance with the ARVO statement for the use of animals in vision research.

Chick maintenance

The use of animals in these experiments was in accordance with the guidelines established by the National Institutes of Health and the Ohio State University. Newly hatched wild type leghorn chickens (*Gallus gallus domesticus*) were obtained from Meyer Hatchery (Polk, Ohio). Postnatal chicks were kept on a cycle of 12 hours light and 12 hours dark (lights on at 8:00 AM). Chicks were housed in a stainless steel brooder at about 25ºC and received water and Purina^{tm} chick starter *ad libitum*.

Fertilized eggs were obtained from the Michigan State, Department of Animal Science Eggs were incubated between 36.6°C and 37.8°C and embryos were staged according to guidelines set by Hamburger and Hamilton [45].

Mice maintenance

CD1 mice were purchased from Charles River Laboratories (Wilmington, MA). *GLASTCreERT2* and *Sun1-sGFP* transgenic mice were provided by Jeremy Nathans (The Johns Hopkins University School of Medicine). *GLASTCreERT2*; *Sun1-sGFP* were generated by breeding and subsequent backcrossing. To induce *Cre* recombination, *GLASTCreERT2*; *Sun1-sGFP* mice at ~2 weeks of age were intraperitoneally injected with Tamoxifen in corn oil for 2 consecutive days at 1mg/dose. *Nfia*$^{lox/lox}$; *Nfib*$^{lox/lox}$, and *Nfix*$^{lox/lox}$ mice were crossed to *GLASTCreERT2*; *Sun1-sGFP* mice. To generate MG-specific loss of function mutants of *Nfia/b/x* genes, 3 week-old *GLASTCreERT2*; *Nfia/bx*$^{lox/lox}$;*Sun1-sGFP* mice were fed with tamoxifen diet for 3 weeks following by 2 weeks with normal diet. All mice were housed in a climate-controlled pathogen free facility on a 14 h-10 h light/dark cycle (07:00 lights on – 19:00 lights off). All experimental procedures were pre-approved by the Institutional Animal Care and Use Committee (IACUC) of the Johns Hopkins University School of Medicine.

Retinal damage paradigms in fish

To induce death of rods and cones, adult *albino*; Tg[*gfap*:EGFP]*nt11* fish were dark adapted for 14 days, then transferred to clear polycarbonate tanks placed between four fluorescent bulbs (20,000 lux) and water temperature maintained at 32°C for up to 72 hours [46,47]. Fish were euthanized by anesthetic overdose of 0.2% 2-phenoxyethanol in system water.

To induce amacrine and ganglion cell death, we injected N-methyl-D-aspartic acid (NMDA) (Sigma-Aldrich; St. Louis, MO) intravitreally [48]. Adult Tg[*gfap:GFP]mi2001* zebrafish were anesthetized in 0.1% 2-phenoxyethanol. A sapphire blade was used to make an incision in the posterior cornea and 0.5 µl of 100 mM NMDA was injected into the intravitreal space of the eye using a 33 gauge Hamilton syringe. The fish were revived and placed in a 32°C dark incubator for up to 36 hours. Fish were euthanized by anesthetic overdose of 0.2% 2-phenoxyethanol in system water.

TNFα and RO4929097-induced proliferation

Recombinant TNFα was expressed in *E. coli* and purified as previously described [21]. TNFα was diluted to a concentration of 1.5 mg/ml in water prior to injection. Undamaged adult Tg[*gfap:GFP]mi2001* fish were anesthetized in 0.1% 2-phenoxyethanol, cut in the posterior cornea with a sapphire blade, and intravitreally injected with 0.5 µl of TNFα using a 33 gauge Hamilton syringe every 12 hours. The fish were also intraperitoneally injected with 50 µl 0.75 mM RO4929097 (Selleckchem; Houston, TX) using a 30 gauge bevelled needle every 12 hours. The fish were revived and placed in a 32°C dark incubator for up to 72 hours before being euthanized by anesthetic overdose of 0.2% 2-phenoxyethanol in system water.

Intraocular injections in chicks

Chickens were anesthetized via inhalation of 2.5% isoflurane in oxygen and intraocular injections performed as described previously [49]. For all experiments, the right eyes of chicks were injected with the “test” compound and the contra-lateral left eyes were injected with vehicle as a control. Compounds were injected in 20 µl sterile saline with 0.05 mg/ml bovine serum albumin added as a carrier. Compounds used in these studies included NMDA (38.5 or 154 g/dose; Sigma-Aldrich), FGF2 (200 ng/dose; R&D systems), insulin (1µg/dose; Sigma-Aldrich), and BMS309430 (500 ng/dose; Tocris). Two µg EdU was injected to label proliferating cells. Injection paradigms are included in each figure.

Mouse N-methyl-D-aspartate (NMDA) and light damage

For NMDA damage, adult mice, either CD1 or *GLASTCreERT2*; *Sun1-sGFP* mice at ~2 months of age, were anesthetized with isoflurane inhalation. A puncture was made just behind the limbus with a 30G needle. Two microliters of 100mM NMDA in PBS was intravitreally injected using a syringe with a 33G blunt-ended needle. Mice were sacrificed, and retinas were collected at the indicated time points.

For NMDA and growth factor treatment of NFI TKO mice, adult *GLASTCreERT2*; *Nfia/bx*$^{lox/lox}$;*Sun1-sGFP* and *GLASTCreERT2*;*Sun1-sGFP* control mice were intravitreally injected with 2 µl of 100mM NMDA, 100ng/µl bFGF and 1µg/µl Insulin, followed by 2 µl of 500ng/µl EdU, 100ng/µl bFGF and 1µg/µl Insulin at 24 and 48hr after the first injection. Contralateral eyes were injected with PBS and EdU as the controls.

Light damage was performed as previously described [37]. In brief, the mice were reared in cyclic 12-hour low light/12-hour dark conditions at the University of Florida animal housing facility. Prior to light damage, mice were placed in a modified cage equipped with dimmable white light LED strips. Light intensity was measured using a light meter (Thermo Fisher Scientific Inc., Waltham, MA) and set to 2000 lux. Animals

were subjected to the damaging light for 4 hours (6PM-10PM) and were moved back to low light conditions to recover. All animals were kept in ventilated racks for the duration of the experiment and the lighting equipment was approved by Animal Care Services.

Zebrafish retinal dissociation, cell sorting and methanol fixation

Zebrafish were euthanized in 0.2% 2-phenoxyethanol and retinas dissected and placed in Leibowitz medium (Thermo Fisher Scientific) prior to dissociation. Retinas were subsequently placed in 20U/ml papain (10 retinas per 1ml) (Worthington), and incubated at 28°C for 30 minutes with gentle agitation. Cells were then pelleted and resuspended in PBS containing 0.1mg/ml leupeptin (Sigma-Aldrich) and 10U/ml DNaseI (Roche). Cells were then filtered through a 70 µm filter (Miltenyi Biotec), and kept on ice until sorting and incubated with 5 µg/ml propidium iodide (PI; Life Technologies).

Cells were transferred into a 5ml FACS tube and loaded into a FACS AriaII. Gating was set using dissociated retinas from *albino* zebrafish both unstained and PI stained and unstained Tg[*gfap*:EGFP]*nt11* zebrafish. GFP+ and GFP- cells were sorted directly into Trizol LS (Life Technologies), and immediately flash frozen.

For single-cell RNA-Seq dissociated cells were methanol fixed [50]. Dissociated cells were pelleted and resuspended in D-PBS (Thermo Fisher Scientific). This was repeated for a total of 3 washes, with samples kept on ice throughout. Cells were then resuspended in 100 µl D-PBS in cryovials and 900 µl -20°C methanol added dropwise while slowly vortexing. Cells were left on ice for 15 minutes and then placed directly in -80°C.

Mouse retinal cell dissociation and fluorescence-activated cell sorting (FACS)

Mice were euthanized, and eye globes were removed and kept in ice-cold 1x PBS. The neural retinas were dissected, and cells were dissociated using Papain Dissociation System (LK003150, Worthington). Briefly, retinas were incubated in pre-equilibrated papain/DnaseI mix (20 units/ml papain and 100 units/ml DnaseI, 200ul per retina) in a 37^0C water bath. Retinas were mixed by inverting the tubes every 5 min and gently triturated after first 10 min at every 5 min with 1ml pipette tips. After 25 min of enzyme incubation, undigested retinal tissues were removed by running through a 50um cell filter. Dissociated cell mixtures were treated with DnaseI and subjected to density gradient centrifugation to remove the cell debris by following the manufacturer instructions. For FACS sorting experiments, dissociated cell pellets were resuspended in ice-cold PBS containing 2% heat-inactivated FBS and 5 units/ml DnaseI).

For scRNA-Seq experiments, cell pellets were rinsed twice without disturbing the cell pellet to completely remove any trace of DnaseI. Dissociated cells were then re-suspended with ice-cold PBS containing 0.04% BSA and 0.5U/µll RNase inhibitor to obtain a final cell concentration of 0.5-1.5x10^6 cells/ml. Dissociated cells were filtered through a 50um filter, and cell count and viability were assessed by Trypan blue staining.

FACS experiments were performed using Sony SH800S Cell Sorter. Retinal cells from non-transgenic mice were used to set the gating threshold for GFP-positive cells. Cells were flow-sorted into GFP-positive and GFP-negative fractions into ice-cold PBS containing a final concentration of 10% heat-activated FBS. For ATAC-Seq, cells were

kept on ice until use. For bulk RNA-Seq, cells were centrifuged at 500xg for 5min, resuspended in 700µl QIAzol lysis reagent (miRNAeasy Mini Kit) and stored at -80°C until RNA extraction.

Immunohistochemistry and imaging in mice

Mouse eye globes were collected and fixed in 4% PFA in PBS for 2hr at room temperature. Eyes were washed with PBS, and retina were carefully dissected and put in 30% sucrose overnight at 4°C. Retina were then embedded, cryosectioned at 20 um thickness and stored at -80°C. Sections were dried at 37°C for 20 min and washed 3 x 5 min with PBS prior to incubating with a blocking buffer (0.2% Triton, 5% horse serum in PBS) for 2 hr at room temperature. For CRX, HuC/D and NeuN immunostaining, retinal sections were treated with 100 mM sodium citrate pH 6.0 at -80°C for 30 min for antigen retrieval before the blocking step. Sections were then incubated with primary antibodies at indicated concentrations in the blocking buffer (Table S2) overnight at 4°C. Sections were washed 3x 10min in PBST (0.1% Triton in PBS) and incubated with secondary antibodies in the blocking buffer for 2hr at room temperature. Sections were counterstained with DAPI, washed 3 x 10 min in PBST and mounted in Vectashield with DAPI. EdU staining was performed by following the manufacturer instructions. Images were acquired by using confocal Zeiss LSM 700. EdU+/GFP+ cells were counted and averaged from > 6 random whole sections for each retina. CRX+/GFP+ and HuCD;NeuN+/GFP+ cells were counted from 4 images taken across > 4 random sections per retina. Fraction of CRX+/GFP+ and HuCD;NeuN+/GFP+ cells were calculated from total GFP+ cells. Each data point in the quantification bar graphs was calculated from individual retina.

Injection and electroporation of morpholinos in zebrafish

Morpholino mediated knockdown was performed as previously described [51,52]. Morpholinos (Gene Tools) were reconstituted in 100 µl nuclease free water to yield a 3mM solution. Tg[*gfap*:EGFP]*nt11* fish were anesthetized in 0.1% 2-phenoxyethanol and corneas were removed from the eye using tweezers, 0.5 µl of morpholino solution was intravitreally injected into the eye. Platinum plate electrode tweezers (Protech International Inc.) were used to deliver two 50 ms pulses (75 V with a 1 s pause between pulses) to the eye using a CUY21 Square Wave Electroporator (Protech International Inc.). Fish were then allowed to recover, and then anaesthetized and intravitreally injected with NMDA. Fish were then placed in a 32°C dark incubator for 36 hours before being euthanized by anesthetic overdose of 0.2% 2-phenoxyethanol in system water. The following previously validated morpholinos were used in this study: Standard Control: 5'-CCTCTTACCTCAGTTACAATTTATA-3' (Gene Tools), *pcna*: 5'-TGAACCAGACGTGCCTCAAACATTG-3' [51], *myb:* 5'-GCCGCCTCGCCATCCCGCTGTTCG-3' [53], *smarca5*: 5'-CTTCTTCCCGCTGCTGCTCCATGCT-3' [54], *hmga1a*: 5'-CTGTGTCCTTGCCAGAATCACTCAT-3' [55], *yap1*: 5'-CTCTTCTTTCTATCCAACAGAAACC-3' [56].

Verteporfin injection in zebrafish

Verteporfin was resuspended at 3.48 mM in 100% DMSO and then diluted with water to the desired working concentration. Dark-adapted *albino*; Tg[*gfap*:EGFP]*nt11* fish were anesthetized in 0.1% 2-phenoxyethanol and 0.5 µl of either verteporfin or DMSO was intravitreally injected into the eye and the fish were exposed to constant light damage. Every 12 hours, the intravitreal injections were repeated until 36 hours, when the fish were euthanized and retinal sections were analyzed by immunohistochemistry and confocal microscopy.

Immunohistochemistry and imaging in zebrafish

Immunohistochemistry was performed as described previously [12,21]. Fish were euthanized in 0.2% 2-phenoxyethanol. Eyes were collected and fixed in 9:1 ethanolic formaldehyde. After fixation, eyes were rehydrated through an ethanol gradient (90%, 80%, 70%, 50% v/v ethanol in water), then washed for 15 minutes in 5% sucrose in PBS, and incubated overnight in 30% sucrose in PBS at 4°C. Eyes were then transferred into a 2:1 mixture of tissue freezing media (TFM) (Triangle Biomedical Sciences; Durham, NC) and 30% sucrose in PBS and incubated at 4°C overnight. Eyes were embedded in 100% TFM and frozen at -80°C. 14 µm cryosections were prepared and stored at -80°C.

Sections were dried at 50°C for 20 minutes and the sections were surrounded using a liquid PAP pen (Ted Pella; Redding, CA). Sections were then rehydrated in PBS for 20 minutes prior to blocking for 1 hour (1% Tween-20, 1% Triton-X, 2% normal goat serum, 2% DMSO in PBS). Primary antibodies (Table S2) were diluted in blocking solution and slides incubated in primary antibody solution at room temperature overnight. Sections were washed 3 x 10 min in PBST (1x PBS with 0.05% Tween-20), and then incubated for 1 hour at room temperature with secondary antibodies in blocking solution at 1:1000, and counterstained with DAPI at 1:1000. Sections were washed 3 x 10 min in PBST, and No 1.5 coverslips (VWR) were mounted using Prolong Gold (Life Technologies). Confocal imaging was performed using a Nikon A1R laser scanning confocal microscope. Images were obtained from the center of the dorsal retina using only sections containing the optic stalk. Quantification was performed using 6.5 µm z-stacks, and cell counts normalized to 300 µm. Data was statistically analyzed using one-way ANOVA with Tukey's post-hoc test if more than two groups were compared, otherwise using Student's t-test with unequal variances.

Immunohistochemistry and imaging in chick

Tissues were fixed, sectioned, and immunolabeled as described previously [11,57]. Working dilutions and sources of antibodies were listed in Table S2. None of the observed labeling was due to non-specific labeling of secondary antibodies or autofluorescence because sections labeled with secondary antibodies alone were devoid of fluorescence. Secondary antibodies were diluted to 1:1000 in PBS plus 0.2% Triton X-100. Draq5 (10 uM; Thermofisher) was added to the secondary antibody solution to label nuclei.

For EdU-labeling, immunolabeled sections were fixed in 4% formaldehyde in PBS for 5 minutes at room temperature, washed for 5 minutes with PBS, permeabilized with 0.5% Triton X-100 in PBS for 1 minute at room temperature, and washed twice for 5 minutes in PBS. Sections were incubated for 30 minutes at room temperature in 2M

Tris, 50 mM $CuSO_4$, Alexa Fluor 568 Azide (Thermo Fisher Scientific), and 0.5M ascorbic acid in dH_2O.

Digital photomicrographs were obtained using Leica DM5000B microscopes equipped with epifluorescence and Leica DC500 digital camera. Confocal images were obtained using a Leica SP8 imaging. Images were optimized for color, brightness and contrast, multiple channels overlaid and figures constructed by using Adobe Photoshop. Cell counts were performed on representative images. To avoid the possibility of region-specific differences within the retina, cell counts were consistently made from the same region of retina for each data set.

RNA extraction and RT-qPCR

RNA was extracted from both GFP-positive and GFP-negative cell fractions using miRNAeasy Mini Kit (#217004, Qiagen). To test for MG enrichment of sorted samples, expression of some MG markers was measured by RT-qPCR. Briefly, RNA samples were first reverse transcribed into cDNA using random primers and Superscript IV reverse transcriptase (#18091050, ThermoFisher) according to the manufacturer's instructions. The qPCR assays were performed on the cDNA using GoTaq Green Master Mix (#M7122, Promega) using a StepOnePlus Real-time instrument (ThermoFisher). Intron-spanning primers were designed to specifically quantify targeted mRNA transcripts. Glyceraldehyde 3-phosphate dehydrogenase (*Gapdh*) expression was used as the endogenous control (Table S3). PCR specificity was monitored by determining the product melting temperature and by checking for the presence of a single DNA band on agarose gel analysis of the qRT-PCR products

Bulk RNA-Seq

Flow-sorted RNA samples were sent to the Deep Sequencing and Microarray Core (Johns Hopkins University) for library preparation and sequencing. Briefly, ribosomal RNA was depleted, and total RNA was captured from the RNA samples using Illumina TruSeq Stranded RNA LT kit Ribo-Zero™ Gold (# 15032619, Illumina). Around 8-10 libraries were pooled and sequenced for paired-end 75 cycles using the NextSeq 500 system with ~400-500 million reads per run, resulting in ~45-55 million reads per library.

ScRNA-Seq

For fresh samples (NMDA mouse damage, chick, zebrafish development and some zebrafish damaged samples), cells were resuspended with ice-cold PBS containing 0.04% BSA and 0.5U/UL RNase inhibitor. For methanol-fixed samples (mouse LD damage and most of zebrafish damage samples), fixed cells were processed by following 10x genomic protocol for methanol-fixed cells. Briefly, methanol-fixed cells were centrifuged at 3000xg for 10 min at 4^0C and washed 3 times in a iced-cold re-suspension buffer containing PBS, 1% BSA and 0.5U/UL RNAse inhibitors. Dissociated cells were resuspended in resuspension buffer and filtered through a 50ul cell filter. Cells (~10k) were loaded into a 10x Genomics Chromium Single Cell system (10x Genomics, CA, USA) using v2 chemistry following the manufacturer's instructions. Libraries were pooled and sequenced on Illumina NextSeq with ~200 million reads per

library. Sequencing results were processed through the Cell Ranger 2.1.1 pipeline (10x Genomics) with *--expect cell* =10,000 parameter.

ATAC-Seq

Sorted cells (~50k-75k) from GFP-positive samples were used for ATAC library preparation. In our experiences, there are two main factors that are critical for a successful ATAC-Seq library preparation. First, cell nuclei need to be intact during nuclei isolation process by optimizing IGEPAL concentration and minimal pipetting. Second, enzyme concentration and tagmentation time need to be optimized depending on cell types. Nuclear isolation and all centrifugation steps were carried out on ice in a 4^0C cold room. Briefly, cells were spun down at 500xg for 5 min. Cells were rinsed with ice-cold PBS without disturbing cell pellet and centrifuged again at 500xg for 5 min. Cells were lysed by adding 50UL of ice-cold cell lysis buffer (10 mM Tris.Cl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) containing 0.03% IGEPAL and protease inhibitors (1 tablet per 7ml of lysis buffer) and mixing 3 times by pipetting. Cells were then immediately spun down at 500xg for 10 min and washed with 150UL of ice-cold lysis buffer without IGEPAL and protease inhibitors. For tagmentation, cell nuclei were incubated with 2.5 UL enzyme in 50UL total volume at 37^0C in a thermocycler (Illumina Nextera DNA library prep kit, #FC1211030). DNA was cleaned up using MinElute PCR purification kit (#28006, Qiagen) and eluted in 10UL of EB buffer.

ATAC-Seq DNA was amplified, and the number of PCR cycles were calculated by following previously described protocol [58]. PCR products (10UL) were run on a 1.5% agarose gel for expected DNA pattern. PCR products were then cleaned up by double-sized selection by using Ampure beads (Agencout AMPure XP, Beckman Couler, #A63880) to remove large and small DNA fragments. This was performed by using 1:0.5 and 1:1.6 ratio of sample to Ampure beads (v/v). Completed ATAC-Seq libraries (8-10) were then analyzed by Fragment Bioanalyzer and sequenced for paired-end 75 cycles using the NextSeq 500 system with ~400-500 million reads per run.

Bulk RNA-Seq data analysis to identify differentially expressed genes

Using Illumina HiSeq with a paired-end read of 100bp (RNA-Seq), we measured gene expression profiles of 50 GFP+ (GFP-positive) and 50 GFP- (GFP-negative) samples (Table S1; Data S1). Raw data from mouse and zebrafish were separately mapped to the GRCm38/mm10 and GRCz10/danRer10 genome assembly using STAR [59]. Raw counts of genes were further used to calculate FPKM (Fragments Per Kilobase Of Exon Per Million) and identify differentially expressed genes through EdgeR [60].

To identify GFP+ (Müller glia cells, MG) or GFP- (non-Müller glia cells, non-MG) specific genes, we performed t-test of paired GFP+/GFP- samples from NMDA treatment and light damage models. Genes with fold change > 2 and FDR (false discovery rate) < 0.01 were chosen as GFP+ or GFP- specific expressed genes (Data S2). Furthermore, we identified species-highly expressed genes from GFP+ specific expressed genes. In each species and each GFP+ sample, we calculated the rank of gene expression. According to the rank of gene expression, we performed Wilcox rank test between two species to identify species-highly expressed genes in GFP+ samples.

We compared each time point of treated samples with control samples to identify temporally changed genes in GFP+ samples. If fold change > 1.5 and FDR < 0.05, the

gene was defined as significantly differentially expressed across different time points of treatment (Data S3). To identify consistently changed genes between NMDA treatment and light damage model, we calculated the Pearson's correlation coefficient at corresponding treatment timepoint in each species. Meanwhile, we shuffled the expression for each gene and then calculated the correlation as the control. Genes were regarded as the correlated expression in bulk RNA-Seq if the correlation > 0.5 between two models. In the mouse, the correlated and differentially expressed genes (cDEGs) are those > 0.5 correlation and significantly differential expression in both NMDA treatment and light damage model. In the zebrafish, cDEGs are those > 0.5 correlation and significantly differential expression in NMDA treatment or light damage model (Data S4).

To compare differentially or highly expressed genes between two species, we identified orthologs between mouse and zebrafish based on vertebrate ortholog information from MGI database (http://www.informatics.jax.org/). We then separated gene orthologs into six categories between mouse and zebrafish, which are mouse gene mapping to zero ortholog in zebrafish (1 vs. 0), zebrafish gene mapping to zero ortholog in mouse (0 vs. 1), one mouse/zebrafish gene mapping to one ortholog in the other species (1 vs.1), one mouse gene mapping to multiple zebrafish orthologs (1 vs. N), multiple mouse genes mapping to one zebrafish ortholog (N vs. 1), and multiple mouse/zebrafish genes mapping to multiple orthologs in the other species (N vs. N). For multiple orthologs were mapped, we chose the most significantly differentially expressed genes. If there is no significantly differential expression, we selected the most correlated and expressed genes between two models. In RNA-Seq data analysis of species-specific changed or highly expressed genes, we divided differentially or highly expressed genes into seven categories (four broad categories), including mouse-specific changed genes without orthologs in the zebrafish, zebrafish-specific changed genes without orthologs in the mouse, mouse-specific changed genes with orthologs, zebrafish-specific changed genes with orthologs, genes changed in both mouse and zebrafish, mouse highly expressed genes, and zebrafish highly expressed genes (Data S5).

In addition, we performed K-means of GFP+ samples to separate cDEGs into three groups, including injury-repressed genes, rapidly induced genes and slowly induced genes. Function enrichment analysis on genes are conducted in R package 'clusterProfiler' [61]. Hierarchical clustering of all RNA-Seq samples, and principal component analysis (PCA) are also processed in R platform.

Single-cell RNA sequencing data analysis

Mapping and quality control of single-cell RNA-Seq data

We analyzed 19 mouse libraries, 12 chick libraries and 27 zebrafish libraries using single-cell RNA-Seq (scRNA-Seq, Data S1). Mouse, chick and zebrafish raw reads from scRNA-Seq were separately mapped to GRCm38/mm10, Gallus_gallus-5.0/galGal5 and GRCz10/danRer10 genome using STAR [62]. The gene-cell matrices were generated using Cell Ranger from 10X Genomics. We further calculated the number of genes and the number of UMIs (Unique molecular identifiers) for each cell. In the chick, we removed low-quality cells which have < 200 genes or < 500 UMIs. In mouse and zebrafish, cells with > 200 genes and > 1000 UMIs were included for further

analysis. Meanwhile, for each sample we also removed the outlier cells which have > 2-fold of average number of genes, or > 2-fold of average number of UMIs across all cells in each library.

Single-cell clustering

For all single cells from each treatment, we first identified variable genes which have 0.1~8 expression level and > 1 dispersion. Based on the variable genes, we performed multiple alignment and clustering of single cells across different time points of treatment using Seurat [63]. Single-cell clusters were identified through k-nearest neighbors and a shared nearest neighbor modularity optimization. Then single-cell clustering was visualized through t-SNE [64]. In analysis, top 5000 variable genes were used in mouse NMDA and light damage treatment, and zebrafish T+R treatment. In the zebrafish NMDA and light damage, we separately used top 8000 and 6000 variable genes. In the chick, top 4000 variable genes were used. No obvious batch effects were observed after performing multiple alignments of single cells.

Identification of cell doublets in single-cell RNA sequencing

In order to identify cell doublets in our scRNA-Seq data, we developed a new algorithm. First, we performed single-cell clustering and identified marker genes for each cluster (named cluster markers). For each cluster marker, we then used AUC (area under the receiver operating characteristic curve of gene expression) to calculate the gene power ($GP_{i,j}$) which indicates the capability of marker *j* to distinguish cluster *i* and the other clusters. For each cell and each cluster, we next calculated summed powers (*SP*) following the formula:

$$\mathrm{SP}_i = \sum_{j=1}^{j=N_i} GP_{i,j}$$

Where N_i is the number of markers in cluster *i*. Meanwhile, we calculated the threshold of *SP* in each cluster indicating whether the cell expresses markers in that cluster. The threshold is obtained by performing statistics on *SP* of all cells for each cluster.

Based on *SP* of the cell and threshold, we determined cell doublets. If cell simultaneously expressed markers from more than one cluster, the cell is regarded as a doublet. In each scRNA-Seq clustering, we identified 5~10% cell doublets. Cell doublets were removed for further analysis.

Identification of cell types and markers for each cluster

In analysis, we used known specific marker genes to identify cell type for each cluster of cells (Data S7). Most of known marker genes were reported in the mouse. For the chick and zebrafish, we identified orthologues of known mouse marker genes, and used those markers to determine cell types of clusters. Meanwhile, we identified all potential markers of each cluster (or cell type) in each injury model and species by setting > 0.3 gene power (Data S8).

To identify species-specific or conserved markers for each cell type, we calculated the powers of markers in the treatment models of each species. The markers

of the same cell types or sub-cell types identified in any model were used to perform the comparison across species. If the same markers (derived from the orthologues) were identified in different species, we regarded these markers are species-conserved (Data S9).

Hierarchical tree of the similarity of cell types across species

After identifying cell types, we further used marker genes to identify the relationship among cell types from different species. We chose markers in each cell type by setting gene power (*GP*) > 0.1. We then calculated a relative power (*RP*) as the similarity of two cell types following the formula:

$$RP = \frac{\sum_{t=1}^{T} GP_{I,t} + \sum_{t=1}^{T} GP_{II,t}}{\sum_{n=1}^{N} GP_{I,n} + \sum_{m=1}^{M} GP_{II,m}}$$

Where *T* is the number of shared markers between cell type I and cell type II, *N* is the number of markers from cell type I, and *M* is the number of markers from cell type II.

Based on *RP*, we performed hierarchical clustering to determine the evolutionary distances among cell types from different species. Meanwhile, we compared chick and zebrafish cell types with mouse cell types using the same measurement.

Trajectory and pseudo-time analysis of single cells

After identifying cell types of single-cell clustering, we constructed the trajectory of MG in response to retinal injury and performed pseudo-time analysis using Monocle [65]. Variable genes were used to construct the trajectory of MG. Pseudo-time of cells was calculated by setting unstimulated control cells as the root.

Furthermore, we identified significantly pseudo-temporally changed genes (PCGs), which are differentially expressed as a function of pseudo-time (Data S12). We used the following criteria for PCGs: fraction of expressed cells > 0.01, single-cell expression difference > 0.1 and q-value < 0.001. Single-cell expression difference was calculated through the formula:

Single-cell expression difference = Q95-expression – Q5-expression.

Where, Q95-expression and Q5-expression separately represent 0.95 and 0.05 quantile of log-transformed single-cell expression values across all bins of pseudo-time.

In order to calculate single-cell expression difference, we separated single-cell expression profiles across pseudo-time into 50 bins with equal pseudo-time interval. Single-cell expression profiles were obtained by averaging all cells in each bin (bin-derived expression). Then, quantile values of single-cell expression values were calculated across all 50 bins.

For each gene, we also used bin-derived expression to calculate the single-cell expression correlation between two treatment models. For each species, PCGs in all measured models were further selected according to single-cell expression correlation. The cutoff of single-cell expression correlation for analysis was set to 0.2 based on the distribution of the correlation of shuffled single-cell expression. In the zebrafish, correlated PCGs were defined if expression correlation > 0.2 in comparison of NMDA treatment with each of light damage and T+R treatment (Data S13).

To compare the genes across species, we also identified orthologues in the chick through similar analysis to identify orthologs between mouse and zebrafish. According

to significantly pseudo-temporal change of gene expression in three species, correlated PCGs were divided into seven categories, three of which are species-specific correlated PCGs, three are correlated and changed in two species, and one are shared by three species (Data S14).

In the trajectories of chick and zebrafish MG, we found three obvious branches. Using monocle, we used single-cell expression as a function of branch state to identify branch-specific expressed genes (BEGs). BEGs were defined as the fold change of expression between one branch and the others > 0.05 and q-value < 0.001 (Data S11, S18).

We aligned MG trajectories from two species through cellAlign [66]. In alignment, we separated MG based on branch state and then sorted MG using pseudo-time. Correlated PCGs shared by two species were used in trajectory alignment.

Comparative analysis of MG from zebrafish development and injury models

We first used Seurat to identify cell subpopulations of zebrafish development cells (Data S16). Single-cell clustering was visualized through UMAP [67]. Based on known marker genes of retinal cell types (Data S7), we identified cell types for each cell subpopulation. We then selected 12,680 primary RPC (retinal progenitor cells), neurogenic RPC cells and MG to construct a trajectory and identify pseudo-temporally changed genes (PCGs) during MG development (Data S17). To compare each injury model with retinal development in the zebrafish, we combined 12,680 developmental cells with those MG cells following NMDA treatment, light damage and T+R treatment. Meanwhile, we identified the overlapped PCGs between retinal development and injury model. Using overlapped PCGs, we reconstructed MG trajectory for each injury model combining retinal development. This analysis was performed using Monocle software [65].

For the trajectories with obvious branches, we identified branch-specific expressed genes (BEGs) through a similar method performed in the chick. The same criteria (fold-change of expression between branches > 0.05 and q-value < 0.001) were used to select BEGs (Data S18).

In addition, we used AUCell to identify cells with active gene expression in the trajectory [68]. Active gene expression of single cells was weighted through AUC (Area Under the Curve) score, which indicates the significance of cells expressing a set of genes. The cutoff of AUC score was used to identify actively expressed cells.

ATAC-Seq data analysis and reconstructing regulatory network

Mapping and normalization of ATAC-Seq

In this study, we measured 20 zebrafish GFP+ and 20 mouse GFP+ samples using ATAC (Assay for Transposase-Accessible Chromatin) sequencing technology (Table S1; Data S1). After removing adaptors using cutadapt [69], 50bp paired-end ATAC-Seq reads from zebrafish and mouse were separately aligned to GRCz10/danRer10 and GRCm38/mm10 reference genome using Bowtie2 with default parameters [70]. We used a similar analysis as our previous publication on ATAC-Seq analysis [71]. Briefly, we filtered reads from chromosome M and Y, and included high mapping-quality reads (MAPQ score > 10) for further analysis. Duplicate reads were removed using Picard tools MarkDuplicates program (http://broadinstitute.github.io/picard/). ATAC-Seq peak regions were called using MACS2 with parameters --nomodel --shift -100 --extsize 200

[72]. The blacklisted regions in mouse were excluded from peak regions (https://www.encodeproject.org/annotations/ENCSR636HFF/). We further merged the ATAC-Seq peaks from each sample to get a union set of peaks (or open chromatin regions). Totally we identified 248,761 peaks in the zebrafish and 119,718 peaks in the mouse (Data S19). We counted the raw fragments (C_R) for each peak using HTSeq [73]. Normalized fragments (C_N) were calculated through the formula:

$$C_N = log_2(\frac{C_R}{S_L} \times 100{,}000{,}000 + 1)$$

Where S_L is the sequencing library size of the sample.

Identification of ATAC-Seq footprints and differentially accessible regions

In open chromatin regions, we identified the footprints using DNase2TF (Data S20). Meanwhile, we identified differentially accessible regions (DARs) from open chromatin regions. To identify DARs, we included 150,148 peaks in zebrafish which have > 6 normalized fragments, and 104,442 peaks in mouse which have > 5 normalized fragments in average across all samples. EdgeR was used to identify DARs in comparison of each time point of treatment with control ATAC-Seq samples [60]. In the zebrafish, we chose fold change > 1.5 and FDR < 0.05 as the criteria of DARs. In the mouse, the criteria of DARs is fold change > 2 and FDR < 0.01. Totally, we identified 56,939 DARs in the zebrafish and 77,447 DARs in the mouse (Data S21).

To identify correlated DARs, we calculated the correlation of chromatin accessibility between NMDA treatment and light damage models. Shuffled chromatin accessibility for each peak was used to obtain the random correlations, and > 0.5 correlation was chosen to the criteria of correlated DARs. From all DARs, we identified 24,515 correlated DARs in the zebrafish and 52,297 correlated DARs in the mouse (Data S22).

Reconstructing an integrated regulatory network

In order to integrate bulk RNA-Seq, scRNA-Seq and ATAC-Seq data to reconstruct regulatory network, we developed a new method named ***I***ntegrated ***Re***gulatory ***N***etwork ***A***nalysis (IReNA) (Figure S16A).

First, we identified candidate transcription factors (TFs) following the criteria: (a) Expression level > 5 FPKM for bulk RNA-Seq, and > 0.1 fraction of expressed cells for scRNA-Seq; (b) Expression correlation between NMDA and light damage model > 0.5 for bulk RNA-Seq, and > 0.2 for scRNA-Seq; (c) Differentially expressed in treatment vs. control (identified from above RNA-Seq and scRNA-Seq analysis) or highly expressed in MG. We identified 192 candidate TFs in the zebrafish and 212 candidate TFs in the mouse (Data S25). Corresponding motif information of TFs was extracted from TRANSFAC 2018.3 version.

Next, we separated all differentially or highly expressed genes including candidate TFs (candidate genes) into ten modules through K-means clustering in each species. There are two groups of candidate genes, one is differentially or highly expressed in scRNA-Seq, the other is RNA-Seq specific. We separated the former group into seven modules and divided the latter group into three modules.

For each gene in the modules, we used FIMO to identify DNA motifs within its footprints of correlated DARs (P value < 1.0E-04) [74]. Meanwhile, we calculated footprint

occupancy score (FOS) for each footprint in each ATAC-Seq sample. The formula of FOS is the same as previously described [71],

$$\mathrm{FOS} = \min\left(-\log_2 \frac{N_\mathrm{C}+1}{N_\mathrm{L}+1}, -\log_2 \frac{N_\mathrm{C}+1}{N_\mathrm{R}+1}\right)$$

where N_C indicates the counts of ATAC-Seq inserts in central region of TF motif. The size of the central region is equal to the length of TF motif. N_L and N_R are separately 1/3 of the counts of ATAC-Seq inserts in the left and right flanking regions of TF motif, as 3 times of the size of the central region were chosen as the size of flanking region. To further refine regulatory relationship between TFs and targets, we included binding motifs whose FOS > 1 in any one of ATAC-Seq samples. We obtained regulatory pairs between genes and corresponding TFs of binding motifs.

Then we used expression correlation to determine active (positive correlation) or repressive (negative correlation) regulations of TFs to target genes. If both TF and target gene of one regulatory pair are RNA-Seq specific, the correlation was calculated using bulk RNA-Seq data, otherwise single-cell expression correlation was used. For expression correlation calculating from bulk RNA-Seq data, we chose 0.8 as the cutoff of significant regulations of TFs to target genes. For single-cell correlation, > 0.5 and < -0.5 were selected to active and repressive regulations, respectively. We identified 26,083 regulations in the zebrafish and 30,177 regulations in the mouse (Data S24).

Enrichment of TFs regulating each gene module

For each gene module, we identified significantly regulated TFs using hypergeometric test. We calculated the probability of the TF regulating module A following the formula:

$$\Pr(x = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Where N is the total number of regulations, K is the number of all regulations targeting module A, n is the number of regulations from the TF, k is the number of regulations from the TF to module A, and $\binom{a}{b}$ is a binomial coefficient. We separately calculated p-value for active and repressive regulations. FDR was further calculated based on p-value.

Using FDR < 10E-6 in the zebrafish and FDR < 10E-5 in the mouse, we separately identified 97 zebrafish TFs and mouse 157 TFs which significantly regulated at least one of gene modules (Data S27). Furthermore, we extracted regulatory network consisting of enriched TFs from regulatory network of candidate genes (Figure S18 for zebrafish, Figure S19 for mouse, and Data S26).

Constructing an intramodular regulatory network

To simplify the regulation between any two modules of enriched TFs, we calculated the significance of regulations from module A to module B using hypergeometric test.

$$\Pr(x = q) = \frac{\binom{Q}{q}\binom{M-Q}{m-q}}{\binom{M}{m}}$$

Where M is the total number of regulations, Q is the number of all regulations from module A, m is the number of all regulations targeting module B, q is the number of regulations from module A to module B, and $\binom{a}{b}$ is a binomial coefficient. We assessed the significance of active and repressive regulations, respectively. We set FDR < 0.02 in zebrafish and FDR < 0.001 in mouse as the cutoff to obtain significant regulations between modules (Figure 6D-E).

Features of enriched TFs in regulatory network

We calculated regulation degrees of each enriched TF in regulatory network. The regulation degrees include total degree (number of regulations connecting with TFs), active indegree (number of regulations activated by TFs), repressive indegree (number of regulations repressed by TFs), active outdegree (number of regulations activating TFs), repressive outdegree (number of regulations repressing TFs). For regulatory network of candidate genes as well as regulatory network of enriched TFs, we separately calculated regulation degrees (D_R). Normalized regulation degrees (D_N) were calculated by the formula:

$$D_N = \frac{D_R}{D_T} \times 1000$$

Where D_T is the total number of regulations in gene regulatory network or the network of enriched TFs.

We also included other features of enriched TFs, such as significantly regulated modules in gene or TF regulatory network. Bulk and single-cell expression of enriched TFs is shown in Data S27 as well.

Supplemental Tables:

Table S1: Statistics of bulk RNA-Seq and ATAC-Seq data.

Table S2. List of antibodies used in the study.

Table S3. List of qPCR primers used in the study.

Supplemental Datasets:

Data S1. Information on all samples and libraries measured by RNA-Seq, ATAC-Seq and scRNA-Seq. Tab 1 (RNAseq): Information of RNA-Seq samples; Tab 2 (ATACseq): Information of ATAC-Seq samples; Tab 3 (scRNAseq): Information of scRNA-Seq samples.

Data S2. Gene expression data for all RNA-Seq samples, and list of GFP+ or GFP- specific expressed genes. Tab 1 (Mouse): List of genes specifically expressed in GFP+ and GFP- samples in mouse; Tab 2 (Zebrafish): List of genes specifically expressed in GFP+ and GFP- samples in zebrafish. FPKM (Fragments Per Kilobase Of Exon Per Million) values of gene expression are given in the tables starting from column G.

Data S3. List of differentially expressed genes (DEGs) following NMDA treatment and light damage in mouse and zebrafish. Tab 1 (Mouse): List of DEGs in mouse NMDA and light damage (LD) samples; Tab 2 (Zebrafish): List of DEGs in zebrafish NMDA and light damage samples. Statistics (log.fold.change, P.value and FDR) of each time point

of treatment vs. control are given. FPKM values of gene expression are given in the tables starting from column AC.

Data S4. List of correlated and differentially expressed genes (cDEGs) in mouse and zebrafish. Tab 1 (Mouse): List of cDEGs in mouse; Tab 2 (Zebrafish): List of cDEGs in zebrafish. FPKM values of gene expression are given starting from column E.

Data S5. List of genes showing species-specific changes in expression pattern or level. Tab 1 (Differential.expressed.genes): List of genes showing species-specific changes in expression pattern; FPKM values of gene expression in GFP+ samples are given in the table starting from column F. Tab 2 (Highly.expressed.genes): List of genes showing species-specific changes in baseline expression. Ratio of FPKM in GFP+ vs. corresponding GFP- samples are given in the table starting from column F.

Data S6. tSNE clustering of retinal cells in mouse, chicken and zebrafish. Tab 1 (Mouse.NMDA): tSNE clustering of retinal cells in mouse NMDA treatment; Tab 2 (Mouse.LD): tSNE clustering of retinal cells in mouse light damage (LD); Tab 3 (Chick.NMDA.GF): tSNE clustering of retinal cells in chick NMDA or growth factor (GF) treatment; Tab 4 (Zebrafish.NMDA): tSNE clustering of retinal cells in zebrafish NMDA treatment; Tab 5 (Zebrafish.LD): tSNE clustering of retinal cells in zebrafish light damage; Tab 6 (Zebrafish.T+R): tSNE clustering of retinal cells in zebrafish TNFa+RO4929097 (T+R) treatment.

Data S7. List of known specific marker genes for retinal cell types. Tab 1 (Mouse): Known specific marker genes for mouse retinal cell types; Tab 2 (Chick): Known specific marker genes for chick retinal cell types; Tab 3 (Zebrafish): Known specific marker genes for zebrafish retinal cell types; Tab 4 (Abbreviation.of.cell.types): Full name of retinal cell types in abbreviation.

Data S8. Specific marker genes of retinal cell types from mouse, chicken and zebrafish. Tab 1 (Mouse): Specific marker genes of retinal cell types from mouse NMDA treatment and light damage (LD); Tab 2 (Chick): Specific marker genes of retinal cell types from chick NMDA/growth factor (GF) treatment; Tab 3 (Zebrafish): Specific marker genes of retinal cell types from zebrafish NMDA treatment, light damage (LD) and TNFα+RO4929097 (T+R) treatment.

Data S9. Species-conserved marker genes of each retinal cell type. Tab 1 (Shared.by.mouse.and.chick): Conserved marker genes of retinal cell types shared by mouse and chick; Tab 2 (Shared.by.mouse.and.zebrafish): Conserved marker genes of retinal cell types shared by mouse and zebrafish; Tab 3 (Shared.by.chick.and.zebrafish): Conserved marker genes of retinal cell types shared by chick and zebrafish; Tab 3 (Shared.by.three.species): Conserved marker genes of retinal cell types shared by mouse, chick and zebrafish.

Data S10. Trajectory and pseudotime of mouse, chick and zebrafish MG. Tab 1 (Mouse): Trajectory and pseudotime of mouse MG; Tab 2 (Chick): Trajectory and pseudotime of chick MG; Tab 3 (Zebrafish): Trajectory and pseudotime of zebrafish MG.

Data S11. List of branch-specific genes in chick MG. Differential proportion of expressed cells in current branch was calculated by subtracting the proportion of expressed cells in the other branches.

Data S12. List of pseudo-temporally changed genes (PCGs) in mouse, chicken and zebrafish MG. Tab 1 (Mouse): List of PCGs in mouse MG; Tab 2 (Chick): List of PCGs in chick MG; Tab 3 (Zebrafish): List of PCGs in zebrafish MG. Fractions of expressed cells are given in Tab 1 starting from column K, in Tab 2 starting from column K, in Tab 3 starting from column P.

Data S13. List of correlated and pseudo-temporally changed genes (cPCGs) in mouse, chicken and zebrafish MG. Tab 1 (Mouse): List of cPCGs in mouse MG; Tab 2 (Chick): List of cPCGs in chick MG; Tab 3 (Zebrafish): List of cPCGs in zebrafish MG. Fractions of expressed cells are given in Tab 1 starting from column K, in Tab 2 starting from column K, in Tab 3 starting from column P.

Data S14. List of genes showing species-specific changes in single-cell expression pattern or level in mouse, chick and zebrafish MG. Tab 1 (Pseudo-temporally.changed.genes): List of genes showing species-specific changes in single-cell expression pattern; Tab 2 (Highly.expressed.genes): List of genes showing species-specific changes in single-cell baseline expression. Fractions of expressed cells are given in the tables starting from column AC.

Data S15. List of correlated and differentially expressed genes identified by RNA-Seq and scRNA-Seq data analysis. Tab 1 (Mouse): List of correlated and differentially expressed genes identified by RNA-Seq and scRNA-Seq data analysis in mouse; Tab 2 (Zebrafish): List of correlated and differentially expressed genes identified by RNA-Seq and scRNA-Seq data analysis in zebrafish. FPKM values and fractions of expressed cells are given separately for RNA-Seq and scRNA-Seq samples.

Data S16. UMAP clustering of retinal cells in zebrafish development. Cell type and UMAP coordinates (UMAP.1, UMAP.2 and UMAP.3) of each retinal cell in zebrafish development are shown.

Data S17. List of pseudo-temporally changed genes in zebrafish MG development. Fractions of expressed cells are given starting from column G.

Data S18. List of branch-specific expressed genes (BEGs) in zebrafish MG following NMDA, light damage (LD) and TNFα+RO4929097 (T+R) treatment. Tab 1 (NMDA): List of BEGs in zebrafish MG following NMDA treatment; Tab 2 (LD): List of BEGs in zebrafish MG following light damage; Tab 3 (T+R): List of BEGs in zebrafish MG following T+R treatment. Differential proportion of expressed cells in current branch was calculated by minusing the proportion of expressed cells in the other branches.

Data S19. Counts of open chromatin regions (OCRs) in zebrafish and mouse MG identified by ATAC-Seq analysis. Tab 1 (Zebrafish): Counts of OCRs in zebrafish MG; Tab 2 (Mouse): Counts of OCRs in mouse MG. Raw counts of each peak are given starting from column H.

Data S20. Transcription factor footprints detected in open chromatin regions (OCRs) of zebrafish and mouse MG. Tab 1 (Zebrafish): Transcription factor footprints detected in OCRs of zebrafish MG; Tab 2 (Mouse): Transcription factor footprints detected in OCRs of mouse MG. The column log10.p.value represents the significance of TF footprint as predicted by Dnase2tf.

Data S21. Differentially accessible chromatin regions (DARs) in zebrafish and mouse MG. Tab 1 (Zebrafish): DARs in zebrafish MG; Tab 2 (Mouse): DARs in mouse MG. Normalized counts are given in Tab 1 starting from column Z, and in Tab 2 starting from column AB.

Data S22. Correlated and differentially accessible regions (cDARs) in zebrafish and mouse. Tab 1 (Zebrafish): cDARs in zebrafish MG; Tab 2 (Mouse): cDARs in mouse MG. Normalized counts are given in Tab 1 starting from column Z, and in Tab 2 starting from column AB.

Data S23. Gene list of ten modules in zebrafish and mouse MG. Tab 1 (Zebrafish): Gene list of ten modules in zebrafish MG; Tab 2 (Mouse): Gene list of ten modules in mouse MG. Fractions of expressed cells and FPKM values are given separately for scRNA-Seq and RNA-Seq samples.

Data S24. Regulatory network of regulators and genes in zebrafish and mouse MG. Tab 1 (Zebrafish): Regulatory network of regulators and genes in zebrafish MG; Tab 2 (Mouse): Regulatory network of regulators and genes in mouse MG. The last column contains information of TF binding sites in target gene. Motif ID from TRANSFAC database. FOS, Footprint occupancy score.

Data S25. Significance of candidate regulators regulating gene modules in zebrafish and mouse MG. Tab 1 (Zebrafish): Significance of candidate regulators regulating gene modules in zebrafish MG; Tab 2 (Mouse): Significance of candidate regulators regulating gene modules in mouse MG.

Data S26. Regulatory network of enriched regulators in zebrafish and mouse MG. Tab 1 (Zebrafish): Regulatory network of enriched regulators in zebrafish MG; Tab 2 (Mouse): Regulatory network of enriched regulators in mouse MG. The last column contains information of TF binding sites in target gene. Motif ID from TRANSFAC database. FOS, Footprint occupancy score.

Data S27. Features of enriched regulators in zebrafish and mouse MG. Tab 1 (Zebrafish): Features of enriched regulators in zebrafish MG; Tab 2 (Mouse): Features of enriched regulators in mouse MG. Tab 3 (Zebrafish.vs.Mouse): Features of enriched regulators across species. Gene module/network indicates module/network consisting of candidate genes (including TFs). TF module/network means module/network only consisting of enriched TFs. Indegree is the number of regulations from TFs. Outdegree is the number of regulations targeting TFs. Degree of TFs regulatory network is normalized by dividing the total number of regulations. TF expression is given using fractions of expressed cells for scRNA-Seq and FPKM values for RNA-Seq samples.

References

1. Wan, J. & Goldman, D. Retina regeneration in zebrafish. *Curr. Opin. Genet. Dev.* **40**, 41–47 (2016).

2. Goldman, D. Müller glial cell reprogramming and retina regeneration. *Nat. Rev. Neurosci.* **15**, 431–442 (2014).
3. Wilken, M. S. & Reh, T. A. Retinal regeneration in birds and mice. *Curr. Opin. Genet. Dev.* **40**, 57–64 (2016).
4. Gallina, D., Todd, L. & Fischer, A. J. A comparative analysis of Müller glia-mediated regeneration in the vertebrate retina. *Exp. Eye Res.* **123**, 121–130 (2014).
5. Hamon, A. *et al.* Linking YAP to Müller Glia Quiescence Exit in the Degenerative Retina. *Cell Rep.* **27**, 1712–1725.e6 (2019).
6. Todd, L., Palazzo, I., Squires, N., Mendonca, N. & Fischer, A. J. BMP- and TGFβ-signaling regulate the formation of Müller glia-derived progenitor cells in the avian retina. *Glia* **65**, 1640–1655 (2017).
7. Todd, L., Squires, N., Suarez, L. & Fischer, A. J. Jak/Stat signaling regulates the proliferation and neurogenic potential of Müller glia-derived progenitor cells in the avian retina. *Sci. Rep.* **6**, 35703 (2016).
8. Gallina, D., Palazzo, I., Steffenson, L., Todd, L. & Fischer, A. J. Wnt/β-catenin-signaling and the formation of Müller glia-derived progenitors in the chick retina. *Dev. Neurobiol.* **76**, 983–1002 (2016).
9. Ramachandran, R., Zhao, X.-F. & Goldman, D. Ascl1a/Dkk/beta-catenin signaling pathway is necessary and glycogen synthase kinase-3beta inhibition is sufficient for zebrafish retina regeneration. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15858–15863 (2011).
10. Fischer, A. J., Scott, M. A., Ritchey, E. R. & Sherwood, P. Mitogen-activated protein kinase-signaling regulates the ability of Müller glia to proliferate and protect retinal

neurons against excitotoxicity. *Glia* **57**, 1538–1552 (2009).

11. Fischer, A. J., Scott, M. A. & Tuten, W. Mitogen-activated protein kinase-signaling stimulates Müller glia to proliferate in acutely damaged chicken retina. *Glia* **57**, 166–181 (2009).
12. Nelson, C. M. *et al.* Tumor necrosis factor-alpha is produced by dying retinal neurons and is required for Muller glia proliferation during zebrafish retinal regeneration. *J. Neurosci.* **33**, 6524–6539 (2013).
13. Kassen, S. C. *et al.* CNTF induces photoreceptor neuroprotection and Müller glial cell proliferation through two different signaling pathways in the adult zebrafish retina. *Exp. Eye Res.* **88**, 1051–1064 (2009).
14. Wan, J., Zhao, X.-F., Vojtek, A. & Goldman, D. Retinal injury, growth factors, and cytokines converge on β-catenin and pStat3 signaling to stimulate retina regeneration. *Cell Rep.* **9**, 285–297 (2014).
15. Fausett, B. V., Gumerson, J. D. & Goldman, D. The proneural basic helix-loop-helix gene ascl1a is required for retina regeneration. *J. Neurosci.* **28**, 1109–1117 (2008).
16. Ramachandran, R., Fausett, B. V. & Goldman, D. Ascl1a regulates Müller glia dedifferentiation and retinal regeneration through a Lin-28-dependent, let-7 microRNA signalling pathway. *Nat. Cell Biol.* **12**, 1101–1107 (2010).
17. Jorstad, N. L. *et al.* Stimulation of functional neuronal regeneration from Müller glia in adult mice. *Nature* **548**, 103–107 (2017).
18. Rueda, E. M. *et al.* The Hippo Pathway Blocks Mammalian Retinal Müller Glial Cell Reprogramming. *Cell Rep.* **27**, 1637–1649.e6 (2019).
19. Kassen, S. C. *et al.* Time course analysis of gene expression during light-induced

photoreceptor cell death and regeneration in albino zebrafish. *Dev. Neurobiol.* **67**, 1009–1031 (2007).

20. Mo, A. *et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86**, 1369–1384 (2015).
21. Conner, C., Ackerman, K. M., Lahne, M., Hobgood, J. S. & Hyde, D. R. Repressing notch signaling and expressing TNFα are sufficient to mimic retinal regeneration by inducing Müller glial proliferation to generate committed progenitor cells. *J. Neurosci.* **34**, 14403–14419 (2014).
22. Nagashima, M., Barthel, L. K. & Raymond, P. A. A self-renewing division of zebrafish Müller glial cells generates neuronal progenitors that require N-cadherin to regenerate retinal neurons. *Development* **140**, 4510–4521 (2013).
23. Gorsuch, R. A. *et al.* Sox2 regulates Müller glia reprogramming and proliferation in the regenerating zebrafish retina via Lin28 and Ascl1a. *Exp. Eye Res.* **161**, 174–192 (2017).
24. Lahne, M., Li, J., Marton, R. M. & Hyde, D. R. Actin-Cytoskeleton- and Rock-Mediated INM Are Required for Photoreceptor Regeneration in the Adult Zebrafish Retina. *J. Neurosci.* **35**, 15612–15634 (2015).
25. Fischer, A. J., Zelinka, C. & Scott, M. A. Heterogeneity of glia in the retina and optic nerve of birds and mammals. *PLoS One* **5**, e10774 (2010).
26. Fischer, A. J., Scott, M. A., Zelinka, C. & Sherwood, P. A novel type of glial cell in the retina is stimulated by insulin-like growth factor 1 and may exacerbate damage to neurons and Müller glia. *Glia* **58**, 633–649 (2010).
27. Lenkowski, J. R. & Raymond, P. A. Müller glia: Stem cells for generation and

regeneration of retinal neurons in teleost fish. *Prog. Retin. Eye Res.* **40**, 94–123 (2014).

28. Clark, B. S. *et al.* Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. *Neuron* **102**, 1111–1126.e5 (2019).
29. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
30. Boekhoorn, K. *et al.* The microtubule destabilizing protein stathmin controls the transition from dividing neuronal precursors to postmitotic neurons during adult hippocampal neurogenesis. *Dev. Neurobiol.* **74**, 1226–1242 (2014).
31. Liu, S. *et al.* Assessing the model transferability for prediction of transcription factor binding sites based on chromatin accessibility. *BMC Bioinformatics* **18**, (2017).
32. Lenkowski, J. R. *et al.* Retinal regeneration in adult zebrafish requires regulation of TGFβ signaling. *Glia* **61**, 1687–1697 (2013).
33. Esain, V., Postlethwait, J. H., Charnay, P. & Ghislain, J. FGF-receptor signalling controls neural cell diversity in the zebrafish hindbrain by regulating olig2 and sox9. *Development* **137**, 33–42 (2010).
34. Taylor, S. M., Giuffre, E., Moseley, P. & Hitchcock, P. F. The MicroRNA, miR-18a, Regulates NeuroD and Photoreceptor Differentiation in the Retina of Zebrafish. *Dev. Neurobiol.* **79**, 202–219 (2019).
35. Vitorino, M. *et al.* Vsx2 in the zebrafish retina: restricted lineages through derepression. *Neural Dev.* **4**, 14 (2009).
36. de Melo, J. *et al.* Injury-independent induction of reactive gliosis in retina by loss of

function of the LIM homeodomain transcription factor Lhx2. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4657–4662 (2012).

37. Ueki, Y., Wang, J., Chollangi, S. & Ash, J. D. STAT3 activation in photoreceptors by leukemia inhibitory factor is associated with protection from light damage. *J. Neurochem.* **105**, 784–796 (2008).
38. Wang, X., Shu, Q., Ni, Y. & Xu, G. CRISPR-mediated SOX9 knockout inhibits GFAP expression in retinal glial (Müller) cells. *Neuroreport* **29**, 1504–1508 (2018).
39. Bergsland, M., Werme, M., Malewicz, M., Perlmann, T. & Muhr, J. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev.* **20**, 3475–3486 (2006).
40. Liu-Chittenden, Y. *et al.* Genetic and pharmacological disruption of the TEAD-YAP complex suppresses the oncogenic activity of YAP. *Genes Dev.* **26**, 1300–1305 (2012).
41. Zhao, X.-F. *et al.* Leptin and IL-6 family cytokines synergize to stimulate Müller glia reprogramming and retina regeneration. *Cell Rep.* **9**, 272–284 (2014).
42. Ueki, Y. *et al.* Transgenic expression of the proneural transcription factor Ascl1 in Müller glia stimulates retinal regeneration in young mice. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13717–13722 (2015).
43. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
44. Bernardos, R. L. & Raymond, P. A. GFAP transgenic zebrafish. *Gene Expr. Patterns* **6**, 1007–1013 (2006).

45. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. 1951. *Dev. Dyn.* **195**, 231–272 (1992).

46. Vihtelic, T. S., Soverly, J. E., Kassen, S. C. & Hyde, D. R. Retinal regional differences in photoreceptor cell death and regeneration in light-lesioned albino zebrafish. *Exp. Eye Res.* **82**, 558–575 (2006).

47. Vihtelic, T. S. & Hyde, D. R. Light-induced rod and cone cell death and regeneration in the adultalbino zebrafish (Danio rerio) retina. *Journal of Neurobiology* **44**, 289–307 (2000).

48. Powell, C., Cornblath, E., Elsaeidi, F., Wan, J. & Goldman, D. Zebrafish Müller glia-derived progenitors are multipotent, exhibit proliferative biases and regenerate excess neurons. *Sci. Rep.* **6**, 24851 (2016).

49. Fischer, A. J., Seltner, R. L., Poon, J. & Stell, W. K. Immunocytochemical characterization of quisqualic acid- and N-methyl-D-aspartate-induced excitotoxicity in the retina of chicks. *J. Comp. Neurol.* **393**, 1–15 (1998).

50. Alles, J. *et al.* Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* **15**, 44 (2017).

51. Thummel, R. *et al.* Characterization of Müller glia and neuronal progenitors during adult zebrafish retinal regeneration. *Exp. Eye Res.* **87**, 433–444 (2008).

52. Thummel, R., Bailey, T. J. & Hyde, D. R. In vivo electroporation of morpholinos into the adult zebrafish retina. *J. Vis. Exp.* e3603 (2011).

53. Lin, Y.-C. *et al.* c-Myb is an evolutionary conserved miR-150 target and miR-150/c-Myb interaction is important for embryonic development. *Mol. Biol. Evol.* **25**, 2189–2198 (2008).

54. Aguirre, A. *et al.* In vivo activation of a conserved microRNA program induces mammalian heart regeneration. *Cell Stem Cell* **15**, 589–604 (2014).

55. Nik, S. H. M. *et al.* Alzheimer's disease-related peptide PS2V plays ancient, conserved roles in suppression of the unfolded protein response under hypoxia and stimulation of γ-secretase activity. *Human Molecular Genetics* **24**, 3662–3678 (2015).

56. Jiang, Q. *et al.* yap is required for the development of brain, eyes, and neural crest in zebrafish. *Biochemical and Biophysical Research Communications* **384**, 114–119 (2009).

57. Fischer, A. J., Foster, S., Scott, M. A. & Sherwood, P. Transient expression of LIM-domain transcription factors is coincident with delayed maturation of photoreceptors in the chicken retina. *J. Comp. Neurol.* **506**, 584–603 (2008).

58. Zibetti, C., Liu, S., Wan, J., Qian, J. & Blackshaw, S. Epigenomic profiling of retinal progenitors reveals LHX2 is required for developmental regulation of open chromatin. *Commun Biol* **2**, 142 (2019).

59. Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinformatics* **51**, 11.14.1–19 (2015).

60. 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).

61. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* **16**, 284–287 (2012).

62. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

63. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

64. Linderman, G. C., Rachh, M., Hoskins, J. G., Steinerberger, S. & Kluger, Y. Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. *Nat. Methods* **16**, 243–245 (2019).

65. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).

66. Alpert, A., Moore, L. S., Dubovik, T. & Shen-Orr, S. S. Alignment of single-cell trajectories to compare cellular expression dynamics. *Nat. Methods* **15**, 267–270 (2018).

67. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4314

68. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

69. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

70. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

71. Wang, J. *et al.* ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration. *Nat. Commun.* **9**, 1364 (2018).

72. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).

73. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

74. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

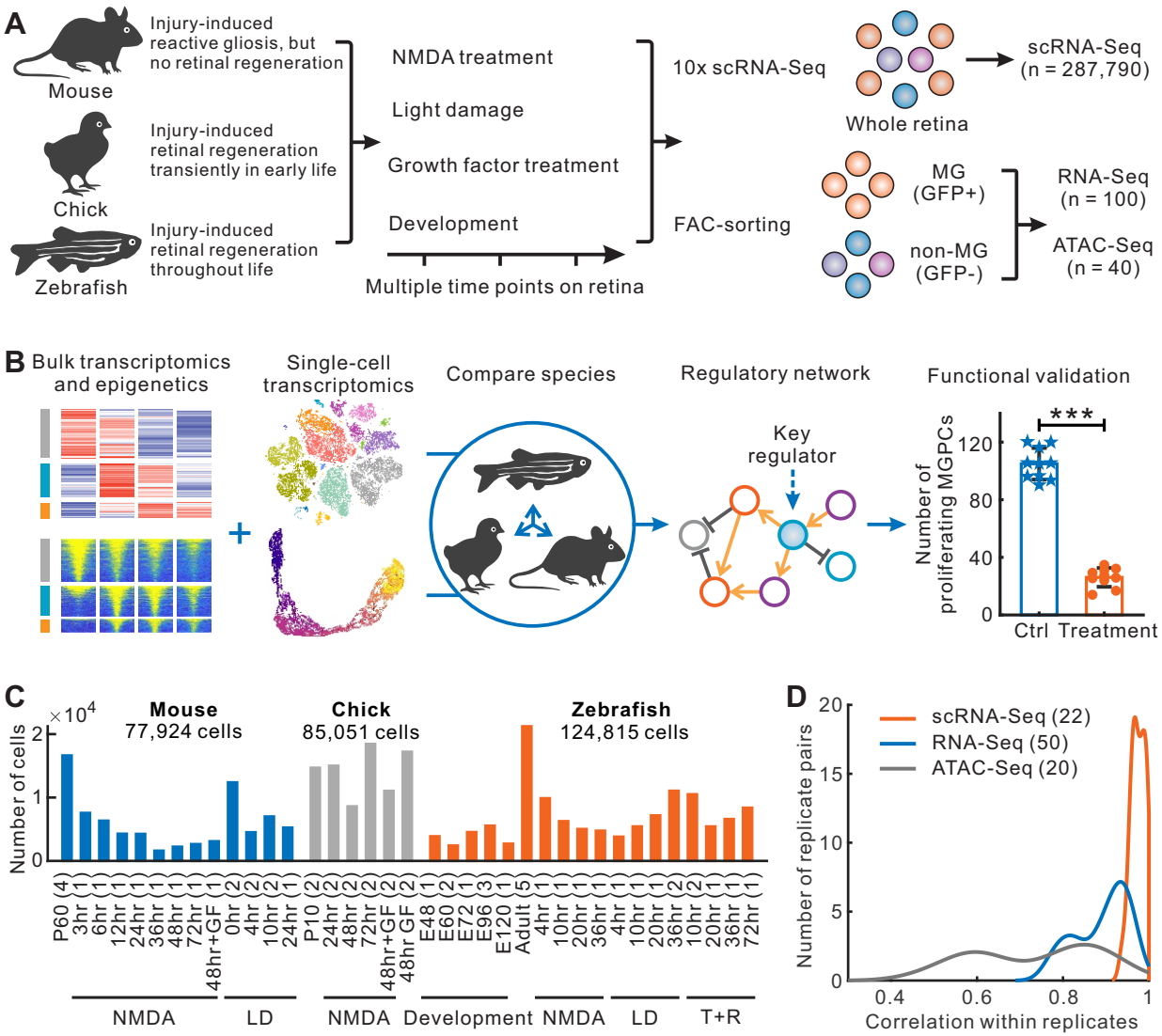

Figure 1, Hoang et al. Overview of study. (A) Schematics of the project. We performed single-cell RNA-Seq (scRNA-Seq) for thewhole retina, bulk RNA-Seq and ATAC-Seq for FAC-sorting (fluorescence activated cell sorting) Müller glia (MG)/non-MG from multiple treatment time-points in three species. (B) Schematic of data analysis and functional validation. MGPCs, Müller glia-derived progenitor cells. (C) Number of single cells from scRNA-Seq. The number in brackets indicates the number of sequenced libraries. LD, light damage. GF, growth factor (insulin+FGF). T+R, TNFα and RO4929097 treatment. (D) Distribution of the correlations for all replications from scRNA-Seq, RNA-Seq and ATAC-Seq. The number in brackets of the legend indicates the number of replicates analyzed.

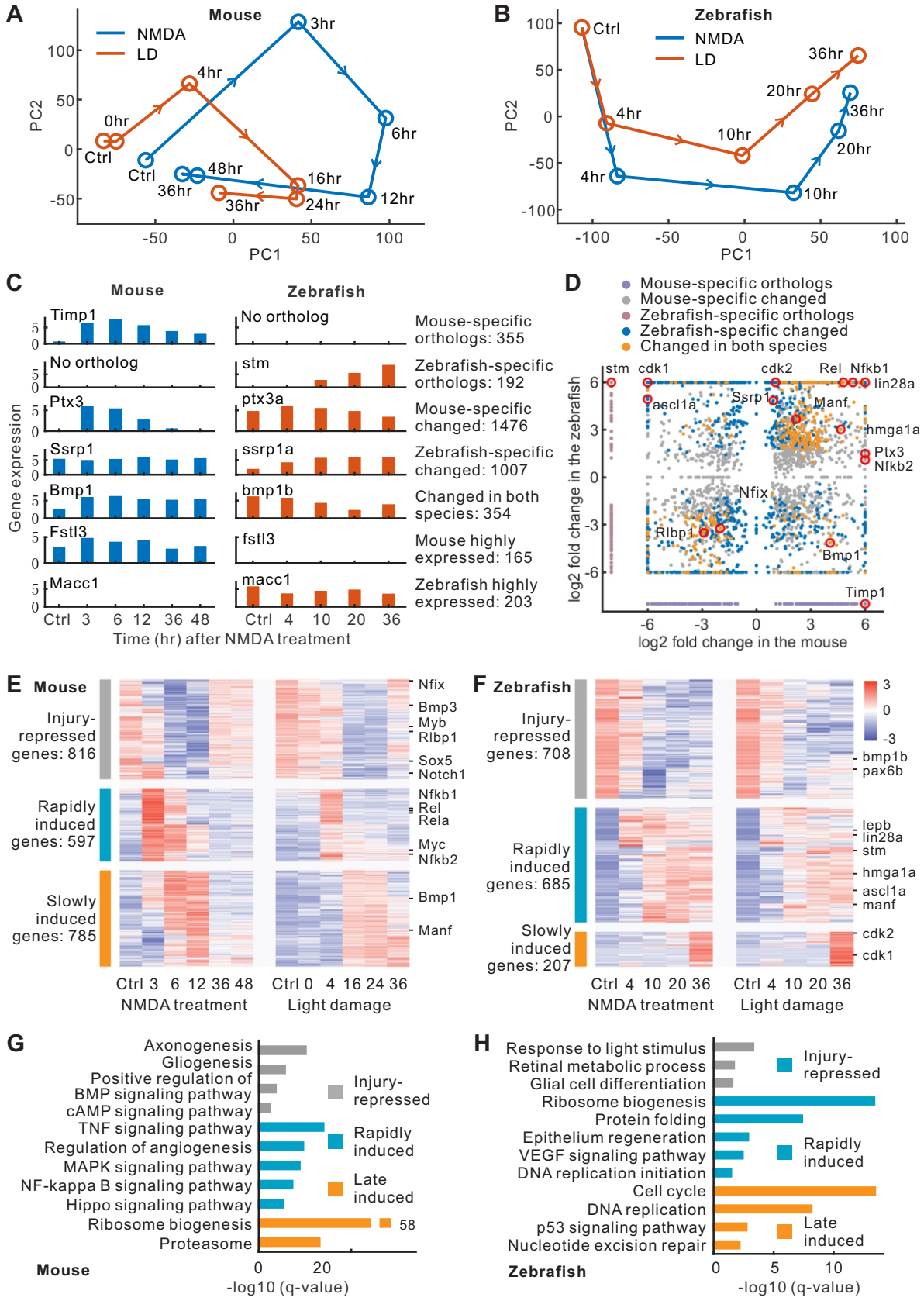

Figure 2, Hoang et al. Dynamic bulk transcriptomes in mouse and zebrafish MG. (A-B) Principal component (PC) analysis of MG expression from retinal injury models in mouse and zebrafish. (C) Examples of genes showing species-specific changes or differences in baseline expression. (D) Comparison of correlated and differentially expressed genes (cDEG) between mouse and zebrafish. The maximal fold-change between the treatment and the control was used. (E-F) Heatmap of cDEGs in mouse and zebrafish. Scaled expression of GFP+ samples is shown in color bar. (G-H) Most highly enriched functions of cDEGs.

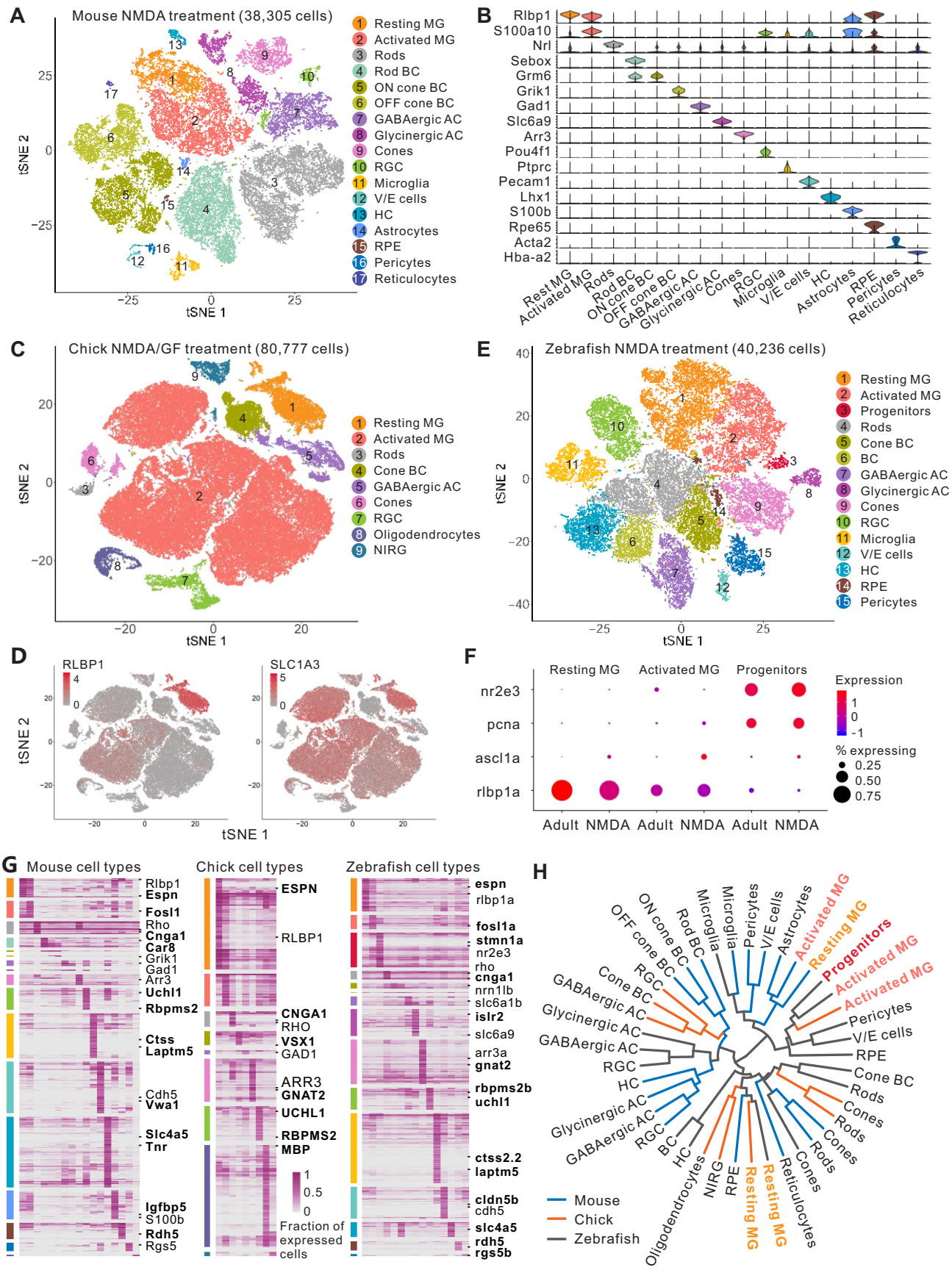

Figure 3, Hoang et al. Single-cell transcriptomes from mouse, chick and zebrafish retinas. (A) tSNE view of retinal cells from mouse NMDA treatment. (B) Violin plot showing expression of known marker genes in mouse retinal cells. (C) tSNE view of chick retinal cells. (D) Single-cell expression of MG marker genes in the chick. (E) tSNE view of zebrafish retinal cells. (F) Gene expression of MG and progenitors in adult control and NMDA-treated samples. (G) Known and new (in bold) cell-type specific marker genes in all three species. Each column represents one cell type. (H) Hierarchical tree measuring transcriptional similarity of cell types in mouse, chick and zebrafish. The tree is constructed based on the relative power of shared markers between cell types. MG, Müller glia cells; BC, bipolar cells; AC, amacrine cells; RGC, retinal ganglion cells; V/E cells, vascular/endothelial cells; HC, horizontal cells; RPE, retinal pigment epithelium; NIRG, non-astrocytic inner retinal glial cells.

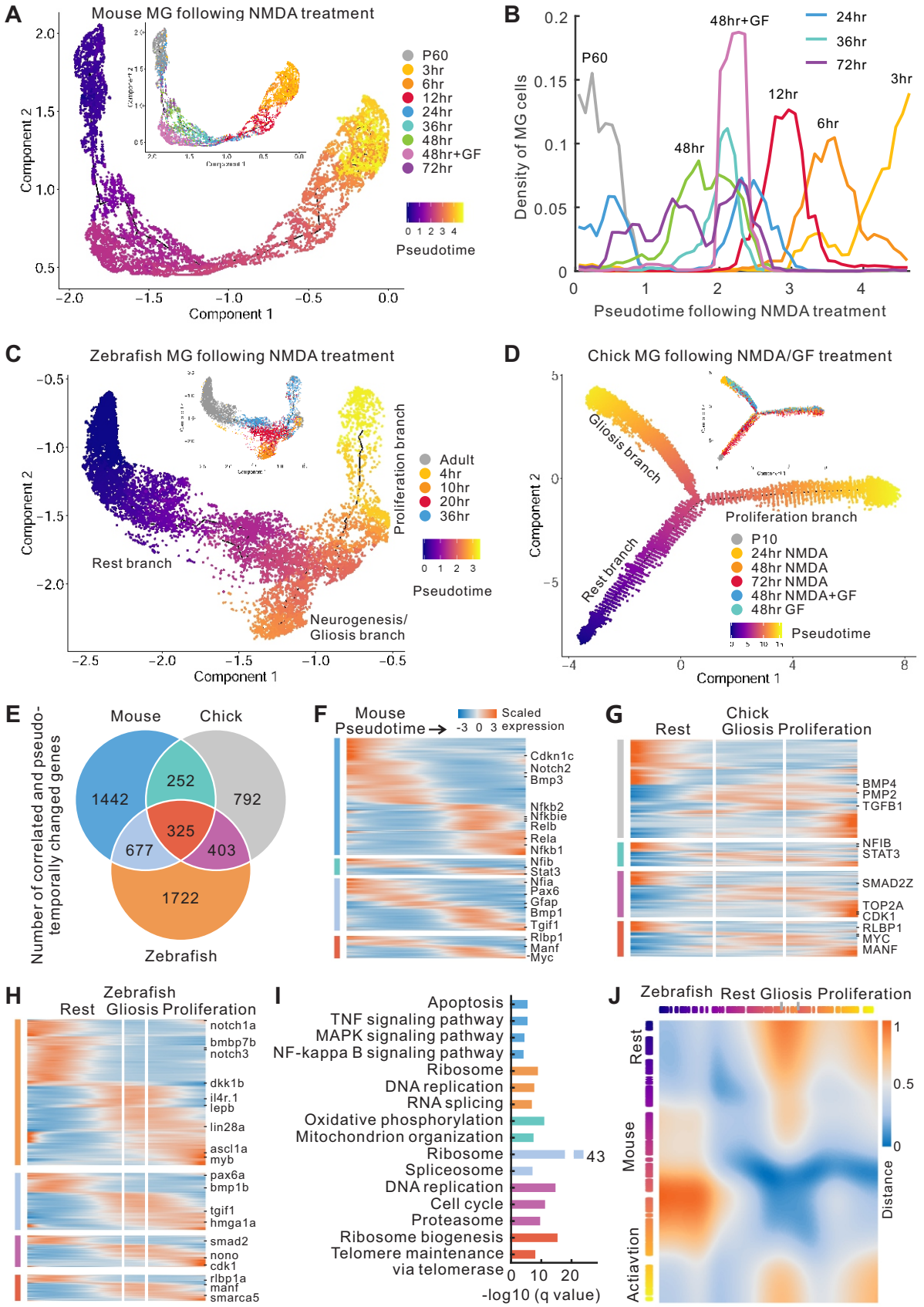

Figure 4, Hoang et al. Trajectory and pseudotime analysis of mouse, chick and zebrafish MG. (A) Trajectory of mouse MG following NMDA treatment. Inserted figure includes real time of MG measured by scRNA-Seq. GF, growth factor (insulin+FGF) treatment. (B) The density of mouse MG across the pseudotime. (C) Trajectory of zebrafish MG following NMDA treatment. (D) Trajectory of chick MG. (E) Venn diagram of the number of correlated and pseudo-temporally changed genes (cPCGs) in three species. (F-H) Heatmap of cPCGs from three species. Color bars on the left correspond to the categories of cPCGs in Figure 4E. Genes expressed in scRNA-Seq data are labeled on the right. (I) Top enriched functions of cPCGs. Colors represent the categories of species-specific or shared cPCGs corresponding to Figure 4E. (J) Alignment of two MG trajectories from mouse and zebrafish. The trajectory is colored to indicate pseudotime. The color of the heatmap represents the cell dissimilarity.

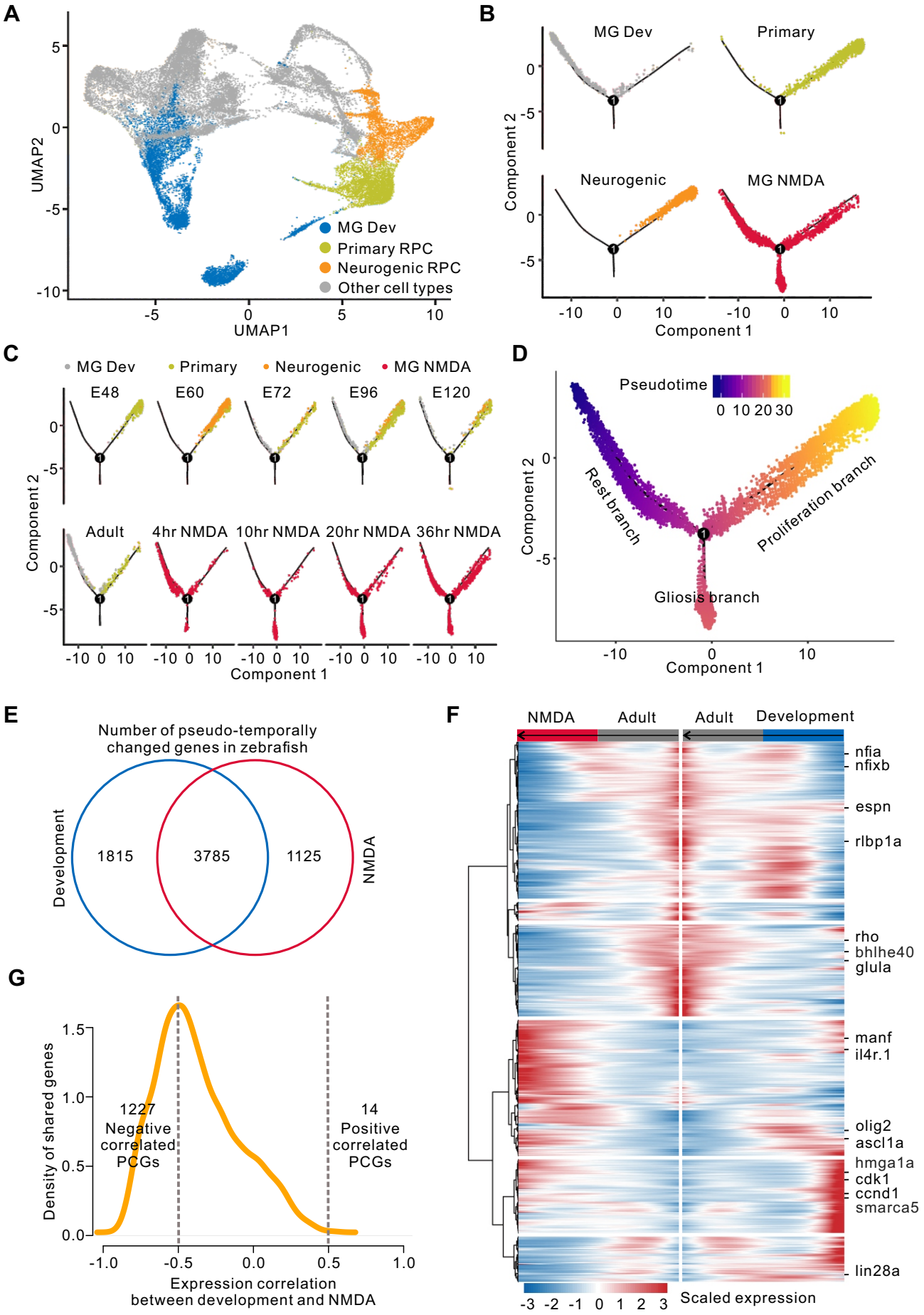

Figure 5, Hoang et al. Comparison of developing MG/RPCs and NMDA-treated MG in zebrafish. (A) UMAP view of cell types from zebrafish development. (B) MG/RPC trajectory from zebrafish development and NMDA treatment, separated by cell types. (C) Trajectory of zebrafish MG, separated by real time. (D) Pseudotime of zebrafish MG. (E) Venn diagram of the number of pseudo-temporally changed genes (PCGs) from zebrafish development and NMDA treatment. (F) Heatmap of single-cell expression of PCGs shared by developing MG and NMDA-treated MG. Scaled expression profile for each gene is shown in the heatmap. The column represents single cells ordered by pseudotime. (G) Density of expression correlation for overlapping PCGs.

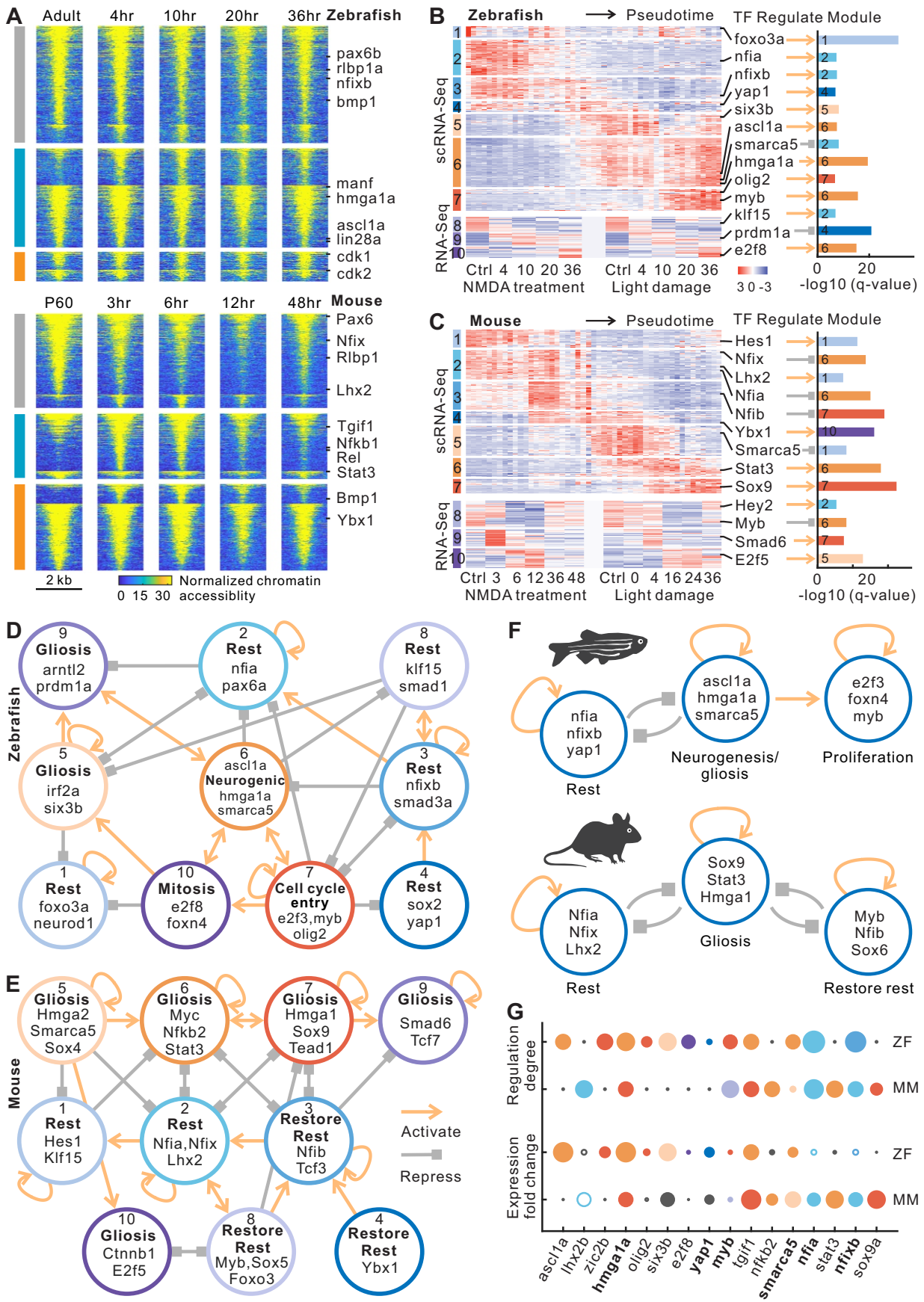

Figure 6, Hoang et al. MG regulatory network response to retinal injury. (A) Changes in chromatin accessibility of correlated differentially expressed genes (cDEGs) following NMDA treatment in zebrafish and mouse. Each line indicates one correlated and differentially accessible regions (cDARs) which are related to injury-repressed (grey), rapidly induced (cyan) and slowly induced (orange) cDEGs from Figure 2E-F. In each group, cDARs are separated by their positive and negative correlations with expression, and then ranked by the intensity of chromatin accessibility. (B-C) Ten modules are derived from differentially and highly expressed genes in zebrafish and mouse. In each module (color bars on the left), enriched TFs and their most significantly regulated modules are shown. (D-E) Intramodular regulatory network in zebrafish and mouse. Colored circles represent the modules. The connection indicates significant regulation among modules. (F) Super regulatory network in the zebrafish and mouse. (G) Comparison of TF features between zebrafish (ZF) and mouse (MM). TFs are ordered by the difference of expression fold changes between zebrafish and mouse. Color represents TF module. Black circle indicates the TF is not an enriched TF in corresponding species. Circle size indicates fold change of single-cell expression, or the number of TF regulations. Hollow circles indicate decreased fold change in treatment vs. control.

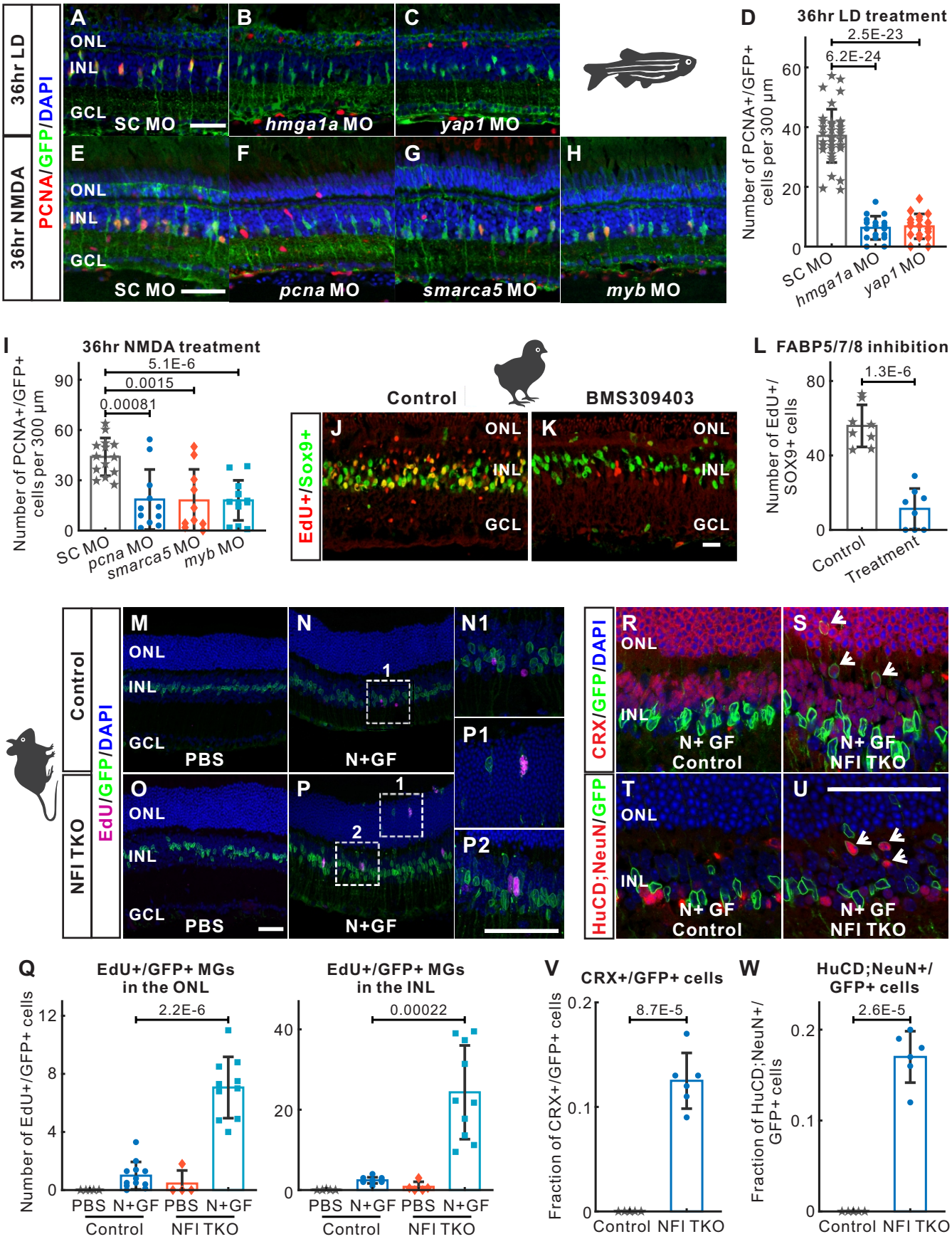

Figure 7, Hoang et al. Functional validation of key regulators of proliferative and neurogenic competence in zebrafish, chick and mouse. (A-D) In zebrafish, morpholino-mediated knockdown of hmga1a and yap1 inhibited MG proliferation after 36hr light damage. (E-I) Knockdown of smarca5 and myb significantly decreased MG proliferation after NMDA damage, as compared with the control morpholino. Pcna morpholinos were used as a positive control. (J-L) In chick, pharmacological inhibition of FABP5/7/8 using BMS309403 significantly inhibited MGPC formation 3 days after NMDA injury. (M-P) In mice, intravitreal injection of NMDA and growth factor significantly increased MG proliferation in both ONL and INL layers (Q) in the NFI TKO retina as compared with the control at 3 days after injection. (N1, P1, P2) Higher magnifications of (Q) and (S), respectively. (R-U) CRX+ and HuCD;NeuN+ cells derived from MG were found at 10 days after injection (white arrow head). (V-W) Quantification of CRX+/GFP+ and HuCD;NeuN+/GFP+ cells from total GFP+ cells, respectively. ONL, outer nuclear layer; INL, inner nuclear layer; GCL, ganglion cell layer. Scale bar, 20 µm.