Supplementary Methods

Sample Preparation and Collection

Retinal dissection and dissociation protocol
Eyes were removed from dark-adapted (≥ 2 hours) light-lesioned adult transgenic Tg(gfap:EGFP)mi2002 zebrafish at 8 and 16 hours post-lesion (hpl) or unlesioned control Tg(gfap:EGFP)mi2002 fish (0 hpl). The anterior chamber was removed with microscissors, retinas were detached from the retinal pigmented epithelium, the optic nerve head was cut, and the neural retina was removed from the eyecup and minced with a single-edged razor blade. Retinal fragments were digested (6 retinas in 1 mL) in StemPro Accutase (LifeTechnologies) and 100 µg/ml DNAseI (Roche) at 28.5°C for 10 minutes then triturated with polished glass Pasteur pipettes, followed by three more rounds of incubation and trituration. Dissociated cells were placed on ice. Wild-type fish were dissected and dissociated for cell sorting controls.

Fluorescence-Activated Cell Sorting (FACS) protocol
Dissociated retinal cells were transferred to 5 ml FACS tubes (BD Biosciences), diluted in phosphate-buffered saline (PBS), pH 7.4, and incubated in 10 µg/mL propidium iodide (PI) (LifeTechnologies) for 15 minutes at room temperature. The samples were then diluted with 2 mL of PBS, pH 7.4. Green fluorescent protein-positive (GFP+) and GFP-negative (GFP-) cells were isolated using a BD FACSAriaIII cell sorter (BD Biosciences). Gating was set with wild type unstained retinal cells, wild type retinal cells stained with PI, and mi2002 (GFP+) unstained samples. The final gating step was set at a stringency of 99.99% (tolerates 1 non-GFP+ cell per 1 x10⁴ cells sorted) for each sample. The actual purity, as calculated by re-sorting of previously sorted (GFP+) cells, was 98% for initial samples tested. For each sample 5 x10⁵ cells were sorted directly into TRIzol LS (Life Technologies) and immediately flash frozen. The light lesion, dissociation, and FACS steps were repeated three times for 0, 8, and 16 hpl (independent biological samples). All samples were collected between 11 am and 5 pm.

RNA isolation
Flash frozen FACS cells were thawed and RNA was isolated using TRIzol LS and the manufacturer's (Life Technologies) recommended protocol. Isolated RNA was loaded into a RNA 6000 Nano (Agilent Technologies) chip and measured on a Bioanalyzer 2100 (Agilent Technologies) to assess RNA quality. Samples containing RNA with a RNA Integrity Number (RIN) greater than 7 were considered of high enough quality for RNA-sequencing.
RNA-sequencing and Validation

RNA-seq library preparation protocol
Stranded RNA-seq libraries were generated as previously described and sequenced on an Illumina GAIIx (Illumina) ¹.

Read quality trimming and quality assessments
Trim Galore! (v0.2.7) ² was used to trim adapter sequences and poor quality bases (below Phred of 20) from the reads while removing any reads that were less than 20 nucleotides long, using the default parameters. Trim Galore! makes use of cutadapt (v1.4.2) ³ (-f fastq -e 0.1 -q 20 -O 1 -a AGATCGGAAGAGC file.fq.gz). The quality of the reads was assessed before and after trimming with FastQC (v0.10.1) ⁴.

Read mapping and gene-level quantitation
Quality trimmed and filtered reads were aligned to release 78 of the Zv9 Ensembl genome build with bowtie2 (v2.2.6) ⁵ and gene-level quantitation was performed with RSEM (v1.2.22) ⁶. This was done using the rsem-calculate-expression command from RSEM, which calls bowtie2 (--sensitive --dpad 0 --gbar 99999999 --mp 1,1 --np 1 --score-min L,0,-0.1) and streams reads into RSEM for quantitation.

Differential expression analysis and annotation
The gene-level counts output from RSEM were filtered to remove noise prior to normalization with trimmed means of M (TMM), such that only genes with a FPKM (fragments per kilobase of exon per million reads mapped) value greater than 1 in all replicates of any time-point were retained. Counts per million (CPM) were determined using edgeR (v3.10.2) ⁷, genes with a CPM < 1 in all samples were removed, and remaining counts were TMM normalized. Multidimensional scaling analysis with edgeR showed that one of the replicates at each of the 8 and 16 hpl time-points did not cluster with the others. Limma (v3.24.15) was used to voom transform the filtered count data by empirically deriving and applying quality weights to the samples ⁸,⁹. These weighted values were used to calculate differential expression using limma. Annotations for each gene were added using biomaRt (v2.24.0) ¹⁰,¹¹, including both the Danio rerio Entrez gene identifiers and the corresponding Mus musculus Entrez orthologous gene identifiers (Suppl. Data S1).

Reverse Transcriptase-quantitative Polymerase Chain Reaction (RT-qPCR)
validation of specific RNA-seq data
RNA was isolated from sorted Müller glia as described in the Sample Preparation
and Collection section above. RNA was isolated using TRIzol LS and the manufacturer’s recommended protocol, and DNaseI treated to remove genomic DNA (LifeTechnologies). The SuperScript VILO cDNA Synthesis Kit (LifeTechnologies) was used to create cDNA from 200 ng of phenol-chloroform cleaned RNA, using the manufacturer’s specified protocol. Quantitative PCR was performed using the primers specified in Supplementary Table S1 and PowerUP SYBR Green Master Mix (LifeTechnologies) with the manufacturer’s recommended protocol. Reactions were performed in duplicate for each of the 3 biological replicates at each time-point. The comparative method quantification method ($2^{-\Delta\Delta CT}$) was used to determine relative transcript abundance, normalized to glucose phosphate isomerase a ($gpi\alpha$).

**Gene Ontology and Pathway Analysis**

**Overview of workflow**

Gene Ontology (GO) term enrichment analysis was performed using a log$_2$-fold-change (log$_2$FC) ranked list from limma (no FDR cutoff) as input into clusterProfiler (v2.2.4) $^{12}$. This analysis determines which Molecular Function, Biological Process, or Cellular Component GO terms are positively or negatively enriched at each time point (8 and 16 hpl) compared with unlesioned controls (0 hpl), at a false discovery rate $\leq 0.05$, while taking into account the magnitude and direction of change (Suppl. Data S2). We then collapsed the GO terms into common categories with Revigo $^{13}$.

Two pathway databases, Kyoto Encyclopedia of Genes and Genomes $^{14}$ and Reactome pathway $^{15}$ analyses were used. A log$_2$FC ranked list of all differential gene expression data from limma (no FDR cutoff) was input into clusterProfiler (v2.2.4) $^{12}$ and ReactomePA (1.12.3) $^{16}$, respectively. The output from Reactome and KEGG pathways was largely redundant, so we chose to display the results from the KEGG analysis (with significant FDR $\leq 0.05$) using the Enrichment Map application (v2.1.0) $^{17}$ in Cytoscape (v3.3.0) $^{18}$.

**Correction for incomplete annotation of zebrafish transcriptome data**

The incompleteness of zebrafish gene annotations in the manually curated GO, KEGG, and Reactome databases was a limitation for this analysis. As an example, of the top 20 upregulated and downregulated genes at 8 hpl in our dataset, 11 and 17 genes, respectively, are not annotated with a known or inferred gene name in the GO database (Suppl. Table S2). At 16 hpl, the numbers of unannotated genes in the top 20 upregulated and downregulated genes is 17 and 23, respectively (Suppl. Table S3). This problem of incomplete annotation of zebrafish genes is also an issue for the KEGG and Reactome
pathways. For example, the total number of KEGG pathways is 299 for human (hsa), 295 for mouse (mmu), but only 164 for zebrafish (dre) 14. The number of pathways in the KEGG database for each species was determined by performing an empty search of the KEGG database 14, with each organism selected, which returns all pathways for that organism.

To address this issue of incomplete annotation, we manually annotated all the zebrafish genes in our dataset with their murine orthologs using biomaRt 10, and then performed KEGG and Reactome pathway analyses in two ways: 1) zebrafish gene annotations (from zebrafish differential expression data) and zebrafish pathway annotations, 2) mouse gene annotation (from zebrafish differential expression data) and mouse pathway annotations. By doing this, we were able to identify several biological pathways that do not exist in the zebrafish pathway databases, but were differentially regulated in Müller glia in response to light damage. These analyses were performed separately with zebrafish Entrez gene identifiers and the orthologous mouse Entrez gene identifiers, to determine KEGG or Reactome pathways that are positively or negatively enriched at each time point (8 and 16 hpl) compared with unlesioned control (0 hpl), at a false discovery rate ≤ 0.05 (Suppl. Data S3).

**Analysis of published Müller glia microarray data from mouse degeneration models**

**Microarray data background correction, normalization, and expression quantification**

Previously published 19 raw CEL files of the series GSE35386 were downloaded from the Gene Omnibus Expression (GEO) database using affycoretools (v1.40.5) 20. The GeneChip Robust Multichip Average (GCRMA) method, from the gcrma package (v2.40.0) 21 was used to calculate background-corrected, normalized expression values. Genefilter (v1.50.0) 22 was used to remove uninformative probesets, such as internal controls, probesets with low variance, or probesets with background level expression values.

**Microarray differential expression analysis and annotation**

Principal component analysis (PCA) plots of the normalized, filtered samples showed that there were several outlying samples. In particular, two of the Pde6b<sup>d1/d1</sup> mutant samples clustered with the Pde6b<sup>+/+</sup> control samples. This sample clustering might be due to biological heterogeneity of individual Müller glia cells 23. The limma package (v3.24.15) was therefore used to apply empirically derived array weights to samples 24 and those weights were used in calculating differential expression values 25. Annotations for each gene were
added using biomaRt (v2.24.0)\(^{10,11}\) (Suppl. Data S1).

**Gene ontology and pathway analysis**
Gene ontology (Suppl. Data S2) and pathway analysis (Suppl. Data S3) were performed in the same way as for the zebrafish RNA-seq data described above, with the exception of the significance cutoffs applied to the gene ontology analysis. When applied to gene ontology terms, an FDR cutoff of \(\leq 0.05\) resulted in almost no significant changes in biological processes in single Müller glia cells isolated from degenerating retinas compared with wild-type. Because these data were obtained from a heterogeneous population of individual Müller glial cells, the sample variability is inherently large\(^{19}\). Therefore, a p-value cutoff of 0.05 was applied to identify enriched GO terms. The output from Reactome and KEGG pathways was largely redundant, so we chose to display the results from the KEGG analysis (with significant FDR \(\leq 0.05\)) using the Enrichment Map application (v2.1.0)\(^{17}\) in Cytoscape (v3.3.0)\(^{18}\).

**Cross-contamination analysis of published Müller glia and photoreceptor expression data**

**Data acquisition and transformation**
Normalized expression data from FACS-isolated Müller glia or photoreceptors were downloaded from the GEO database, [http://www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/) (Suppl. Table S4). For the single-cell Müller glia dataset from *Roesch et al.*, normalized expression data were obtained as described above (“Analysis of published Müller glia microarray data from mouse degeneration models”). Sample mean expression values were calculated from biological replicates for each gene or probeset (with the exception of the individual Müller glia data) and the resulting values were log\(_2\) transformed, if not already in this format.

**Plotting expression data**
Expression data for control samples or samples of interest (Suppl. Table S4) were plotted as a histogram (Frequency vs. Log\(_2\) Expression Value). Selected, cell-specific Müller glial, rod photoreceptor, and cone photoreceptor transcripts (Suppl. Table S5) were plotted with a vertical line indicating the expression value of that gene (Suppl. Figs. S3; S4A-F; S5A,B,D,E). Multiple probes for several genes in microarrays resulted in multiple values being plotted for those genes. In order to perform a crude assessment of the relative photoreceptor contamination in the hand-picked Müller glia, the expression level of glial-specific *Rlbp1* in each individual cell was plotted vs the expression level of rod-specific *Rho* (Suppl. Fig. S4I). To show the level of contamination in sorted photoreceptor cell types with transcripts specific to other types of photoreceptors and Müller glia, expression
values for the selected cell-type specific transcripts (Suppl. Table S4) were plotted by sorted cell-type (Suppl. Fig. S5C,F).

Supplementary References

1. Brooks MJ, Rajasimha HK, Swaroop A. Retinal transcriptome profiling by directional next-generation sequencing using 100 ng of total RNA. Methods Mol Biol. 2012;884:319–34.

2. Krueger F. Trim Galore! [Internet]. 2012. Available from: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

3. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011 May 2;17(1):10–2.

4. Andrews S. FastQC: A quality control tool for high throughput sequence data [Internet]. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

5. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Apr;9(4):357–9.

6. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011 Aug 4;12:323.

7. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–40.

8. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.

9. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014 Feb 3;15(2):R29.

10. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005 Aug 15;21(16):3439–40.

11. Kersey PJ, Allen JE, Christensen M, Davis P, Falin LJ, Grabmueller C, et al. Ensembl Genomes 2013: scaling up access to genome-wide data. Nucleic Acids Res. 2014 Jan;42(Database issue):D546–52.
12. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012 May;16(5):284–7.

13. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011 Jul 18;6(7):e21800.

14. KEGG Pathway Database [Internet]. [cited 2016 Jan 9]. Available from: http://www.genome.jp/kegg/pathway.html.

15. Reactome Pathway Database [Internet]. [cited 2016 Jan 9]. Available from: http://www.reactome.org/.

16. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016 Jan 26;12(2):477–9.

17. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One. 2010 Nov 15;5(11):e13984.

18. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498–504.

19. Roesch K, Stadler MB, Cepko CL. Gene expression changes within Müller glial cells in retinitis pigmentosa. Mol Vis. 2012 May 9;18:1197–214.

20. MacDonald JW. affycoretools: Functions useful for those doing repetitive analyses with Affymetrix GeneChips [Internet]. 2008. Available from: http://bioconductor.org/packages/release/bioc/html/affycoretools.html.

21. Wu Z, Gentry J. gcrma: Background Adjustment Using Sequence Information [Internet]. Available from: https://www.bioconductor.org/packages/release/bioc/html/gcrma.html.

22. Gentleman R, Carey V, Huber W, Hahne F. genefilter: methods for filtering genes from microarray experiments [Internet]. Available from: http://bioconductor.org/packages/release/bioc/html/genefilter.html.

23. Roesch K, Jadhav AP, Trimarchi JM, Stadler MB, Roska B, Sun BB, et al. The transcriptome of retinal Müller glial cells. J Comp Neurol. 2008 Jul 10;509(2):225–38.

24. Ritchie ME, Diyagama D, Neilson J, van Laar R, Dobrovic A, Holloway A, et al. Empirical array quality weights in the analysis of microarray data. BMC Bioinformatics. 2006 May 19;7:261.
25. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004 Feb 12;3:Article3.