Identification of transcription factors involved in the specification of photoreceptor subtypes

Juan Angueyra[1,✉], Vincent P. Kunze[1,*], Laura K. Patak[1,*], Hailey Kim[1], Katie S. Kindt[2], and Wei Li[1,✉]

[1]Unit of Retinal Neurophysiology, National Eye Institute, National Institutes of Health, Bethesda, MD, USA
[2]Section on Sensory Cell Development and Function, National Institute on Deafness and other Communication Disorders, National Institutes of Health, Bethesda, MD, USA
[*]Equal contributions

During development, retinal progenitors navigate a complex landscape of fate decisions that culminates with an array of unique cell types that are required for proper vision. Here, we aim to identify factors that are required for fate decisions in photoreceptors. These factors help create a diversity of photoreceptor subtypes that sustain vision in day and night, enable the detection of colors, of prey and predators, and other aspects of vision. To identify these factors, we generate a high-quality and deep transcriptomic profile of each photoreceptor subtype in zebrafish. From these profiles, we focus on transcription factors—key players in cell-fate decisions. We apply CRISPR-FØ screening as a versatile platform to explore the involvement of transcription factors in photoreceptor subtype-specification. We find that three differentially-expressed transcription factors (Foxq2, Tbx2a and Tbx2b) play unique roles in controling the identity of photoreceptor subtypes within the retina. Our results provide novel insights into the function of these factors and how photoreceptors acquire their final identities. Furthermore, we have made our transcriptomic dataset openly available and easy to explore. This dataset and the screening method will be valuable to the scientific community and will enable the exploration of genes involved in many essential aspects of photoreceptor biology.

photoreceptor fate | transcription factor | tbx2 | foxq2 | zebrafish | retina | cone subtypes

Correspondence: ***angueyraLab@gmail.com liwei2@nei.nih.gov***

Introduction

The specification of cell types is a fundamental and complex process that underlies the formation of tissues and organs. Cell specification is especially critical for the primary sensors in sensory systems. Primary sensory cells not only have to acquire distinct and elaborate specializations, to be able to detect and transduce particular physical stimuli, but also have to wire with specific circuits to faithfully relay sensory information. In the retina, for instance, photoreceptor specification must coordinate with the distinct developmental timelines of other retinal cell types to form functional visual circuits (1). In addition, photoreceptors have clearly defined photoreceptor subtypes which differ in spectral sensitivity, morphology, density across the retina, wiring, and gene expression. Comprehensively, the factors required to specify photoreceptor subtypes remains unclear. Our aim is to identify these factors, exploiting technical advantages of the zebrafish model.

Evolutionarily, fish and birds have retained an ancestral and diverse set of photoreceptor subtypes. In zebrafish, there are five photoreceptor subtypes that can be readily distinguished by morphology and opsin expression: rods expressing rhodopsin (*rho*), UV cones (or short single cones) expressing a UV-sensitive opsin (*opn1sw1*), S cones (or long single cones) expressing a short-wavelength-sensitive opsin (*opn1sw2*), M cones (or accessory members of double cones) expressing mid-wavelength-sensitive opsins encoded by genes that have been duplicated twice (*opn1mw1* - *opn1mw4*), and L cones (or principal members of double cones) expressing long-wavelength-sensitive opsins encoded by duplicated genes (*opn1lw1* - *opn1lw2*) (2–5). In comparison, most mammals possess only three photoreceptor subtypes: rods, a short-wavelength sensitive cone (S cone), and a mid/long-wavelength sensitive cone (M/L-cone). The mammalian S and M/L cones are evolutionarily related to the ancestral UV and L cones, respectively, while the ancestral S and M cones were lost. Even more recently in evolution, duplication and subsequent mutations of the L-opsin gene led to the emergence of an additional functional subtype in Old-World Primates, which supports human trichromatic vision (6).

The correct specification of photoreceptor subtypes is critical for proper vision. Foundational work across vertebrates has identified transcription factors that are required for the generation of photoreceptors progenitors during development, before subtypes are clearly defined(7–12). Numerous studies have investigated how photoreceptor subtypes are specified from these progenitors. Studies in mice have shown that the transcription factor NRL and its downstream effector NR2E3 are required for rod specification (13–15). Interestingly, in humans, mutations in NR2E3 cause *enhanced S-cone syndrome*, where failures in rod specification ultimately lead to impaired visual acuity, abnormal color vision, night blindness and retinal degeneration (16). In addition to conserved factors involved in rod specification, THRb is required for L-cone specification in mice (17), zebrafish (18) and, most likely, birds (19). In the absence of these "key gatekeepers"— NRL, NR2E3 or THRb—mouse photoreceptor progenitors acquire an S-cone fate, and S-cone specification is frequently assumed to be a passive process (12). Yet, mounting evidence derived mainly from work in non-mouse species challenges this simplistic model. For example, while Nr2e3 is also required to generate rods in other vertebrates (20), Nrl is dispensable (21, 22). In addition,

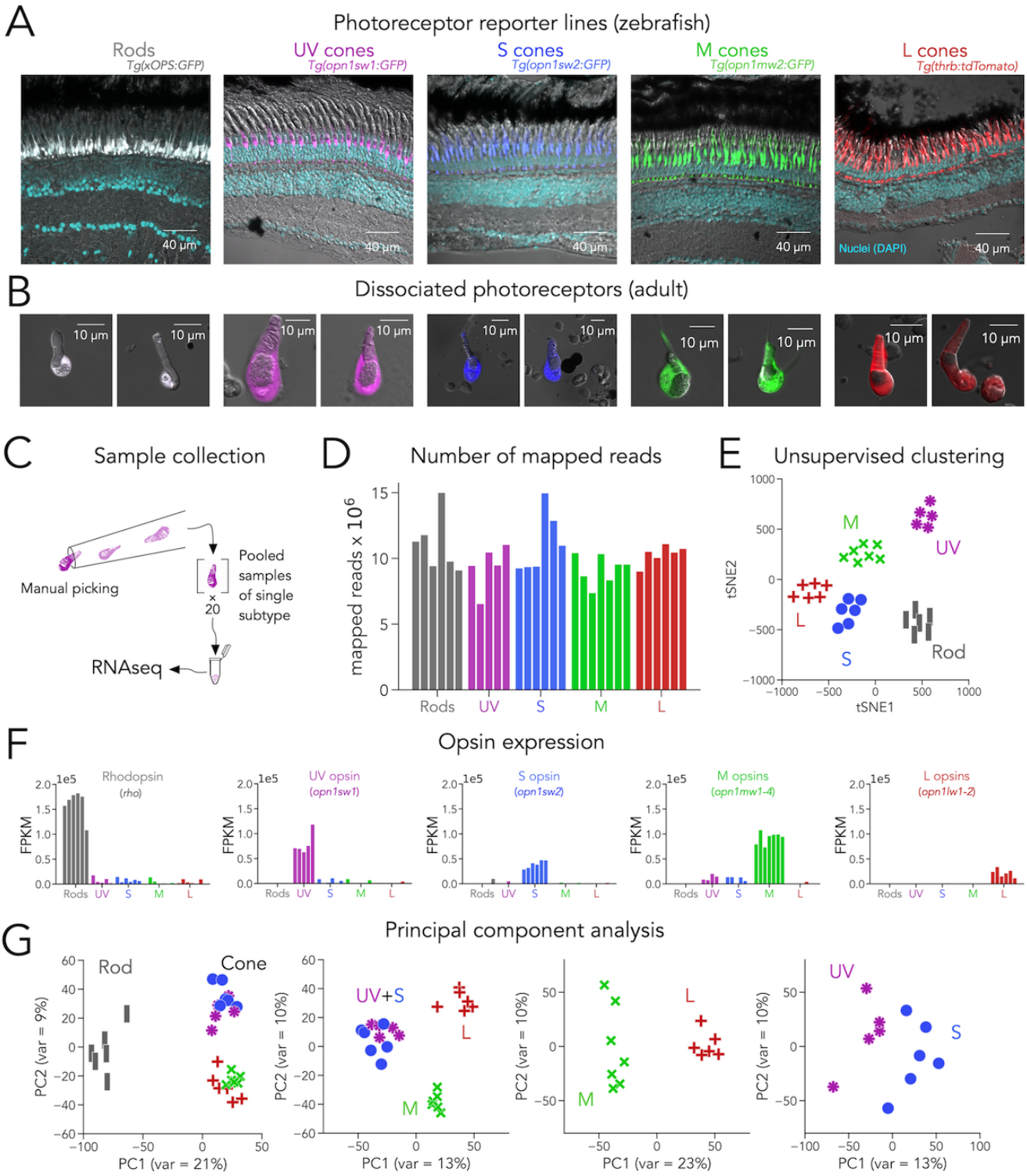

Figure. 1. Transcriptomic profiling (RNAseq) of zebrafish photoreceptors. (A) Confocal images of fixed adult zebrafish retinal cross-sections, from transgenic reporters used to identify photoreceptor subtypes. Reporter expression is exclusive to the outer retina, and each line labels a single photoreceptor subtype with unique morphology, including rods (gray), UV cones (magenta), S cones (blue), M cones (green) and L cones (red). The inner retinal layers can be distinguished in the overlayed nuclear stain (DAPI, cyan) and transmitted DIC image (grey). **(B)** Confocal images of dissociated and live photoreceptors of each subtype, identified by fluorescent reporter expression. Photoreceptors have preserved outer segments and identifiable mitochondrial bundles. **(C)** Sample collection method. After dissociation, 20 healthy photoreceptors of a single subtype were identified by fluorescence and manually picked with a glass micropipette and pooled as a single RNAseq sample. **(D)** High transcriptome depth shown by the number of reads successfully mapped to the zebrafish genome (*GRCz11*) for each RNAseq sample. **(E)** Clustering using t-distributed stochastic neighbor embedding correctly separates samples by their original subtype. **(F)** Plots of opsin expression show high counts for the appropriate opsin in each sample (in fragments per kilobase per million reds or FPKM) and low-to-negligible counts of other opsins. For M-opsin quantification, we calculated the sum of counts for *opn1mw1*, *opn1mw2*, *opn1mw3* and *opn1mw4* and for L-opsin quantification, we summed counts for *opn1lw1* and *opn1lw1*. **(G)** Iterative principal component analysis (PCA) shows that differences in gene expression separate rods and cones (first panel), and UV/S cones from M/L cones (second panel). M and L cones can also be distinguished by a single principal component (third panel), while separation of UV and S cones is more difficult (fourth panel).

UV- and S-cone specification in zebrafish is far from passive and requires the action of key transcription factors—Tbx2b and Foxq2 (23, 24). Moreover, factors that regulate opsin expression—*e.g.* Rora, Rorb or Rxrg in mouse (25–27)— or survival of retinal progenitors—*six6*, *six7* and *gdf6a* in zebrafish (28–33) —can also cause subtype-specific alterations in photoreceptor development. Together these studies highlight that our understanding of how photoreceptor subtypes are specified remains incomplete.

Our study seeks to identify additional genes involved in this complex process and is divided into three sections. First, we obtain a deep and high-quality transcriptomic profile (RNA-seq) of the five zebrafish photoreceptor subtypes. Second, we explore this RNA-seq dataset and identify multiple transcription factors that are potentially involved in the specification of photoreceptor subtypes. Third, we use a CRISPR-based FØ-screening approach (34, 35) to directly explore the role of three transcription factors—Foxq2, Tbx2a and Tbx2b. We demonstrate that Foxq2 is required for S-cone specification exclusively. UV-cone specification requires both Tbx2a and Tbx2b. Additionally, Tbx2a and Tbx2b respectively maintain the identity of L cones and S cones by repressing M-cone cell fate.

In the future, our dataset can be applied to explore other important aspects of photoreceptor biology that also differ between subtypes (phototransduction, metabolism, synaptic wiring, *etc.*). To facilitate future studies, we provide open and easy access to our transcriptomic dataset and analysis (https://github.com/angueyraNIH/drRNAseq/). This dataset can be used to further our understanding of how photoreceptors acquire their final identities. This knowledge can be used to inform strategies to control the photoreceptor differentiation in organoids—a potential gateway for cell-replacement therapies in retinal degenerations.

Results

Transcriptomic analysis of adult zebrafish photoreceptors.

Identifying the transcription factors required to specify photoreceptor subtypes is critical to understand the normal development of the retina and to inform cell-replacement therapies to restore vision. RNA-seq is a powerful way to identify novel genes expressed in particular cell subtypes. Although RNA-seq approaches have been used to identify genes differentially expressed between photoreceptor subtypes in many species, the limited transcriptome depth derived from single-cell techniques (36) constitutes a barrier in the reliable detection of transcription factors, which are frequently expressed at low levels (37). To obtain a deep, high-quality RNA-seq dataset from zebrafish photoreceptors, we manually collected pools of photoreceptors of a single subtype (38). We identified photoreceptors using well-characterized transgenic lines that express fluorescent proteins in each subtype with high specificity, including rods—*Tg(xOPS:GFP),* UV cones—*Tg(opn1sw1:GFP)*, S cones—*Tg(opn1sw2:GFP)*, M cones—*Tg(opn1mw2:GFP)* and L cones—*Tg(thrb:tdTomato)* (18, 39–42) (Figure 1A). We collected dissociated photoreceptors from these transgenic lines under epifluorescence. Manual collection allowed us to focus on fluorescent and healthy photoreceptors, with intact outer segments, cell bodies and mitochondrial bundles and to avoid cellular debris and other contaminants (Figure 1B).

For each sample, we collected pools of 20 photoreceptors of a single subtype derived from a single adult retina. After collection, we isolated mRNA and generated cDNA libraries for sequencing using SMART-seq2 technology (Figure 1C). In total, we acquired 6 rod samples and 5 UV-, 6 S-, 7 M- and 6 L-cone samples. On average, we were able to map 86.4% of reads to the zebrafish genome (GRCz11; range: 76.3% - 90.4%), corresponding to 10.19 million ± 1.77 million mapped reads per sample (mean ± s.d.) (Figure 1D). This quantity of reads demonstrates that our technique provides substantially deep transcriptomes—especially when compared to single-cell droplet-based techniques where the number of reads per cell is largely 20000 or less (36, 43, 44). Using unsupervised clustering (t-distributed Stochastic Neighbor Embedding or tSNE), we found that samples correctly clustered by the subtype they were derived from. Proper clustering provides evidence that differences in gene expression captured in our RNA-seq data stem mainly from distinctions between photoreceptor subtypes (Figure 1E).

The expression of opsin genes is unique between photoreceptor subtypes and under normal conditions it is a reliable marker of subtype identity. Consistent with this idea, each sample had a high number of reads for the appropriate opsin. Furthermore, there were very low reads for other opsins, corroborating the purity of our samples (Figure 1F). We also found that reads for phototransduction genes were high and consistent with the known differences in gene expression between rods and cones (*e.g.*, *gnat1*, rod transducin, had high reads only in rod samples while *gnat2*, cone transducin, had high reads in all cone samples) and between cone subtypes (*e.g.*, expression of *arr3a* in M and L cones and *arr3b* in UV and S cones) (Figure 1 – Figure supplement 1) (44, 45). To expand our analysis to other genes, we first used principal component analysis (PCA) as an unbiased approach to determine how variability in gene expression defines photoreceptor subtypes. PCA revealed that most of the differences in gene expression were between rods and cones. When cones were considered separately, the biggest differences in gene expression arose from two groupings: UV and S cones vs. M and L cones. Subsequent analysis revealed a clear separation of M and L cones, with UV and S cones showing the least differences (Figure 1G). Guided by this analysis, we performed differential gene-expression analysis by making pairwise comparisons following the directions of the principal components, revealing a diverse set of differentially expressed genes (Figure 1 – Figure Supplement 2 and Supplementary Data 1).

In summary, our manual, cell-type specific, SMART-seq2-based approach yielded high-quality zebrafish photoreceptor transcriptomes, with ~2000-fold more depth than published single-cell RNA-seq studies in the retina (36, 43, 44, 46).

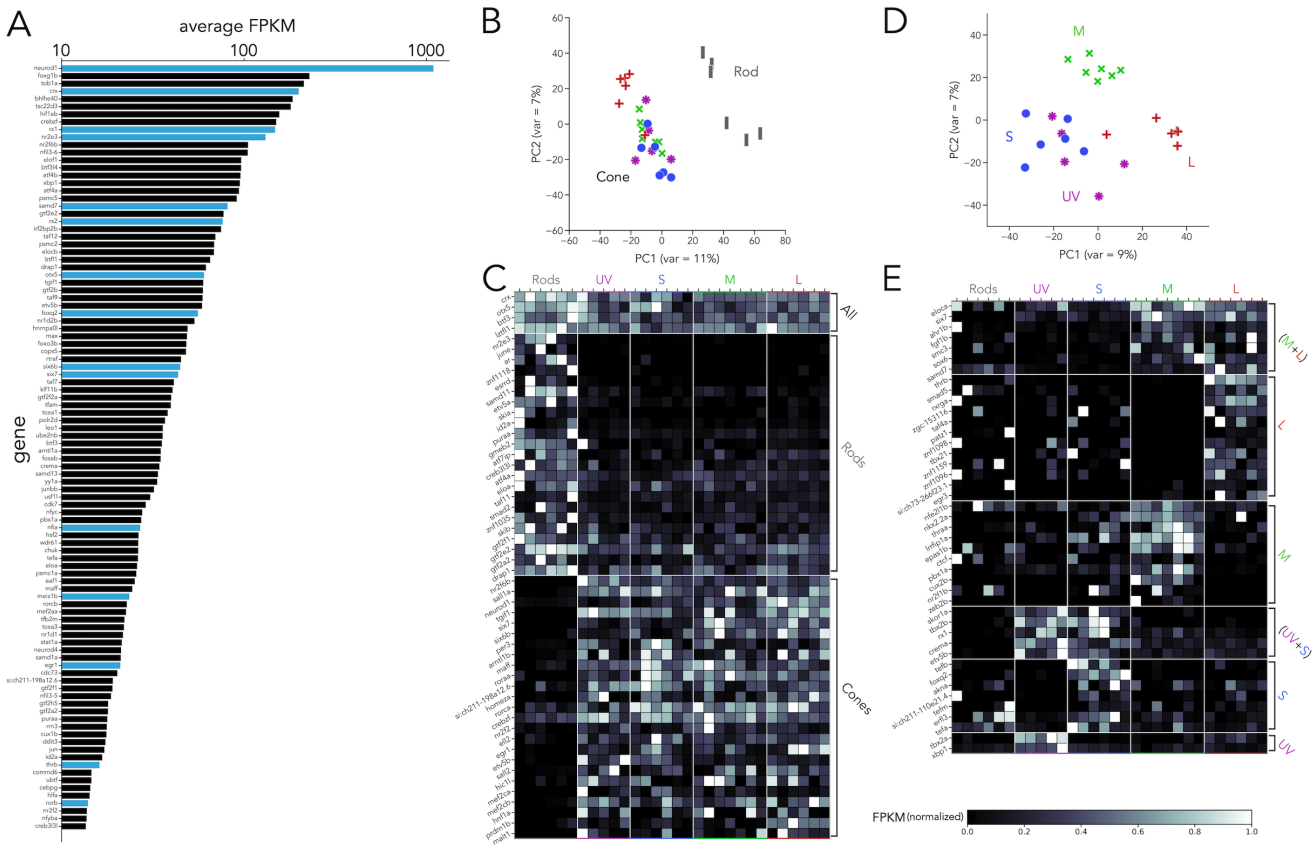

Figure. 2. Transcription factor expression in zebrafish adult photoreceptors. (A) Top 100 transcription factors ranked by average expression across all samples, displayed on a log-scale. Genes highlighted in blue are known to be expressed in photoreceptor progenitors or during photoreceptor development. **(B)** Principal component analysis (PCA) of transcription factors shows that differences between rods and cones are the main source of variance. **(C)** Heatmap showing differential expression of transcription factors between rods and cones and divided into three groups: consistently expressed by all photoreceptors, enriched in rod samples and enriched in cones samples. Grey values indicate expression level normalized in each row by the maximal value. Rod- and cone-enriched genes have been arranged by degree of enrichment. **(D)** PCA of cone samples shows that the largest differences in expression separate L cones (PC1) and M cones (PC2), while separation of UV and S cones is more difficult. **(E)** Heatmap of transcription factors differentially expressed in cone subtypes, divided into 6 relevant groups. Full list of differentially-expressed transcription factors available in Supplementary Data 2 and 3.

Our RNA-seq data is in good agreement with current knowledge of photoreceptor-expressed genes, but importantly also uncovered novel differences in gene expression between photoreceptor subtypes. This RNA-seq dataset constitutes a useful resource to explore genes that are generally or differentially expressed by photoreceptor subtypes which could be involved in multiple aspects of photoreceptor biology. Our subsequent analyses center on transcription factors.

Expression of transcription factors in zebrafish photoreceptors.

Transcription factors play a critical role in photoreceptor specification. Therefore, we isolated all RNA-seq reads that could be mapped to transcription factors. When ranked by average expression levels across all samples (in Fragments Per Kilobase of transcript per Million mapped reads or FPKM), *neurod1* was revealed as the transcription factor with the highest expression (~5-fold). High expression in adult photoreceptors suggests that *neurod1* plays a role not only during development or regeneration but also in the mature retina (10, 47, 48). Among the 100 most highly expressed transcription factors, we identified many genes known to known to be important during photoreceptor development or specification including *crx, otx5, rx1, rx2, nr2e3, six6b, six7, meis1b, egr1* and *thrb* (Figure 2A) (7–9, 28, 32, 49–52).

Next, we used PCA to understand how transcription-factor expression differs between photoreceptor subtypes. Like our whole-transcriptome analysis, we found that most of the differences in transcription-factor expression can be attributed to differences between rods and cones (Figure 2B). By performing pairwise comparisons of transcription factors based on rod vs. cone expression, we identified three relevant groups: (1) consistently expressed across all subtypes, (2) rod-enriched and (3) cone-enriched (Figure 2C). Consistent with previous studies, expression of *crx* and *otx5* was similar across subtypes (7, 8), *nr2e3, samd7* and *samd11* showed clear rod-enrichment (15, 53, 54), while expression of *nrl*, which is only transiently required to specify rods in zebrafish (22), was low and only detected in a third of the rod samples. Lastly, our results verified that *six6a*, *six6b, six7,* and *neurod1* were cone enriched transcription factors (10, 28, 32). By expanding our analysis beyond previously characterized genes, our dataset revealed a total of 75 transcription factors with significant differential expression between rods and cones (Figure 2C and Supplementary Data 2).

We next examined the variance in transcription-factor ex-

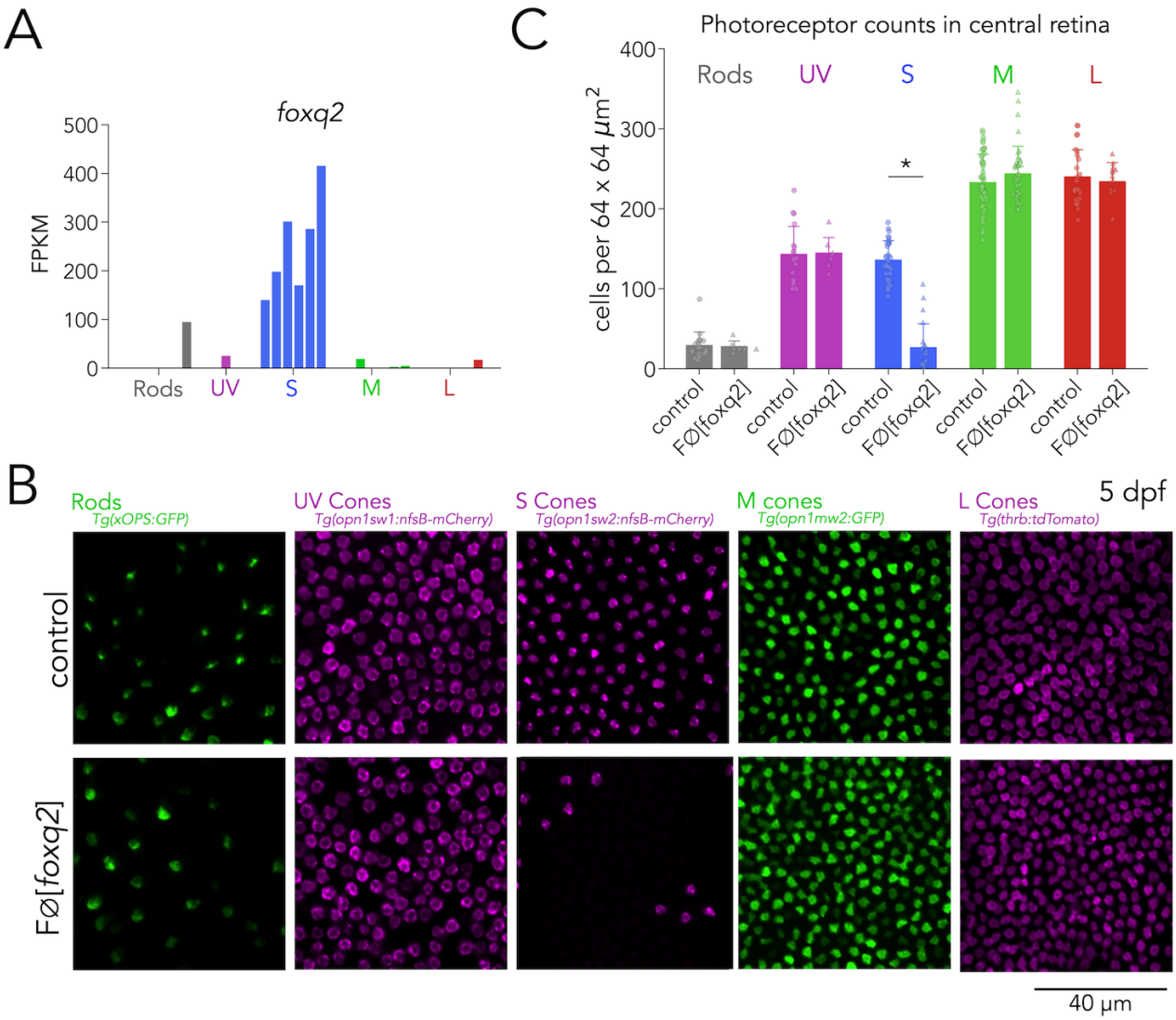

Figure. 3. Foxq2 is required for S-cone specification. (A) Expression of *foxq2* across photoreceptors shows clear S-cone specificity. **(B)** Mutations in *foxq2* cause a specific loss of S cones. Representative confocal images of the central retina of control (top row) and FØ[*foxq2*] (bottom row) larvae at 5 dpf. Each column corresponds to a transgenic line that labels a unique photoreceptor subtype (green corresponds to GFP expression, magenta to mCherry or tdTomato expression). **(C)** Quantification of photoreceptors in control and FØ[*foxq2*] larvae. Bars represent averages, error bars correspond to standard deviations, and markers correspond to individual retinas. There is a significant reduction in S cones in FØ[*foxq2*] compared to controls (Mann-Whitney U = 709, $p = 1 \times 10^{-9}$, $n_{wt} = 34$, $n_{FØ[foxq2]} = 21$) without significant differences in the number of rods (Mann-Whitney U = 61, p = 0.93, $n_{wt} = 18$, $n_{FØ[foxq2]} = 7$), UV cones (Mann-Whitney U = 57, p = 0.74, $n_{wt} = 18$, $n_{FØ[foxq2]} = 7$), M cones (Mann-Whitney U = 870, p = 0.37, $n_{wt} = 49$, $n_{FØ[foxq2]} = 40$) or L cones (Mann-Whitney U = 161, p = 0.98, $n_{wt} = 25$, $n_{FØ[foxq2]} = 13$).

pression between cone subtypes. PCA revealed that both L and M cones could be distinguished by differences in transcription-factor expression alone, while UV and S cones again showed the fewest differences in transcription factor expression (Figure 2D). By analyzing the differentially expressed transcription factors between cones subtypes, we found a group of seven transcription factors significantly enriched in both L and M cones compared to UV and S cones (Figure 2E). These included *ahr1b, a gene* associated with Retinitis Pigmentosa in humans (55) and *six7*, known to be involved in cone progenitor development and survival (28). We found twelve L-cone enriched transcription factors (Figure 2E); these included *thrb*, known to be critical for L-cone identity across vertebrates (17, 18) and *rxrga*, a regulator of L-opsin expression in mouse (25). Amongst the ten M-cone enriched transcription factors, we identified *thraa*, another thyroid hormone receptor, known to be expressed by photoreceptors (56). A small group of just five genes was enriched in both UV and S cones compared to L and M cones and included *tbx2b*. Seven transcription factor were enriched in S cones—including *foxq2*—and two were enriched in UV cones—including *tbx2a* (Figure 2E and Supplementary Data 3).

In summary, our analysis has revealed novel patterns of expression of transcription factors between photoreceptor subtypes. Notably, a considerable fraction of these transcription factors has no known function in photoreceptors in zebrafish or in other species, making them clear targets for follow-up studies aimed at understanding photoreceptor specification and other subtype-specific functions.

FØ screening as a platform to explore the involvement of transcription factors in photoreceptor identity.

Our dataset revealed an extensive collection of transcription factors potentially involved in photoreceptor differentiation. Therefore, we next established methods to identify which transcription factors are involved in subtype specification. We focused on a small subset of these transcription factors, and evaluated their function using a highly efficient CRISPR-Cas9-guided mutagenesis approach that enables FØ screening (34, 35). In these FØ screens, single-cell embryos are injected with Cas9 and multiple guide RNAs (gRNAs) targeting a gene of interest. Just days after injection, in the FØ generation, injected larvae can be assessed phenotypically. Although these FØ larvae are genetic mosaics (some cells may not carry mutations and mutations between cells are not identical), each larva can be genotyped to establish a robust link between gene function and phenotype. In addition, FØ larvae can be created in the context of any combination of existing transgenic or mutant lines. We used this FØ-screening approach to study the role of three transcription factors in the determination of photoreceptor identity: Foxq2, Tbx2a and Tbx2b. The results presented below provide novel insights into the function of these genes and help establish this approach as a versatile method for future screens.

Foxq2 is required for S-cone specification

First, we explored the function of Foxq2. Our RNA-seq data revealed that *foxq2* is expressed at relatively high levels, ranking 33rd among transcription factors (Figure 2A), and specifically enriched in S-cones, with negligible expression in other photoreceptor subtypes (Figure 3A). Based on this S-cone specific expression, we used the FØ-screening platform to test whether *foxq2* was involved in S-cone specification.

For our FØ analyis, we designed two gRNAs targeted against the DNA-binding forkhead domain of *foxq2* (57). Preliminary experiments established that we were able to produce mutations in *foxq2* in ~95% (21/22) of injected FØ larvae. All analyses presented here correspond to FØ larvae that have been genotyped to confirm mutations in *foxq2* (FØ[*foxq2*]). To evaluate changes in photoreceptors, we used subtype-specific photoreceptor reporter lines and evaluated changes in photoreceptor densities in the central retina of larvae at 5 days post-fertilization (5 dpf). Compared to wild-type controls, FØ[*foxq2*] larvae displayed a dramatic decrease in the density of S cones. We were able to clearly identify a decrease in S cones through their mCherry expression in *Tg(opn1sw2:mCherry)* (Figure 3B), or by the lack of S-opsin antibody labeling (Figure 3 – figure supplement 1A). The density of other photoreceptor subtypes remained unchanged, with a decrease in S cones being the only significant change (Figure 3C). This demonstrates that the loss of S cones in FØ[*foxq2*] larvae is selective and not compensated by an increase in another photoreceptor subtype. Our results are close in agreement with a recent study that characterized a complete *foxq2* loss-of-function (see discussion) (24), demonstrating the power of the FØ-screening approach. In conclusion, *foxq2* is a S-cone specific gene that is required for the early specification of S cones but not of other photoreceptor subtypes.

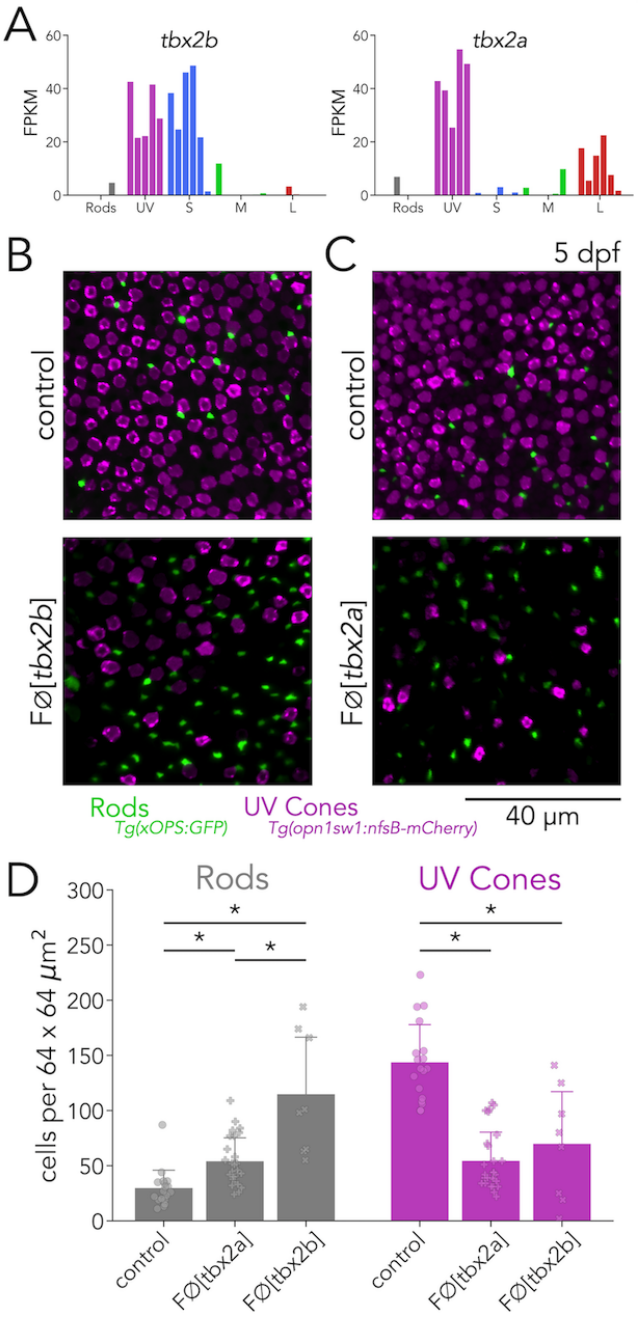

Figure. 4. Tbx2a and Tbx2b are independently required for UV-cone specification. (A) *tbx2b* is expressed by both UV and S cones (left), while *tbx2a* is expressed by both UV and L cones (right). **(B)** Mutations in *tbx2b* cause a loss of UV cones and an increase in rods. Representative confocal images of the central retina of control and FØ[*tbx2b*] at 5 dpf, in double transgenic larvae with labeled UV cones (magenta) and rods (green). **(C)** Mutations in *tbx2a* also cause a loss of UV cones and an increase in rods. Representative confocal images of the central retina of control and FØ[*tbx2b*] at 5 dpf, in the same double transgenic larvae. **(D)** Quantification of rods and UV cones in control, FØ[*tbx2b*] and FØ[*tbx2a*] larvae. Bars represent averages, error bars correspond to standard deviations, and markers correspond to individual retinas. Compared to control (n_{wt} = 18), there is a significant increase in rods and a marked decrease in UV cones in both FØ[*tbx2b*] (Mann-Whitney U_{UV} = 128, p_{UV} = 0.002, U_{Rods} = 3, p_{Rods} = 0.0001 $n_{FØ[tbx2b]}$= 7) and in FØ[*tbx2a*] (Mann-Whitney U_{UV} = 478, p_{UV} = 5 × 10^{-8}, U_{Rods} = 69, p_{Rods} = 6 × 10^{-5}, $n_{FØ[tbx2a]}$= 27). The increase in rods is significantly higher in FØ[*tbx2b*] compared to FØ[*tbx2a*], while the loss of UV cones is similar (Mann-Whitney U_{Rods} = 17.5, p_{Rods} = 0.01, U_{UV} = 24.5, p_{UV} = 1.0).

Tbx2a and Tbx2b are independently required for UV-cone specification

To further expand our FØ analysis, we explored the role of Tbx2 in photoreceptor differentiation. As a teleost duplicated gene, there are two versions of *tbx2* in the zebrafish genome: *tbx2a* and *tbx2b*. We chose to evaluate the function of Tbx2 for three main reasons. First, our RNA-seq data revealed that both *tbx2a* and *tbx2b* show high expression in UV cones (Figure 4A). Second, Tbx2 is known to be differentially expressed in cones of many species, including cichlids (58), chickens (59) and primates (46). Third, Tbx2b in particular, has been shown to be involved in the determination of UV cones in zebrafish (23).

For our FØ analysis, we injected embryos with Cas9 protein and 3 gRNAs targeting exon 3 of *tbx2b* or 3 gRNAs targeting exon 3 of *tbx2a*. In both genes, exon 3 contains critical DNA-binding residues that are completely conserved across vertebrates (60). Initial experiments showed that, using this approach, we were able to produce mutations in *tbx2b* in ~80% (19/23) of injected larvae and in *tbx2a* in ~90% (30/33) of injected larvae. Our gRNAs were selective for their respective gene, and we did not observe any mutations in the non-targeted paralogue (0/48 larvae tested). Moving forward, we restricted all subsequent analysis to larvae with confirmed mutations in *tbx2b* (FØ[*tbx2b*]) or in *tbx2a* (FØ[*tbx2a*]). As *tbx2a* and *tbx2b* are enriched in UV-cones, we first assessed changes in UV-cone fate, using a UV-cone reporter line, *Tg(opn1sw1:nfsB-mCherry)*, (61). We used this UV-cone reporter line along with a rod reporter line, *Tg(xOPS:GFP)* (39) as loss of Tbx2b function increases rod density (23). In wild-type control larvae at 5 dpf, UV cones are numerous and densely distributed across the retina, while overall rod density is low, with most rods concentrated in the ventral retina and with lowest density in the central retina (23, 62, 63). Using these transgenic lines, we first quantified UV cone and rod density in the central retina of control and (FØ[*tbx2b*])larvae. In agreement with previous studies, loss of Tbx2b function produced a marked decrease in UV cones and an increase in rod density (Figure 4B and D) (23). The robust phenotype in FØ[*tbx2b*] larvae further demonstrates the efficiency and flexibility of our approach. After replicating the described loss-of-function phenotypes of *tbx2b* mutants in FØ[*tbx2b*] larvae, we examined FØ[*tbx2a*] larvae. Surprisingly, we found that FØ[*tbx2a*] also displayed the same phenotype as FØ[*tbx2b*]—a marked loss of UV cones and an increase in rods, albeit this increase in rods was lower in FØ[*tbx2a*] than in FØ[*tbx2b*] (Figure 4C and D).

To confirm the phenotypes of *tbx2* mutants revealed through imaging of reporter lines, we quantified opsin expression using real-time quantitative PCR (qPCR). We found that, in comparison to controls, FØ[*tbx2b*] showed a clear decrease in UV-opsin expression and a significant increase in rhodopsin expression. FØ[*tbx2a*] also showed a clear decrease in UV-opsin expression, but without significant changes in rhodopsin expression (Figure 4 - figure supplement 1). Together our reporter lines and qPCR analyses suggest that, despite 87% protein-sequence similarity and expression of the two genes in the same cell, both Tbx2a and Tbx2b are independently required for the specification of zebrafish UV cones. Loss-of-function of either gene leads to a decrease in UV cones and a concomitant routing of photoreceptor progenitors towards a rod fate. In FØ[*tbx2b*] routing towards a rod fate appears to be stronger than in FØ[*tbx2a*].

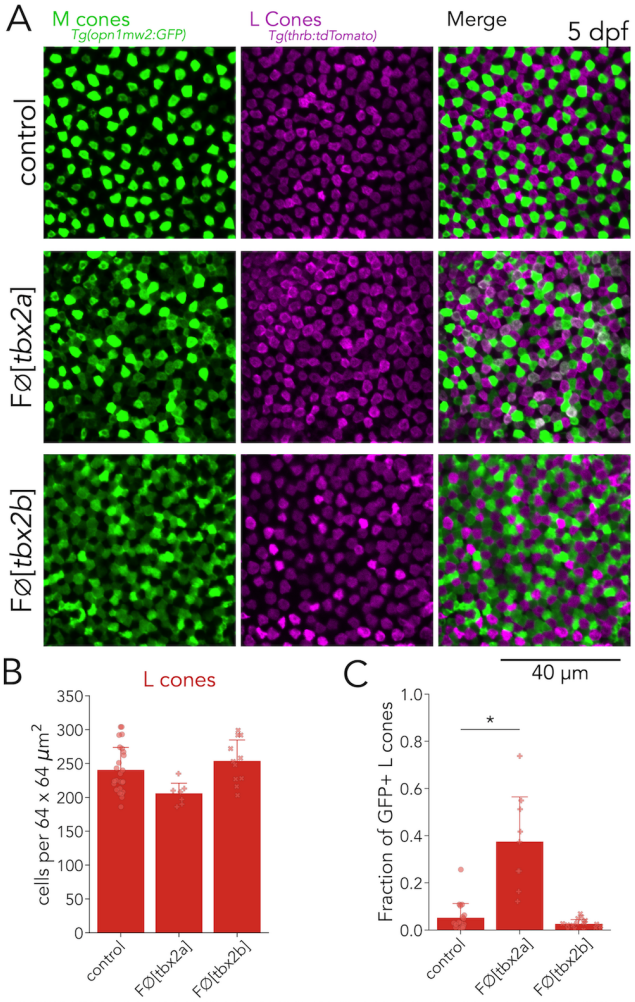

Figure. 5. Tbx2a inhibits M-opsin expression in L cones. (A) Representative confocal images of the central retina of control, FØ[*tbx2a*] and FØ[*tbx2b*] at 5 dpf, in double transgenic larvae that label M cones—or M-opsin expressing cells—with GFP (green) and L cones with tdTomato (magenta). FØ[*tbx2a*] and FØ[*tbx2b*] display an increase in GFP-positive cells. In FØ[*tbx2a*], increase in GFP signal is restricted to tdTomato-positive cells which appear as double-positive (white) in merged images, while in FØ[*tbx2b*], increase in GFP signal is excluded from tdTomato-positive cells, producing a decrease in the space without fluorescence. **(B)** Quantification of L cones in the central retina shows no significant changes in FØ[*tbx2a*] ($n_{FØ[tbx2a]}$=7) compared to control (n_{wt} = 25, Mann-Whitney U = 143, p = 0.012), or in FØ[*tbx2b*] ($n_{FØ[tbx2b]}$=11, Mann-Whitney U = 104.5, p = 0.26). **(C)** Quantification of the fraction of GFP-positive L cones (double positive cells in A) reveals a significant increase only in FØ[*tbx2a*] (Mann-Whitney U = 4, p = 8 × 10^{-5}), and not in FØ[*tbx2b*] (Mann-Whitney U = 186.5, p = 0.27)

Tbx2a inhibits M-opsin expression in L cones

After ascertaining the requirement of Tbx2a and Tbx2b in UV-cone specification, we examined whether either of these transcription factors impacted the specification of other photoreceptor subtypes. In addition to expression in UV cones, we detected significant enrichment of *tbx2a* in L cones, albeit with expression levels lower than in UV cones (Fig-

ure 4B). Furthermore, our qPCR quantification of opsins revealed a significant increase in M-opsin expression in *tbx2a* mutants—specifically of *opn1mws2*—along with a small decrease in L-opsin expression—specifically of *opn1lw2* (Figure 3 – Figure Supplement 1). Based on these results, we tested whether Tbx2a is involved in M-cone or L-cone specification.

To examine M and L cones, we assessed FØ[*tbx2a*] larvae using an M-cone reporter line, where GFP expression is under direct control of the M-opsin promoter, *Tg(opn1mws2:GFP)*, in combination with an L-cone reporter line, *Tg*(*thrb:tdTomato*). In control larvae, the expression of GFP and tdTomato is non-overlapping, reflecting the distinct fate of M cones and L cones (Figure 5A, top). In FØ[*tbx2a*], we found no significant changes in the number of L cones (identified by their tdTomato expression) (Figure 5B), and a dramatic increase in the number of GFP-positive cells (presumptive M cones). Interestingly, in FØ[*tbx2a*], many GFP-positive cells co-express tdTomato (L-cone marker)—a phenotype which is not present in control larvae (Figure 5A, middle). We quantified the fraction of tdTomato-positive L cones with significant GFP expression in control larvae (see methods) and found that, only a small fraction of L cones are classified as double positive (mean ± s.d.: 5.2% ± 6.0). In comparison, in FØ[*tbx2a*], this fraction is significantly higher (mean ± s.d.: 37.5% ± 18.9%, $p < 0.01$) (Figure 5C). This abnormal expression of GFP in L cones in FØ[*tbx2a*] larvae suggests a loss of inhibitory control over the M-opsin promoter, and is supported by the increase in M-opsin expression identified through qPCR. As a control, we repeated this M-cone and L-cone assessment in FØ[*tbx2b*] larvae—despite no detectable expression of *tbx2b* in M or L cones (Figure 3B). While FØ[*tbx2b*] also had a significant increase in the number of GFP-positive cells (see next section) (Figure 5A, bottom), there were no changes in the number of tdTomato-positive L cones (Figure 5B) or in the fraction of L cones with significant GFP expression (mean ± s.d.: 2.6% ± 1.8%, $p = 0.27$) (Figure 5C).

These results suggest that Tbx2a, but not Tbx2b, is important for L-cone identity. While not required for their initial specification, upon maturation, Tbx2a helps preserve L cone identity. Without Tbx2a, L cones are unable to suppress M-opsin expression (58). Overall, analysis of FØ[*tbx2a*] larvae revealed that Tbx2a is important for both UV-cone specification and maintaining L-cone identity.

Tbx2b inhibits M-opsin expression in S cones

After identifying an additional role for Tbx2a, we turned our analysis to Tbx2b. Our RNA-seq revealed that in addition to UV cones, *tbx2b* is expressed in S cones (Figure 4A). Furthermore, our qPCR quantification also showed an increase in M-opsin expression in FØ[*tbx2b*]—specifically of *opn1mw1* and *opn1mw2* (Figure 4 - figure supplement 1)—and in S-opsin expression. Based on these results, we tested whether Tbx2b is involved in S- or M-cone specification.

For our analysis of S and M cones in FØ[*tbx2b*] larvae, we used the M-opsin reporter line, *Tg(opn1mws2:GFP)*, in combination with an S-cone reporter line, *Tg(opn1sw2:nfsB-mCherry)*. In control larvae, expression of the reporter proteins is largely non-overlapping, except for a small fraction of S cones that consistently expresses GFP (Figure 6A, top) (42). In FØ[*tbx2b*], we did not find significant changes in the number of S cones (identified by mCherry expression) (Figure 6B). As described above, in FØ[*tbx2b*], we observed a clear increase in the number of GFP-positive cells (presumptive M cones). Furthermore, in FØ[*tbx2b*] larvae, we found that this increase in GFP expression was restricted to S cones, which are double-positive for GFP and mCherry expression (Figure 6A, bottom). We quantified the fraction of mCherry-positive S cones with significant GFP expression, and found that, in control larvae, this fraction is low (mean ± s.d.: 9.2% ± 10.2%). In comparison, in FØ[*tbx2b*] this fraction is signif-

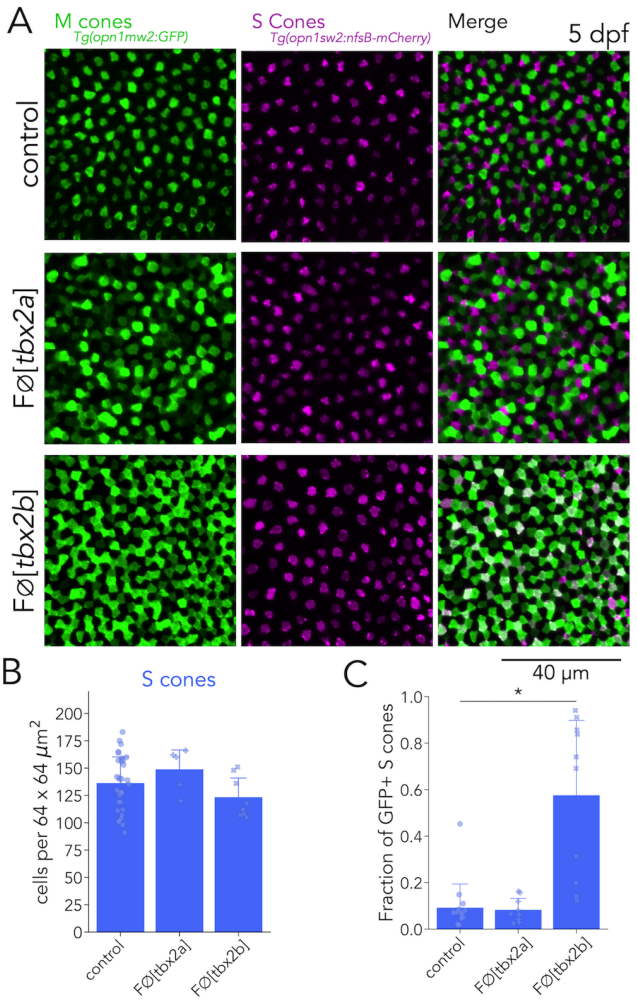

Figure. 6. Tbx2b inhibits M-opsin expression in S cones. (A) Representative confocal images of the central retina of control, FØ[*tbx2a*] and FØ[*tbx2b*] at 5 dpf, in double transgenic larvae that label M cones—or M-opsin expressing cells— with GFP (green) and S cones with mCherry (magenta). FØ[*tbx2a*] and FØ[*tbx2b*] display an increase in GFP-positive cells. In FØ[*tbx2a*], increase in GFP signal is excluded from mCherry-positive cells, producing a decrease in the space without fluorescence, while in FØ[*tbx2b*], increase in GFP signal is restricted to mCherry-positive cells, which appear as double positive (white) in the merged images. **(B)** Quantification of S cones in the central retina shows no significant changes in either FØ[*tbx2a*] ($n_{FØ[tbx2a]}=5$, Mann-Whitney U = 55, $p = 0.21$) or FØ[*tbx2b*] ($n_{FØ[tbx2b]}=8$, Mann-Whitney U = 179, $p = 0.17$) compared to control ($n_{wt} = 34$). **(C)** Quantification of the fraction of GFP-positive S cones (double positive cells in A) reveals a significant increase only in FØ[*tbx2b*] (Mann-Whitney U = 6, $p = 14 \times 10^{-5}$), and not in FØ[*tbx2a*] (Mann-Whitney U = 63, $p = 0.81$).

icantly higher (mean ± s.d.: 57.6% ± 32.2%, $p < 0.01$) (Figure 6C). This abnormal increase in GFP expression in S cones in FØ[*tbx2b*], combined with the increase in M-opsin expression found in our qPCR analysis, indicates a loss of inhibitory control over the M-opsin promoter. As a control, we repeated this S- and M-cone assessment in FØ[*tbx2a*]. We again observed an increase in GFP-positive cells in FØ[*tbx2a*] but without any significant changes in the number of S cones (Figure 6B) or in the fraction of mCherry-positive S cones with significant GFP expression (mean ± s.d.: 8.3% ± 4.9, $p = 0.81$) (Figure 6C). These results again corroborate our RNA-seq data showing that Tbx2a is not expressed by S cones and therefore is not involved in S-cone determination.

In summary, further analysis of FØ[*tbx2b*] revealed that, while not involved in the initial specification of S cones, upon maturation, Tbx2b helps to maintain S-cone identity. Without Tbx2b, S cones are unable to suppress the expression of M-opsin. Overall, analysis of FØ[*tbx2b*] larvae revealed that Tbx2b is important for both UV-cone specification and maintaining S-cone identity.

Discussion

Our work provides a valuable resource to accelerate discovery in photoreceptor and retinal biology. We have generated transcriptomic profiles from photoreceptors with unmatched depth and purity. These transcriptomes can be used to explore gene expression differences across photoreceptor subtypes. Importantly, we are able to reliably identify transcription factors that play a central role in controlling fate decisions during photoreceptor development. We also demonstrate how FØ screening can be applied as a rapid, efficient, and flexible platform to create and study loss-of-function phenotypes relevant to photoreceptor specification. In our study, we apply FØ-screening to investigate loss-of-function phenotypes, associated with the transcription factors Foxq2, Tbx2a and Tbx2b, and find a strong correlation between phenotype and expression pattern (Figure 7). Together these methods provide an excellent *in vivo* setting to discover the function of other novel genes identified in our RNA-seq dataset.

Relation to other transcriptomic datasets

Three recent studies have derived transcriptomes from zebrafish retinal cells and contain information from adult photoreceptors that provide an excellent resource to benchmark the quality of transcriptomes presented here. In our study, we derived samples using manual collection for a cell-type specific, SMART-seq2-based approach (38). The other recent studies used a variety of methods to segregate cell types in the retina. Rod transcriptomes were obtained by fluorescent-activated cell sorting (FACS) (Sun *et al*., 2018) (64). Retinal-cell transcriptomes were obtained using a single-cell droplet-based (dropSeq) approach in adults and at several time points during development (Hoang *et al*., 2020) (43). Finally, transcriptomes from adult zebrafish photoreceptors were obtained by enrichment through FACS followed by dropSeq (Ogawa and Corbo, 2021) (44).

We find that there are general consistencies across these datasets, which can be exemplified by focusing on phototransduction genes: we identify rod-enrichement in 26 out of the 27 phototransduction genes that are known to be rod-specific (the exception, *cngb1b*, previously named *si:dkey-44k1.5,* was found as expressed in all photoreceptors), and 22 are also identified as rod-enriched in Sun *et al*., 2018. Our datset also identifies cone-enrichement in 31 out of the 35 phototransduction genes known to be cone specific, with sparse or null counts for *rcvrnb*, *gnb3a*, *pde6hb* and *opn1mw4*. We did not detect *opn1mw4* because we used *Tg(opn1mw2:GFP)* to collect M cones. Out of the total 62 known phototransduction genes, 58 display very similar subtype-specific expression patterns in Ogawa and Corbo, 2021. We found that these subtype-specific expression patterns are obscured in Hoang *et al*., 2020 due to contamination with rod transcripts in all the retinal cells derived from adults. In this study many known rod-specific genes are present in all photoreceptor subtypes (Figure 1 - figure supplement 3A). Rods are the predominant cell type in the zebrafish adult retina—constituting ~40% of all photoreceptors (39). In our experience, rods are fragile during dissociation and rod contamination presents a challenge to obtaining pure, subtype-specific datasets. Rhodopsin (*rho*) detection in non-rod samples is a simple way to assess contamination. We find that samples in Sun *et al*., 2018 and in Hoang *et al*., 2019 have significant rod contamination (> 15%), while in Ogawa and Corbo, 2021 and in the data presented here, the rod contamination is low (< 5%) (Figure 1 – figure supplement 3B). Transcriptome depth, quantified as the number of reads per sample for our study and for Sun *et al.*, 2018 or reads per cell

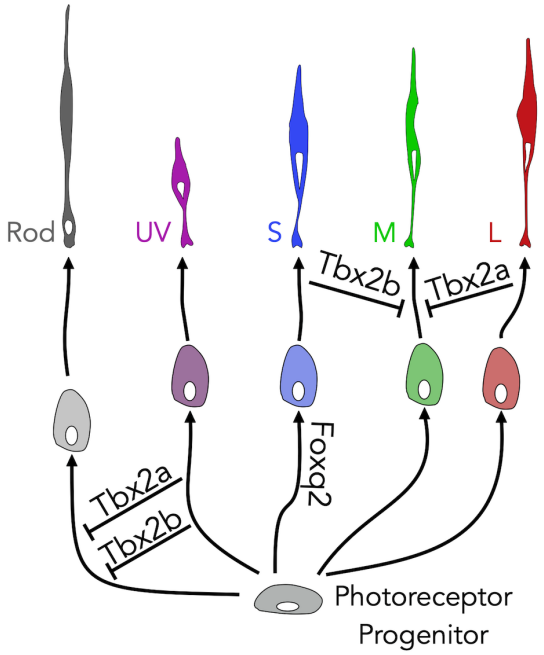

Figure. 7. Summary diagram. The specification of zebrafish UV cones requires an early decision in photoreceptor progenitors in which Tbx2a and Tbx2b are required to repress rod fate. The specification of zebrafish S cones also requires an early decision in which Foxq2 is required to activate S-cone fate. After their initial specification, S cones require Tbx2b to repress M-cone fate, while L cones require Tbx2a to repress M-cone fate.

for the dropSeq studies, was considerably higher in our study (Figure 1 – figure supplement 3C).

We then focused on expression of the targets of our FØ-screen—*foxq2, tbx2a*, and *tbx2b*. The phenotypes of our targets relate directly to the expression patterns derived from our transcriptomes. We find that similar to our rod samples, transcriptomes from FACS-rods have low counts for these three transcription factors. The transcriptomes obtained from dropSeq show enrichment of: *foxq2* in S cones (but also significant expression in UV cones), *tbx2a* in UV cones (but not clearly in L cones), *tbx2b* in S cones (but not clearly in UV cones). Finally, transcriptomes from FACS-dropSeq photoreceptors matched our data for *foxq2* and *tbx2b*, but the expression of *tbx2a* in L cones was missed (Figure 1 – figure supplement 3D).

In summary, we find that the methods presented in this study are especially useful to generate high-quality transcriptomes of targeted cells. High depth and low contamination increase the statistical confidence and allow the detection of genes expressed at relatively low levels (*e.g.*, *tbx2a* expression in L cones). Our method nicely complements dropSeq approaches which sample many more cells, which is especially advantageous for discovering new cell types or tracking developmental trajectories. In our view, these techniques are complementary and integration across datasets is important. To facilitate such comparisons, we have created an interactive plotter that integrates the reanalysis across the datasets outlined here. Our interactive plotter is openly available for the community and allows easy exploration and direct comparisons of all datasets (https://github.com/angueyraNIH/drRNAseq/), including code and data needed to replicate our analyses.

Reliability and efficiency of FØ screening

The RNAseq data and analyses presented here provide a substantial number of candidate genes potentially involved in the specification of photoreceptor subtypes. The generation of loss-of-function mutants remains a cornerstone in the study of gene function. Nevertheless, creating true mutants for all candidate genes would require excessive effort and resources, especially given the need for crosses with the relevant reporter lines. We propose that using an FØ screen in this context is an advantageous method to accelerate discovery and will enable future resources to be focused adequately. To validate this method and to quantitatively benchmark its reliability and efficiency, we first focused on *foxq2*, an S-cone specific transcription factor with high expression and reliable detection in S cones across datasets (Figure 1 - figure suppelment 3C). A recent report, published as we completed this study, independently identified *foxq2* as a cone-enriched transcription factor and characterized a *foxq2* full knockout in zebrafish (24), allowing us to make direct comparisons between true germline mutants and genetic mosaic FØ[*foxq2*] larvae. In germline *foxq2* mutants, there is a negligible expression of S opsin and a complete loss of S cones, without any apparent change in the number of any other photoreceptor subtype (Figure 7). In our FØ[*foxq2*] larvae, we found an average loss of S cones of 80%, without changes in the densities of the other photoreceptor subtypes. This demonstrates that the methods presented here are able to create mutations in a targeted gene with high efficiency and, to reproduce the phenotype of a full knockout without noticeably affecting related cell types. Moreover, the ability to evaluate the effect of gene mutations in less than a week, in any genetic background—transgenic lines or other mutant lines to investigate genetic pathways and networks—makes FØ screening a versatile technique to identify other transcription factors that may be involved in photoreceptor development and subtype specification.

Tbx2 is a master regulator of photoreceptor fate

After the success of uncovering phenotypes in FØ[*foxq2*], we explored the effects of *tbx2* mutations in photoreceptor identity. Our analyses revealed that Tbx2 is connected to the fate of all photoreceptor subtypes in zebrafish (Figure 7).

First, we showed that *tbx2a* and *tbx2b* are both expressed in UV cones, and the loss of either gene impairs UV-cone specification. The high conservation in the amino acid sequence of TBX2 across vertebrates and the specific expression in evolutionarily related cone subtypes (*opn1sw1*-expressing photoreceptors in zebrafish, chicken and primates) (46, 59) suggests that TBX2 may play a similar role across vertebrate species. We find that loss of UV cones in either *tbx2a* or *tbx2b* mutants is associated with an increase in the number of rods during development. The switch in fate from UV cone to rod suggests that Tbx2a and Tbx2b play a role in an early fate decision in photoreceptor progenitors, allowing the acquisition of UV-cone identity by actively repressing the expression of rod genes. Interestingly the increase in rods (or rhodopsin expression) was not equal between FØ[*tbx2a*] and FØ[*tbx2b*], suggesting that the two transcription factors regulate downstream targets differently. In addition, *in vitro* experiments have shown that Tbx2 binds to DNA as a monomer (60), which makes the possibility of Tbx2a/Tbx2b dimers unlikely. It still remains a mystery why the specification of UV cones in zebrafish would require a "two-factor authentification" system that relies on two highly-homologous but independent transcription factors.

Second, we show that *tbx2a* and *tbx2b* are expressed in L cones and S cones, respectively. Further, Tbx2a and Tbx2b help maintain L-cone and S-cone identity by repressing the expression of M opsin *in vivo*. A recent study in cichlids demonstrated that Tbx2a can bind and directly regulate the M-opsin promoter *in vitro* (58). This work also found that expression of *tbx2a* correlated strongly with the relative expression of M and L opsins, which cichlid species use to adjust their overall spectral sensitivity and match the requirements imposed by their habitats. These findings highlight that Tbx2 plays a late role in L-cone and S-cone fate, helping to maintain their identity after specification.

Outlook

While conducting the experiments described in this paper, we learned a few lessons worth highlighting. First, we find that manual picking targeted cell types allowed us to focus on collecting healthy cells and generate transcriptomes of high depth and quality. Another important advantage of this

method is that barriers imposed by a cell type with a low density can be largely ignored, as long as the targeted cell types can be recognized. For this reason we think it would be interesting to apply this technique to fully understand further subdivisions of each photoreceptor subtype including the differences between *opn1lw1*- and *opn1lw2*-expressing L cones (65) or between *opn1mw4*-expressing M cones and other M cones in zebrafish (44). Furthermore it would be interesting to explore regional specializations across the retina like the one proposed for UV cones in the acute zone (66), and for fovea *vs.* periphery differences in primates (46). This technique is likely to also be useful beyond photoreceptors to dissect differences between subtypes of other retinal cells.

Second, we find it is critical to create fast and easy access to multiple transcriptomic datasets. Eliminating technological barriers is important to ensure data can be accessed by all users. By ensuring proper access, new hypotheses pertaining to factors involved in photoreceptor development and other aspects of photoreceptor biology can be more readily explored. For example, many orthologs of human genes associated with retinal degenerations show high expression in zebrafish photoreceptors. For these reasons, we have taken a special effort to provide an interactive plotter that allows open exploration of all datasets in a single place. We hope that this tool is valuable to the scientific community.

Third, the results of our FØ-screen highlight some important features of how photoreceptors acquire their final identity. The process of specification seems to require several stages: defects in early stages can lead to a loss of subtypes (*e.g.* S cones in *foxq2* mutants) or to a change in identity (*e.g.* rods and UV cones in *tbx2* mutants), while defects in later stages can lead to specification deficits without loss of cells (*e.g.* misexpression of M opsin in *tbx2* mutants). In addition, while some transcription factors (like Foxq2 but also Thrb for example) mainly play a role in activating a particular fate, it is clear that others (like Tbx2, but also shown for Nr2e3 or Prdm1) play mainly a role in inhibiting the fate of other cell types (49, 67). Because of its conserved sequence and expression, TBX2 may play a similar role in mammalian S cones—actively repressing the fate of rods and M/L cones. Such active repression is most likely a fundamental mechanism to maintain subtype identity throughout the life span of an organism. These mechanisms of cell identity echo beyond photoreceptors into the context of specification of any cell type.

Finally, although we focused on the differential expression of transcription factors because of their central role in controlling fate decisions, similar approaches can be used to identify and study the function of genes involved in phototransduction, metabolism, ciliary transport, synaptic machinery, or any other aspect of photoreceptor biology. The dataset and methods described here are an excellent resource to propose hypotheses, to generate an initial list of candidate genes and to perform efficient screening.

Materials and Methods

Animals.

We grew zebrafish larvae at 28°C in E3 embryo media (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl2, and 0.33 mM MgSO4) and kept them under a 12 h:12 h light-dark cycle. At 1 dpf, we added 0.003% 1-phenyl-2-thiourea (PTU) to the embryo media to block melanogenesis. All work performed at the National Institute of Health was approved by the NIH Animal Use Committee under animal study protocol #1362-13. For RNAseq samples with adult zebrafish, animals of both sexes were used. For FØ, larvae were examined at 5 dpf. At these ages, sex cannot be predicted or determined, and therefore sex of the animals was not considered. Transgenic lines used in this study are listed in Table 1.

Table 1. Zebrafish transgenic lines

Label	Transgenic line	Reference
Rods	*Tg(xOPS:EGFP)*^{fl1Tg}	(39)
UV cones	*Tg(-5.5opn1sw1:EGFP)*^{kj9Tg}	(40)
UV cones	*Tg(opn1sw1:nfsB-mCherry)*^{q28Tg}	(61)
S cones	*Tg(-3.5opn1sw2:EGFP)*^{kj11Tg}	(41)
S cones	*Tg(opn1sw2:nfsB-mCherry)*^{q30Tg}	(68)
M cones	*Tg(opn1mw2:EGFP)*^{kj4Tg}	(42)
L cones	*Tg(thrb:tdTomato)*^{q22Tg}	(18)

RNA-seq sample collection.

We euthanized adult zebrafish by immersion in ice-cold water (below 4°C) followed by decapitation. We pierced the cornea with a 30-gauge needle and removed the cornea and lens before performing enucleation. Once the eye was isolated, we gently separated the retina from sclera and RPE using fine forceps or electrically-sharpened tungsten electrodes (69) and immediately started incubation in papain solution (5 U/mL papain Calbiochem#5125, 5.5 mM L-Cysteine, 1 mM EDTA in divalent-free Hank's balanced salt solution) for 35 minutes at 28°C. After a brief wash in DMEM supplemented with 5% bovine serum albumin, we performed mechanical trituration of the retina with the tip of a 1 mL pipette and used a cell-strainer polystyrene tube to obtain a single-cell suspension. After spin-down (2000x G for 2 min), we resuspended cells in 500 µL of enzyme-free fresh DMEM and diluted the cell suspension into three serial 10-fold dilutions before plating in glass-bottom petri dishes. The dilutions ensured that we could find a preparation where the density of cells and debris was low and most photoreceptors were truly isolated. We inspected the cell suspension using an epifluorescence microscope (Invitrogen EVOS cell-imaging system) and collected 20 photoreceptors per retina based on their fluorescence and morphology (prioritizing cells that looked healthy, had intact outer segments, visible mitochondrial bundles and undamaged cell membranes) using an oil-based microinjector system (Eppendorf CellTram 4R) and glass pipettes with a 15 µm opening (Eppendorf TransferTip-ES). After collection, we resuspended photoreceptors in 1 µL of fresh PBS, reinspected cells for fluorescence, collected them in a PCR tube containing 8 µL of lysis buffer of the RNA kit and kept the tube on ice until cDNA libraries were prepared. We used the SMART-seq v4 ultra-low input RNA kit for sequencing

(Takara #634897) using the manufacturer's instructions for single-cell samples, followed by the Low Input Library Prep Kit v2 (Takara #634899). We pooled up to 12 samples (with different barcodes) in one lane of a flow cell for sequencing (Illumina HiSeq 2500) and used a 150 bp paired-end read configuration. The first sequencing batch contained 4 UV-cone and 4 S-cone samples in a single flow cell, and the second sequencing batch contained the rest of the samples divided across 2 flow cells (6 rod, 1 UV-cone, 2 S-cones, 7 M-cones and 6 L-cone samples).

RNA-seq data analysis.

After an initial quality control and trimming of primer and adapters sequences using Trimmomatic (70), we aligned reads to the zebrafish genome (Danio rerio.*GRCz11*) using HiSat2 (71) and we assembled and quantified transcripts using Stringtie (72). We performed differential expression analysis using DEseq2 and pcaExplorer for initial visualizations (73, 74). Genes were considered as differentially expressed if |fold-enrichment|> 1.5, p-value < 0.01 and the estimated false positive rate or p-adjusted < 0.1. In addition, genes were required to have positive reads in > 50% of the enriched samples. To be able to detect differences that relied on expression on just 1 or 2 cone subtypes, we removed the requirement on fold-enrichment in rod *vs.* cone comparisons. To further explore the data, we transformed read numbers into fragments per kilobase per million reads (FPKM) (Supplementary Data 01) and developed custom routines in Python for plotting. We subselected transcription factors by selecting genes identified with "DNA-binding transcription factor activity" in ZFIN (75) and repeated principal component and differential expression analyses (Supplementary Data 02 and 03). To ensure broad access to our transcriptomic data, we provide access to the raw data (GEO accession number GSE188560), and after analysis in several formats including as a plain csv file, as a Seurat object for easy integration with dropSeq datasets (76), and finally as an interactive database for easy browsing and visualization (https://angueyraNIH.github.io/drRNAseq/). In order to make direct comparisons between our data and other RNAseq studies, we have integrated visualizations that use their publicly available data. For rod transcriptomes obtained using FACS we used the provided analyzed data, which includes gene log counts per million (cpm) for 4 rod samples (GFP-positive cells) and 4 non-rod samples (GFP-negative retinal cells) (GSE100062) (64). For transcriptomes from adult photoreceptors obtained using FACS followed by dropSeq (44), we used the Seurat object provided by the authors (GSE175929) and we used custom scripts in R using RStudio and Seurat (76–78) to export average expression values and percent of cells with positive counts of each gene for each cluster. For trancriptomes of retinal cells obtained using dropSeq (43), we used the Seurat object for zebrafish development provided by the authors (http://bioinfo.wilmer.jhu.edu/jiewang/scRNAseq/), and we updated the object to Seurat v03 (79), extracted cells that corresponded to adult rods and cones, performed clustering and used the expression of opsins and other markers to identify cone subtypes (including *arr3a* for L and M cones, *arr3b* and *tbx2b* for UV and S cones, *thrb* and *si:busm1-57f23.1* for L cones and *foxq2* for S cones). All results and scripts necessary to recreate this analyses are also provided openly (https://github.com/angueyraNIH/drRNAseq).

Quantitative PCR (qPCR).

We euthanized groups of 20 to 30 zebrafish larvae by immersion in ice-cold water (below 4°C) and immediately performed RNA extraction using the RNeasy Mini Kit (Qiagen) and reverse transcription using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher), which relies on random primers. Samples were kept frozen (-20°C) until use. For qPCR assays we used the PowerUp SYBR Green Master Mix (Thermo Fisher) and a 96-well system (CF96, Biorad) following manufacturer's protocols. We estimated expression levels using the relative standard curve method, using 5 serial standard dilutions of cDNA obtained from wild-type larvae. To calculate fold differences in gene expression, we normalized transcripts levels to the levels of β-actin 2 (*actb2*), and all measurements were repeated in triplicates. We performed statistical testing using Mann-Whitney rank sum tests, with a p-value < 0.05 required for significance. All primers used for qPCR are provided in Table 2.

Table 2. qPCR primers

Gene	Primer (5' to 3')
rho F	TCCGAGACCACACAGCG
rho R	CTGCTTGTTCATGCAGATG
opn1sw1 F	ATGGTCCTTGGCTGTTCTGG
opn1sw1 R	CCTCGGGAATGTATCTGCTCC
opn1sw2 F	GGAGGAATGGTGAGTTTGTG
opn1sw2 R	GGTCTTGAAGGTAAAGTTCC
opn1mw1 F	CAGCCCAGCACAAGAAACTC
opn1mw1 R	AGAGCAACCTGACCTCCAAGT
opn1mw2 F	TTTTTGGCTGGTCCCGATACA
opn1mw2 R	CAGGAACGCAGAAATGACAGC
opn1mw3 F	TGCTTTCGCTGGGATTGGATT
opn1mw3 R	CCCTCTGGAATATACCTTGACCA
opn1mw4 F	CACGCTTTCGCAGGATGC
opn1mw4 R	CGGAATATACCTGGACCAAC
opn1lw1 F	CCCACACTGCATCTCGACAA
opn1lw1 R	AAGGTATTCCCCATCACTCCAA
opn1lw2 F	AGAGGGAAGAACTGGACTTTCAGA
opn1lw2 R	TTCAGAGGAGTTTTGCCTACATATGT
actb2 F	GTACCACCAGACAATACAGT
actb2 R	CTTCTTGGGTATGGAATCTTGC

FØ-CRISPR screening.

We designed guide RNAs (gRNAs) using the online resource CHOPCHOP (80). We picked guides that targeted exons that encode the DNA-binding domains of transcription factors, had no self-complementarity, and that had 3 or more mismatches with other regions of the zebrafish genome. We used purified Cas9 protein (Alt-R® S.p. Cas9 nuclease, v.3) and chemically synthesized AltR-modified crRNA and tracrRNA (Integrated DNA technologies) for injections (35) (Ta-

ble 3). We prepared 1 µL aliquots of a 25 µM stock solution of Cas9 protein diluted in 20 mM HEPES-NaOH (pH 7.5), 350 mM KCl and 20% glycerol, and stored them at -80°C until use. We diluted each target-specific crRNA and the common tracrRNA using the provided duplex buffer as a 100 µM stock solution and stored them at -20°C. We prepared a 50 µM crRNA:tracrRNA duplex solution by mixing equal volumes of the stock solutions followed by annealing in a PCR machine (95°C, 5 min; cooling 0.1°C /s to 25°C; 25°C for 5 min; rapid cooling to 4°C), then we used the duplex buffer to obtain a 25 µM stock solution, before mixing equal volumes of the guides targeted to a single gene (3 guides for *tbx2a* and *tbx2b*, 2 guides for *foxq2*), making aliquots (3 µL for *tbx2a* and *tbx2b*, 2 µL for *foxq2*) and storing at -20°C until use. Prior to microinjection, we prepared 5 µM RNP complex solutions by mixing 1 µL of 25 µM Cas9, 1 µL of 0.25% phenol red and 3 µL of the *tbx2a* or *tbx2b* duplex solution, or 1 µL of pure water and 2 µL of the *foxq2* duplex solution . We incubated the RNP solution at 37°C for 5 minutes and kept at room temperature for use in the following 2 – 3 hours. We injected ~1 nL of the 5 µM RNP complex solution into the cytoplasm of one-cell stage zebrafish embryos.

Table 3. crRNA Sequences

Gene	gRNA
foxq2	TCATTTCTGGGCAATTCACCCGG
foxq2	CCCATCCGTTATTGTGCTTCCCG
tbx2a	TAACGATATGAAACCTGGGTTGG
tbx2a	GACAGCTATAAAATCGGTCTCGG
tbx2a	GGCTCTAACGATATGAAACCTGG
tbx2b	TATCGTTGGCTCTCACAATATGG
tbx2b	CAAGGTATGTACCCATATTTTGG
tbx2b	CGGAAGCTTCAGAATATCGTTGG

Genotyping.

We extracted DNA from the bodies of larvae (5 dpf) after enucleation by placing them in 25 µL of 25 mM NaOH with 0.2 mM EDTA, heating to 95°C for 30 minutes, and cooling to 4°C. Then we neutralized the solution by adding 25 µL of 40 mM Tris-HCl and vortexed the samples. For genotyping, we used a fluorescent PCR method (81), and all primers and expected sizes are provided in Table 4. We added the M13F adapter sequence (5'-TGTAAAACGACGGCCAGT-3') to forward primers and the PIGtail adapter sequence (5'-GTGTCTT-3') to reverse primers and used incorporation of fluorescent M13F-6FAM for detection. Our PCR mixture (1x), for a 20 µL reaction, contained forward primer (0.158 µM), reverse primer (0.316 µM), M13-FAM (0.316 µM, IDT), Phusion HF PCR Master Mix (1x, BioLabs), water (6.42 µL), and 2 µL of DNA. We used the following PCR protocol: 1) 98°C denaturation for 30 seconds, 2) 34 cycles of 98°C for 10 seconds, 64-67°C for 20 seconds, 72°C for 20 seconds 3) final extension at 72°C for 10 minutes, 4) hold at 4°C.

Immunohistochemistry.

We fixed zebrafish larvae at 5 dpf in 4% paraformaldehyde in phosphate buffered saline (PBS) for 1 hour at room temperature, followed by washes with 1% Triton X-100 PBS (3 x 10 min). We incubated larvae in primary antibodies diluted in 2% normal donkey serum (Jackson ImmunoResearch) and 1% Triton X-100 PBS for five days at 4°C with continuous and gentle shaking. To label S cones, we used a rabbit polyclonal anti-blue opsin (Kerafast EJH012) in a 1:200 dilution. After incubation with primary antibodies, we performed washes with 1% Triton X-100 PBS (3 x 15 min). We incubated larvae in donkey polyclonal secondary antibodies labeled with Cy5 (Jackson ImmunoResearch) in 1% Triton X-100 PBS overnight at 4°C with continuous and gentle shaking and performed washes in 1% Triton X-100 PBS (3 x 15 min) before mounting.

Table 4. Primer Sequences

Gene	Primer (5' – 3')	Product Size
foxq2 F	TGCTCTTCAAACAGGACAAGAA	
foxq2 R	TTCCAGCACATGCAGAAATAAT	406 bp
tbx2a F	CGTTCATTCGAATTCATTGTGT	
tbx2a R	TGTTTTGATGTCGCTGATTTTC	462 bp
tbx2b F	TGACGAGCACTAATGTCTTCCT	
tbx2a R	GCATCGCAGAACGAAAGTAGAT	309 bp

Imaging.

Sample Preparation and Image Acquisition For imaging, we enucleated eyes from fixed larvae using electrically sharpened tungsten wires (69). We placed isolated eyes on a coverslip and oriented eyes to place photoreceptors closest to the coverslip before using a small drop of 1.5% low-melting point agarose to fix them in place. Upon solidification, we added a polyvinyl-based mounting medium (10% polyvinyl alcohol type II, 5% glycerol and 25 mM Tris buffer, pH 8.7) and placed the coverslip on a glass slide, separated by a spacer (Grace Biolabs and/or duct tape) to avoid compression. We used the bodies of the larvae for genotyping and imaged the corresponding larval retinas using a Nikon A1R resonant-confocal microscope with a 25x, 1.10 NA water-immersion objective. We acquired z-stack images from a 64 µm x 64 µm square area of the central retina (dorsal to optic nerve) for photoreceptor quantification every 0.4 – 0.5 µm at a 1024 x 1024 pixel resolution.

Image Analysis Photoreceptor quantification: We imported confocal z-stacks of the central region of the retina (64 µm x 64 µm) into Napari (82). We created maximum intensity projections (MIPs) using a small subset of the z-stack (2 – 10 planes) that ensured that we captured all photoreceptor cells in the region into a single image. We then used the Napari plugin of Cellpose, a machine-learning based segmentation algorithm, to segment photoreceptors in each image, using the *cyto2* model (83). Finally, we manually corrected the segmentation to ensure all photoreceptors were properly counted. We performed statistical comparisons for counts of each photoreceptor subtype between *wt* and FØ larvae using

Mann-Whitney rank sum tests, with a p-value < 0.01 required for statistical significance.

Identification of double-positive cells in *tbx2* mutants: The increase in the *Tg(opn1mws2:GFP)*+ cells in FØ[*tbx2a*] and FØ[*tbx2b*] larvae made segmentation of the green channel difficult and unreliable, as these additional cells did not conform to the normal spatial separation between M cones. For this reason, we exploited the more accurate segmentation of L cones and S cones using the red channel, when imaging *Tg(thrb:tdTomato)* or *Tg(opn1sw2:nfsB-mCherry)*, respectively, and used it to create masks for the gren channel. We normalized the GFP signal across the whole image to span a $0-1$ range (to be able to make comparison between images) and used a 10-pixel erosion (to avoid effects due to optical blurring during imaging of the GFP signal) before calculating the average normalized GFP signal contained within each S-cone or L-cone. By plotting the distribution of GFP signal in L cones, we were able to establish a threshold of 0.195 that was exceeded by only 5.2% of L cones in *wt* larvae and used it to classify L cones as GFP+ in both *wt* and FØ larvae. In the original work that established the *Tg(opn1mws2:GFP)* line, it was noted that a subset of S cones in *wt* larvae are GFP+ (42). We were able to identify these cells using a GFP signal threshold of 0.255 (9.2% of *wt* S cones), and again we used this same threshold to quantify the fraction of GFP+ S cones in both *wt* and FØ larvae.

Statistical analyses.

We performed statistical analyses and data plots using Python in Jupyter notebooks (84). Values of data and error bars in figures correspond to averages and standard deviations, and for statistical comparisons we used Mann-Whitney U tests. Samples sizes and significant levels are stated in the figure captions. No randomization, blinding, or masking was used for our animal studies, and all replicates are biological. For RNAseq, we performed an initial sequencing run after collecting dissociated photoreceptors in squirrel (38) and zebrafish and established that a minimum of four samples per subtype were required to establish reliable statistical significance in differential gene-expression analysis. For FØ screening, our initial experiments were aimed at replicating the loss of UV cones and the increase in rods reported for *tbx2b* mutants (23), and we established that a minimum of 7 injected larvae per group were needed to provide enough statistical power in photoreceptor quantifications in FØ larvae. Injected larvae that had normal (wild-type) genotypes—a sign that CRISPR mutagenesis was not successful—were excluded from analysis.

Supplemental Materials

Supplementary data 01: Differential gene expression in zebrafish photoreceptors. Collection of CSV files containing output of differential gene expression analysis using DeSEQ2, along relevant directions (rods *vs.* cones; (UV+S) *vs.* (M+L); M *vs.* L; UV *vs.* S) and including counts (in FPKM) for all detected genes.

Supplementary data 02: Differential transcription factor expression in zebrafish photoreceptors: rods vs. cones. CSV file containing transcription factors with significant differential expression between rod and cone samples.

Supplementary data 03: Differential transcription factor expression in zebrafish photoreceptors: cone subtypes. Collection of CSV files containing transcription factors with significant differential expression between cone subtypes.

Data Availability

Sequencing data have been deposited in GEO under accession GSE188560. Data can be visualized at https://github.com/angueyraNIH/drRNAseq.

Acknowledgements

This work was supported by the National Eye Institute Intramural Research Program (W.L.), the National Institute on Deafness and Other Communication Disorders Intramural Research Program (*1ZIADC000085-01*, K.S.K), and National Eye Institute Pathway to Independence Award (*K99EY030144-01*, J.A.). We would like to thank Jamie Sexton, Alisha Beirl and Katherine Pinter for all the animal care and technical support, members of the Kindt Lab and the Li Lab for useful discussions, and Matthew Brooks, Linn Gieser and Anand Swaroop for sequencing services. We are very grateful to Rachel Wong, Takeshi Yoshimatsu, Ralph Nelson, James Fadool, Steven Leach, Brian Perkins and Xiangyun Wei for providing the transgenic zebrafish lines used in this study.

Declaration of Interests

The authors declare no competing or financial interests

Bibliography

1. Sanes, J. R. and Zipursky, S. L. Design principles of insect and vertebrate visual systems. *Neuron*, 66(1):15–36, April 2010. ISSN 1097-4199. doi: 10.1016/j.neuron.2010.01.018.
2. Nawrocki, L., BreMiller, R., Streisinger, G., and Kaplan, M. Larval and adult visual pigments of the zebrafish, Brachydanio rerio. *Vision Research*, 25(11):1569–1576, 1985. ISSN 0042-6989. doi: 10.1016/0042-6989(85)90127-0.
3. Engström, K. Cone Types and Cone Arrangement in the Retina of Some Cyprinids. *Acta Zoologica*, 41:277 – 295, 1960. doi: 10.1111/j.1463-6395.1960.tb00481.x.
4. Raymond, P. A., Barthel, L. K., Rounsifer, M. E., Sullivan, S. A., and Knight, J. K. Expression of rod and cone visual pigments in goldfish and zebrafish: a rhodopsin-like gene is expressed in cones. *Neuron*, 10(6):1161–1174, June 1993. ISSN 0896-6273. doi: 10.1016/0896-6273(93)90064-x.
5. Robinson, J., Schmitt, E. A., Hárosi, F. I., Reece, R. J., and Dowling, J. E. Zebrafish ultraviolet visual pigment: absorption spectrum, sequence, and localization. *Proceedings of the National Academy of Sciences of the United States of America*, 90(13):6009–6012, July 1993. ISSN 0027-8424. doi: 10.1073/pnas.90.13.6009.
6. Baden, T. and Osorio, D. The Retinal Basis of Vertebrate Color Vision. *Annual Review of Vision Science*, 5:177–200, September 2019. ISSN 2374-4650. doi: 10.1146/annurev-vision-091718-014926.

7. Furukawa, T., Morrow, E. M., and Cepko, C. L. Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell*, 91(4):531–541, November 1997. ISSN 0092-8674. doi: 10.1016/s0092-8674(00)80439-0.
8. Shen, Y.-c. and Raymond, P. A. Zebrafish cone-rod (crx) homeobox gene promotes retinogenesis. *Developmental Biology*, 269(1):237–251, May 2004. ISSN 0012-1606. doi: 10.1016/j.ydbio.2004.01.037.
9. Ji, D., Wang, S., Li, M., Zhang, S., and Li, H. Involvement of Lypge in the formation of eye and pineal gland in zebrafish. *Gene*, 642: 491–497, February 2018. ISSN 1879-0038. doi: 10.1016/j.gene.2017.11.062.
10. Ochocinska, M. and Hitchcock, P. NeuroD regulates proliferation of photoreceptor progenitors in the retina of the zebrafish. *Mechanisms of development*, 126(3-4):128–141, 2009. ISSN 0925-4773. doi: 10.1016/j.mod.2008.11.009.
11. Liu, H., Etter, P., Hayes, S., Jones, I., Nelson, B., Hartman, B., Forrest, D., and Reh, T. A. NeuroD1 regulates expression of thyroid hormone receptor 2 and cone opsins in the developing mouse retina. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 28(3):749–756, January 2008. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.4832-07.2008.
12. Forrest, D. and Swaroop, A. Minireview: the role of nuclear receptors in photoreceptor differentiation and disease. *Molecular Endocrinology (Baltimore, Md.)*, 26(6):905–915, June 2012. ISSN 1944-9917. doi: 10.1210/me.2012-1010.
13. Mears, A. J., Kondo, M., Swain, P. K., Takada, Y., Bush, R. A., Saunders, T. L., Sieving, P. A., and Swaroop, A. Nrl is required for rod photoreceptor development. *Nature Genetics*, 29(4):447–452, December 2001. ISSN 1061-4036. doi: 10.1038/ng774.
14. Cheng, H., Aleman, T. S., Cideciyan, A. V., Khanna, R., Jacobson, S. G., and Swaroop, A. In vivo function of the orphan nuclear receptor NR2E3 in establishing photoreceptor identity during mammalian retinal development. *Human Molecular Genetics*, 15(17):2588–2602, September 2006. ISSN 0964-6906. doi: 10.1093/hmg/ddl185.
15. Oh, E. C. T., Cheng, H., Hao, H., Jia, L., Khan, N. W., and Swaroop, A. Rod differentiation factor NRL activates the expression of nuclear receptor NR2E3 to suppress the development of cone photoreceptors. *Brain Research*, 1236:16–29, October 2008. ISSN 0006-8993. doi: 10.1016/j.brainres.2008.01.028.
16. Haider, N. B., Jacobson, S. G., Cideciyan, A. V., Swiderski, R., Streb, L. M., Searby, C., Beck, G., Hockey, R., Hanna, D. B., Gorman, S., Duhl, D., Carmi, R., Bennett, J., Weleber, R. G., Fishman, G. A., Wright, A. F., Stone, E. M., and Sheffield, V. C. Mutation of a nuclear receptor gene, NR2E3, causes enhanced S cone syndrome, a disorder of retinal cell fate. *Nature Genetics*, 24(2):127–131, February 2000. ISSN 1061-4036. doi: 10.1038/72777.
17. Ng, L., Hurley, J. B., Dierks, B., Srinivas, M., Saltó, C., Vennström, B., Reh, T. A., and Forrest, D. A thyroid hormone receptor that is required for the development of green cone photoreceptors. *Nature Genetics*, 27(1):94–98, January 2001. ISSN 1061-4036. doi: 10.1038/83829.
18. Suzuki, S. C., Bleckert, A., Williams, P. R., Takechi, M., Kawamura, S., and Wong, R. O. L. Cone photoreceptor types in zebrafish are generated by symmetric terminal divisions of dedicated precursors. *Proceedings of the National Academy of Sciences of the United States of America*, 110(37):15109–15114, September 2013. ISSN 1091-6490. doi: 10.1073/pnas.1303551110.
19. Enright, J. M., Lawrence, K. A., Hadzic, T., and Corbo, J. C. Transcriptome profiling of developing photoreceptor subtypes reveals candidate genes involved in avian photoreceptor diversification. *The Journal of Comparative Neurology*, 523(4):649–668, March 2015. ISSN 1096-9861. doi: 10.1002/cne.23702.
20. Xie, S., Han, S., Qu, Z., Liu, F., Li, J., Yu, S., Reilly, J., Tu, J., Liu, X., Lu, Z., Hu, X., Yimer, T. A., Qin, Y., Huang, Y., Lv, Y., Jiang, T., Shu, X., Tang, Z., Jia, H., Wong, F., and Liu, M. Knockout of Nr2e3 prevents rod photoreceptor differentiation and leads to selective L-/M-cone photoreceptor degeneration in zebrafish. *Biochimica Et Biophysica Acta. Molecular Basis of Disease*, 1865(6):1273–1283, June 2019. ISSN 1879-260X. doi: 10.1016/j.bbadis.2019.01.022.
21. Ochi, H., Sakagami, K., Ishii, A., Morita, N., Nishiuchi, M., Ogino, H., and Yasuda, K. Temporal expression of L-Maf and RaxL in developing chicken retina are arranged into mosaic pattern. *Gene expression patterns: GEP*, 4(5):489–494, September 2004. ISSN 1567-133X. doi: 10.1016/j.modgep.2004.03.005.
22. Oel, A. P., Neil, G. J., Dong, E. M., Balay, S. D., Collett, K., and Allison, W. T. Nrl Is Dispensable for Specification of Rod Photoreceptors in Adult Zebrafish Despite Its Deeply Conserved Requirement Earlier in Ontogeny. *iScience*, 23(12):101805, December 2020. ISSN 2589-0042. doi: 10.1016/j.isci.2020.101805.
23. Alvarez-Delfin, K., Morris, A. C., Snelson, C. D., Gamse, J. T., Gupta, T., Marlow, F. L., Mullins, M. C., Burgess, H. A., Granato, M., and Fadool, J. M. Tbx2b is required for ultraviolet photoreceptor cell specification during zebrafish retinal development. *Proceedings of the National Academy of Sciences of the United States of America*, 106(6):2023–2028, February 2009. ISSN 1091-6490. doi: 10.1073/pnas.0809439106.
24. Ogawa, Y., Shiraki, T., Fukada, Y., and Kojima, D. Foxq2 determines blue cone identity in zebrafish. *Science Advances*, 7(41):eabi9784, October 2021. ISSN 2375-2548. doi: 10.1126/sciadv.abi9784.
25. Roberts, M. R., Hendrickson, A., McGuire, C. R., and Reh, T. A. Retinoid X receptor (gamma) is necessary to establish the S-opsin gradient in cone photoreceptors of the developing mouse retina. *Investigative Ophthalmology & Visual Science*, 46(8):2897–2904, August 2005. ISSN 0146-0404. doi: 10.1167/iovs.05-0093.
26. Fujieda, H., Bremner, R., Mears, A. J., and Sasaki, H. Retinoic acid receptor-related orphan receptor alpha regulates a subset of cone genes during mouse retinal development. *Journal of Neurochemistry*, 108(1):91–101, January 2009. ISSN 1471-4159. doi: 10.1111/j.1471-4159.2008.05739.x.
27. Srinivas, M., Ng, L., Liu, H., Jia, L., and Forrest, D. Activation of the blue opsin gene in cone photoreceptor development by retinoid-related orphan receptor beta. *Molecular Endocrinology (Baltimore, Md.)*, 20(8):1728–1741, August 2006. ISSN 0888-8809. doi: 10.1210/me.2005-0505.
28. Ogawa, Y., Shiraki, T., Kojima, D., and Fukada, Y. Homeobox transcription factor Six7 governs expression of green opsin genes in zebrafish. *Proceedings. Biological Sciences*, 282(1812):20150659, August 2015. ISSN 1471-2954. doi: 10.1098/rspb.2015.0659.
29. Sotolongo-Lopez, M., Alvarez-Delfin, K., Saade, C. J., Vera, D. L., and Fadool, J. M. Genetic Dissection of Dual Roles for the Transcription Factor six7 in Photoreceptor Development and Patterning in Zebrafish. *PLoS genetics*, 12(4):e1005968, April 2016. ISSN 1553-7404. doi: 10.1371/journal.pgen.1005968.
30. Asai-Coakwell, M., March, L., Dai, X. H., Duval, M., Lopez, I., French, C. R., Famulski, J., De Baere, E., Francis, P. J., Sundaresan, P., Sauvé, Y., Koenekoop, R. K., Berry, F. B., Allison, W. T., Waskiewicz, A. J., and Lehmann, O. J. Contribution of growth differentiation factor 6-dependent cell survival to early-onset retinal dystrophies. *Human Molecular Genetics*, 22(7):1432–1442, April 2013. ISSN 1460-2083. doi: 10.1093/hmg/dds560.
31. Nadolski, N. J., Balay, S. D., Wong, C. X. L., Waskiewicz, A. J., and Hocking, J. C. Abnormal Cone and Rod Photoreceptor Morphogenesis in gdf6a Mutant Zebrafish. *Investigative Ophthalmology & Visual Science*, 61(4):9, April 2020. ISSN 1552-5783. doi: 10.1167/iovs.61.4.9.
32. Ogawa, Y., Shiraki, T., Asano, Y., Muto, A., Kawakami, K., Suzuki, Y., Kojima, D., and Fukada, Y. Six6 and Six7 coordinately regulate expression of middle-wavelength opsins in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America*, 116(10):4651–4660, March 2019. ISSN 1091-6490. doi: 10.1073/pnas.1812884116.
33. Duval, M. G., Oel, A. P., and Allison, W. T. gdf6a is required for cone photoreceptor subtype differentiation and for the actions of tbx2b in determining rod versus cone photoreceptor fate. *PloS One*, 9(3): e92991, 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0092991.
34. Kroll, F., Powell, G. T., Ghosh, M., Gestri, G., Antinucci, P., Hearn, T. J., Tunbak, H., Lim, S., Dennis, H. W., Fernandez, J. M., Whitmore, D., Dreosti, E., Wilson, S. W., Hoffman, E. J., and Rihel, J.

A simple and effective F0 knockout method for rapid screening of behaviour and other complex phenotypes. *eLife*, 10, January 2021. ISSN 2050-084X. doi: 10.7554/eLife.59683.

35. Hoshijima, K., Jurynec, M. J., Klatt Shaw, D., Jacobi, A. M., Behlke, M. A., and Grunwald, D. J. Highly Efficient CRISPR-Cas9-Based Methods for Generating Deletion Mutations and F0 Embryos that Lack Gene Function in Zebrafish. *Developmental Cell*, 51(5):645–657.e4, December 2019. ISSN 1878-1551. doi: 10.1016/j.devcel.2019.10.004.
36. Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., and McCarroll, S. A. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, 161(5):1202–1214, May 2015. ISSN 1097-4172. doi: 10.1016/j.cell.2015.05.002.
37. Wang, X., He, Y., Zhang, Q., Ren, X., and Zhang, Z. Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2. *Genomics, Proteomics & Bioinformatics*, pages S1672–0229(21)00048–6, March 2021. ISSN 2210-3244. doi: 10.1016/j.gpb.2020.02.005.
38. Kunze, V. P. *Genetic profiling of a cone-dominated retina and cone photoreceptor subtypes*. phd, Universität Oldenburg, 2017.
39. Fadool, J. M. Development of a rod photoreceptor mosaic revealed in transgenic zebrafish. *Developmental Biology*, 258(2):277–290, June 2003. ISSN 0012-1606. doi: 10.1016/s0012-1606(03)00125-8.
40. Takechi, M., Hamaoka, T., and Kawamura, S. Fluorescence visualization of ultraviolet-sensitive cone photoreceptor development in living zebrafish. *FEBS letters*, 553(1-2):90–94, October 2003. ISSN 0014-5793. doi: 10.1016/s0014-5793(03)00977-3.
41. Takechi, M., Seno, S., and Kawamura, S. Identification of cis-acting elements repressing blue opsin expression in zebrafish UV cones and pineal cells. *The Journal of Biological Chemistry*, 283(46): 31625–31632, November 2008. ISSN 0021-9258. doi: 10.1074/jbc.M806226200.
42. Tsujimura, T., Chinen, A., and Kawamura, S. Identification of a locus control region for quadruplicated green-sensitive opsin genes in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America*, 104(31):12813–12818, July 2007. ISSN 0027-8424. doi: 10.1073/pnas.0704061104.
43. Hoang, T., Wang, J., Boyd, P., Wang, F., Santiago, C., Jiang, L., Yoo, S., Lahne, M., Todd, L. J., Jia, M., Saez, C., Keuthan, C., Palazzo, I., Squires, N., Campbell, W. A., Rajaii, F., Parayil, T., Trinh, V., Kim, D. W., Wang, G., Campbell, L. J., Ash, J., Fischer, A. J., Hyde, D. R., Qian, J., and Blackshaw, S. Gene regulatory networks controlling vertebrate retinal regeneration. *Science (New York, N.Y.)*, 370(6519), November 2020. ISSN 1095-9203. doi: 10.1126/science.abb8598.
44. Ogawa, Y. and Corbo, J. C. Partitioning of gene expression among zebrafish photoreceptor subtypes. *Scientific Reports*, 11(1):17340, August 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-96837-z.
45. Renninger, S. L., Gesemann, M., and Neuhauss, S. C. F. Cone arrestin confers cone vision of high temporal resolution in zebrafish larvae. *The European Journal of Neuroscience*, 33(4):658–667, February 2011. ISSN 1460-9568. doi: 10.1111/j.1460-9568.2010.07574.x.
46. Peng, Y.-R., Shekhar, K., Yan, W., Herrmann, D., Sappington, A., Bryman, G. S., van Zyl, T., Do, M. T. H., Regev, A., and Sanes, J. R. Molecular Classification and Comparative Taxonomics of Foveal and Peripheral Cells in Primate Retina. *Cell*, 176(5):1222–1237.e22, 2019. ISSN 1097-4172. doi: 10.1016/j.cell.2019.01.004.
47. Taylor, S. M., Alvarez-Delfin, K., Saade, C. J., Thomas, J. L., Thummel, R., Fadool, J. M., and Hitchcock, P. F. The bHLH Transcription Factor NeuroD Governs Photoreceptor Genesis and Regeneration Through Delta-Notch Signaling. *Investigative Ophthalmology & Visual Science*, 56(12):7496–7515, November 2015. ISSN 1552-5783. doi: 10.1167/iovs.15-17616.
48. Thomas, J. L., Ochocinska, M. J., Hitchcock, P. F., and Thummel, R. Using the Tg(nrd:egfp)/albino zebrafish line to characterize in vivo expression of neurod. *PloS One*, 7(1):e29128, 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0029128.
49. Swaroop, A., Kim, D., and Forrest, D. Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nature Reviews. Neuroscience*, 11(8):563–576, August 2010. ISSN 1471-0048. doi: 10.1038/nrn2880.
50. Erickson, T., French, C. R., and Waskiewicz, A. J. Meis1 specifies positional information in the retina and tectum to organize the zebrafish visual system. *Neural Development*, 5:22, September 2010. ISSN 1749-8104. doi: 10.1186/1749-8104-5-22.
51. Heine, P., Dohle, E., Bumsted-O'Brien, K., Engelkamp, D., and Schulte, D. Evidence for an evolutionary conserved role of homothorax/Meis1/2 during vertebrate retina development. *Development (Cambridge, England)*, 135(5):805–811, March 2008. ISSN 0950-1991. doi: 10.1242/dev.012088.
52. Zhang, L., Cho, J., Ptak, D., and Leung, Y. F. The role of egr1 in early zebrafish retinogenesis. *PloS One*, 8(2):e56108, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0056108.
53. Omori, Y., Kubo, S., Kon, T., Furuhashi, M., Narita, H., Kominami, T., Ueno, A., Tsutsumi, R., Chaya, T., Yamamoto, H., Suetake, I., Ueno, S., Koseki, H., Nakagawa, A., and Furukawa, T. Samd7 is a cell type-specific PRC1 component essential for establishing retinal rod photoreceptor identity. *Proceedings of the National Academy of Sciences of the United States of America*, 114(39):E8264–E8273, September 2017. ISSN 1091-6490. doi: 10.1073/pnas.1707021114.
54. Kubo, S., Yamamoto, H., Kajimura, N., Omori, Y., Maeda, Y., Chaya, T., and Furukawa, T. Functional analysis of Samd11, a retinal photoreceptor PRC1 component, in establishing rod photoreceptor identity. *Scientific Reports*, 11(1):4180, February 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-83781-1.
55. Zhou, Y., Li, S., Huang, L., Yang, Y., Zhang, L., Yang, M., Liu, W., Ramasamy, K., Jiang, Z., Sundaresan, P., Zhu, X., and Yang, Z. A splicing mutation in aryl hydrocarbon receptor associated with retinitis pigmentosa. *Human Molecular Genetics*, 27(14):2563–2572, July 2018. ISSN 1460-2083. doi: 10.1093/hmg/ddy165.
56. Volkov, L. I., Kim-Han, J. S., Saunders, L. M., Poria, D., Hughes, A. E. O., Kefalov, V. J., Parichy, D. M., and Corbo, J. C. Thyroid hormone receptors mediate two distinct mechanisms of long-wavelength vision. *Proceedings of the National Academy of Sciences of the United States of America*, 117(26):15262–15269, June 2020. ISSN 1091-6490. doi: 10.1073/pnas.1920086117.
57. Yu, J.-K., Holland, N. D., and Holland, L. Z. AmphiFoxQ2, a novel winged helix/forkhead gene, exclusively marks the anterior end of the amphioxus embryo. *Development Genes and Evolution*, 213(2):102–105, March 2003. ISSN 0949-944X. doi: 10.1007/s00427-003-0302-3.
58. Sandkam, B. A., Campello, L., O'Brien, C., Nandamuri, S. P., Gammerdinger, W. J., Conte, M. A., Swaroop, A., and Carleton, K. L. Tbx2a Modulates Switching of RH2 and LWS Opsin Gene Expression. *Molecular Biology and Evolution*, 37(7):2002–2014, July 2020. ISSN 1537-1719. doi: 10.1093/molbev/msaa062.
59. Yamagata, M., Yan, W., and Sanes, J. R. A cell atlas of the chick retina based on single-cell transcriptomics. *eLife*, 10, January 2021. ISSN 2050-084X. doi: 10.7554/eLife.63907.
60. Sinha, S., Abraham, S., Gronostajski, R. M., and Campbell, C. E. Differential DNA binding and transcription modulation by three T-box proteins, T, TBX1 and TBX2. *Gene*, 258(1-2):15–29, November 2000. ISSN 0378-1119. doi: 10.1016/s0378-1119(00)00417-0.
61. Yoshimatsu, T., D'Orazi, F. D., Gamlin, C. R., Suzuki, S. C., Suli, A., Kimelman, D., Raible, D. W., and Wong, R. O. Presynaptic partner selection during retinal circuit reassembly varies with timing of neuronal regeneration in vivo. *Nature Communications*, 7:10590, February 2016. ISSN 2041-1723. doi: 10.1038/ncomms10590.
62. Yoshimatsu, T., Williams, P. R., D'Orazi, F. D., Suzuki, S. C., Fadool, J. M., Allison, W. T., Raymond, P. A., and Wong, R. O. Transmission from the dominant input shapes the stereotypic ratio of photoreceptor inputs onto horizontal cells. *Nature Communications*, 5:3699, May 2014. ISSN 2041-1723. doi: 10.1038/ncomms4699.
63. Raymond, P. A., Colvin, S. M., Jabeen, Z., Nagashima, M., Barthel, L. K., Hadidjojo, J., Popova, L., Pejaver, V. R., and Lubensky, D. K.

Patterning the cone mosaic array in zebrafish retina requires specification of ultraviolet-sensitive cones. *PloS One*, 9(1):e85325, 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0085325.

64. Sun, C., Galicia, C., and Stenkamp, D. L. Transcripts within rod photoreceptors of the Zebrafish retina. *BMC genomics*, 19(1):127, February 2018. ISSN 1471-2164. doi: 10.1186/s12864-018-4499-y.
65. Mackin, R. D., Frey, R. A., Gutierrez, C., Farre, A. A., Kawamura, S., Mitchell, D. M., and Stenkamp, D. L. Endocrine regulation of multichromatic color vision. *Proceedings of the National Academy of Sciences of the United States of America*, 116(34):16882–16891, August 2019. ISSN 1091-6490. doi: 10.1073/pnas.1904783116.
66. Yoshimatsu, T., Schröder, C., Nevala, N. E., Berens, P., and Baden, T. Fovea-like Photoreceptor Specializations Underlie Single UV Cone Driven Prey-Capture Behavior in Zebrafish. *Neuron*, 107(2): 320–337.e6, July 2020. ISSN 1097-4199. doi: 10.1016/j.neuron.2020.04.021.
67. Brzezinski, J. A., Lamba, D. A., and Reh, T. A. Blimp1 controls photoreceptor versus bipolar cell fate choice during retinal development. *Development (Cambridge, England)*, 137(4):619–629, February 2010. ISSN 1477-9129. doi: 10.1242/dev.043968.
68. Salbreux, G., Barthel, L. K., Raymond, P. A., and Lubensky, D. K. Coupling mechanical deformations and planar cell polarity to create regular patterns in the zebrafish retina. *PLoS computational biology*, 8(8):e1002618, 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002618.
69. Protocols, C. S. H. Sharpened Tungsten Needles. *Cold Spring Harbor Protocols*, 2012(4):pdb.rec069468, April 2012. ISSN 1940-3402, 1559-6095. doi: 10.1101/pdb.rec069468. Publisher: Cold Spring Harbor Laboratory Press.
70. Bolger, A. M., Lohse, M., and Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15):2114–2120, August 2014. ISSN 1367-4811. doi: 10.1093/bioinformatics/btu170.
71. Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8):907–915, August 2019. ISSN 1546-1696. doi: 10.1038/s41587-019-0201-4.
72. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, 11(9):1650–1667, September 2016. ISSN 1750-2799. doi: 10.1038/nprot.2016.095.
73. Love, M. I., Huber, W., and Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8.
74. Marini, F. and Binder, H. pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components. *BMC bioinformatics*, 20(1):331, June 2019. ISSN 1471-2105. doi: 10.1186/s12859-019-2879-1.
75. Bradford, Y. M., Slyke, C. E. V., Singer, A., Paddock, H., Eagle, A., Fashena, D., Howe, D. G., Frazer, K., Martin, R., Pich, C., Ramachandran, S., Ruzicka, L., and Westerfield, M. Zebrafish Information Network, the knowledgebase for Danio rerio research. Technical report, September 2021. Company: Cold Spring Harbor Laboratory Distributor: Cold Spring Harbor Laboratory Label: Cold Spring Harbor Laboratory Section: New Results Type: article.
76. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, 33(5):495–502, May 2015. ISSN 1546-1696. doi: 10.1038/nbt.3192.
77. RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA, 2020.
78. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
79. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. Comprehensive Integration of Single-Cell Data. *Cell*, 177(7):1888–1902.e21, June 2019. ISSN 1097-4172. doi: 10.1016/j.cell.2019.05.031.
80. Labun, K., Krause, M., Torres Cleuren, Y., and Valen, E. CRISPR Genome Editing Made Easy Through the CHOPCHOP Website. *Current Protocols*, 1(4):e46, April 2021. ISSN 2691-1299. doi: 10.1002/cpz1.46.
81. Carrington, B., Varshney, G. K., Burgess, S. M., and Sood, R. CRISPR-STAT: an easy and reliable PCR-based method to evaluate target-specific sgRNA activity. *Nucleic Acids Research*, 43(22): e157, December 2015. ISSN 1362-4962. doi: 10.1093/nar/gkv802.
82. contributors, n. napari: a multi-dimensional image viewer for python, 2019.
83. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, January 2021. ISSN 1548-7105. doi: 10.1038/s41592-020-01018-x.
84. Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., and Willing, C. Jupyter Notebooks – a publishing format for reproducible computational workflows. In Loizides, F. and Schmidt, B., editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87 – 90. IOS Press, 2016.

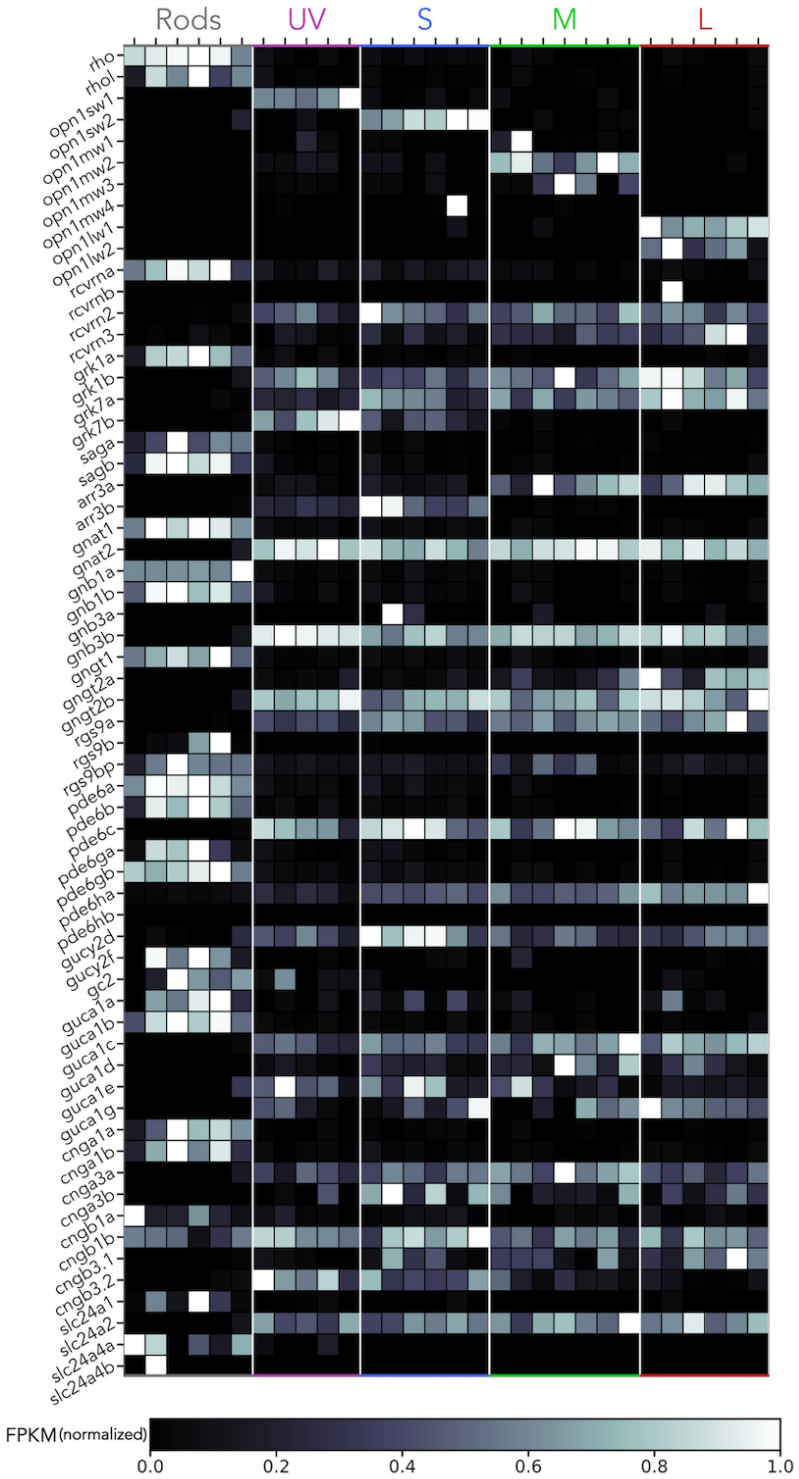

Figure 1 - figure supplement 1 . Expression of phototransduction genes. Heatmap showing expression of phototransduction genes for each RNAseq sample. Grey values indicate expression level normalized in each row by the maximal value. Genes have been arranged by functional family.

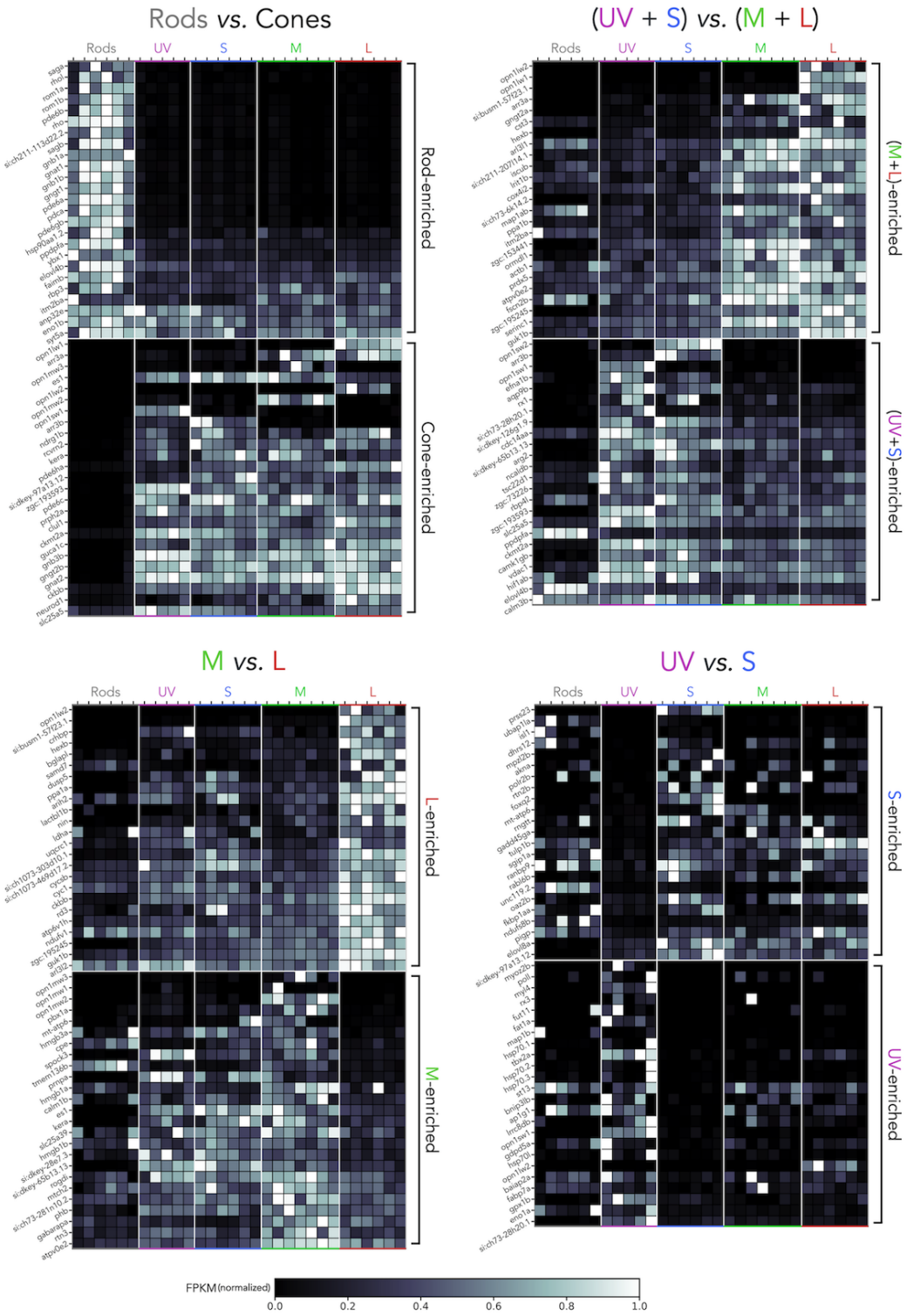

Figure 1 - figure supplement 2 . Differentially-expressed genes between photoreceptor subtypes. Heatmaps showing the top 50 differentially-expressed genes, identified using pairwise comparisons that follow the direction of principal components shown in Figure 1G, and indicated in the titles of each panel. Grey values indicate expression level normalized in each row by the maximal value. Genes have been arranged by degree of enrichment.

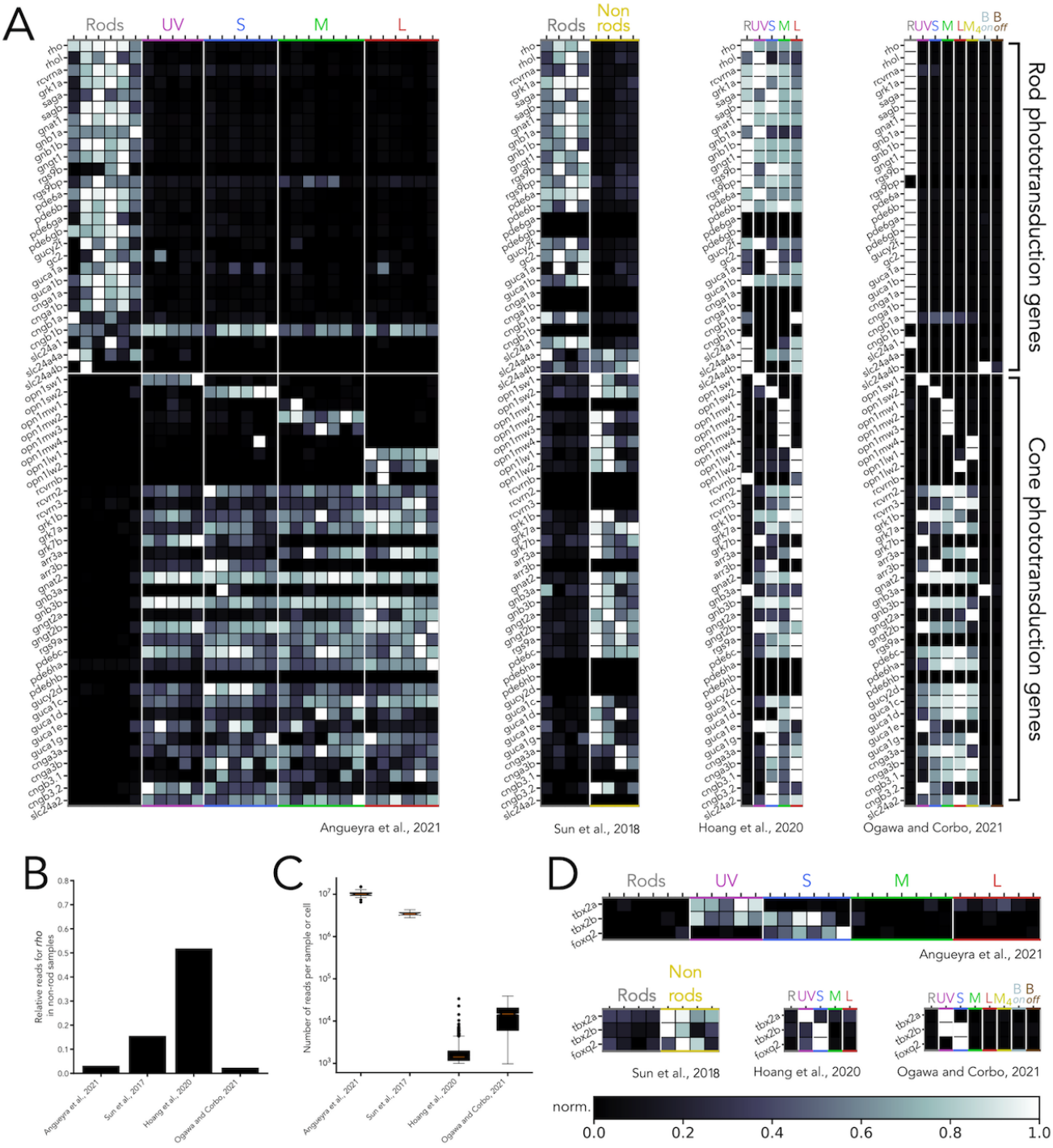

Figure 1 - figure supplement 3 . Comparison of RNAseq datasets across studies. (A) Heatmaps showing expression of phototransduction genes across four RNAseq studies containing transcriptomes of adult zebrafish photoreceptors. Grey values indicate expression level normalized in each row by the maximal value, and genes have been first arranged by their known expression in rods or cones and subsequently by functional family. **(B)** Degree of contamination in each study, measured as expression of rhodopsin (*rho*) in non-rod samples or cells, normalized to expression in rod samples or cells. **(C)** Transcriptome depth in each study, measured as the number of reads per sample (Angueyra *et al., 2021*, Sun *et al., 2018*), or number of reads per cell (Hoang *et al., 2020*, Ogawa *et al., 2021*), shown as a box and whiskers plot in log scale (red line corresponds to mean, boxes correspont to inter-quartile ranges, whiskers correspond to extremes and symbols correspond to outliers) **(D)** Heatmaps showing expression of the three transcription factors explored in this study, normalized by row. In heatmaps, R = rods, B_{on} = on bipolar cells, B_{off} = off bipolar cells.

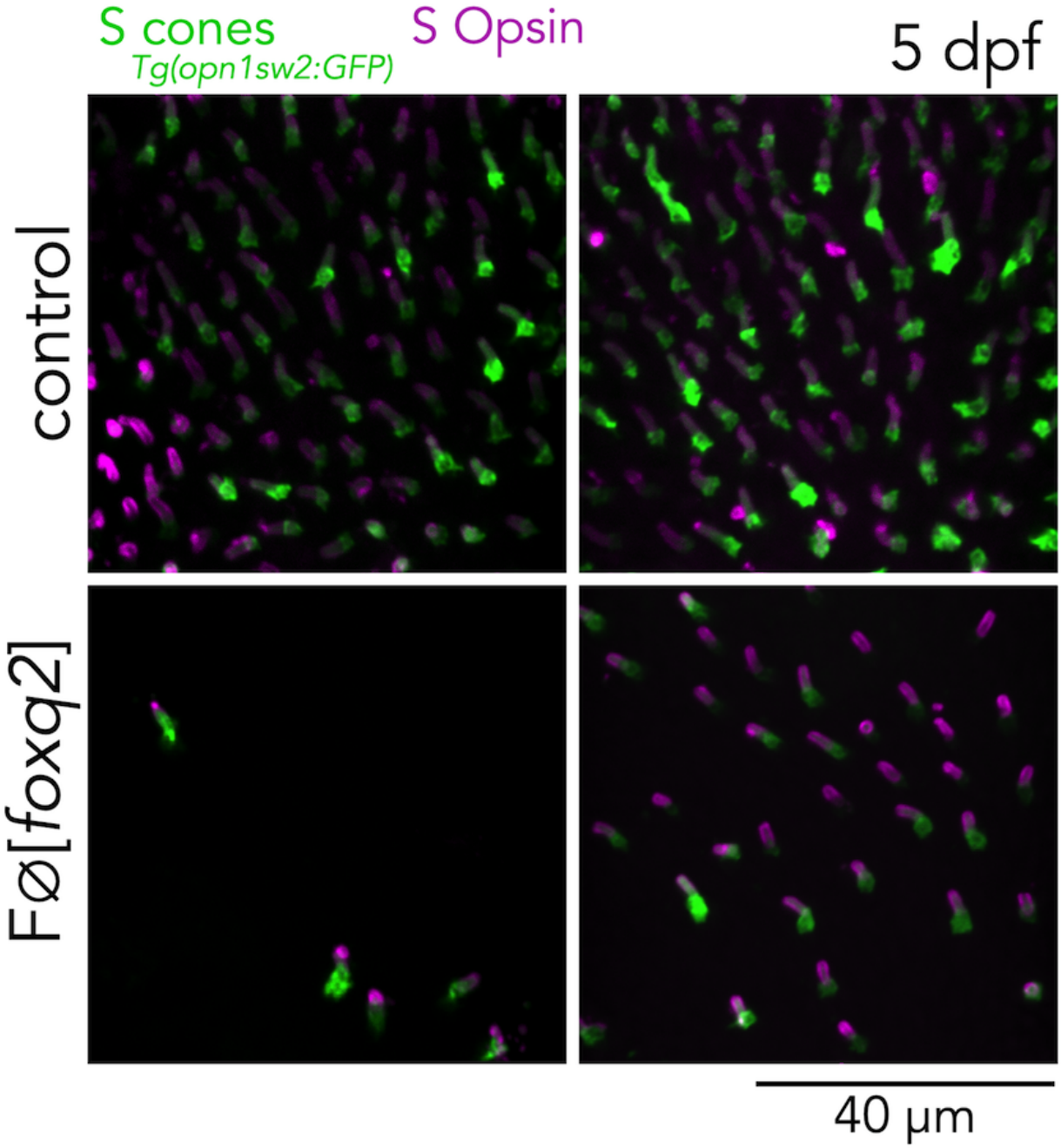

Figure 3 - figure supplement 1 . Mutations in *foxq2* cause a decrease in S opsin positive photoreceptors. Confocal images of the central retina of 2 control and 2 FØ[*foxq2*] larvae at 5 dpf, labeled with an S opsin antibody (magenta), in an S-cone reporter line (green). In the control images, all GFP-positive cells are also positive for S opsin antibody labeling (the curvature of the larval eye causes some apparent uneven labeling). In FØ[*foxq2*] larvae there is a significant loss of both GFP and S opsin antibody labeling, and remaining GFP-positive cells are also positive for S opsin.

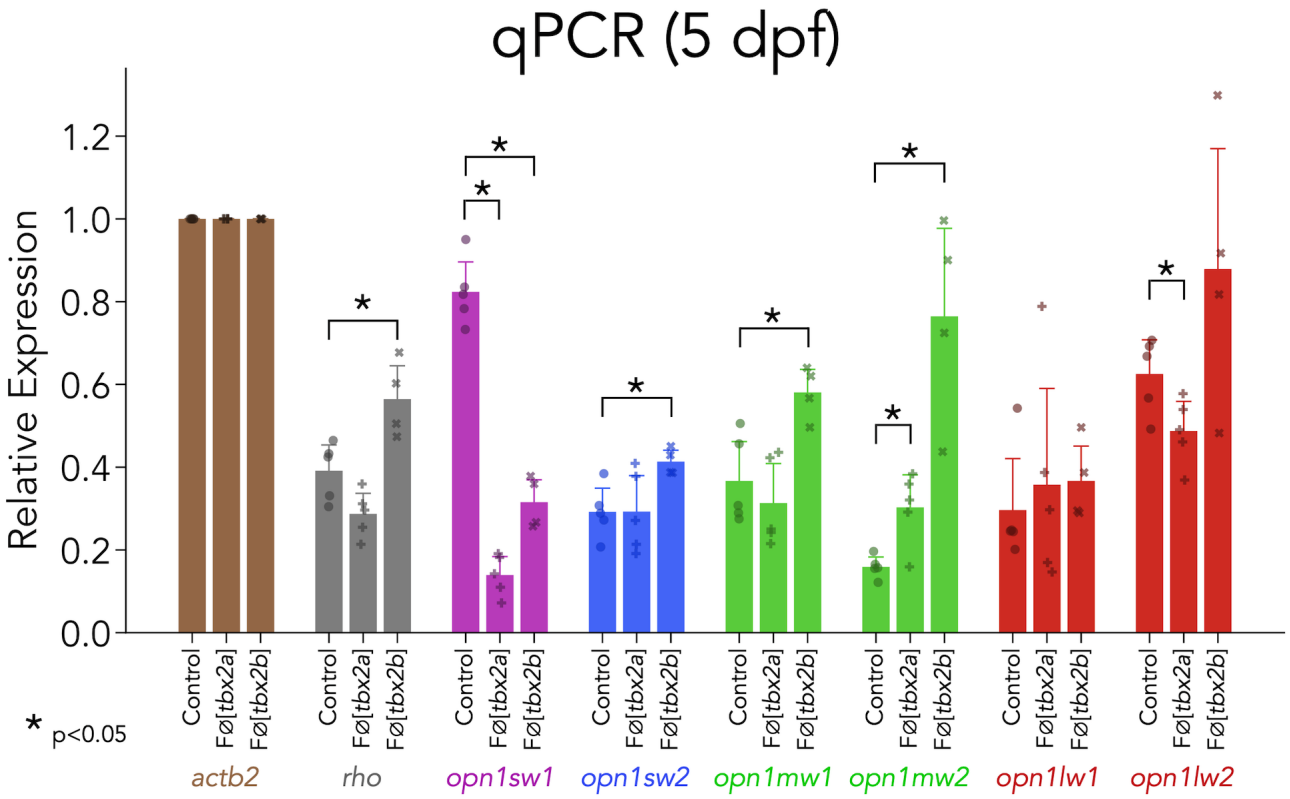

Figure 4 - figure supplement 1 . Mutations in *tbx2a* and *tbx2b* cause multiple changes in opsin expression. Quantification of opsin expression using real-time quantitative PCR, using the standard curve method and relative to expression of β-actin 2 (*actb2*). Compared to controls (n_{wt} = 5), FØ[*tbx2a*] ($n_{FØ[tbx2a]}$=5) showed significant changes in the expression of UV opsin (*opn1sw1*, Mann-Whitney U = 25, p = 0.008), M opsin (*opn1mw2*, Mann-Whitney U = 2.0, p = 0.03) and L opsin (*opn1lw2*, Mann-Whitney U = 22, p = 0.05). FØ[*tbx2b*] ($n_{FØ[tbx2a]}$=4) showed significant changes in the expression of Rhodopsin (*rho*, Mann-Whitney U = 0, p = 0.01), UV opsin (*opn1sw1*, Mann-Whitney U = 20, p = 0.01), S opsin (*opn1sw2*, Mann-Whitney U = 0, p = 0.01) and M opsin (*opn1mw1*, Mann-Whitney U = 1, p = 0.03; *opn1mw2*, Mann-Whitney U = 0, p = 0.01). We ran all measurements in triplicate. We were not able to detect expression of *opn1mw3* or *opn1mw4*.