Identification of genes with enriched expression in early developing mouse cone photoreceptors

Diego F. Buenaventura[1,2,+], Adrianne Corseri[1], Mark M. Emerson[1,2*]

[1] Department of Biology, The City College of New York, City University of New York, New York, NY, 10031

[2] Biology Ph.D. Program, Graduate Center, City University of New York, New York, NY, 10031

[+] Current Affiliation: Department of Biology, New York University, New York, NY 10003

*Corresponding author: memerson@ccny.cuny.edu

Abstract

Cone photoreceptors are the critical first cells that mediate high acuity vision. Despite their importance and their potential use as cell-based therapies for retinal diseases, there is a lack of knowledge about the early developmental stages of these cells. Here we characterize the expression of the homeobox transcription factor Lhx4 as an early and enriched cone photoreceptor expressed gene in both chicken and mouse. A Lhx4 GFP reporter mouse was found to recapitulate this early cone photoreceptor expression and was used to purify and profile embryonic mouse cone photoreceptors by single cell RNA sequencing. This enrichment in cone photoreceptors allowed for the robust identification of genes associated with the early cone transcriptome and also identified subpopulations of cones. To identify early cone genes not previously identified through analysis of the cone transcriptional repressor Nrl, the dataset generated here was filtered through previously reported datasets of Nrl dysregulated genes and early rod photoreceptors. The resulting set of genes was compared to other datasets if early human cones, late embryonic cones, and adult photoreceptors to generate a set of cone photoreceptor-enriched genes that are not regulated by Nrl.

INTRODUCTION

Cone and rod photoreceptors are the photosensitive cells of the retina that contribute to image formation. Cones mediate color discrimination while rods provide

photosensitivity in low- light conditions. Given the importance of cones in high acuity and color vision, deficiency in this cell type as a result of conditions such as retinitis pigmentosa or macular degeneration, lead to a debilitating loss of vision (Mitchell and Bradley, 2006). Because of their critical role, photoreceptors are currently the focus of a NIH-directed effort targeting the possibility of regeneration and replacement. As such, it is imperative that more is known about the required developmental steps in photoreceptor production and that new tools for this purpose are identified to produce essential developmental benchmarks for proper regeneration efforts.

While much is known about photoreceptors in the adult, we are still far from a complete picture during development, especially in the earlier stages of cell determination. Currently, only a few transcription factors have been shown to be part of gene regulatory networks that determines the fate of a cell into a photoreceptor. Most is known about rod photoreceptors, which are currently thought to be largely determined by NRL, or Neural Retina Leucine Zipper, the earliest and most upstream gene known in rod photoreceptor determination in mammals (Swaroop et al., 1992, 2010).

The regulation of cone production hasn't been as fully explored, but certain factors, such as retinoid x receptor gamma (RXRG) (Hoover et al., 1998) and, Thyroid Hormone Receptor Beta (THRB), are known to be involved. The THRB gene, for example, is crucial for the production of M cones and has also been proposed to be the earliest specific marker for cone photoreceptors in the developing retina (Applebury et al., 2007; Eldred et al., 2018; Ng et al., 2001). The identification of markers such as these is critical because a temporal progression of specific factors is usually required for proper cell differentiation

and, to assess true regeneration, it will be necessary to evaluate whether this temporal progression is recapitulated.

Thus, there is a requirement for new tools for the purpose of delineating cone photoreceptor development. One study has identified a reporter for postmitotic cone photoreceptors during development and adulthood (Smiley et al., 2016) based on a gene trap insertion of EGFP into the CCDC136 gene. However, this reporter seems to covary with S-opsin expression thus not labeling all cones, though quantification of cone expression is lacking during development as well as in the adult. In addition, the reporter is also active in bipolar cells (BCs) in the adult. Thus, there is a need for the development and validation of genetic tools to study cone photoreceptors, including the identification of new specific markers for cones.

With the advent of high-throughput sequencing, finding specific markers for a population is now more practical, as long as one is able to identify and sequence the correct cells. As markers may be expressed at different timepoints, it becomes a question of using the correct markers for the appropriate time. One of the earliest reported retinal populations that is biased to cone photoreceptor and horizontal cells (HCs) is identified by the activity of the ThrbCRM1 enhancer (Emerson et al., 2013). While this is a dividing progenitor population, it has limited mitotic potential and cell fate choices in its lineage, including cone photoreceptors. The transcriptome of these cells was highly enriched in known cone and HC markers, which suggested this dataset could serve as a source of novel cone-related genes in early chick development.

With this in mind, we set out to find highly enriched novel transcription factors in the ThrbCRM1 population and, consequently, in cones. We identified the Lim Homeobox Protein 4 (LHX4) gene as enriched in cones during early chick retinal development, in addition to BCs. In the mouse, LHX4 remains a reliable marker for cones during early retinal development but with expanded expression in late embryonic development and eventual expression in BCs. A LHX4::GFP BAC transgenic line (Heintz, 2004) recapitulates the same pattern of expression as the endogenous LHX4 protein and was used to generate a single cell dataset enriched in cone photoreceptors, which provides an in-depth look at the molecular profile of these cells in the earliest stages of differentiation in the mouse.

RESULTS

LHX4 is present in early cone photoreceptors in the chicken retina

A recent study examined the transcriptional profile of a class of restricted progenitors biased towards the cone and horizontal cell fate in the early chick retina (Buenaventura et al., 2018). Using the same dataset, we screened for potential cone-enriched transcripts that could play a role in development or serve as markers for early cones. After establishing a criterion for >1.5-fold change score between our populations, we selected for transcription factors enriched in the cone/HC progenitor population (Fig 1 A). Expectedly, many of the transcription factors that are important for this population, like OTX2, OC1 and THRB, are part of this group (Buenaventura et al., 2018). In addition, we identified the gene LIM homeobox 4 (LHX4) as highly enriched. This transcript has significant fold change (b = 3.3) and a very low number of reads in the other RPC population, which suggests that it has high specificity towards the cone/HC progenitor population (Fig 1 B).

Interestingly, a previous report has examined the presence of LIM-domain factors in early chick photoreceptor development in the chick (Fischer et al., 2008). Antibody staining of LHX3 suggested that this protein was abundantly present in in the apical portion of the retina and localized to photoreceptors once the ONL is clearly distinguished. Because our data indicates that LHX4 expression is prominent in cone progenitors at early stages while LHX3 transcript presence is marginal (Supp. Fig. 1 A), we suspected that these previous studies might be detecting LHX4 at the earlier timepoints. To test this hypothesis, we electroporated a CAG::mLHX4 plasmid alongside a CAG::nucB-gal construct into the chick E6 retina, cultured it for 2 days, and detected for LHX4 using the

LHX3 Developmental Studies Hybridoma Bank (DSHB) antibody. As predicted, we observed robust immunoreactivity of Bgal+ electroporated cells with the LHX3 antibody (Supp. Fig. 1 B), suggesting that this antibody is also capable of detecting mouse LHX4, and thus likely chicken Lhx4 as well.

With the use of this LHX3/4 antibody and a rabbit polyclonal antibody, we confirmed LHX4 presence during embryonic development of chick at E6 and E10. Expression at E6 is restricted to the scleral portion of the retina, where photoreceptors are located. As the LHX4 transcript was highly enriched in cone/HC progenitors, which were collected through use of a cis-regulatory region that requires OTX2 for activity, we expected LHX4 to be present predominantly in the OTX2+ population at E6. Indeed, LHX4 is detected in a large percentage (but not all) of OTX2-positive cells (Fig. 1 C).

At this developmental time point, the major class of photoreceptors that are being produced are cones, as the earliest known rd photoreceptor marker in the chicken retina, L-Maf, is not detected until E9 (Ochi et al., 2004). As the LHX4 pattern at this stage is localized in the apical portion of the retina, we examined LHX4+ cells for co-expression with the cone marker RXRG. Many of the most apically located LHX4 cells were indeed positive for RXRG (Fig. 1 C).

It is unclear to what extent and in which cells LHX3, LHX4 or both proteins are present in later timepoints, especially at the stage of developing BCs. In one report (Fischer et al., 2008), antibody staining using DSHB anti-LHX3 shows strong nuclear Outer Nuclear Layer (ONL) signal at E10.

At this timepoint the ONL is exclusively composed of photoreceptors, having stratified from the Inner Nuclear Layer (INL), where BCs, HCs, Amacrine Cells, and Muller

Glia are located. In contrast to their antibody staining, a RNA in situ hybridization at E12 shows that LHX3 transcript is detected strongly in the INL, but with only distal signal in the ONL along the edge of the section, which could potentially be background signal.

Our data is in agreement with the immunostaining, as we detect signal in the ONL with the LHX3 antibody. However, the rabbit anti-LHX4 antibody (Proteintech) detects a similar pattern with strikingly strong signal in the ONL and weaker in the INL (Suppl. Fig. 1 C). The absence of LHX3 transcript taken together with the cross-labeling of LHX4 by the DSHB anti-LHX3 antibody, suggests that LHX4 and not LHX3 is present in chicken photoreceptors, while both are likely present in developing BCs, as described in previous reports.

Thus, at E10, LHX4 (or LHX3) is strongly expressed in the cells located in the basal portion of the ONL with weaker expression in more apically located photoreceptor cells and in the INL (Fig 1 D,E). As mentioned, those cells positive in the INL are likely BCs, as one of the previous reports showed immunoreactivity to VSX2 at this stage in the majority of INL LHX3/4 positive cells, using the mouse DSHB antibody (Edqvist et al., 2009).

When compared to RAXL expression, a marker for cone photoreceptors (Ochi et al., 2004), all RAXL-positive cells were also positive for LHX4 (Fig 1 D). A smaller percentage of cells in the upper part of the ONL were not RAXL-positive. As the ONL contains both cones and rods, we used an antibody to MAFA (Benkhelifa et al., 2001; Ochi et al., 2004), the earliest dedicated rod marker, to determine if these cells were rod photoreceptors. No overlap between MAFA and LHX4 was detected at E10 in the ONL (Fig 1 E). This data confirms the presence of LHX4 in developing cone photoreceptors in the chicken retina.

LHX4 is present in early cone photoreceptors in the mouse retina

While the transcriptome and LHX4 immunostaining analyses were performed in chick, we sought to establish if this protein was also present in early photoreceptors of the mouse retina. At E14.5, we found that LHX4 is present in the scleral portion of the retina, where developing photoreceptors are located (Fig 2 A). To confirm that LHX4 was present in cone photoreceptors, we used RXRG to mark early cone precursors and OTX2 to label restricted progenitors. A large percentage of the population of LHX4 cells are also positive for RXRG and OTX2 at E14.5, in a similar pattern to the chick retina.

Because RXRG is currently the most dedicated marker for early cones, we determined the overlap with LHX4 protein expression. At E14.5, we find that 91.1+-1.9% of cells positive for LHX4 are also positive for RXRG (Fig 2 B), signifying that the majority of the LHX4+ population at this timepoint is composed of early cones. In fact, nearly all cells positive for RXRG at E14.5 are also positive for LHX4 (99+-1%, Fig 2 C).

When examined at a later embryonic stage (E17.5) we found that LHX4 cells still co-localize with RXRG protein but at a lower percentage, while LHX4 cells remain positive for OTX2 (Fig 2 A). At this timepoint, only 67.8+-1.1% of LHX4+ cells are positive for RXRG, suggesting that LHX4 is active in other emerging cell populations in addition to early cones (Fig 2 B).

These data indicate that the majority of cones retain LHX4 expression during development, as 95.1+-1% of RXRG+ cells are positive for LHX4 at e17.5. The rb-LHX4 polyclonal antibody had a reduced quality of staining at this timepoint, so it's possible the small fraction of cones that are not LHX4+ had undetectable LHX4 signal under these conditions or could point to a small subpopulation of cones that do not express LHX4 or cells that have downregulated this protein (Fig 2 C).

Additionally, we tested if LHX4 was present in post-mitotic cone photoreceptors or in dividing cells. E14.5 retinas exposed to a 2hr EdU pulse showed no overlap between EdU and LHX4 or RXRG (Supp. Fig. 2), which is consistent with LHX4 expression beginning after cell cycle exit.

We conclude that LHX4 is a relatively specific marker for post-mitotic cone photoreceptors at early stages in the mouse retina but is also present in other cell populations at later developmental stages.

LHX4 is expressed in several developing and adult cell types.

Previous reports indicate that LHX4 protein is present in adult cone BCs (Balasubramanian et al., 2014), although little is known about its expression pattern in the developing mouse retina. Our results are in agreement with the adult data, as we observed strong LHX4 labeling in the upper portion of the INL in cells positive for OTX2, an adult marker for BCs (Supp. Fig. 3). Interestingly, we also noticed sparse but positive staining in the ONL. This LHX4 signal is located in some but not all RXRG+ cones with varying degrees of strength (Supp. Fig. 3).

As noted above, previous reports (Edqvist et al., 2009; Fischer et al., 2008) suggest that in the chick retina, LHX4+ cells in the INL are likely developing BCs. In the mouse, our data indicates that LHX4 is initially in developing cones. However, at E17.5, RXRG cones no longer represent the near totality of LHX4+ cells. Since the peak of BCs production is not seen until ~P3 (Morrow et al., 2008), the identity of the remaining cells remains intriguing. While BCs are produced at earlier timepoints (including E17.5), we still sought to ascertain if only LHX4+ BCs were being produced at this time, or if this LHX4 expression is present in another cell type. As rods are the most common cell type in the mouse retina and are also produced at this timepoint, we aimed to see if these cells expressed LHX4 during development.

Newborn mice (P0) were injected with EdU to mark cells undergoing S-phase at the time of injection and 3 days later the retinas were harvested. This length of time was chosen to allow some newly produced rods enough time to produce NR2E3, a well-known marker for rod fate (Cheng et al., 2004). It has been previously suggested before that

NR2E3 is also present in cone photoreceptors transiently (Chen et al., 2005; Haider et al., 2006; Jean-Charles et al., 2018), but as other many reports indicate (Carter-Dawson and Lavail, 1979), no cones are produced at P0. Therefore, the presence of EdU and NR2E3 simultaneously should reliably mark other cell types than cones, likely rod photoreceptors. NR2E3 expression has not been reported in BCs but given the similarities in molecular profiles between BCs and photoreceptors, at this time we cannot exclude this possibility.

When detecting for LHX4 at the same time, we find LHX4+ cells that are positive for NR2E3 and EdU (Fig 2 D). This indicates that while LHX4 is present in developing and adult BCs and cones, it is also possible that is transiently expressed in some rod photoreceptors.

LHX4 BAC-GFP transgenic line has GFP activity in photoreceptors during development.

After confirming the presence of LHX4 in developing photoreceptors, specifically cones at the peak of this cell type's birth, in chick and mouse, we set out to establish if the a previously reported LHX4 BAC-EGFP line (Heintz, 2004) could be a reliable tracer for developing cones early in development. We found that the LHX4::GFP line had robust expression in the retina throughout development (Fig 3 A). The detected GFP signal was located in the apical section of the retina, where photoreceptors are located during development.

EGFP could be detected during embryonic and postnatal development and also had stable expression in the adult retina (Supp. Fig. 4). This line has been used previously for transcriptomic and electrophysiological analyses (Siegert et al., 2012; Szikra et al., 2011), because it shows reliable activity in cone BCs. Interestingly, a subpopulation of cones is labeled by this line in the adult and this selective expression resembles the sparse presence of LHX4 protein in the adult mouse retina. In order to verify if this sparse labeling was due to LHX4 presence in exclusively S or L/M types of cones, which are located in contrasting gradients in the adult retina, we assessed EGFP expression in the dorsal and ventral retina. We detected no difference in EGFP expression between these two areas, while the EGFP remains in a subset of all Cone Arrestin+ cells throughout the retina (Supp. Fig 5).

We were unable to use the rabbit anti-LHX4 antibody to validate the LHX4::GFP expression as we encountered abnormal detection of LHX4 immunoreactivity in any tissue coming from this line; not nuclear, as expected, but cytoplasmic, mimicking exactly the EGFP expression pattern. As noted in the strain description, EGFP was placed after the ATG of LHX4, to replace the LHX4 coding region and function as a tracer. To examine the genetic insertion on the LHX4::GFP BAC, we amplified and sequenced the coding region downstream of the LHX4 ATG using the reported UTR primer used for genotyping and, surprisingly, we found it contained a 70 bp section of 5' endogenous LHX4 coding sequence immediately followed by a multiple cloning site and the EGFP coding sequence, all in frame to produce a viable LHX4::GFP fusion protein, where the first 34 amino acids are from the sequence of LHX4 and the multiple cloning site (Supp. Fig. 6 A).

To test if this portion of LHX4 was sufficient to produce an epitope for the antibody, we amplified the coding sequence from the LHX4::GFP genomic DNA and cloned it into a misexpression vector and then electroporated into the chick retina (Supp. Fig. 6 B-D). Control retinas had no detectable EGFP and normal staining of LHX4 with the rabbit antibody (Supp. Fig. 6 B). When the CAG::5pLHX4GFP construct was electroporated, electroporated cells were strongly immunoreactive cytoplasmic LHX4, in a pattern that completely overlapped that of the cytomplasmic EGFP signal (Supp. Fig. 6 C). To show that the rabbit-LHX4 is still detecting endogenous LHX4, we also imaged the same retinas at the edge of the electroporation patch with higher gain, which shows endogenous protein expression alongside overexposed electroporation signal (Supp. Fig. 6 D).

Therefore, to assess if EGFP recapitulates LHX4 expression patterns and its activity in early cones, we resorted to checking for RXRG expression in EGFP+ cells at two relevant timepoints of embryonic development. At E14.5, the EGFP reporter faithfully recapitulates LHX4 expression as 95.3+-1% of all EGFP+ cells are positive for RXRG (Fig 3 C). Likewise, at E17.5 only 65.7+-1.3% of EGFP cells are RXRG+.

As our data shows, nearly all RXRG+ cones are positive for LHX4 at both E14.5 and E17.5. This is also true in the EGFP+ population, where 94.8+-0.5% and 99.5+-0.5% of all RXRG cells, respectively, were positive for the EGFP reporter (Fig 3 D). Taken together, this data suggests that the LHX4 BAC-EGFP reliable recapitulates LHX4 protein expression and is a dependable marker for cone photoreceptors during early retinal development.

Single cell sequencing of LHX4::GFP cells in the E14.5 developing mouse retina.

Having verified that the LHX4::GFP reporter is a marker for cone photoreceptors in the early stages of mouse retinal development, we took advantage of this system to examine more closely the molecular profile of early cone photoreceptors. LHX4::GFP E14.5 littermates were screened for GFP expression and positive retinas were pooled and dissociated in preparation for FACS sorting (Fig 4 A). GFP+ and GFP- cells were collected and approx. 4000 cells were sequenced per condition in the 10x Chromium platform.

Using the Seurat program (Satija et al., 2015), we performed an unsupervised clustering analysis on the combined cell transcriptomes from both conditions that passed standard 10x QC (GFP+: 3728, GFP-:4444). TSNE projections revealed that the GFP+ and GFP- populations clearly segregated, consistent with LHX4::GFP cells being a molecularly distinct population (Fig 4 B). The unsupervised analysis separated the cells into 9 distinct clusters which were identified by established markers and show hallmarks of known cell classes in the developing retina (Fig 4 C-D; Supplementary File 1).

Two cell clusters were assigned to multipotent RPCs as they showed a molecular profile of dividing retinal progenitor cells, with expression of VSX2, HES1, NR2E1, among others. When looking at known cell cycle genes, it's possible to assign a signature of a particular phase to each of the cell transcriptomes (Buettner et al., 2015) (Fig 4 E-F). This analysis confirmed our assignment as the multipotent RPC clusters were populated by dividing cells, although each cluster has different proportion of cells at distinct phases. The split in the clusters is also marked by the presence of some differential cell cycle markers like MKI67 and GMNN.

One other cluster was comprised entirely of cycling cells, exhibiting markers such as OLIG2, NEUROG2, OTX2 and ATOH7, which identify neurogenic RPCs (Cepko, 2014; Clark et al., 2018) as these markers in cycling cells indicate restricted RPCs close to their final division. A second cluster had a fraction of cycling cells, high ATOH7 levels and expression of POU4F2, but no OTX2 or OLIG2, which we assigned to RGC/AC RPCs and Precursors as they exhibit some markers of likely differentiation to RGCs or ACs.

Of the cell clusters assigned as mostly post-mitotic, one was assigned as RGCs as it displayed known markers RBPMS and GAP43. Another cluster assigned was AC/HCs exhibited markers for both these fates, like TFAP2B and PTF1A. As previously reported (Clark et al., 2018), while ACs and HCs are distinct fates, it is difficult to separate them by their transcriptomes in early development and without an appropriate amount of sequenced cells for proper resolution.

Three clusters were assigned as cone photoreceptors. As expected from our previous data, these consisted of the near totality of GFP cells and exhibit markers of cone photoreceptors: THRB, RXRG, GNGT2 and GNB3 (Fig 4 C-D), as well as LHX4. The cell cycle phase of these clusters is consistent with our previous data suggesting LHX4 is in post-mitotic cones. As a result of our sorting strategy, we enriched for cone photoreceptor transcriptomes, which allowed the clustering analysis to separate these cells into subpopulations. While all cone clusters had high levels of established markers, two subpopulations differed in expression of novel genes. The highest differential marker for one of the populations was FABP7, a previously characterized marker for cones in the adult murine retina (Su et al., 2016) with reported expression in the developing retina (Blackshaw et al., 2004) but not at this early stage. Meanwhile, a second subpopulation

had increased levels of solute-carrier genes like SLC7A3 and SLC7A5 (Fig 4 C). Our analysis indicates that we successfully sorted and sequenced a developing E14.5 retina and identified its cell populations while enriching for cone photoreceptors.

Early cone marker identification in LHX4::GFP cells

Having obtained the LHX4::GFP+ and GFP- transcriptomes, we sought to use this dataset to find new markers for early cone photoreceptors. Using the Seurat package, we performed a differential expression analysis, comparing the GFP+ population, which we established as cones, with the GFP- population, which should encompass the rest of the developing retina. Additionally, for visualization purposes and as a proxy for average population levels of transcript expression we calculated the expression of an average GFP+ and a GFP- cell and used this in combination with the differential expression results. We found over 898 significantly differentially expressed transcripts. As expected, we identified known cone markers as highly enriched in the GFP+ population (Fig 5 A, Supplementary File 2). We then looked for novel transcripts enriched in this population (Fig 5 B). A heatmap for the top 100 cone-enriched transcripts displaying the transcript expression of an average cell in every individual clusters are included in Fig. 5 C..

This above analysis identifies genes that are enriched in early cone photoreceptors compared to the other cell types present at that time. We first compared these results to other studies that have reported cone transcriptome analyses in the mouse or human (Supp Fig 7 A-B). Only one study has reported a transcriptome from early developing cones, similar to this one (Welby et al.). Through the use of human fetal explants infected with cone opsin promoter reporter viruses, transcriptome analysis of early and late phases of gene expression in labeled cells versus non-labeled cells was determined. A comparison to these datasets shows that a number of genes are enriched in both datasets (165 genes in early and 211 in late fetal cones), suggesting that the mouse is a potential model for investigating the function of these genes in cone photoreceptors. (Supp. Fig 7, Supp. File 3 and 4).

Additional datasets have purified adult rods and cones from the mouse to identify differential transcripts between these two types of photoreceptors (Mo et al., 2016). Comparison of our developing cone dataset to the Mo et al. dataset identified that 251 of the genes in our dataset were enriched in adult cones and 121 were enriched in adult rods (Supp. Fig 7, Supp. File 3 and 5). Adult cone-enriched genes present in this dataset support the cell-specificity of known early markers detected in our dataset, such as GNGT2 or RXRG, as well as many genes not explored in detail before, such as QSOX1 or LHX4.

We next were interested in determining whether the genes associated with early cone genesis were negatively regulated by the rod photoreceptor factor Nrl. A prevalent model for Nrl function is that it serves as a fate switch in photoreceptor precursors, with Nrl-negative precursors becoming cones and Nrl-positive ones becoming rods. In Nrl

mutants, the photoreceptors that normally would become rods are found to undergo a morphological and gene expression change that suggests a rod to cone fate switch. However, it has been noted that these cells are morphologically distinguishable from normal cones and, in addition, known early cone genes, such as Thrb and Rxrg, are unaltered in these cells. As there has been a lack of known early cone genes, we sought to determine whether there were other cone photoreceptor genes in addition to Thrb and Rxrg with Nrl-independent regulation. We first identified those genes in our dataset that were dysregulated either positively or negatively in Nrl knockout photoreceptors. The Nrl dataset used contained multiple isoforms of genes, but to apply a stringent criteria, any isoform that showed dysregulation at either P2 or P28 in Nrl mutants led to that gene being removed, leaving a total of 259 Nrl-independent genes (Fig 6 A). Interestingly, many of the cone-enriched genes are only enriched at P28 and not also at P2 when a large number of rods would be forming (see discussion).

As some of these genes likely represent pan-photoreceptor genes that wouldn't be expected to be under the transcriptional control of Nrl, we identified those genes also expressed in developing rods. To do so, we analyzed a recently generated single cell dataset from E18.5 whole retinas (Clark et al., 2018). The data from 2 replicates of E18 was extracted and we performed an unbiased clustering analysis using Seurat (Satija et al., 2015). (Supp Fig 8). The detected 9 clusters were assigned to known cell types through marker expression. Within those, two clusters were positive for CRX+ and, thus, determined to be photoreceptors. First, all DE transcripts for photoreceptors at e18 were determined (Supp. Table 7, Supp. Fig 8) and then these cells were re-analyzed for further sub-clustering. We were able to determine a cluster with cone signature, rod

signature and 2 precursor clusters. From those, we determined all DE transcripts between photoreceptor clusters (Supp. Table 8, Supp Fig 8). Interestingly, many of the genes detected as enriched in early cone photoreceptors in this study have been identified as differentially expressed across retinal pseudotime by Clark, Stein O'Brien et al, 2018 (Supp. Table 9)

Removal of those genes with expression in any of the non-cone clusters at e18, produced 198 Nrl-independent cone-enriched genes (Fig 6 A-B, Supp. Table 10). To provide some measure of confirmation of the cone/photoreceptor-associated expression of these genes, we identified those genes with cone-enriched expression or specifically enriched in cone cluster 2 that were also identified in the previously mentioned datasets (Fig 6 C,D, Supp. Table 10). This left a set of genes with cone-enriched expression or specifically enriched in cone cluster 2 that have only been identified to date in this study as cone associated (Fig 6 E,F).

DISCUSSION

Our data establishes LHX4 as a novel and specific marker of cone photoreceptors in the earliest stages of development and the LHX4 BAC-GFP mouse as a tool to detect these cells in the developing retina. The LHX4 reporter can be detected as early as E12.5 and can reliably mark post-mitotic cones, usually located in the apical side of the retina. Cone activity remains throughout development, although at later stages of embryonic development it begins to be active transiently in rods and eventually remains active in cone bipolar cells. During early stages, the specificity for cones was confirmed by quantification of the LHX4 protein and reporter overlap with RXRG and by single-cell sequencing of these cells.

Cell type specific reporter lines are crucial for surveying and targeting specific cells. The LHX4::GFP line has been used before for electrophysiology and transcriptomic analyses in the adult retina (Siegert et al., 2012; Szikra et al., 2011). We show in this report that the LHX4::GFP line can also be used developmentally, particularly suited for early cone labeling and possibly a tool for exploring early post-mitotic photoreceptor or bipolar cell populations in postnatal development. The cytoplasmic localization of the GFP also allows for visualization at embryonic stages of processes in new cones projecting towards the basal retina (Fig 3-6 A), which have been reported before (Johnson et al., 1999, 2001) but with no known physiological function as of yet.

One previous report has established a reporter line that works specifically in cones (Smiley et al., 2016). This line has very specific activity in developing cones through a gene trap insertion of GFP to the locus of the CCDC136 gene. Interestingly, in addition to cones, this line also shows activity in rod bipolar cells in the adult, as opposed to the LHX4::GFP line which is active in cone bipolar cells. A drawback of this line is that the

activity in cones in the adult is similar to that of S-opsin and follows the same dorsal-ventral gradient. One advantage, however, is the activity might be specific to cones and not rods, although this isn't completely verified as there was no quantification to this effect.

In the case of the LHX4::GFP line, there is activity in cones, transiently in rods and eventually in cone bipolar cells. One clear advantage of the LHX4::GFP line is that all RXRG cells turn on this reporter, as our quantification shows in the embryonic retina. At this time, it is not clear if all cones turn on the CCDC136-GFP reporter during development. Each line may be uniquely suited for different applications, according to experimental needs.

There are a few drawbacks to the LHX4 reporter line, such as the relatively weak strength of the reporter. This might be of importance depending on the intended application, as it may affect approaches such as in vivo imaging, although it is more than sufficient for FACs sorting. These mice were maintained and analyzed as heterozygotes, as the insertion point for the BAC construct is unknown, in order to prevent any unintended genetic disruption. If processed as homozygous, the relative strength of the reporter may be increased, but care should be taken to check for abnormalities.

A second drawback is the interesting phenotype of the sparse labeling of cones in the adult LHX4::GFP mouse. We verified that this phenotype does not follow a dorsal-ventral gradient, as it does in the CCDC136 mouse. This makes it less than adequate for adult identification of cones, but, at the same time, may be indicative of a subpopulation of cones yet to be described.

Our data therefore indicated that the LHX4::GFP line was particularly suited for the collection and sequencing of early cone precursors. It allowed for collection of all post-

mitotic cones at E14.5 in quantities that would not be practically possible from wildtype retinas. The number of usable cone transcriptomes recovered from this experiment (3728 cells after QC) provides enough power to recognize subgroups in a large population of developing cones which might not have been resolved otherwise. All cones are still marked by known markers such as RXRG and THRB, but certain cones upregulate novel genes that have not been reported before at this time point.

One distinct cluster upregulates the SLC7A3 and SLC7A5 genes. SLC7A5 is a known thyroid hormone transporter for its T3/T4 states (Scalise et al., 2018) that has been reported in human fetal retinas and organoids (Eldred et al., 2018; Hoshino et al., 2017). Because these transcriptomic analyses were done in whole retinas it wasn't clear if this expression was specific to any particular cell population, but our data suggests that mouse cone photoreceptors specifically express these genes. SLC7A3 has not been reported in the retina but likely plays a similar role as its expression has been linked to T3 administration in the brain (Grijota-Martínez et al., 2011). Thyroid hormone and its targets are known to affect M-cone differentiation (Applebury et al., 2007; Billings et al., 2010; Eldred et al., 2018; Ng et al., 2001, 2009). We found several other thyroid-related genes throughout the retina, but only THRB was exclusive to cones at this time. For example, there was almost no detectable amount of DIO2 in any cell but many more counts of DIO3, although these were present in the multipotent RPC clusters.

Some of the other markers in this cluster such as ATF3 and CHAC1 have been reportedly associated with retinal damage and apoptosis (Crawford et al., 2015). Since we collected a large amount of a presumably very homogenous population, having passed through dissociation and FACS sorting, it would not be abnormal to have a fraction

of cells that have upregulated apoptotic or stress markers that can be identified as a different population by clustering algorithms. Since ATF3 is present in the retina during development (Aldiri et al., 2017), confirming the presence of these transcripts in the photoreceptor layer through *in situ* hybridizations at E14.5, without dissociation, would be advantageous to verify if they are normally expressed in cones or only as a stress reaction.

Exclusive markers between the other two cone clusters were few, the likeliest explanation being that these clusters differentiate in the quantity of transcripts associated with them rather than the presence of exclusive markers. One of the exclusive markers for one cluster was FABP7, which has been reported in the adult murine retina. Interestingly, in situ hybridizations for this gene at E14 did not produce any detectable signal (Blackshaw et al., 2004), but this could be due to different sensitivities between techniques. What differential presence of this transcript entails for cone photoreceptors is still unknown and might require more detailed loss and gain-of-function studies.

The last cone cluster contained the vast majority of cones and, accordingly, different fractions of these cells shared most markers with the other cone clusters. One marker that separated this cluster from the FABP7 cluster is the NFIB gene. This gene and its family members have been recently implicated in regulation of BC and MG specification (Clark et al., 2018) but is also detected in photoreceptor precursors, which presumably includes cones, in their dataset and in the present work.

Materials & Methods

Animals

All experimental procedures were carried out in accordance with the City College of New York, CUNY animal care protocols under IACUC protocol 965. CD-1 mice were used and provided by Charles River. LHX4::GFP strain (strain: Tg(Lhx4-EGFP)KN199Gsat/Mmucd, RRID:MMRRC_030699-UCD) were obtained from MMRRC. LHX4::GFP mice were kept and used experimentally only as a heterozygous. Fertilized chick eggs were from Charles River, stored in a 16°C room for 0-10 days and incubated in a 38°C humidified incubator. All experiments that used animals were not sex-biased.

Genotyping

Genotyping was performed as specified per strain in the MMRRC strain description page referenced.

Cloning and DNA electroporation

Misexpression plasmids were created by PCR amplifying the coding sequence for each gene with added AgeI/NheI or EcorI sites. They were subsequently cloned into a CAG backbone, from which the GFP had been removed from Stagia3 (Billings et al., 2010; Emerson and Cepko, 2011) and the CAG promoter had been cloned upstream. For mouse LHX4 the primers were designed against the annotated mRNAs in NCBI galgal5 RNA libraries and with an added Kozak sequence 5' (ACC). For 5pLHX4GFP, the 5' primer from genotyping for LHX4::GFP mice was combined with a 3' primer to coding region of GFP and the fusion gene was amplified from genomic DNA extracted from LHX4::GFP+ mice.

To deliver the plasmids after retinal dissection, ex vivo electroporation experiments were carried out as detailed in (Emerson and Cepko, 2011). All experiments used a Nepagene electroporator. Ex vivo retinas were cultured 24 hours on all experiments. The DNA mixtures used for the ex vivo electroporations were diluted in sterile 1X PBS to a final concentration of 0.16 µg/µl for reporter plasmids and 0.1 µg/µl for the co-electroporation control plasmids and misexpression plasmids.

Retinal cells dissociations and Florescence Activated Flow Sorting (FACS)

Retinal pigment epithelium and vitreal material was removed from the retinae in HBSS (GIBCO, 14170112) and dissociated using a papain-based procedure (Worthington, L5003126) (Matsuda and Cepko, 2004). Samples were dissociated and resuspended in DMEM/10%FBS for sorting and cells were also collected in DMEM/10%FBS. Dissociated cells were collected separately into FACS tubes (BD Falcon, 352054) upon being filtered through 40µm cell strainers to ensure suspension of individual cells (Biologix, 15-1040). Fluorescent activated cell sorting (FACS) for GFP-based sorting was carried out with a BD FACS Aria II machine, using the 488nm laser for LHX4-eGFP detection.

Immunohistochemistry and EdU labeling

Retinae processed for immunohistochemistry were fixed in 4% paraformaldehyde for 30 minutes at room temperature, sunk in 30% sucrose/0.5X PBS and then snap-frozen in OCT (Sakura Tissue-Tek, 4583). 20 µm vertical sections were obtained using a Leica Cryostat.

All immunofluorescence experiments were performed as in Emerson and Cepko, 2011 and Buenaventura et al, 2018. After a minimum of 1 h at room temperature in

blocking solution, the slides were incubated at 4°C overnight with a mix of the primary antibodies, 5% serum and 0.1% Tween. Secondary antibodies appropriate for multiple labeling were from Jackson ImmunoResearch. Alexa 488 and 647 conjugated secondary antibodies were used at 1:400 and Cy3 at 1:250. A solution of 1:10,000 DAPI was applied on sections prior to 3 final washes of 15 minutes at room temperature in 1X PBT and slides were mounted in Fluoromount (Southern Biotech, 0100-01) with 34x60mm cover slips (VWR, 48393 106).

For E14.5 mouse retinas, EdU labeling was performed by injecting pregnant dams with 150ul of 10mM EdU resuspended in 1X PBS. EdU detection was performed with a Click-iT EdU Alexa Fluor 647 imaging or flow cytometry kit (C10340, Invitrogen).

Microscopy

Confocal images were acquired using a Zeiss LSM880 confocal microscope using ZEN Black 2015 2.1 SP2 software and images were converted into picture format using the FIJI version of ImageJ (Schindelin et al., 2012). Figures were assembled using Affinity Designer vector editor. Images were adjusted uniformly with regards to brightness and contrast.

10X single cell experiment

Live cells were collected as mentioned previously. Samples were extracted from 2 litters of E14.5 LHX4::GFP mice, pooled into one sorting solution during dissociation. For each sample (GFP+ and GFP-), approx. 4000 cells were individually lysed, and their RNA transcribed and sequenced. Filtered count matrixes provided by 10X Cell Ranger pipeline using mm10 genome were used for downstream analysis.

Single-cell clustering analysis

Unsupervised clustering analysis was performed using Seurat (Butler et al., 2018; Satija et al., 2015). Single cell transcriptomes from both LHX4::GFP E14.5 samples were analyzed in concert to produce cluster and TSNE analysis. Only cells with over 200 genes detected and only genes detected in >3 cells were used for analysis. Cell cycle was scored and was regressed out as an unwanted source of variation according to Seurat's guidelines (Buettner et al., 2015).

Data comparisons

For the comparison of retinal datasets, differential expression output data was obtained from each report (Kim et al., 2016; Mo et al., 2016; Welby et al.). The data was cross-referenced to the genes obtained in the analysis from this paper and are detailed in Supp. Files 3, 4, 5 and 6. In the case of Kim et al, 2016, genes are reported with PPDE as opposed to p-val-FDR or q-val as probability of differential expression, and were filtered to show only those with PPDE>0.95

Data availability

The data will be deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and will be accessible through GEO Series accession number GSExxxxxx.

Quantitative analysis of markers in retina sections

One z-stack from the central portion of each retina of four biological replicates was counted and used for mean and SEM calculations in each condition. Cells were counted using Cell Counter plugin in ImageJ or Fiji (Schindelin et al., 2012; Schneider et al., 2012).

Acknowledgments

Support was provided by NIH National Eye Institute grant R01EY024982 (to M.E.) and 3G12MD007603-30S2 (CCNY). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Eye Institute, the National Institute On Minority Health and Health Disparities or the National Institutes of Health. Excellent technical support was provided by Brandon Webley, Jeffrey Walker, and Jorge Morales. The MafA antibody was kindly provided by Celio Pouponnot. We thank Brian Clark and Seth Blackshaw for granting access to their recently generated single cell datasets.

References

Aldiri, I., Xu, B., Wang, L., Chen, X., Hiler, D., Griffiths, L., Valentine, M., Shirinifard, A., Thiagarajan, S., Sablauer, A., et al. (2017). The Dynamic Epigenetic Landscape of the Retina During Development, Reprogramming, and Tumorigenesis. Neuron *94*, 550-568.e10.

Applebury, M.L., Farhangfar, F., Glösmann, M., Hashimoto, K., Kage, K., Robbins, J.T., Shibusawa, N., Wondisford, F.E., and Zhang, H. (2007). Transient expression of thyroid hormone nuclear receptor TRbeta2 sets S opsin patterning during cone photoreceptor genesis. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *236*, 1203–1212.

Balasubramanian, R., Bui, A., Ding, Q., and Gan, L. (2014). Expression of LIM-homeodomain transcription factors in the developing and mature mouse retina. Gene Expr. Patterns GEP *14*, 1–8.

Benkhelifa, S., Provot, S., Nabais, E., Eychène, A., Calothy, G., and Felder-Schmittbuhl, M.P. (2001). Phosphorylation of MafA is essential for its transcriptional and biological properties. Mol. Cell. Biol. *21*, 4441–4452.

Billings, N.A., Emerson, M.M., and Cepko, C.L. (2010). Analysis of thyroid response element activity during retinal development. PloS One *5*, e13739.

Blackshaw, S., Harpavat, S., Trimarchi, J., Cai, L., Huang, H., Kuo, W.P., Weber, G., Lee, K., Fraioli, R.E., Cho, S.-H., et al. (2004). Genomic analysis of mouse retinal development. PLoS Biol. *2*, E247.

Buenaventura, D.F., Ghinia-Tegla, M.G., and Emerson, M.M. (2018). Fate-restricted retinal progenitor cells adopt a molecular profile and spatial position distinct from multipotent progenitor cells. Dev. Biol. *443*, 35–49.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat. Biotechnol. *33*, 155–160.

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

Carter-Dawson, L.D., and Lavail, M.M. (1979). Rods and cones in the mouse retina. II. Autoradiographic analysis of cell generation using tritiated thymidine. J. Comp. Neurol. *188*, 263–272.

Cepko, C. (2014). Intrinsically different retinal progenitor cells produce specific types of progeny. Nat. Rev. Neurosci. *15*, 615–627.

Chen, J., Rattner, A., and Nathans, J. (2005). The Rod Photoreceptor-Specific Nuclear Receptor Nr2e3 Represses Transcription of Multiple Cone-Specific Genes. J. Neurosci. *25*, 118–129.

Cheng, H., Khanna, H., Oh, E.C.T., Hicks, D., Mitton, K.P., and Swaroop, A. (2004). Photoreceptor-specific nuclear receptor NR2E3 functions as a transcriptional activator in rod photoreceptors. Hum. Mol. Genet. *13*, 1563–1575.

Clark, B., Stein-O'Brien, G., Shiau, F., Cannon, G., Davis, E., Sherman, T., Rajaii, F., James-Esposito, R., Gronostajski, R., Fertig, E., et al. (2018). Comprehensive analysis of retinal development at single cell resolution identifies NFI factors as essential for mitotic exit and specification of late-born cells. BioRxiv.

Crawford, R.R., Prescott, E.T., Sylvester, C.F., Higdon, A.N., Shan, J., Kilberg, M.S., and Mungrue, I.N. (2015). Human CHAC1 Protein Degrades Glutathione, and mRNA Induction Is Regulated by the Transcription Factors ATF4 and ATF3 and a Bipartite ATF/CRE Regulatory Element. J. Biol. Chem. *290*, 15878–15891.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210.

Edqvist, P.H.D., Myers, S.M., and Hallböök, F. (2009). Early identification of retinal subtypes in the developing, pre-laminated chick retina using the transcription factors Prox1, Lim1, Ap2a, Pax6, Isl1, Isl2, Lim3 and Chx10. Eur. J. Histochem. *50*, 147–154.

Eldred, K.C., Hadyniak, S.E., Hussey, K.A., Brenerman, B., Zhang, P.-W., Chamling, X., Sluch, V.M., Welsbie, D.S., Hattar, S., Taylor, J., et al. (2018). Thyroid hormone signaling specifies cone subtypes in human retinal organoids. Science *362*.

Emerson, M.M., and Cepko, C.L. (2011). Identification of a retina-specific Otx2 enhancer element active in immature developing photoreceptors. Dev. Biol. *360*, 241–255.

Emerson, M.M., Surzenko, N., Goetz, J.J., Trimarchi, J., and Cepko, C.L. (2013). Otx2 and Onecut1 Promote the Fates of Cone Photoreceptors and Horizontal Cells and Repress Rod Photoreceptors. Dev. Cell *26*, 59–72.

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

Grijota-Martínez, C., Díez, D., Morreale de Escobar, G., Bernal, J., and Morte, B. (2011). Lack of action of exogenously administered T3 on the fetal rat brain despite expression of the monocarboxylate transporter 8. Endocrinology *152*, 1713–1721.

Haider, N.B., Demarco, P., Nystuen, A.M., Huang, X., Smith, R.S., Mccall, M.A., Naggert, J.K., and Nishina, P.M. (2006). The transcription factor Nr2e3 functions in retinal progenitors to suppress cone cell generation. Vis. Neurosci. *23*, 917–929.

Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). Nat. Neurosci. *7*, 483.

Hoover, F., Seleiro, E.A., Kielland, A., Brickell, P.M., and Glover, J.C. (1998). Retinoid X receptor gamma gene transcripts are expressed by a subset of early generated retinal cells and eventually restricted to photoreceptors. J. Comp. Neurol. *391*, 204–213.

Hoshino, A., Ratnapriya, R., Brooks, M.J., Chaitankar, V., Wilken, M.S., Zhang, C., Starostik, M.R., Gieser, L., La Torre, A., Nishio, M., et al. (2017). Molecular Anatomy of the Developing Human Retina. Dev. Cell *43*, 763-779.e4.

Jean-Charles, N., Buenaventura, D.F., and Emerson, M.M. (2018). Identification and characterization of early photoreceptor cis-regulatory elements and their relation to Onecut1. Neural Develop. *13*.

Johnson, P.T., Williams, R.R., Cusato, K., and Reese, B.E. (1999). Rods and cones project to the inner plexiform layer during development. J. Comp. Neurol. *414*, 1–12.

Johnson, P.T., Raven, M.A., and Reese, B.E. (2001). Disruption of transient photoreceptor targeting within the inner plexiform layer following early ablation of cholinergic amacrine cells in the ferret. Vis. Neurosci. *18*, 741–751.

Kim, J.-W., Yang, H.-J., Brooks, M.J., Zelinger, L., Karakülah, G., Gotoh, N., Boleda, A., Gieser, L., Giuste, F., Whitaker, D.T., et al. (2016). NRL-Regulated Transcriptome Dynamics of Developing Rod Photoreceptors. Cell Rep. *17*, 2460–2473.

Matsuda, T., and Cepko, C.L. (2004). Electroporation and RNA interference in the rodent retina in vivo and in vitro. Proc. Natl. Acad. Sci. U. S. A. *101*, 16–22.

Mitchell, J., and Bradley, C. (2006). Quality of life in age-related macular degeneration: a review of the literature. Health Qual. Life Outcomes *4*, 97.

Mo, A., Luo, C., Davis, F.P., Mukamel, E.A., Henry, G.L., Nery, J.R., Urich, M.A., Picard, S., Lister, R., Eddy, S.R., et al. (2016). Epigenomic landscapes of retinal rods and cones. ELife *5*, e11613.

Morrow, E.M., Chen, C.-M.A., and Cepko, C.L. (2008). Temporal order of bipolar cell genesis in the neural retina. Neural Develop. *3*, 2.

Ng, L., Hurley, J.B., Dierks, B., Srinivas, M., Saltó, C., Vennström, B., Reh, T.A., and Forrest, D. (2001). A thyroid hormone receptor that is required for the development of green cone photoreceptors. Nat. Genet. *27*, 94–98.

Ng, L., Ma, M., Curran, T., and Forrest, D. (2009). Developmental expression of thyroid hormone receptor beta2 protein in cone photoreceptors in the mouse. Neuroreport *20*, 627–631.

Ochi, H., Sakagami, K., Ishii, A., Morita, N., Nishiuchi, M., Ogino, H., and Yasuda, K. (2004). Temporal expression of L-Maf and RaxL in developing chicken retina are arranged into mosaic pattern. Gene Expr. Patterns GEP *4*, 489–494.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Scalise, M., Galluccio, M., Console, L., Pochini, L., and Indiveri, C. (2018). The Human SLC7A5 (LAT1): The Intriguing Histidine/Large Neutral Amino Acid Transporter and Its Relevance to Human Health. Front. Chem. *6*.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Siegert, S., Cabuy, E., Scherf, B.G., Kohler, H., Panda, S., Le, Y.-Z., Fehling, H.J., Gaidatzis, D., Stadler, M.B., and Roska, B. (2012). Transcriptional code and disease map for adult retinal cell types. Nat. Neurosci. *15*, 487–495.

Smiley, S., Nickerson, P.E., Comanita, L., Daftarian, N., El-Sehemy, A., Tsai, E.L.S., Matan-Lithwick, S., Yan, K., Thurig, S., Touahri, Y., et al. (2016). Establishment of a cone photoreceptor transplantation platform based on a novel cone-GFP reporter mouse line. Sci. Rep. *6*, 22867.

Su, X., Tan, Q.S.W., Parikh, B.H., Tan, A., Mehta, M.N., Wey, Y.S., Tun, S.B.B., Li, L.-J., Han, X.-Y., Wong, T.Y., et al. (2016). Characterization of Fatty Acid Binding Protein 7 (FABP7) in the Murine Retina. Invest. Ophthalmol. Vis. Sci. *57*, 3397–3408.

Swaroop, A., Xu, J.Z., Pawar, H., Jackson, A., Skolnick, C., and Agarwal, N. (1992). A conserved retina-specific gene encodes a basic motif/leucine zipper domain. Proc. Natl. Acad. Sci. U. S. A. *89*, 266–270.

Swaroop, A., Kim, D., and Forrest, D. (2010). Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. Nat. Rev. Neurosci. *11*, 563–576.

Szikra, T., Farrow, K., and Roska, B. (2011). Cone-mediated Circuit Switch Activates Lateral Inhibition In A Retinal Ganglion Cell. Invest. Ophthalmol. Vis. Sci. *52*, 4569–4569.

Welby, E., Lakowski, J., Di Foggia, V., Budinger, D., Gonzalez-Cordero, A., Lun, A.T.L., Epstein, M., Patel, A., Cuevas, E., Kruczek, K., et al. Isolation and Comparative Transcriptome Analysis of Human Fetal and iPSC-Derived Cone Photoreceptor Cells. Stem Cell Rep.

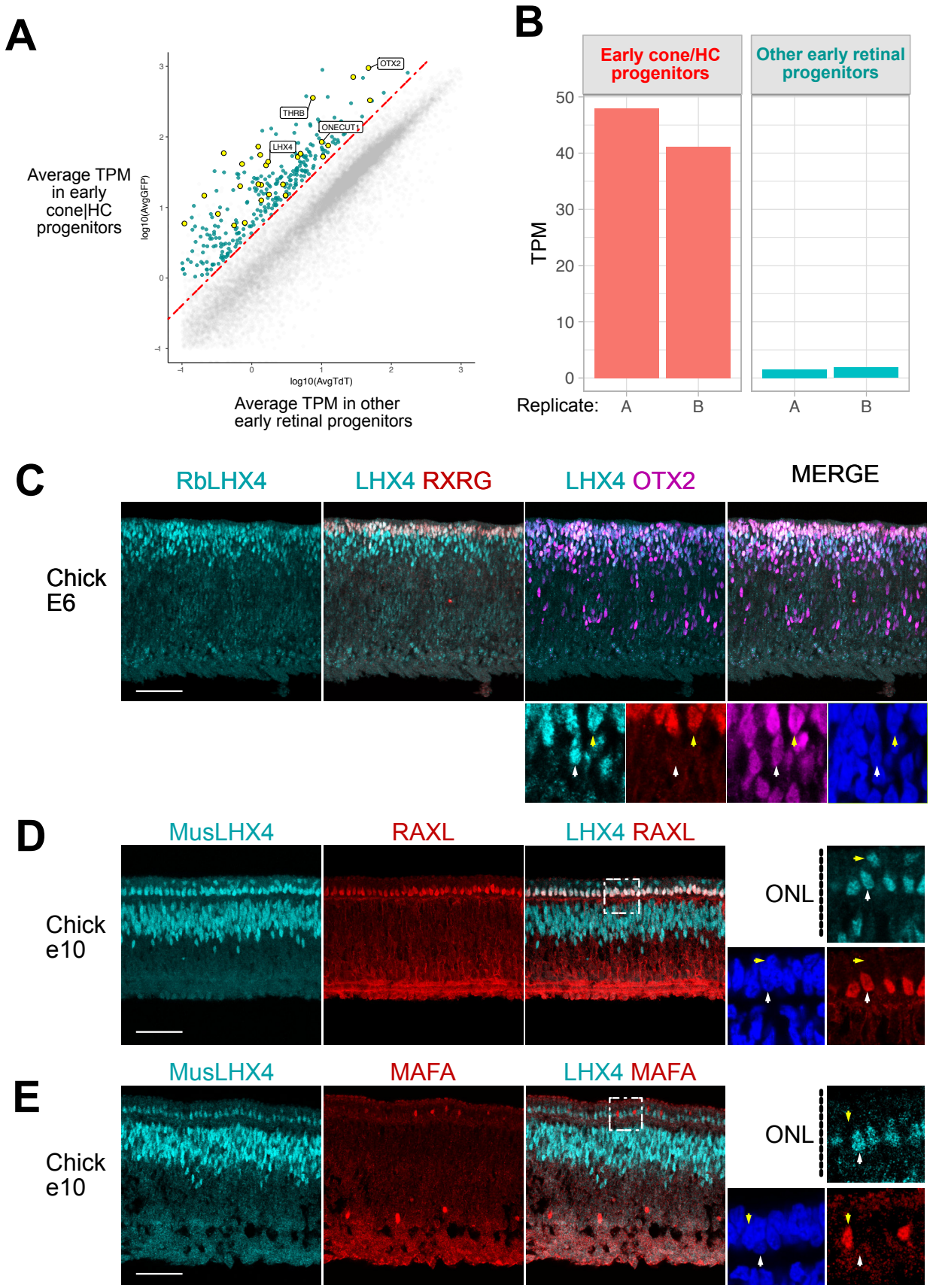


Figure 1 - **LHX4 is present in early cone photoreceptors during chick retinal development.** Large panels are maximum intensity projections Z-stacks and small panels are single planes of the same Z-stacks

(A) Average Transcripts Per Million (TPM) of genes between Cone/HC progenitors and other early retinal progenitors. Highlighted in green are genes 1.5-fold enriched in Cone/HC progenitors and highlighted in yellow, transcription factors.

(B) TPM values for LHX4 in Cone/HC progenitors and other early retinal progenitors.

(C) Cross-section of E6 chick retina imaged for LHX4, RXRG, and OTX2. Higher magnification panels show a single z-plane with marked a LHX4+/RAXL+ (white arrow) or LHX4+/OTX2+/RXRG+ cell (yellow arrow).

(D) Cross-section of E10 chick retina imaged for LHX4 and RAXL. Higher magnification panels show a single z-plane with marked a LHX4+/RAXL+ (white arrow) or LHX4+/RAXL- cell (yellow arrow).

(E) Cross-section of E10 chick retina imaged for LHX4 and MAFA. Higher magnification panels show a single z-plane with marked a LHX4+/MAFA- (white arrow) or LHX4-/MAFA+ cell (yellow arrow). Scale bar represents 50 µm

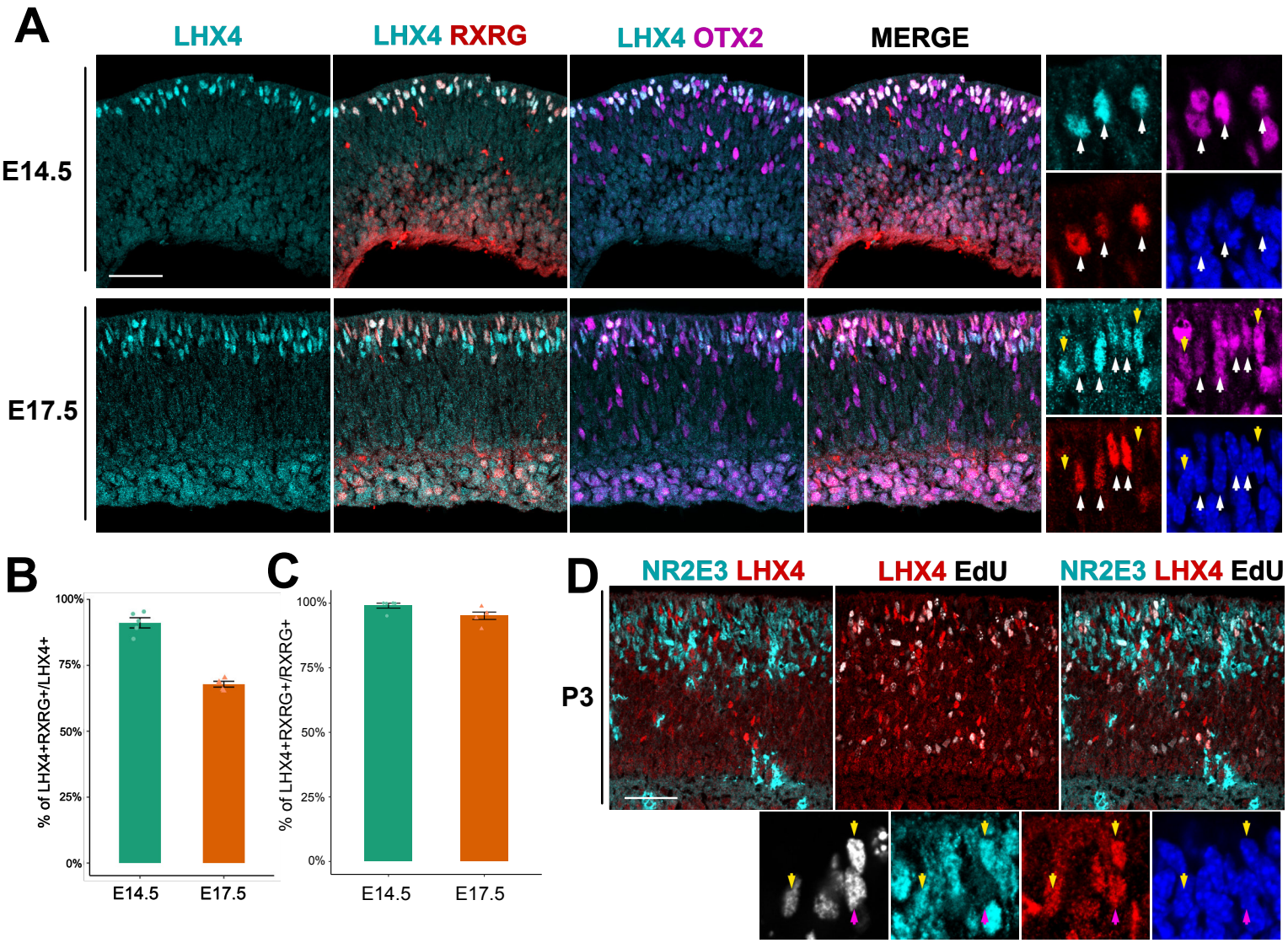


Figure 2 - **LHX4 protein is present in early photoreceptors precursors during mouse retinal development.** Large panels are maximum intensity projections Z-stacks and small panels are single planes of the same Z-stacks

A. Cross-section of embryonic mouse retinas imaged for LHX4, RXRG, and OTX2 at the designated timepoints. Higher magnification panels show a single z-plane with marked LHX4 cells positive (white arrows) or negative (yellow arrows) for RXRG and OTX2.

B. Quantification of the percentage of LHX4+ cells that are also RXRG+ at E14.5 and E17.5.

C. Quantification of the percentage of RXRG+ cells that are also LHX4+ at E14.5 and E17.5.

D. Cross-section of a P3 retina after a P0 EdU injection imaged for EdU, NR2E3 and LHX4. Higher magnification panels show a single z-plane with marked LHX4+/EdU+ cells positive (yellow arrows) or negative (purple arrows) for NR2E3. Scale bar represents 50 µm

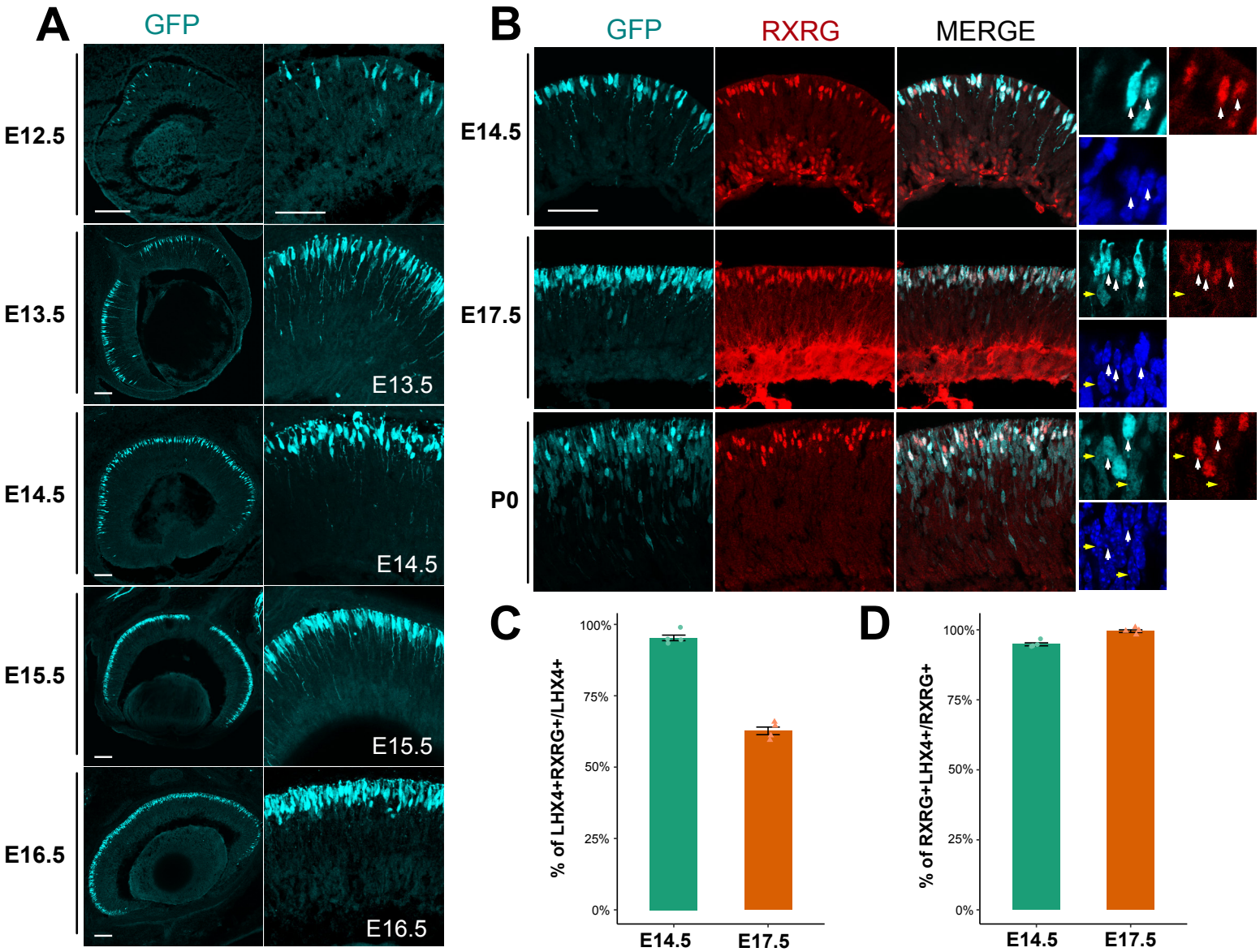


Figure 3 - **LHX4 BAC-EGFP is a reliable reporter of photoreceptors during retinal development.** Large panels are maximum intensity projections Z-stacks and small panels are single planes of the same Z-stacks. Scale bar represents 50 µm

A. Cross-section of embryonic LHX4-GFP mouse retinas imaged for EGFP at the designated timepoints.

B. Cross-section of embryonic LHX4-GFP mouse retinas imaged for EGFP and RXRG at the designated timepoints. Higher magnification panels show a single z-plane with marked EGFP cells positive (white arrows) or negative (yellow arrows) for RXRG.

C. Quantification of the percentage of EGFP+ cells that are also RXRG+ at E14.5 and E17.5.

D. Quantification of the percentage of RXRG+ cells that are also EGFP+ E14.5 and E17.5.

20x
e12.5 is 10x and not het cross

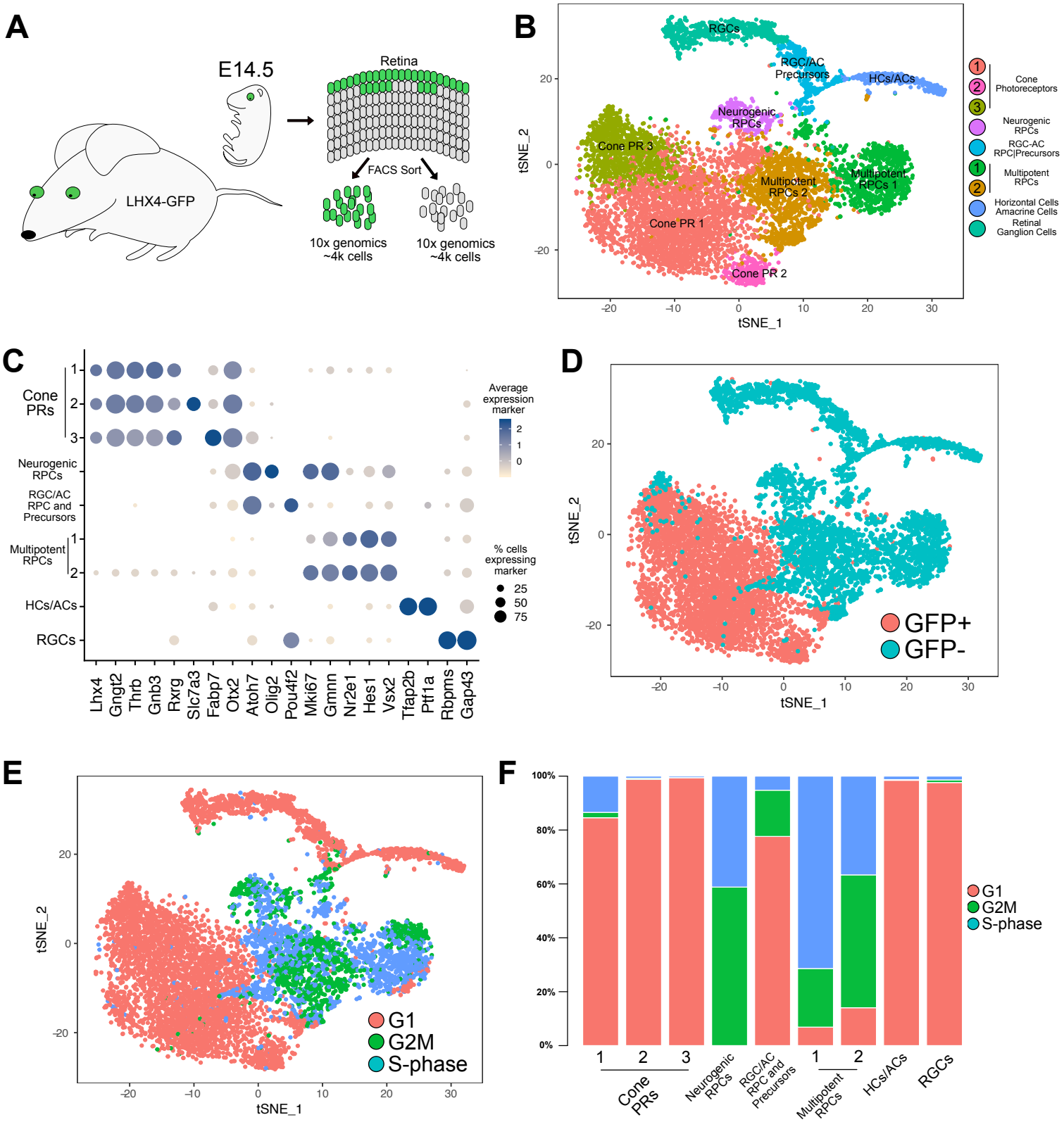


Figure 4 - TSNE - **Single cell sequencing of LHX4-GFP E14.5 developing retina.**

(A) Schematic of experimental design for single cell collection and sequencing.

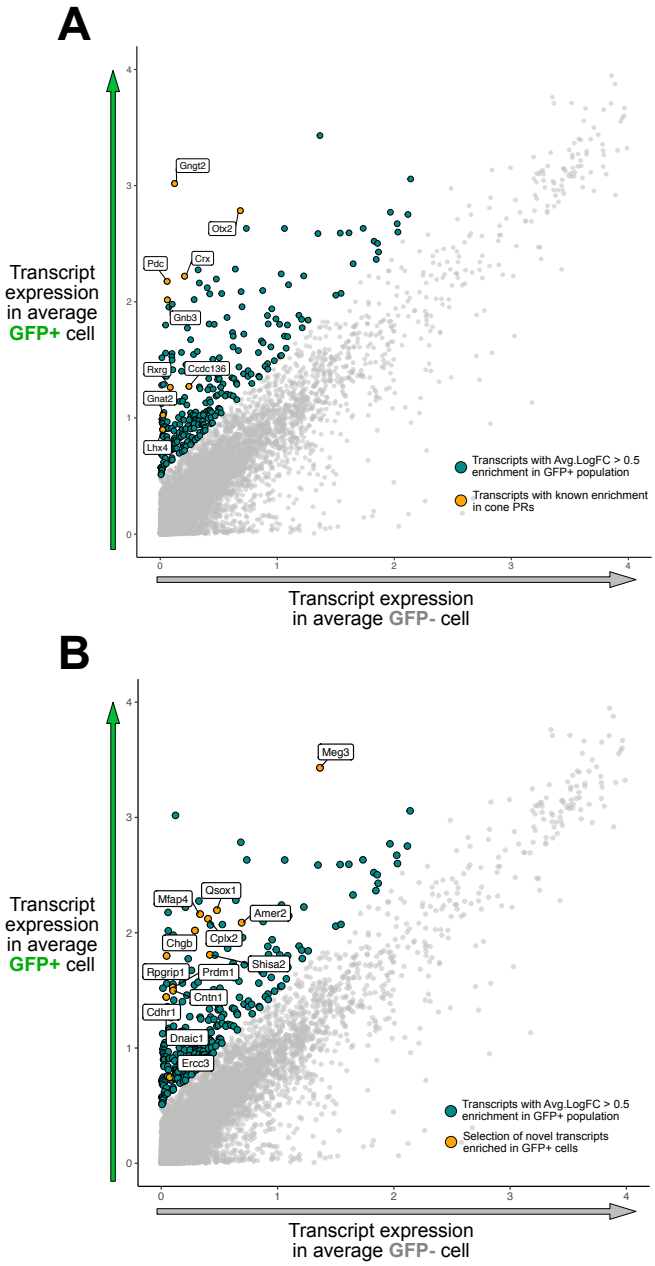
(B) tSNE plot of all GFP+ and GFP- cells collected, analyzed in tandem, and displaying the results of unsupervised cluster analysis with the assigned cell class according to their molecular signature.

(C) Dot plot displaying the average expression and percentage of cells expressing specific markers in each cluster. Markers displayed on the x axis were used for assignment of cell class to each cluster.

(D) Same tSNE plot as (B) displaying the original source of each cell, GFP+ or GFP- samples.

(E) Same tSNE plot as (B) displaying the cell cycle assignment of each cell.

(F) Percentage of cells in each cluster assigned to different cell cycle phases.



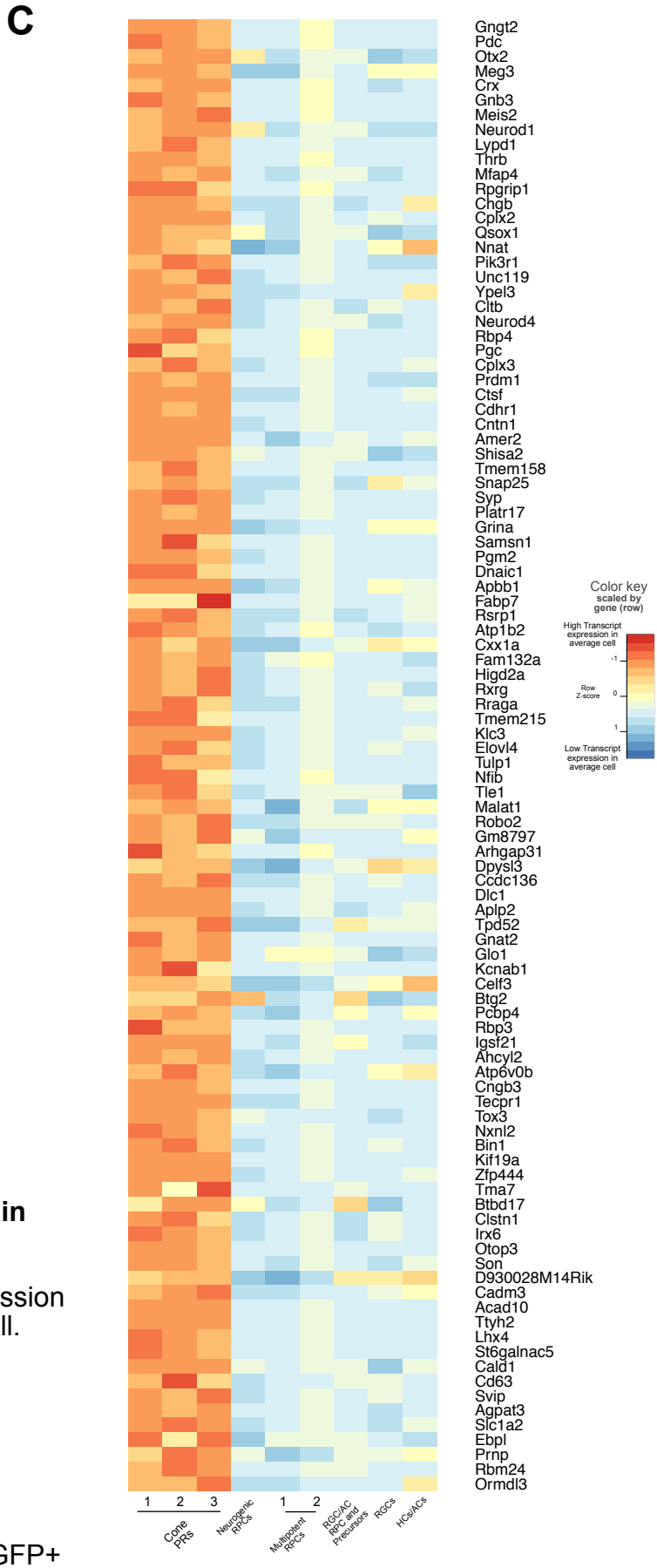


Figure 5 - **Identification of known and novel transcripts enriched in cone photoreceptors in LHX4-GFP+ cells.**

(A) Plot displaying the expected transcript expression in an average GFP+ (y-axis) or GFP- (x-axis) cell. Highlighted are all genes with >0.5 fold change enrichment in GFP+ cells and labelled are a selection of cone-related transcripts.

(B) Same plot as (A) labeling novel transcripts enriched in the same population.

(C) Heatmap of top 100 transcripts enriched in GFP+ cells. Color indicates the expected transcript expression of that gene in an average cell for each cluster identified. Expression values are scaled per gene.

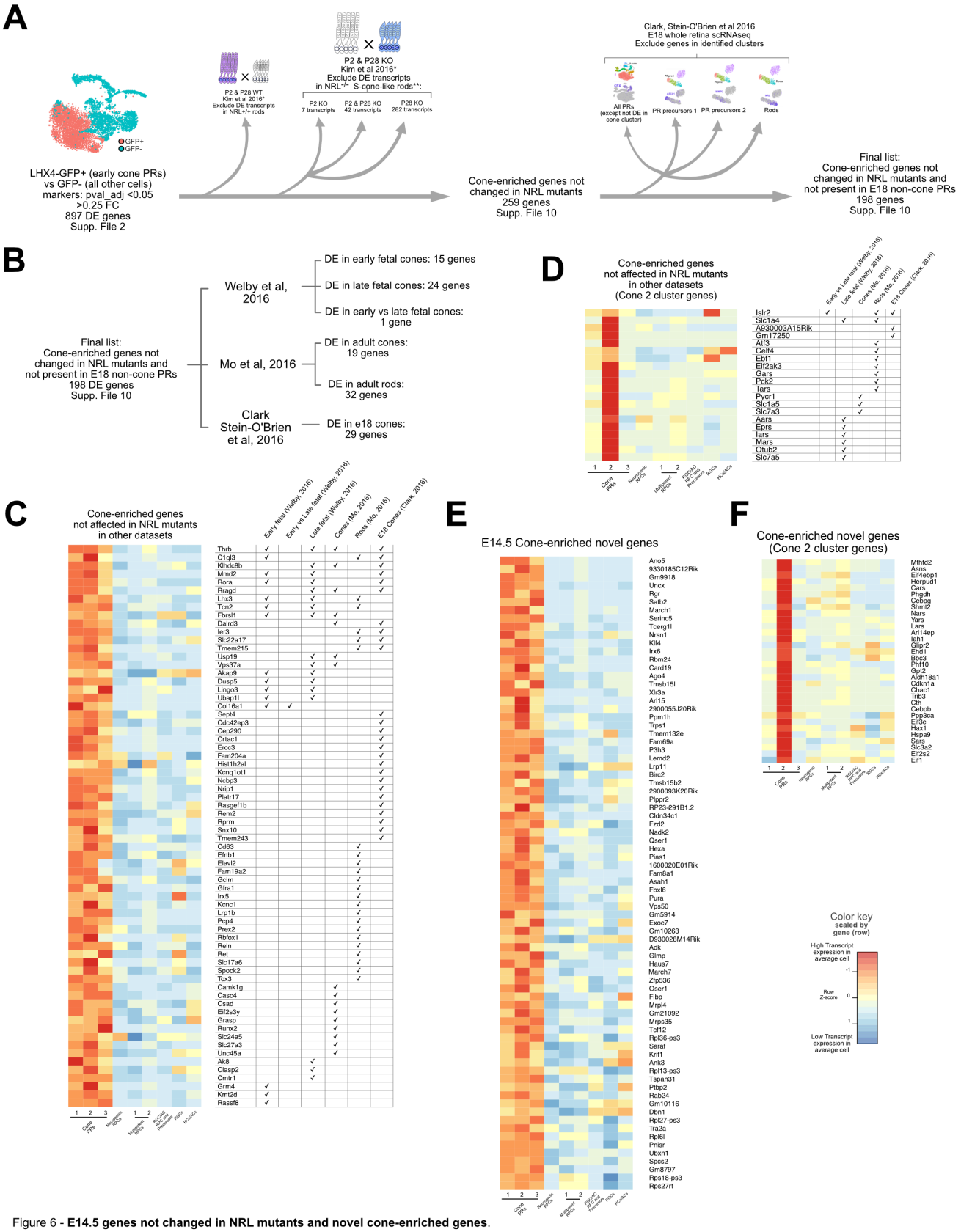


Figure 6 - **E14.5 genes not changed in NRL mutants and novel cone-enriched genes**.

A) Schematic describing the comparison and selection of genes enriched in E14.5 cones that are not changed in the NRL mutant dataset and not present in early rods or photoreceptor precursors at E18 (Clark, Stein O'Brien 2018)

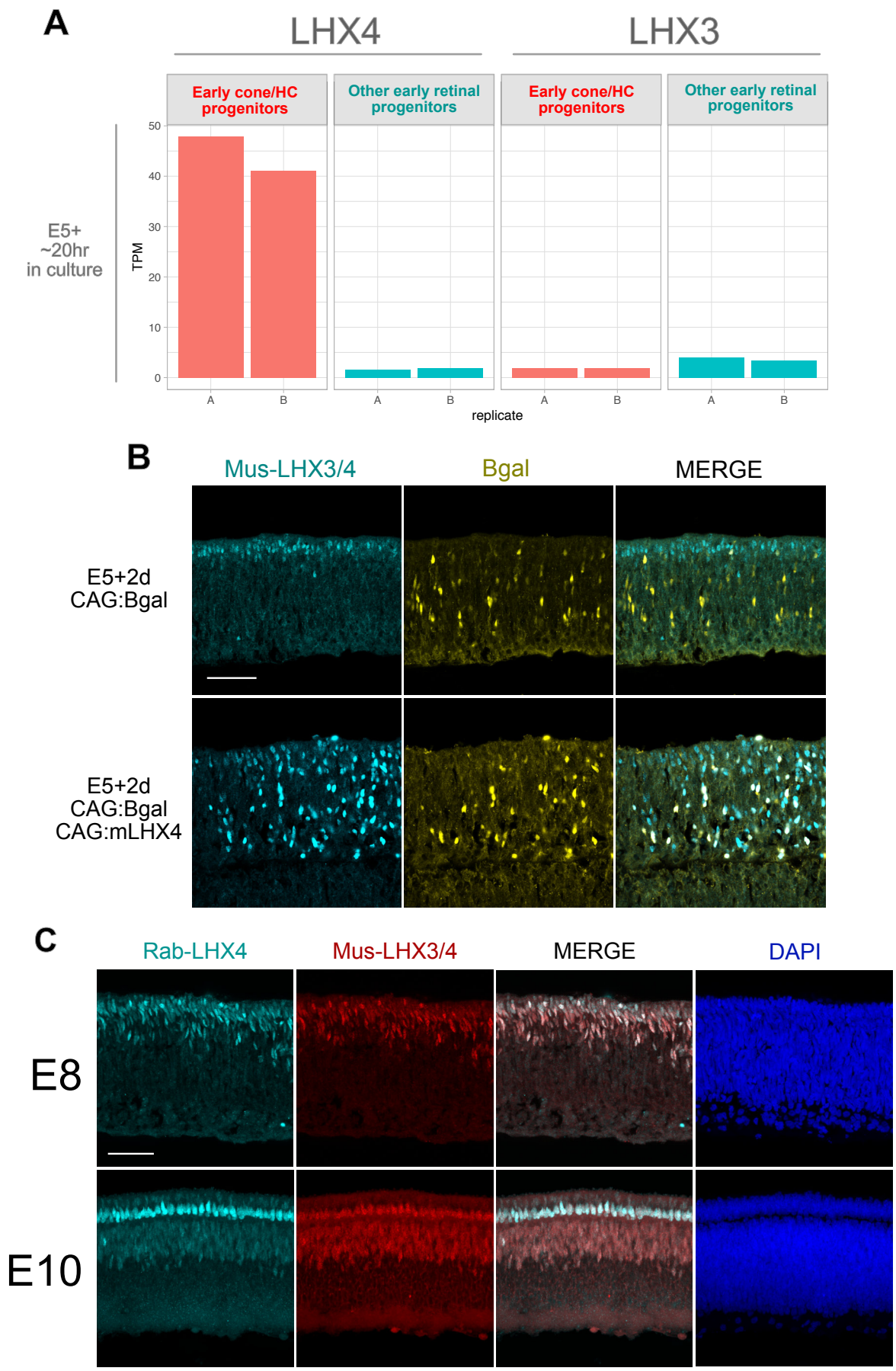
B) Amount of genes found overlapping in the displayed datasets from the final list of genes.

C) Heatmap of transcripts enriched in E14.5 cones that are not changed in the NRL mutant dataset and not present in early rods or photoreceptor precursors at E18. The genes displayed are also present in the datasets analyzed, and their presence is indicated by a checkmark in the right table. Color indicates the expected transcript expression of that gene in an average cell for each cluster identified. Expression values are scaled per gene.

D) Same as C but for genes found in the Cone 2 cluster.

E) Heatmap of novel genes found in this study to be enriched in E14.5 cones but not present in any of the other datasets.Color indicates the expected transcript expression of that gene in an average cell for each cluster identified. Expression values are scaled per gene.

F) Same as E but for novel genes in the Cone 2 cluster.

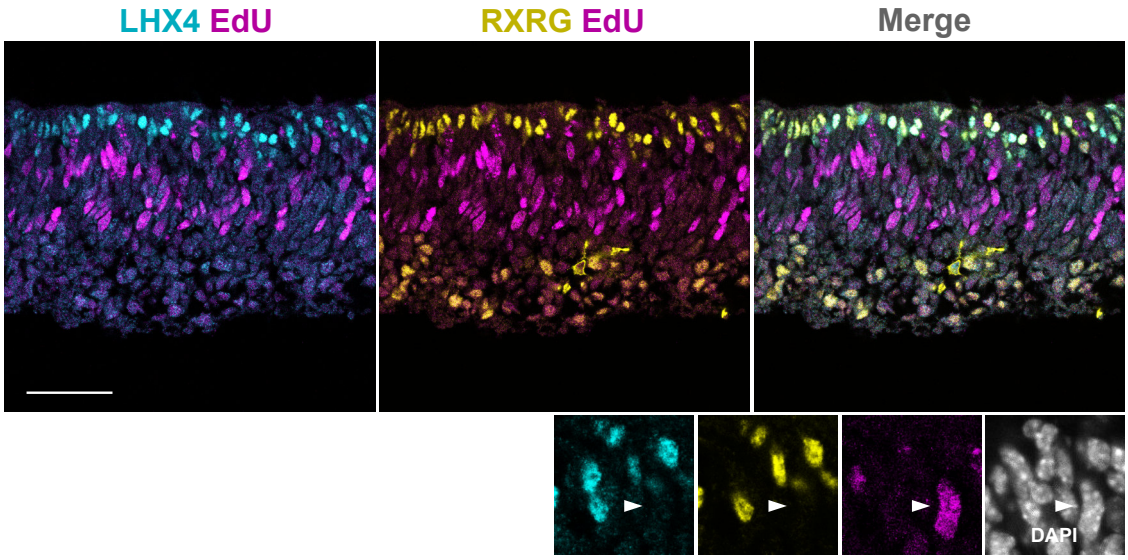


Supplemental Figure 1 - **LHX4 and not LHX3 is detected in early chick retina development.**

(A) TPM values for LHX4 and LHX3 in Cone/HC progenitors and other early retinal progenitors.

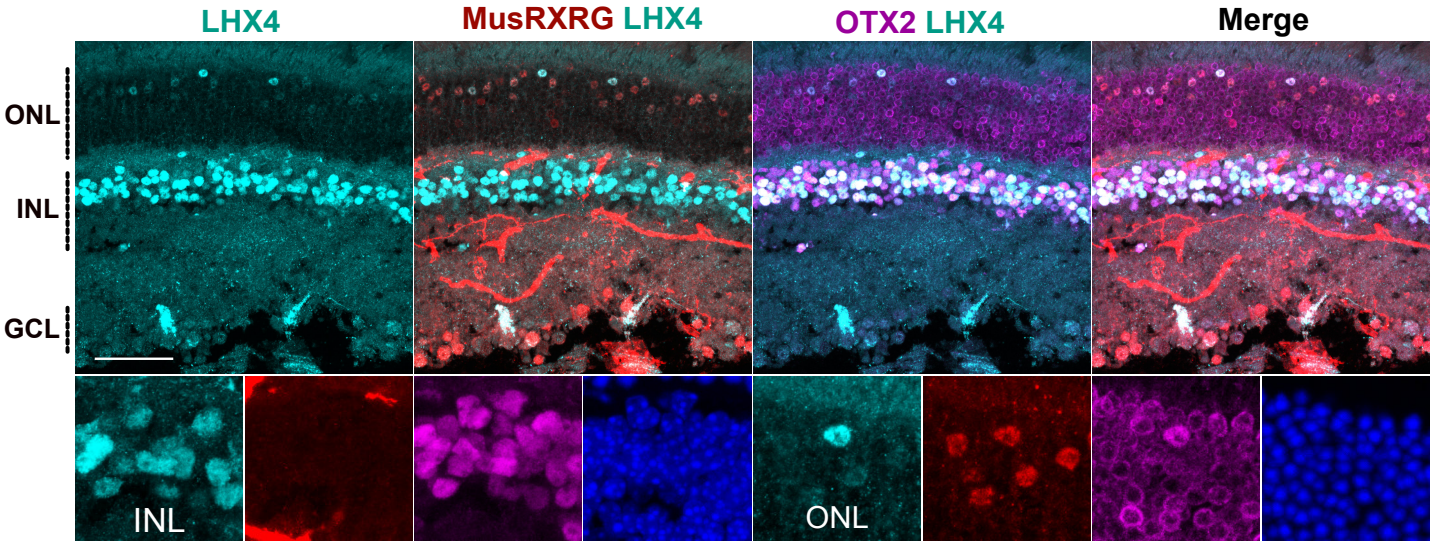
(B) Cross-section of a chicken retina electroporated at E5 and cultured for ~20hrs. Electroporated constructs are denoted on the left side. Retinas were imaged for Bgal and LHX3/4 with mouse antiLHX3 (DSHB).

(C) Maximum intensity projections of cross-section in chicken retinas at designated timepoints imaged for LHX4 with mouse (DSHB) and rabbit (Proteintech) antibodies.S cale bar represents 50 µm



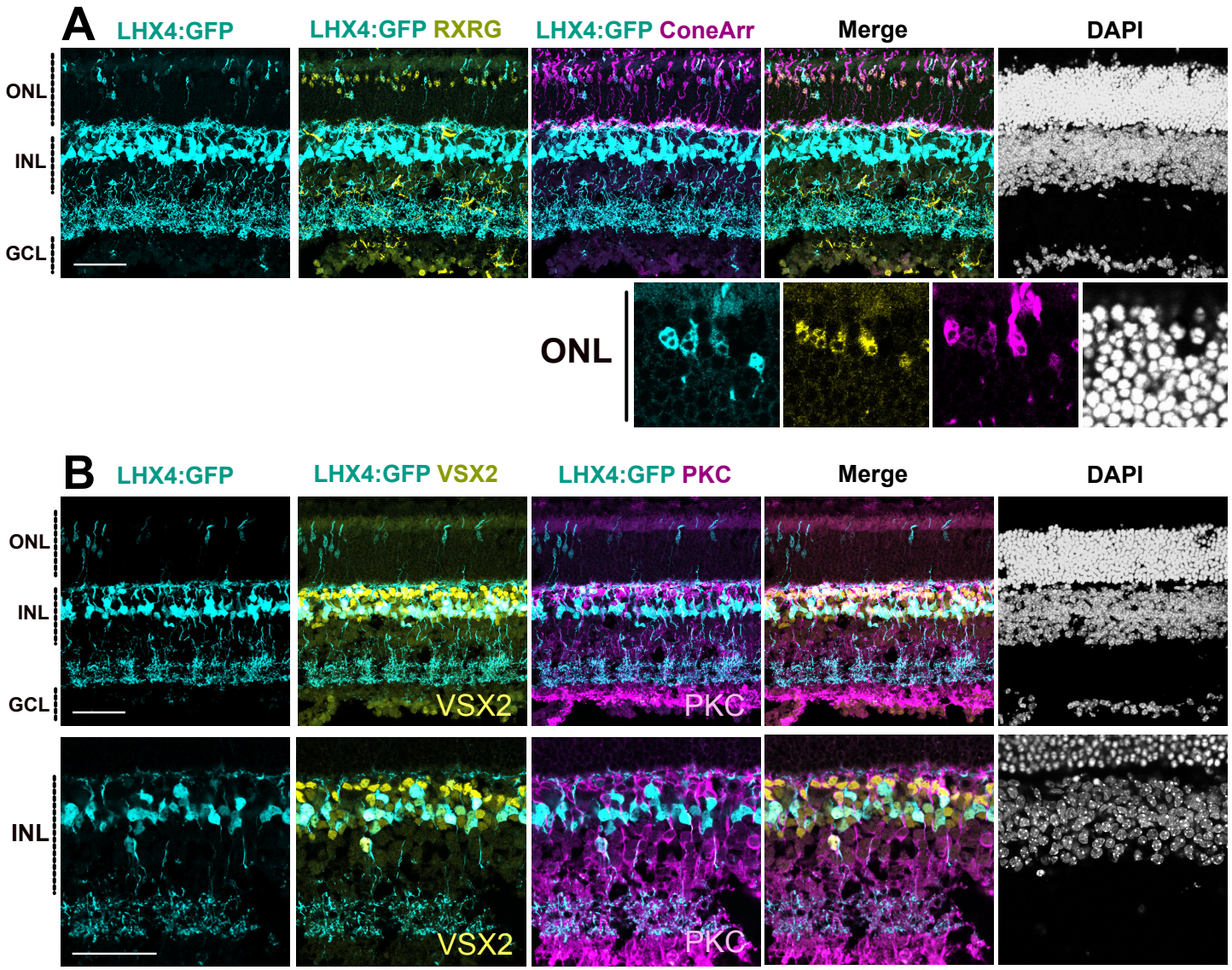
Supplemental Figure 2 - **LHX4 expression starts in post-mitotic cones**

Cross-section of a E14.5 mouse retina, developed for EdU (2hr pulse) and imaged for LHX4 and RXRG. Small panels are digitally zoomed. Arrow points at representative EdU cell negative for LHX4 and RXRG Scale bar represents 50 µm .



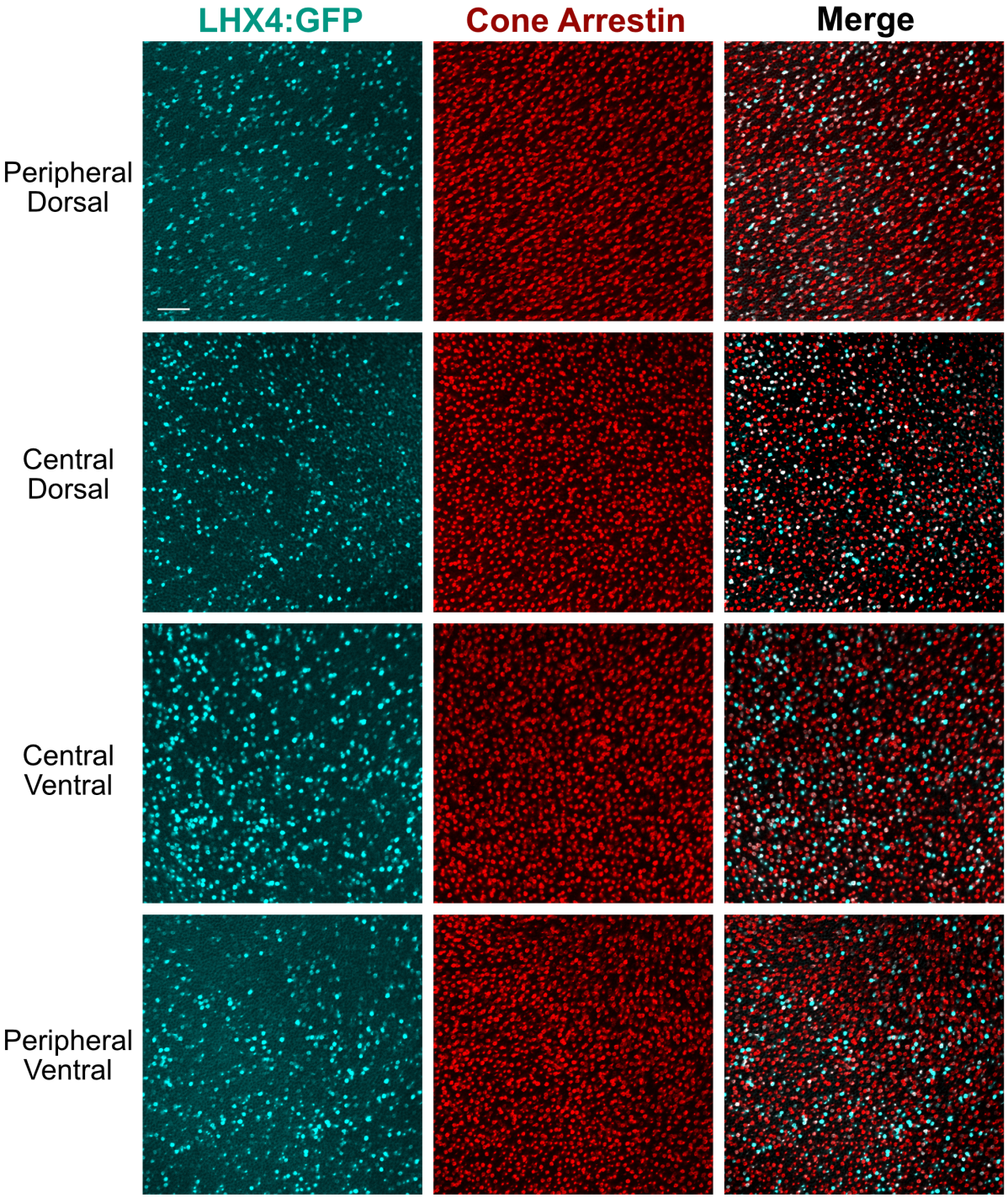
Supplemental Figure 3 - **LHX4 is expressed in cone bipolar cells and a subpopulation of cones in the adult mouse retina.**

Cross-section of a P27 mouse retina imaged for LHX4, RXRG, and OTX2. Higher magnification panels in the ONL and INL as specified. Large panels are maximum intensity projection Z-stack and small panels are single planes of the same Z-stack. Scale bar represents 50 μm



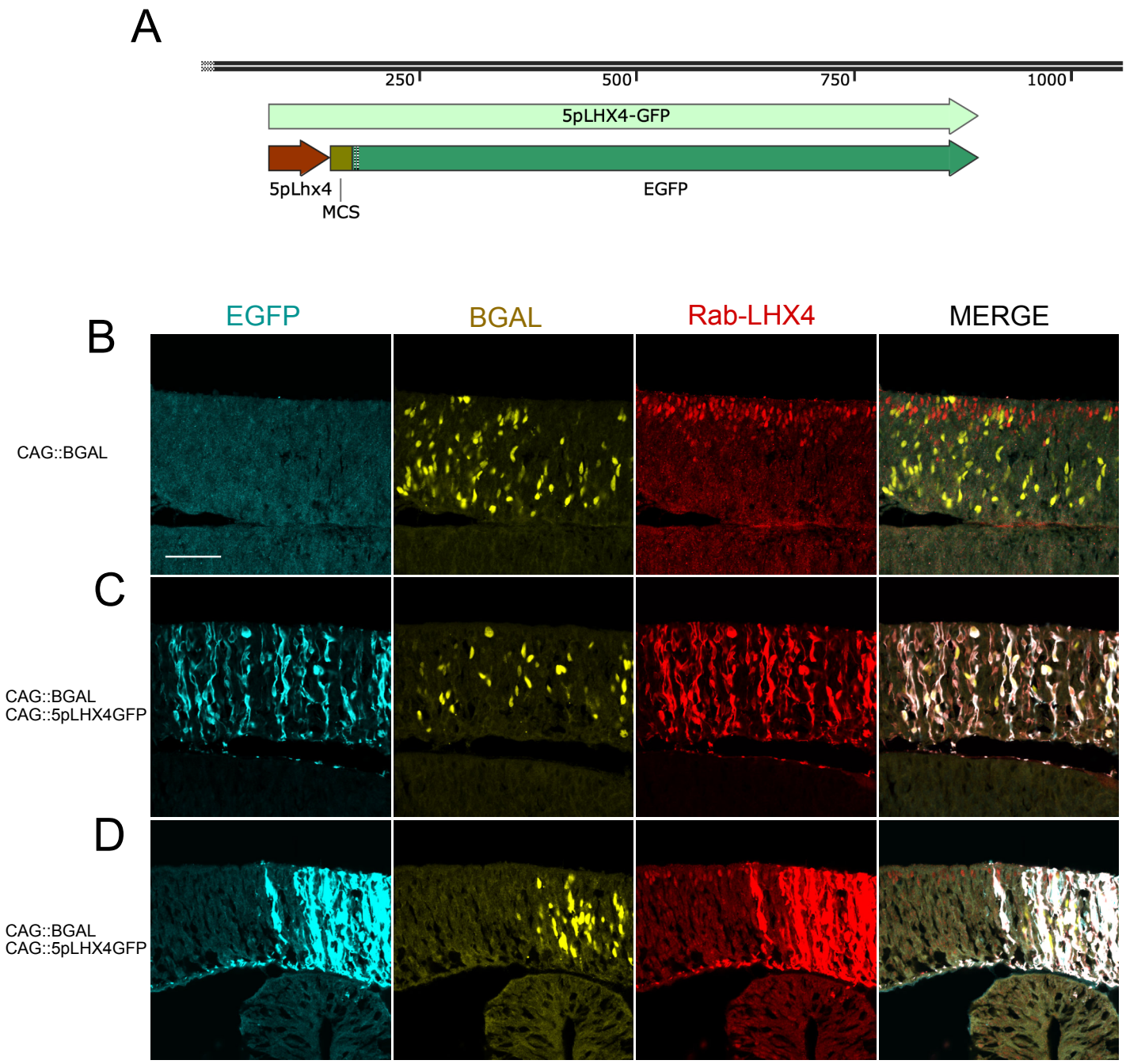
Supplemental Figure 4 - **LHX4-GFP reporter is active in cone bipolar cells and a subset of cones in the adult mouse retina.**

Cross-section of P27 mouse retinas imaged for EGFP, RXRG, and Cone Arrestin in (A), and EGFP, VSX2 and PKC in (B). Higher magnification panels in (A) show a single z-plane in the ONL. Large panels are maximum intensity projections Z-stacks and small panels are single planes of the same Z-stacks. Scale bar represents 50 µm



Supplemental Figure 5 - **LHX4-GFP reporter in the adult is active is not subject to dorsal-ventral gradient.**

Whole mount of a P27 LHX4-GFP mouse retina imaged for EGFP and Cone Arrestin imaged at the ONL. Scale bar represents 50 µm

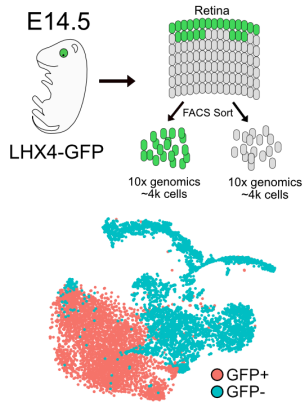


Supplemental Figure 6 - **LHX4-GFP insertion retains part of the LHX4 5' coding sequence and produces a N-terminusLHX4-GFP fusion.**

(A) Schematic of coding region amplified from LHX4-GFP locus containing a portion of the N-terminus of LHX4.

(B-D) Cross-section of chicken retinas after electroporation and ~20hr incubation. Retinas are imaged for EGFP, RbLHX4 and Bgal. In the case of (C), imaged at the edge of the electroporated area with increased exposure to show endogenous LHX4 signal. Scale bar represents 50 μm

A



LHX4-GFP+ (early cone PRs)
vs GFP- (all other cells)
markers: pval_adj <0.05
>0.25 FC
897 DE genes
Supp File 2

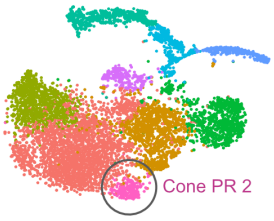
77 novel E14.5 genes

Supp. File 3

Current dataset	Dataset | Experimental design	# of genes (or transcripts) overlap
Welby et al 2017 Human fetal retina explants AAV2/9.pR2.1::GFP sort L-M Opsin gene promoter pval_adj <0.05; >0.5 FC Supp. File 4	Early fetal GFP+ Samples: 10wks, 11wks,12wks, 14wks GFP+ x GFP- cells 1145 enriched genes in GFP+	165
Welby et al 2017 Human fetal retina explants AAV2/9.pR2.1::GFP sort L-M Opsin gene promoter pval_adj <0.05; >0.5 FC Supp. File 4	Late fetal GFP+ Samples: 17wks, 19wks(1),19wks(2), 20wks GFP+ x GFP- cells 1721 enriched genes in GFP+	211
Welby et al 2017 Human fetal retina explants AAV2/9.pR2.1::GFP sort L-M Opsin gene promoter pval_adj <0.05; >0.5 FC Supp. File 4	Early vs Late Fetal GFP+ Samples: 10wks, 11wks,12wks, 14wks GFP+ x GFP- cells 1145 enriched genes in GFP+	15
Mo et al 2016 Mouse Cone/Rod adult bulk RNAseq Rods: Lmopc::Cre (Le et al,2006) Cones: HRGP::Cre (Le et al, 2004) PPDE >0.95 ; >0.5 FC Supp. File 5	Adult cones vs adult rods 2959 DE genes in cones	251
Mo et al 2016 Mouse Cone/Rod adult bulk RNAseq Rods: Lmopc::Cre (Le et al,2006) Cones: HRGP::Cre (Le et al, 2004) PPDE >0.95 ; >0.5 FC Supp. File 5	Adult cones vs adult rods 2253 DE genes in rods	121
Kim et al 2016* Mouse NRL::GFP Rods bulk RNAseq NRL+/+ vs NRL-/- pval_adj <0.05; >0.5 FC *transcript-level analysis Supp. File 6	P2 NRL+/+:NRL-GFP rods vs NRL-/-:NRL-GFP S-cone-like rods 688 DE transcripts in NRL+/+ rods	51
Kim et al 2016* Mouse NRL::GFP Rods bulk RNAseq NRL+/+ vs NRL-/- pval_adj <0.05; >0.5 FC *transcript-level analysis Supp. File 6	P2 NRL+/+:NRL-GFP rods vs NRL-/-:NRL-GFP S-cone-like rods 308 DE transcripts in NRL-/- S-cone-like rods	67
Kim et al 2016* Mouse NRL::GFP Rods bulk RNAseq NRL+/+ vs NRL-/- pval_adj <0.05; >0.5 FC *transcript-level analysis Supp. File 6	P28 NRL+/+:NRL-GFP rods vs NRL-/-:NRL-GFP S-cone-like rods 5171 DE transcripts in NRL+/+ rods	341
Kim et al 2016* Mouse NRL::GFP Rods bulk RNAseq NRL+/+ vs NRL-/- pval_adj <0.05; >0.5 FC *transcript-level analysis Supp. File 6	P28 NRL+/+:NRL-GFP rods vs NRL-/-:NRL-GFP S-cone-like rods 10680 DE transcripts in NRL-/- S-cone-like rods	508
Clark, Stein-O'Brien et al 2016 Mouse E18 whole retina scRNAseq 2x replicates pval_adj <0.05; >0.25 FC Supp. File 7	All photoreceptor clusters vs all retinal cells 524 marker genes in photoreceptors at E18	412
Clark, Stein-O'Brien et al 2016 Mouse E18 PR subclustering scRNAseq 2x replicates pval_adj <0.05; >0.25 FC Supp. File 8	PR precursors 1 (ASCL1+) vs all other PRs 272 marker genes in photoreceptors at E18	17
Clark, Stein-O'Brien et al 2016 Mouse E18 PR subclustering scRNAseq 2x replicates pval_adj <0.05; >0.25 FC Supp. File 8	PR precursors 2 (MMP2+) vs all other PRs 155 marker genes in photoreceptors at E18	59
Clark, Stein-O'Brien et al 2016 Mouse E18 PR subclustering scRNAseq 2x replicates pval_adj <0.05; >0.25 FC Supp. File 8	Rod cluster (NRL+) vs all other PRs 138 marker genes in photoreceptors at E18	78
Clark, Stein-O'Brien et al 2016 Mouse E18 PR subclustering scRNAseq 2x replicates pval_adj <0.05; >0.25 FC Supp. File 8	Cone cluster (OPN1SW+) vs all other PRs 363 marker genes in photoreceptors at E18	163

B

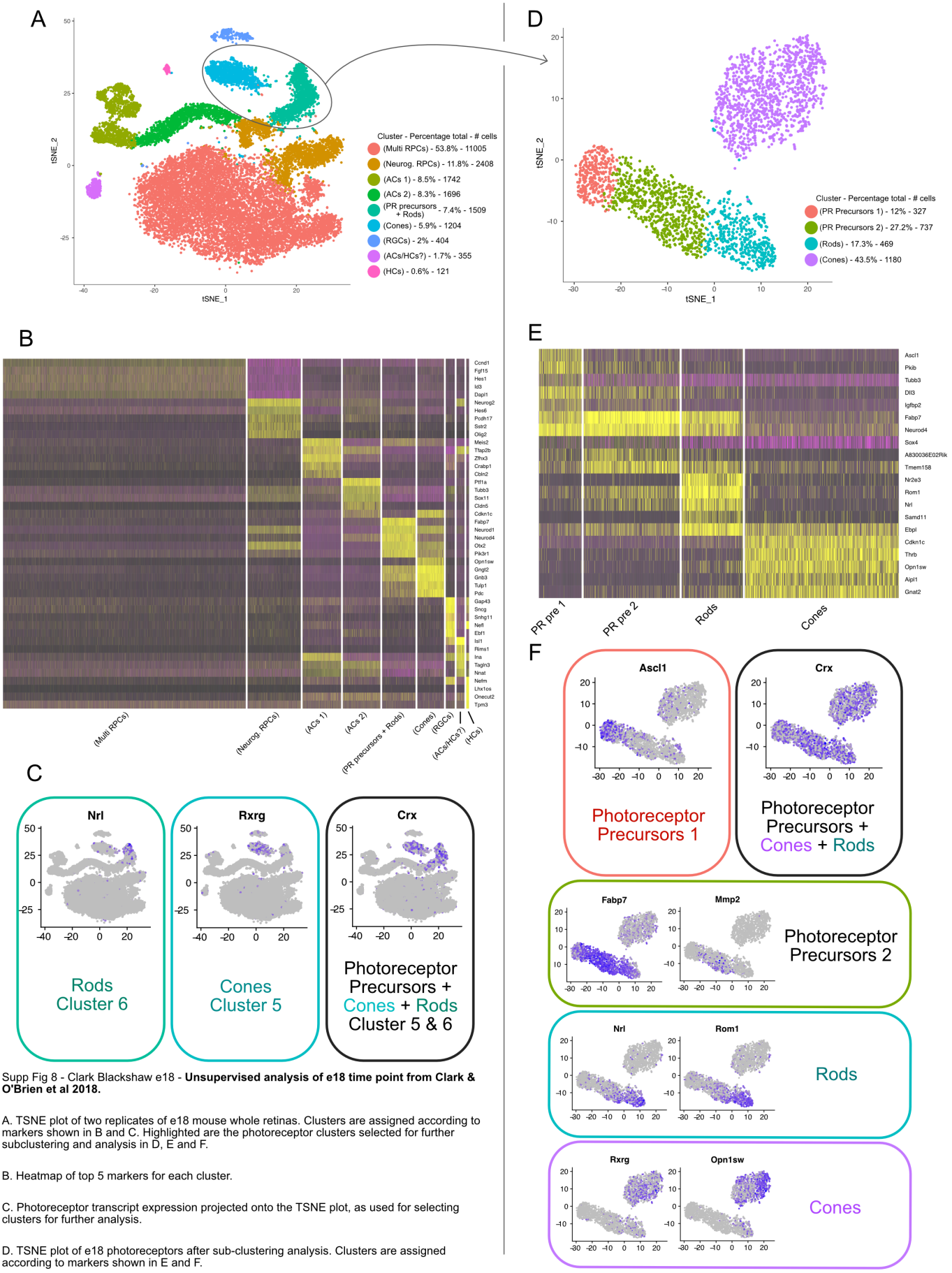
Cone PR2 cluster (SLC7A3+)
vs all other clusters
markers: pval_adj <0.05
>0.5 FC
55 (unique) DE genes
Supp. File 1



33 novel E14.5 genes

Supp. File 3

Supplemental Figure 7 - **Summary of datasets compared with E14.5 cone-enriched genes from LHX4-GFP current datasets.**
A) Table describing the datasets used for comparison and the amount of overlapping genes with E14.5 cone-enriched genes. For each report (Welby, Mo, Kim and Clark), datasets are divided in subsections depending on enrichment in a particular cell type, as denoted.
B) Summary of comparison with Cone 2 cluster enriched genes.



Supp Fig 8 - Clark Blackshaw e18 - **Unsupervised analysis of e18 time point from Clark & O'Brien et al 2018.**

A. TSNE plot of two replicates of e18 mouse whole retinas. Clusters are assigned according to markers shown in B and C. Highlighted are the photoreceptor clusters selected for further subclustering and analysis in D, E and F.

B. Heatmap of top 5 markers for each cluster.

C. Photoreceptor transcript expression projected onto the TSNE plot, as used for selecting clusters for further analysis.

D. TSNE plot of e18 photoreceptors after sub-clustering analysis. Clusters are assigned according to markers shown in E and F.

E. Heatmap of top 5 markers for each photoreceptor cluster.

F. Photoreceptor transcript expression projected onto the TSNE plot, as used for cluster assignment..