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 in 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 these cells. A comparison to previously reported datasets allowed the classification of genes according to developmental timing, cell type specificity, and whether they were regulated by the rod transcription factor Nrl. This analysis has extended the set of known early cone enriched genes and identified those that are regulated independently of Nrl. This report furthers our knowledge of the transcriptional events that occur in early cone photoreceptors.

INTRODUCTION

Cone and rod photoreceptors are the photosensitive cells of the retina that contribute to image formation. Cones mediate color discrimination and high acuity vision 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 [1]. As such, the development of cell-based therapeutic strategies based on the formation of new cone photoreceptors is a promising strategy. [2]However, there is presently a gap in our knowledge of the gene regulatory networks that control the genesis of these cells as well as the early steps in their differentiation. Thus, the design of informed strategies to direct cone production and how to appropriately benchmark the differentiation of de novo generated cells is lacking.

Two main strategies have been used to investigate the early gene regulation programs of cone photoreceptors. One has been to develop reagents to label developing cone cells or the RPCs that generate them and use high-throughput methods to identify the genes with enriched expression in these cells compared to other cells present at the time [3,4]. Such methods have also been used at later differentiation timepoints [5–7]. While these methods provide critical information, they also rely on transcriptional reporters that may have expression that is broader than just cone photoreceptors and do not provide cellular resolution of these gene expression patterns. The second strategy to examine cone photoreceptor gene expression has been through the use of the Nrl mouse model. Nrl is a critical rod-expressed transcription factor that is necessary to promote rod gene expression and repress cone gene expression in rod cells [8]. The Nrl mouse knockout leads to a large increase in cone gene expression as

69 the large number of rods in the mouse undergo derepression of cone genes[9–11]. These
70 Nrl knockout rods have been interpreted as either a complete fate switch to cones or a
71 partial conversion to "cods" [8,12,13]. As these cells have been used in a number of studies
72 to model cone photoreceptors, the extent to which these cells are transformed to the
73 cone fate is important for both a justification in using them as a model for endogenous
74 cone cells and to understand photoreceptor diversification.

75 Here, we identified the Lim Homeobox Protein 4 (LHX4) gene as enriched in
76 cones during early chick retinal development, in addition to bipolar cells. In the mouse,
77 LHX4 was also determined to be a reliable marker for cones during early retinal
78 development but with expanded expression in late embryonic development and
79 eventual expression in BCs. A LHX4::GFP transgenic line [14] recapitulated the
80 endogenous LHX4 expression pattern and was used to generate a single cell dataset
81 highly enriched in cone photoreceptors, which provided an in-depth look at the
82 molecular profile of these cells in the earliest stages of differentiation in the mouse. A
83 comparison was made between previous datasets that targeted mammalian cones and
84 photoreceptors, including those made in the Nrl knockout mouse. This led to the
85 identification of unique cone expression signatures not observed in previous datasets,
86 including those of Nrl knockout rods, supporting previous observations that these cells
87 are not completely transformed into cones.

88

89

90 **RESULTS**

91 **LHX4 is present in early cone photoreceptors in the chicken retina**

92 Recently, we established the transcriptional profile of retinal progenitor cells
93 (RPCs), defined by the activity of the ThrbCRM1 element, that are biased towards the
94 cone and horizontal cell (HC) fate in the early chick retina[3]. Using this dataset, we
95 screened for potential cone-enriched transcripts that could serve as markers for early
96 cones. After establishing a criterion for >1.5-fold change score between cone/HC RPCs
97 and other concurrent populations (enriched in "Other early retinal progenitors") we
98 selected for transcription factors (TFs) enriched in the cone/HC RPCs (Fig 1 A). We
99 identified the LIM homeobox 4 (LHX4) gene as highly enriched, along with known TFs in
100 this population such as THRB, ONECUT1, and OTX2. This transcript has significant fold
101 change (b = 3.3) and a low number of reads in the non-ThrbCRM1 active population,
102 suggesting high specificity towards the cone/HC RPC population at this time (Fig 1 B).

103 A previous report has examined the presence of LIM-domain factors in early
104 chick photoreceptor development[15]. This study suggested that LHX3 was abundantly
105 present in the apical portion of the retina and localized to photoreceptors once the ONL
106 is clearly distinguished. As the RNA-Seq data indicated that LHX4 expression is
107 prominent in ThrbCRM1 reporter-positive cells at early stages while LHX3 transcript
108 presence is marginal in all targeted cells (Supp. Fig. 1 A), we suspected that this
109 previous study could have detected LHX4 instead of LHX3 at earlier timepoints. To test
110 this, we electroporated a mouse LHX4 misexpression plasmid (CAG::mLHX4) plasmid
111 alongside a CAG::nucβgal construct into the chick E6 retina, cultured it for 2 days, and
112 detected for LHX4 using the LHX3 Developmental Studies Hybridoma Bank (DSHB)

113 antibody. As predicted, we observed robust immunoreactivity of βgal+ electroporated
114 cells with the LHX3 antibody (Supp. Fig. 1 B), suggesting that this antibody is also
115 capable of detecting mouse LHX4, and thus likely chicken LHX4 as well.

116 With the use of this LHX3/4 antibody and a rabbit polyclonal antibody, we
117 examined LHX4 presence during embryonic development of chick at E6 and E10.
118 Expression at E6 is restricted to the scleral portion of the retina, where photoreceptors
119 are located. As the LHX4 transcript was highly enriched in cells that activate the
120 ThrbCRM1 element, which requires OTX2 for activity, we expected LHX4 to be present
121 predominantly in the OTX2+ population at E6. Indeed, LHX4 is detected in a large
122 percentage (but not all) of OTX2-positive cells (Fig. 1 C).

123 At E6, cones are the major class of photoreceptors that are produced, as the
124 earliest known rod photoreceptor marker in the chicken retina, L-Maf (MAFA), is not
125 detected until E9[16]. As the LHX4 pattern at this stage is localized in the apical portion of
126 the retina, we examined LHX4+ cells for co-expression with the cone marker RXRG.
127 Many of the most apically located LHX4 cells were indeed positive for RXRG (Fig. 1 C).

128 It is unclear to what extent and in which cells LHX3, LHX4 or both proteins are
129 present in later timepoints when the transcriptional status of LHX3 may change. In one
130 report[15], antibody staining using DSHB anti-LHX3 showed strong nuclear Outer Nuclear
131 Layer (ONL) and INL signal at E10, which suggested photoreceptor and bipolar signals.
132 We also detected signal in the ONL and INL with the LHX3 antibody. However, the
133 rabbit anti-LHX4 antibody detected a similar pattern with strikingly strong signal in the
134 ONL and weaker in the INL (Suppl. Fig. 1 C). As clear evidence of LHX3 RNA

135 expression in the ONL was not observed in the previous study, this suggests that LHX4
136 and not LHX3 may continue to be expressed in E10 chicken photoreceptors.

137 To examine the photoreceptor subtype expression more closely, we compared to
138 RAXL expression, a marker for cone photoreceptors[16], and observed that all RAXL-
139 positive cells were also positive for LHX4 (Fig 1 D). A smaller number of cells in the
140 upper part of the ONL were LHX4-positive and not RAXL-positive. As the ONL contains
141 both cones and rods, we used an antibody to MAFA[16,17], the earliest rod marker, to
142 determine if these cells were rod photoreceptors. No overlap between MAFA and LHX4
143 was detected at E10 in the ONL (Fig 1 E). This data suggests that LHX4 is expressed
144 predominantly in developing cone photoreceptors in the chicken retina.

145 **LHX4 is present in early cone photoreceptors in the mouse retina**

146 We sought to establish if this protein was also present in early photoreceptors of the
147 mouse retina. At E14.5, LHX4 immunoreactivity was present in the scleral portion of the
148 retina, where developing photoreceptors are located (Fig 2 A). To confirm that LHX4
149 was present in cone photoreceptors, we used the cone-expressed genes RXRG and
150 OTX2 to identify these cells[18]. RXRG is also expressed in some retinal ganglion cells
151 but these cell types can be readily distinguished by location. Many LHX4 cells are
152 positive for RXRG and OTX2 at E14.5, in a similar pattern to the chick retina. At E14.5,
153 we found that 91.1+-1.9% (mean+-SEM) of cells positive for LHX4 were also positive for
154 RXRG (Fig 2 B), signifying that the majority of the LHX4+ population at this timepoint
155 were composed of early cones. In fact, nearly all cells positive for RXRG at E14.5 were
156 also positive for LHX4 (99+-1%, Fig 2 C).

157 At a later embryonic stage (E17.5) we found that LHX4 cells still co-localize with
158 RXRG protein, but there was an increase in LHX4+/OTX2+ cells that did not express
159 RXRG (Fig 2 A). At this timepoint, only 67.8+-1.1% of LHX4+ cells were positive for
160 RXRG, suggesting that LHX4 was active in other emerging cell populations in addition
161 to early cones (Fig 2 B). However, 95.1+-1% of RXRG+ cells were positive for LHX4 at
162 e17.5, indicating that the majority of cones retained LHX4 expression (Fig 2 C). The
163 rabbit LHX4 polyclonal antibody had a reduced quality of staining at this timepoint, so it
164 is possible the small fraction of cones that are not LHX4+ had undetectable LHX4 signal
165 under these conditions or could point to a small subpopulation of cones that do not
166 express LHX4. Additionally, we tested if LHX4 was present in post-mitotic cone
167 photoreceptors or in dividing cells. E14.5 retinas exposed to a 2 hour 5-ethyny-2'-
168 deoxyuridine (EdU) pulse did not show a qualitative overlap between EdU and LHX4 or
169 RXRG (Supp. Fig. 2), which is consistent with LHX4 expression beginning after cell
170 cycle exit. We conclude that LHX4 is a relatively specific marker for post-mitotic cone
171 photoreceptors at early stages in the mouse retina but is also present in other cell
172 populations at later developmental stages.
173

174 **LHX4 is expressed in several developing and adult cell types.**

175 Previous reports indicated that LHX4 protein was present in adult cone BCs[19],
176 but little is known about its expression pattern in the developing mouse retina. We also
177 observed strong LHX4 labeling in the upper portion of the INL in cells positive for OTX2,
178 an adult marker for BCs (Supp. Fig. 3). Interestingly, we also noticed sparse, but
179 positive staining in the ONL. This LHX4 signal was located in some but not all RXRG+

180 cones with varying degrees of strength (Supp. Fig. 3) suggesting that a sub-population
181 of cones maintain LHX4 expression.

182 As noted above, reports[15,20] suggested that LHX4+ cells in the INL are likely
183 developing BCs. In the mouse, our data indicated that LHX4 is initially in developing
184 cones. However, at E17.5, RXRG cones no longer represent the near totality of LHX4+
185 cells. Since the peak of BC production is not seen until ~P3[21], the identity of the
186 remaining cells is unclear. While BCs can be produced at earlier timepoints (including
187 E17.5), we still sought to ascertain if only LHX4+ BCs were being produced at this time,
188 or if this LHX4 expression is present in another cell type. As rods are another OTX2+
189 cell type produced at this time, we aimed to see if these cells expressed LHX4 during
190 development. Newborn mice (P0) were injected with EdU to mark cells undergoing S-
191 phase at the time of injection and 3 days later the retinas were harvested. This length of
192 time was chosen to allow some newly produced rods enough time to produce NR2E3, a
193 well-known marker for rod fate[22]. It has been previously shown that NR2E3 is also
194 present in cone photoreceptors transiently[23–25], but, as determined previously [26], no
195 cones are produced at P0. Therefore, EdU and NR2E3 co-localization should reliably
196 mark cell types other than cones, likely rod photoreceptors. NR2E3 expression has not
197 been reported in BCs but given the similarities in molecular profiles between BCs and
198 photoreceptors, at this time we cannot exclude this possibility. We observed LHX4+
199 cells that are positive for NR2E3 and EdU at P0 (Fig 2 D). This indicates that while
200 LHX4 is present in developing and adult BCs and cones, it is also possible that is
201 transiently expressed in some rod photoreceptors.

202 **The LHX4::GFP transgenic line recapitulates the endogenous LHX4 expression**
203 **pattern**

204 We set out to establish if a reported LHX4::EGFP mouse transgenic line[14] could
205 be a reliable tracer for early developing cones. The LHX4::GFP line had robust
206 expression in the retina throughout development (Fig 3 A) and GFP signal was located
207 in the apical portion of the retina, where photoreceptors are located during embryonic
208 development. EGFP could be detected during embryonic and postnatal development
209 and also had stable expression in the adult retina (Supp. Fig. 4). This line has been
210 used previously for transcriptomic and electrophysiological analyses[27,28], because it
211 shows reliable activity in cone BCs. Interestingly, a subpopulation of cones is labeled by
212 this line in the adult and this selective expression resembles the sparse presence of
213 LHX4 protein in the adult mouse retina. To verify if this sparse labeling was due to LHX4
214 presence in exclusively S or L/M types of cones, which are located in contrasting
215 gradients in the adult retina, we assessed EGFP expression in the dorsal and ventral
216 retina[29]. We detected no difference in EGFP expression between these two areas and
217 was present in a subset of Cone-Arrestin+ cells regardless of Dorsal-Ventral position
218 (Supp. Fig 5).

219 We were unable to use the rabbit anti-LHX4 antibody to validate the LHX4::GFP
220 expression as we encountered abnormal LHX4 immunoreactivity; not nuclear, as
221 expected, but cytoplasmic and precisely overlapped with EGFP expression. Noted in
222 the strain description, EGFP was placed after the ATG of LHX4, replacing the LHX4
223 coding region so as to function as a tracer. We amplified and sequenced the coding
224 region downstream of the LHX4 ATG using the reported 5’ UTR primer for genotyping
225 and, surprisingly, found it contained a 70 base pair section of 5' endogenous LHX4

226 coding sequence immediately followed by a multiple cloning site and the EGFP coding
227 sequence, all in frame to produce a LHX4::GFP fusion protein. Thus, the first 34 amino
228 acids of this fusion protein are from the sequence of LHX4 and the multiple cloning site
229 (Supp. Fig. 6 A).

230 To test if this portion of LHX4 contained an epitope for the LHX4 antibody, we
231 amplified the coding sequence from the LHX4::GFP genomic DNA, cloned it into a
232 misexpression vector, and electroporated it into the chick retina (Supp. Fig. 6 B-D).
233 Control retinas had no detectable EGFP and normal staining of LHX4 with the rabbit
234 antibody (Supp. Fig. 6 B). In contrast, cells electroporated with the CAG::5pLHX4GFP
235 construct were strongly immunoreactive to the LHX4 antibody in a pattern that
236 completely overlapped that of the cytoplasmic EGFP signal (Supp. Fig. 6 C). The rabbit-
237 LHX4 still detected endogenous LHX4, as we imaged the same retinas at the edge of
238 the electroporation patch with higher gain and observed endogenous protein expression
239 alongside overexposed electroporation signal (Supp. Fig. 6 D).

240 Therefore, to assess if EGFP recapitulates LHX4 expression patterns and its
241 activity in early cones, we resorted to examination of RXRG expression in EGFP+ cells
242 at two relevant timepoints of embryonic development. At E14.5, the EGFP reporter
243 faithfully recapitulated LHX4 expression as 95.3+-1% of all EGFP+ cells were positive
244 for RXRG (Fig 3 C). Likewise, at E17.5 only 65.7+-1.3% of EGFP cells were RXRG+.
245 As the LHX4 immunodetection showed, nearly all RXRG+ cones are positive for LHX4
246 at both E14.5 and E17.5. This was also true in the LHX4::EGFP+ population, where
247 94.8+-0.5% and 99.5+-0.5% of all RXRG cells, respectively, were positive for the EGFP
248 reporter (Fig 3 D). Taken together, this data suggests that the LHX4 BAC-EGFP reliably

249 recapitulates LHX4 protein expression and is a dependable marker for cone
250 photoreceptors during early retinal development.

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

252 Having verified that the LHX4::GFP reporter is a marker for cone photoreceptors
253 in the early stages of mouse retinal development, we took advantage of this system to
254 examine the molecular profile of early cone photoreceptors using single cell RNA
255 sequencing. LHX4::GFP E14.5 littermates were screened for GFP expression and
256 positive retinas were pooled and dissociated in preparation for FACS sorting (Fig 4 A).
257 GFP+ and GFP- cells were collected and approximately 4000 cells were sequenced per
258 condition using the 10x Chromium platform.

259 Using the Seurat program[30], we performed an unsupervised clustering analysis
260 on the combined cell transcriptomes from both conditions that passed standard 10x QC
261 (GFP+: 3728, GFP-:4444). TSNE projections revealed that the GFP+ and GFP-
262 populations clearly segregated, consistent with LHX4::GFP cells being a molecularly
263 distinct population (Fig 4 B). The analysis separated the cells into 9 distinct clusters,
264 which were identified by established markers and bore hallmarks of known cell classes
265 in the developing retina (Fig 4 C-D; Supplementary File 1).

266 Two cell clusters were assigned as multipotent RPCs based on expression of
267 multipotent RPC markers such as VSX2, HES1, NR2E1, among others, as well as a
268 large percentage of cycling cells [31] (Fig 4 E-F). One other cluster was predominantly
269 comprised of cycling cells, with markers such as OLIG2, NEUROG2, OTX2 and
270 ATOH7, which identify these as likely neurogenic RPCs[32,33] with limited mitotic potential.
271 A second cluster had a smaller fraction of cycling cells, high ATOH7 levels and

272 expression of POU4F2, but no OTX2 or OLIG2, which we assigned to RGC/AC RPCs
273 and Precursors as they exhibited markers of likely differentiation to RGCs or ACs.
274 Of the cell clusters assigned as mostly post-mitotic, one was assigned as RGCs
275 as it displayed known markers RBPMS and GAP43[34,35]. Another assigned to AC/HCs,
276 exhibiting markers for both these fates, like TFAP2B and PTF1A[36,37]. As previously
277 reported[33], while ACs and HCs are distinct fates, it is difficult to separate them by their
278 transcriptomes in early development and without an appropriate amount of sequenced
279 cells for proper resolution.
280 Three clusters were assigned as cone photoreceptors. As expected from our
281 previous data, these consisted of the near totality of GFP+ cells and displayed markers
282 of cone photoreceptors: THRB, RXRG, GNGT2 and GNB3 (Fig 4 C-D), as well as
283 LHX4. We did detect some sparse Nrl-positive cells within this population, possibly
284 reflecting some activity of the LHX4 reporter in rod photoreceptors. The cell cycle phase
285 of these clusters is consistent with our previous data suggesting LHX4 is in post-mitotic
286 cones at E14.5. As a result of the sorting strategy, cone representation was high which
287 allowed the clustering analysis to resolve cone subpopulations. All cone clusters had
288 high levels of established markers, but two subpopulations differed in expression of
289 genes. The highest differential marker for one of the populations was FABP7, a
290 previously characterized marker for cones in the adult murine retina[38] with reported
291 expression in the developing retina[39] but not at this early stage. Meanwhile, a second
292 subpopulation had increased levels of solute-carrier genes like SLC7A3 and SLC7A5
293 (Fig 4 C). Our analysis indicates that we successfully sorted and sequenced a

294 developing E14.5 retina and identified its cell populations while enriching for cone
295 photoreceptors.

296 **Early cone marker identification in LHX4::GFP cells**

297 We sought to use this dataset to find new markers for early cone photoreceptors.
298 Using Seurat, we performed a differential expression analysis, comparing the GFP+
299 population, which we established as cones, with the GFP- population, which should
300 encompass the other concurrent populations in the developing retina. Additionally, for
301 visualization purposes and as a proxy for average population levels of transcript
302 expression we calculated the expression of an average GFP+ and a GFP- cell and used
303 this in combination with the differential expression results. We identified over 898
304 significantly differentially expressed transcripts with enrichment in cones. As expected,
305 we identified known cone markers as highly enriched in the GFP+ population (Fig 5 A,
306 Supplementary File 2). We then looked for novel transcripts enriched in this population
307 (Fig 5 B). A heatmap for the top 100 cone-enriched transcripts with the transcript
308 expression of an average cell in every individual cluster is included in Fig. 5 C.

309 The above analysis identified genes enriched in early cone photoreceptors
310 compared to the other cell types present at that time. We first compared these results to
311 other studies that have reported cone transcriptome analyses in the mouse or human
312 (Supp Fig 7 A-B). Only one study has reported a transcriptome from early developing
313 cones, similar to this one[4], through the use of human fetal explants infected with cone
314 opsin promoter reporter viruses, determining early and late phases of gene expression
315 in labeled versus non-labeled cells. A number of genes are enriched in both datasets
316 (165 genes in early and 211 in late fetal cones), suggesting that the mouse is a potential

317 model for investigating the function of these genes in cone photoreceptors. (Supp. Fig
318 7, Supp. File 3 and 4).

319 Additional datasets of purified adult rods and cones from the mouse have been
320 used to identify differential transcripts between these two types of photoreceptors[7].
321 Comparison to this dataset identified that 251 of the genes in our dataset were enriched
322 in adult cones and 121 were enriched in adult rods (Supp. Fig 7, Supp. File 3 and 5).
323 Adult cone-enriched genes present in this dataset support the cell-specificity of known
324 early markers detected in our dataset, such as RXRG, as well as many genes not
325 explored in detail before, such as QSOX1 or LHX4.

326 We next were interested in determining whether the genes associated with early
327 cone genesis were negatively regulated by the rod photoreceptor factor Nrl. A prevalent
328 model for Nrl function is that it serves as a fate switch in photoreceptor precursors, with
329 Nrl-negative precursors becoming cones and Nrl-positive ones becoming rods[8,12,40]. In
330 Nrl mutants, the photoreceptors that normally would become rods are found to undergo
331 a morphological and gene expression change that could suggest a rod to cone fate
332 switch. However, it has been noted that these cells are morphologically distinguishable
333 from normal cones and, in addition, known early cone genes, such as Thrb and Rxrg,
334 are unaltered in newborn photoreceptors, which suggests that Nrl may not be a master
335 regulator of this fate choice [8,12,41]. As there are few known early cone genes, we sought
336 to identify other cone photoreceptor genes, in addition to Thrb and Rxrg, that are
337 regulated independently of Nrl. We first identified those genes in our dataset that were
338 dysregulated either positively or negatively in Nrl knockout photoreceptors. The Nrl
339 dataset used contained multiple isoforms of genes, but to apply a stringent criterion, any

340 isoform that showed dysregulation at either P2 or P28 in Nrl mutants led to that gene
341 being removed, leaving a total of 259 Nrl-independent genes (Fig 6 A). Interestingly,
342 many of the cone-enriched genes present at E14.5 are only altered in Nrl mutants at
343 P28 and not also at P2 when a large number of rods would be generated (see
344 discussion).

345 As some of these genes likely represent pan-photoreceptor genes that wouldn't
346 be expected to be under the transcriptional control of Nrl, we identified those genes also
347 expressed in developing rods. As an identifiable rod cluster was not present in our
348 E14.5 dataset, we analyzed a recently generated single cell dataset from E18.5 whole
349 retinas[33]. The data from 2 replicates of E18 was extracted and we performed an
350 unbiased clustering analysis using Seurat[30]. (Supp Fig 8). The detected 9 clusters were
351 assigned to known cell types through marker expression. Within those, two clusters
352 were positive for CRX+ and, thus, determined to be likely photoreceptors. First, all DE
353 transcripts for photoreceptors at e18 were determined (Supp. Table 7, Supp. Fig 8) and
354 then these cells were re-analyzed for further sub-clustering. We were able to determine
355 a cluster with cone signature, rod signature and 2 precursor clusters. From those, we
356 determined all DE transcripts between photoreceptor clusters (Supp. Table 8, Supp Fig
357 8). Interestingly, many of the genes detected as enriched in early cone photoreceptors
358 in this study have been identified as differentially expressed across retinal pseudotime
359 by Clark, Stein O'Brien et al, 2018 (Supp. Table 9)

360 Removal of genes with expression in any of the non-cone clusters at e18,
361 produced 198 Nrl-independent cone-enriched genes (Fig 6 A-B, Supp. Table 10). To
362 provide some measure of confirmation of the cone/photoreceptor-associated expression

363 of these genes, we identified those genes with cone-enriched expression or specifically
364 enriched in cone cluster 2 that were also identified in the previously mentioned datasets
365 (Fig 6 C,D, Supp. Table 10). This left a set of genes with cone-enriched expression or
366 specifically enriched in cone cluster 2 that have only been identified to date in this study
367 as cone-associated (Fig 6 E,F).
368
369
370

371 **DISCUSSION**

372 Our current knowledge of the molecular and cellular events involved in cone
373 photoreceptor development is incomplete. One missing dataset has been a
374 comprehensive gene expression analysis of endogenously developing cones. Without
375 such knowledge, the rational design of strategies to induce cone formation and
376 adequate benchmarks to assess these cones, such as from stem cell cultures, are not
377 possible. Here we identify LHX4 as a novel and highly specific marker of cone
378 photoreceptors in the early stages of their development and the LHX4::EGFP reporter
379 mouse as a tool to detect and analyze these cells in the developing mouse retina.

380 The single cell transcriptome analysis allowed for an unprecedented molecular
381 examination of early cone photoreceptors. With the increased representation that the
382 LHX4::GFP mouse permitted for the purification of this relatively rare cell type, we were
383 able to identify sub-clusters of cones at this timepoint. The biological significance of
384 these sub-clusters, however, remains, to be determined. They could reflect temporal
385 differences in the differentiation process or spatial effects linked to the dorsal-ventral
386 position of cells, which is known to influence cone opsin expression at later timepoints.
387 One distinct cluster upregulates the SLC7A3 and SLC7A5 genes. SLC7A5 is a known
388 thyroid hormone transporter for T3/T4 states[42] that has been reported in human fetal
389 retinas and organoids[43,44]. Because these transcriptomic analyses were done in whole
390 retinas it wasn't clear if this expression was specific to any particular cell population, but
391 our data suggests that mouse cone photoreceptors specifically express these genes.
392 SLC7A3 has not been reported in the retina but likely plays a similar role as its
393 expression has been linked to T3 administration in the brain[45]. Thyroid hormone and its
394 targets are known to affect M-cone differentiation[43,46–49]. We found several other thyroid-

395 related genes throughout the retina, but only THRB was exclusive to cones at this time.
396 For example, there was almost no detectable amount of DIO2 in any cell but many more
397 counts of DIO3, although these were present in the multipotent RPC clusters. Another
398 observation of cone heterogeneity was in the expression of LHX4 protein and the
399 LHX4::GFP reporter in more mature retinal tissue. Whether this sustained expression of
400 LHX4 in subsets of cones is biologically relevant or is an epigenetic or temporal
401 phenomenon without a functional significance warrants further study.

402 There is a major interest in understanding the molecular differences between
403 cone and rod photoreceptors and how these differences arise during development. This
404 would provide insights into how these two classes of photoreceptors evolved as well as
405 inform the strategic development of methods to specifically produce these cell types.
406 The predominant photoreceptor determination framework is that cones and rods
407 developmentally diverge based solely on whether newborn photoreceptors begin to
408 express the Nrl transcription factor or not. The analysis performed here with multiple
409 previously published datasets and the current one suggests that endogenous newborn
410 cones have a unique molecular signature compared to the induced cones that form in
411 the Nrl knockout. There are caveats that must be considered. For one, the current study
412 used a massive enrichment of cones, which allowed for increased power to detect gene
413 expression differences between cones and other cells. Comparison to other datasets
414 that had less statistical power may have led to the false conclusion that a gene was not
415 differentially expressed. In addition, we compared the profile of embryonic cones to Nrl-
416 dysregulated genes at P2. There could be currently unidentified signaling factors that
417 are temporally different and influence gene expression that could account for these

418 differences. It is important to note that a large number of cones genes are in fact
419 dysregulated in the Nrl mutant. However, it is interesting that many of these genes
420 found in early cones, such as RXRG, are not dysregulated in early rods, but are only
421 significantly changed in much more mature rods. What the cause of this delay is and
422 whether it impacts the differentiation of these transformed cells to a more bona fide
423 cone photoreceptor signature is not known. Regardless, these differences warrant
424 further examination of the utility and validity of using photoreceptors from the Nrl
425 knockout model as a substitute for naturally formed cones.
426

427 **Materials & Methods**

428 **Animals**

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

437 **Genotyping**

438 Genotyping was performed as specified in the MMRRC strain description page
439 referenced (LHX4::GFP strain: Tg(Lhx4-EGFP)KN199Gsat/Mmucd,
440 RRID:MMRRC_030699-UCD).

441 **Cloning and DNA electroporation**

442 Misexpression plasmids were created by PCR amplifying the coding sequence
443 for each gene with added AgeI/NheI or EcoRI sites. They were subsequently cloned into
444 a CAG backbone, from which the GFP had been removed from Stagia3[47,50] and the
445 CAG promoter had been cloned upstream. For mouse LHX4 the primers were designed
446 against the annotated mRNAs in NCBI galgal5 RNA libraries with an added Kozak
447 sequence 5' (ACC). For 5pLHX4GFP, the 5' primer from genotyping for LHX4::GFP
448 mice was combined with a 3' primer to the coding region of GFP and the fusion gene
449 was amplified from genomic DNA extracted from LHX4::GFP+ mice. CAG::nucβgal was
450 developed by the Cepko lab (Harvard Medical School). To deliver the plasmids after
451 retinal dissection, ex vivo electroporation experiments were carried out as detailed in
452 Emerson and Cepko, 2011[50].

453 **Retinal cells dissociations and Florescence Activated Flow Sorting (FACS)**

454 Dissociation of retinal tissue and FACS were performed as in Buenaventura et al,
455 2018[3]

456 **Immunohistochemistry and EdU labeling**

457 All tissue processing and immunofluorescence experiments were performed as in
458 Emerson and Cepko, 2011 and Buenaventura et al, 2018. For E14.5 mouse retinas,

459 EdU labeling was performed by injecting pregnant dams with 150ul of 10mM EdU
460 resuspended in 1X PBS. EdU detection was performed with a Click-iT EdU Alexa Fluor
461 647 imaging kit (C10340, Invitrogen).

462 **Microscopy**

463 Confocal images were acquired using a Zeiss LSM880 confocal microscope
464 using ZEN Black 2015 2.1 SP2 software and images were converted into picture format
465 using the FIJI version of ImageJ[51]. Figures were assembled using Affinity Designer
466 vector editor. Images were adjusted uniformly with regards to brightness and contrast.

467 **10X single cell sample processing**

468 Live cells were collected after dissociation and FACS sorting for GFP signal[3].
469 Samples were extracted from 2 litters of E14.5 LHX4::GFP mice, pooled into one sorting
470 solution during dissociation. For each sample (GFP+ and GFP-), approx. 4000 cells
471 were individually lysed, and their RNA transcribed and sequenced by the Columbia
472 University Single Cell Analysis Core. Filtered count matrixes provided by 10X Cell
473 Ranger pipeline using mm10 genome were used for downstream analysis.

474 **Single-cell clustering analysis**

475 Unsupervised clustering analysis was performed using Seurat[30,52]. Single cell
476 transcriptomes from both LHX4::GFP E14.5 samples were analyzed in concert to
477 produce cluster and TSNE analysis. Only cells with over 200 genes detected and only
478 genes detected in >3 cells were used for initial loading of the cell matrices. Cell cycle
479 was scored and was regressed out as an unwanted source of variation according to
480 Seurat's guidelines [31]. For the LHX4-GFP dataset, cells were filtered for max. 3500 and

481 min. 250 number of genes, and <0.25 percent mitochondrial content following standard
482 guidelines for QC. nUMI, nGene were regressed out in addition to cell cycle. Using the
483 established jackstraw procedure[53] in Seurat, 30 PCs were chosen for clustering and
484 TSNE projections. The 'FindMarkers' function was used with default parameters, only
485 for positive markers, with the populations compared as described in the results section.
486 For Clark, Stein O'Brien et al, 2018, similar procedures were used with the following
487 changes: for the full dataset - 250<nGenes<3500, 0.075<percent mito, 39 PCs and
488 replicate effects were regressed out; for the photoreceptor subclustering – same as full
489 but 6 PCs, data was extracted post-analysis of the full dataset.

490 **Data comparisons**

491 For the comparison of retinal datasets, differential expression output data was
492 obtained from each report[4,7,54]. The data was cross-referenced to the genes obtained in
493 the analysis from this paper and are detailed in Supp. Files 3, 4, 5 and 6. In the case of
494 Kim et al, 2016, genes are reported with PPDE as opposed to p-val-FDR or q-val as
495 probability of differential expression, and were filtered to show only those with
496 PPDE>0.95

497 **Data availability**

498 The data will be deposited in NCBI's Gene Expression Omnibus[55] and will be
499 accessible through GEO Series accession number GSExxxxxx.

500 **Quantitative analysis of markers in retina sections**

501 One z-stack from the central portion of each retina of four biological replicates
502 was counted and used for mean and SEM calculations in each condition. Cells were
503 counted using Cell Counter plugin in ImageJ or Fiji[51,56].

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.

517 **References**

518

519 1. Mitchell, J. & Bradley, C. Quality of life in age-related macular degeneration: a
520 review of the literature. *Health Qual. Life Outcomes* **4**, 97 (2006).
521 2. Gamm, D. M. & Wong, R. Report on the National Eye Institute Audacious Goals
522 Initiative: Photoreceptor Regeneration and Integration Workshop. *Transl. Vis. Sci.*
523 *Technol.* **4**, (2015).
524 3. Buenaventura, D. F., Ghinia-Tegla, M. G. & Emerson, M. M. Fate-restricted retinal
525 progenitor cells adopt a molecular profile and spatial position distinct from
526 multipotent progenitor cells. *Dev. Biol.* **443**, 35–49 (2018).
527 4. Welby, E. *et al.* Isolation and Comparative Transcriptome Analysis of Human Fetal
528 and iPSC-Derived Cone Photoreceptor Cells. *Stem Cell Rep.*
529 doi:10.1016/j.stemcr.2017.10.018
530 5. Enright, J. M., Lawrence, K. A., Hadzic, T. & Corbo, J. C. Transcriptome profiling of
531 developing photoreceptor subtypes reveals candidate genes involved in avian
532 photoreceptor diversification. *J. Comp. Neurol.* **523**, 649–668 (2015).
533 6. Hughes, A. E. O., Enright, J. M., Myers, C. A., Shen, S. Q. & Corbo, J. C. Cell Type-
534 Specific Epigenomic Analysis Reveals a Uniquely Closed Chromatin Architecture in
535 Mouse Rod Photoreceptors. *Sci. Rep.* **7**, 43184 (2017).
536 7. Mo, A. *et al.* Epigenomic landscapes of retinal rods and cones. *eLife* **5**, e11613
537 (2016).
538 8. Mears, A. J. *et al.* Nrl is required for rod photoreceptor development. *Nat. Genet.* **29**,
539 447–452 (2001).

540 9. Hsiau, T. H.-C. *et al.* The cis-regulatory logic of the mammalian photoreceptor
541 transcriptional network. *PloS One* **2**, e643 (2007).
542 10. Yoshida, S. *et al.* Expression profiling of the developing and mature Nrl-/- mouse
543 retina: identification of retinal disease candidates and transcriptional regulatory
544 targets of Nrl. *Hum. Mol. Genet.* **13**, 1487–1503 (2004).
545 11. Akimoto, M. *et al.* Targeting of GFP to newborn rods by Nrl promoter and temporal
546 expression profiling of flow-sorted photoreceptors. *Proc. Natl. Acad. Sci. U. S. A.*
547 **103**, 3890–3895 (2006).
548 12. Daniele, L. L. *et al.* Cone-like morphological, molecular, and electrophysiological
549 features of the photoreceptors of the Nrl knockout mouse. *Invest. Ophthalmol. Vis.*
550 *Sci.* **46**, 2156–2167 (2005).
551 13. Swaroop, A., Kim, D. & Forrest, D. Transcriptional regulation of photoreceptor
552 development and homeostasis in the mammalian retina. *Nat. Rev. Neurosci.* **11**,
553 563–576 (2010).
554 14. Heintz, N. Gene expression nervous system atlas (GENSAT). *Nat. Neurosci.* **7**, 483
555 (2004).
556 15. Fischer, A. J., Foster, S., Scott, M. A. & Sherwood, P. Transient expression of LIM-
557 domain transcription factors is coincident with delayed maturation of photoreceptors
558 in the chicken retina. *J. Comp. Neurol.* **506**, 584–603 (2008).
559 16. Ochi, H. *et al.* Temporal expression of L-Maf and RaxL in developing chicken retina
560 are arranged into mosaic pattern. *Gene Expr. Patterns GEP* **4**, 489–494 (2004).
561 17. Benkhelifa, S. *et al.* Phosphorylation of MafA is essential for its transcriptional and
562 biological properties. *Mol. Cell. Biol.* **21**, 4441–4452 (2001).

563 18. Roberts, M. R., Hendrickson, A., McGuire, C. R. & Reh, T. A. Retinoid X receptor
564 (gamma) is necessary to establish the S-opsin gradient in cone photoreceptors of
565 the developing mouse retina. *Invest. Ophthalmol. Vis. Sci.* **46**, 2897–2904 (2005).
566 19. Balasubramanian, R., Bui, A., Ding, Q. & Gan, L. Expression of LIM-homeodomain
567 transcription factors in the developing and mature mouse retina. *Gene Expr.*
568 *Patterns GEP* **14**, 1–8 (2014).
569 20. Edqvist, P. H. D., Myers, S. M. & Hallböök, F. Early identification of retinal subtypes
570 in the developing, pre-laminated chick retina using the transcription factors Prox1,
571 Lim1, Ap2a, Pax6, Isl1, Isl2, Lim3 and Chx10. *Eur. J. Histochem.* **50**, 147–154
572 (2009).
573 21. Morrow, E. M., Chen, C.-M. A. & Cepko, C. L. Temporal order of bipolar cell genesis
574 in the neural retina. *Neural Develop.* **3**, 2 (2008).
575 22. Cheng, H. *et al.* Photoreceptor-specific nuclear receptor NR2E3 functions as a
576 transcriptional activator in rod photoreceptors. *Hum. Mol. Genet.* **13**, 1563–1575
577 (2004).
578 23. Chen, J., Rattner, A. & Nathans, J. The Rod Photoreceptor-Specific Nuclear
579 Receptor Nr2e3 Represses Transcription of Multiple Cone-Specific Genes. *J.*
580 *Neurosci.* **25**, 118–129 (2005).
581 24. Haider, N. B. *et al.* The transcription factor Nr2e3 functions in retinal progenitors to
582 suppress cone cell generation. *Vis. Neurosci.* **23**, 917–929 (2006).
583 25. Jean-Charles, N., Buenaventura, D. F. & Emerson, M. M. Identification and
584 characterization of early photoreceptor cis-regulatory elements and their relation to
585 Onecut1. *Neural Develop.* **13**, (2018).

586 26. Carter-Dawson, L. D. & Lavail, M. M. Rods and cones in the mouse retina. II.
587 Autoradiographic analysis of cell generation using tritiated thymidine. *J. Comp.*
588 *Neurol.* **188**, 263–272 (1979).
589 27. Siegert, S. *et al.* Transcriptional code and disease map for adult retinal cell types.
590 *Nat. Neurosci.* **15**, 487–495 (2012).
591 28. Szikra, T., Farrow, K. & Roska, B. Cone-mediated Circuit Switch Activates Lateral
592 Inhibition In A Retinal Ganglion Cell. *Invest. Ophthalmol. Vis. Sci.* **52**, 4569–4569
593 (2011).
594 29. Applebury, M. L. *et al.* The murine cone photoreceptor: a single cone type expresses
595 both S and M opsins with retinal spatial patterning. *Neuron* **27**, 513–523 (2000).
596 30. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction
597 of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
598 31. Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell
599 RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**,
600 155–160 (2015).
601 32. Cepko, C. Intrinsically different retinal progenitor cells produce specific types of
602 progeny. *Nat. Rev. Neurosci.* **15**, 615–627 (2014).
603 33. Clark, B. *et al.* Comprehensive analysis of retinal development at single cell
604 resolution identifies NFI factors as essential for mitotic exit and specification of late-
605 born cells. *bioRxiv* (2018). doi:10.1101/378950
606 34. Rodriguez, A. R., Müller, L. P. de S. & Brecha, N. C. The RNA binding protein
607 RBPMS is a selective marker of ganglion cells in the mammalian retina. *J. Comp.*
608 *Neurol.* **522**, 1411–1443 (2014).

609 35. Reh, T. A., Tetzlaff, W., Ertlmaier, A. & Zwiers, H. Developmental study of the
610 expression of B50/GAP-43 in rat retina. *J. Neurobiol.* **24**, 949–958 (1993).
611 36. Bassett, E. A. *et al.* Overlapping expression patterns and redundant roles for AP-2
612 transcription factors in the developing mammalian retina. *Dev. Dyn. Off. Publ. Am.*
613 *Assoc. Anat.* **241**, 814–829 (2012).
614 37. Fujitani, Y. *et al.* Ptf1a determines horizontal and amacrine cell fates during mouse
615 retinal development. *Dev. Camb. Engl.* **133**, 4439–4450 (2006).
616 38. Su, X. *et al.* Characterization of Fatty Acid Binding Protein 7 (FABP7) in the Murine
617 Retina. *Invest. Ophthalmol. Vis. Sci.* **57**, 3397–3408 (2016).
618 39. Blackshaw, S. *et al.* Genomic analysis of mouse retinal development. *PLoS Biol.* **2**,
619 E247 (2004).
620 40. Ng, L. *et al.* Two Transcription Factors Can Direct Three Photoreceptor Outcomes
621 from Rod Precursor Cells in Mouse Retinal Development. *J. Neurosci.* **31**, 11118–
622 11125 (2011).
623 41. Emerson, M. M., Surzenko, N., Goetz, J., Trimarchi, J. M. & Cepko, C. Otx2 and
624 Onecut1 Promote the Fates of Cone Photoreceptors and Horizontal Cells and
625 Repress Rod Photoreceptors. *Dev. Cell* **26**, 59–72 (2013).
626 42. Scalise, M., Galluccio, M., Console, L., Pochini, L. & Indiveri, C. The Human
627 SLC7A5 (LAT1): The Intriguing Histidine/Large Neutral Amino Acid Transporter and
628 Its Relevance to Human Health. *Front. Chem.* **6**, (2018).
629 43. Eldred, K. C. *et al.* Thyroid hormone signaling specifies cone subtypes in human
630 retinal organoids. *Science* **362**, (2018).

631 44. Hoshino, A. *et al.* Molecular Anatomy of the Developing Human Retina. *Dev. Cell*
632 **43**, 763-779.e4 (2017).
633 45. Grijota-Martínez, C., Díez, D., Morreale de Escobar, G., Bernal, J. & Morte, B. Lack
634 of action of exogenously administered T3 on the fetal rat brain despite expression of
635 the monocarboxylate transporter 8. *Endocrinology* **152**, 1713–1721 (2011).
636 46. Applebury, M. L. *et al.* Transient expression of thyroid hormone nuclear receptor
637 TRbeta2 sets S opsin patterning during cone photoreceptor genesis. *Dev. Dyn. Off.*
638 *Publ. Am. Assoc. Anat.* **236**, 1203–1212 (2007).
639 47. Billings, N. A., Emerson, M. M. & Cepko, C. L. Analysis of thyroid response element
640 activity during retinal development. *PloS One* **5**, e13739 (2010).
641 48. Ng, L. *et al.* A thyroid hormone receptor that is required for the development of
642 green cone photoreceptors. *Nat. Genet.* **27**, 94–98 (2001).
643 49. Ng, L., Ma, M., Curran, T. & Forrest, D. Developmental expression of thyroid
644 hormone receptor beta2 protein in cone photoreceptors in the mouse. *Neuroreport*
645 **20**, 627–631 (2009).
646 50. Emerson, M. M. & Cepko, C. L. Identification of a retina-specific Otx2 enhancer
647 element active in immature developing photoreceptors. *Dev. Biol.* **360**, 241–255
648 (2011).
649 51. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat.*
650 *Methods* **9**, 676–682 (2012).
651 52. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell
652 transcriptomic data across different conditions, technologies, and species. *Nat.*
653 *Biotechnol.* (2018). doi:10.1038/nbt.4096

654 53. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual
655 Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
656 54. Kim, J.-W. *et al.* NRL-Regulated Transcriptome Dynamics of Developing Rod
657 Photoreceptors. *Cell Rep.* **17**, 2460–2473 (2016).
658 55. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene
659 expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210
660 (2002).
661 56. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of
662 image analysis. *Nat. Methods* **9**, 671–675 (2012).
663
664
665

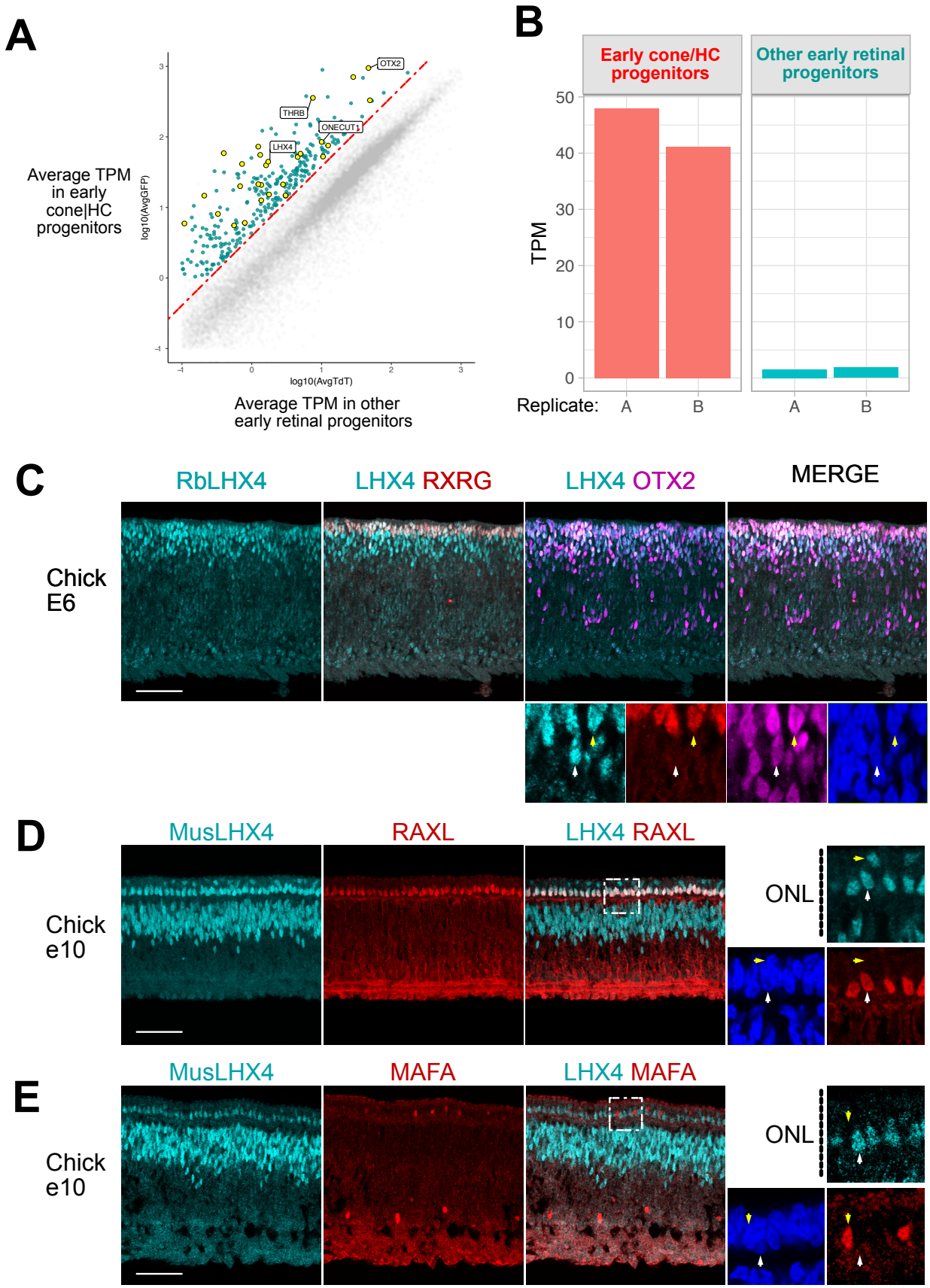


Figure 1 - **LHX4 is present in early cone photoreceptors during chick retinal development.** Large panels are maximum intensity projections Z-stacks and small panels are single planes of the same Z-stacks

(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

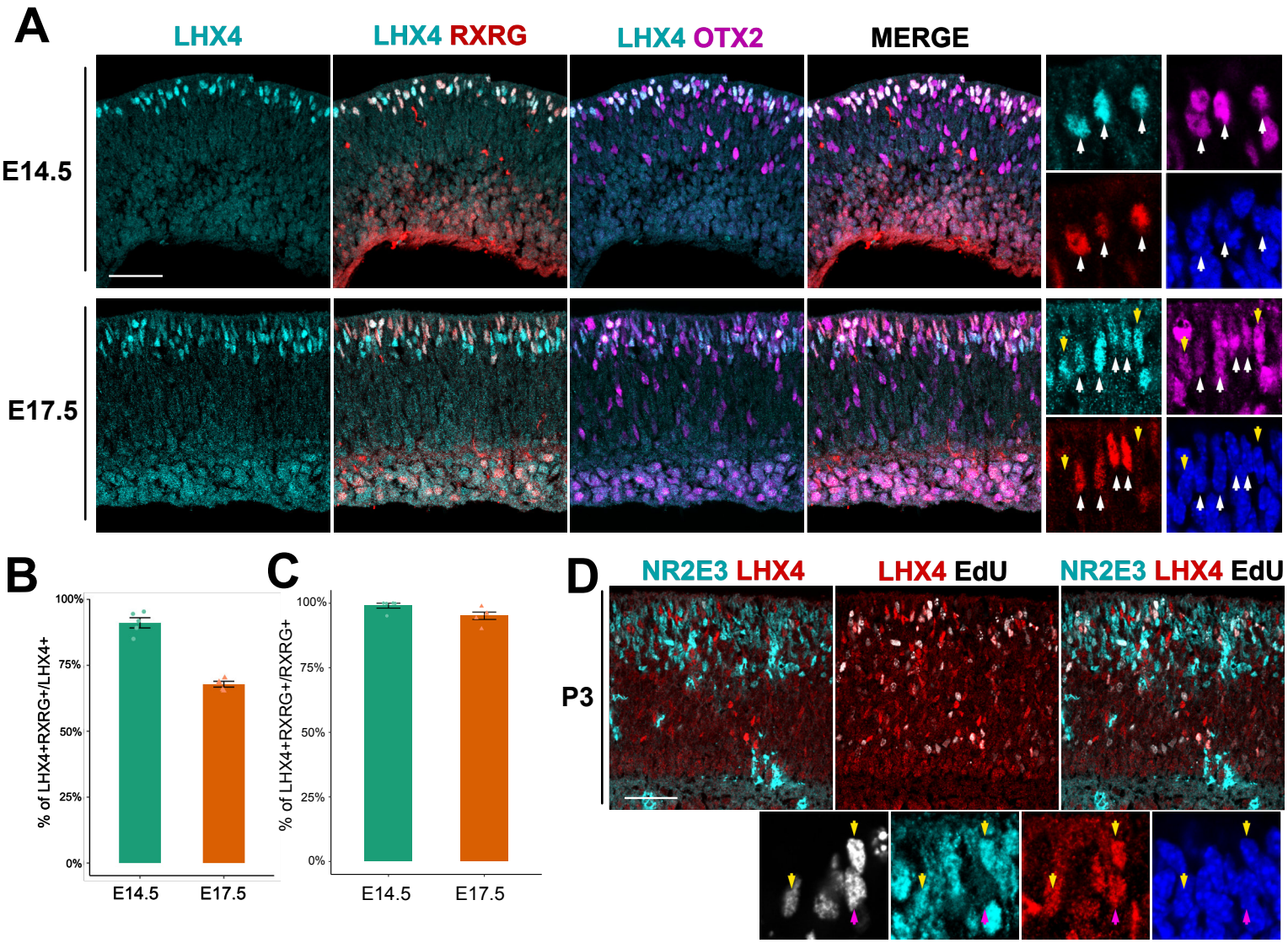


Figure 2 - **LHX4 protein is present in early photoreceptors precursors during mouse retinal development.** Large panels are maximum intensity projections Z-stacks and small panels are single planes of the same Z-stacks

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

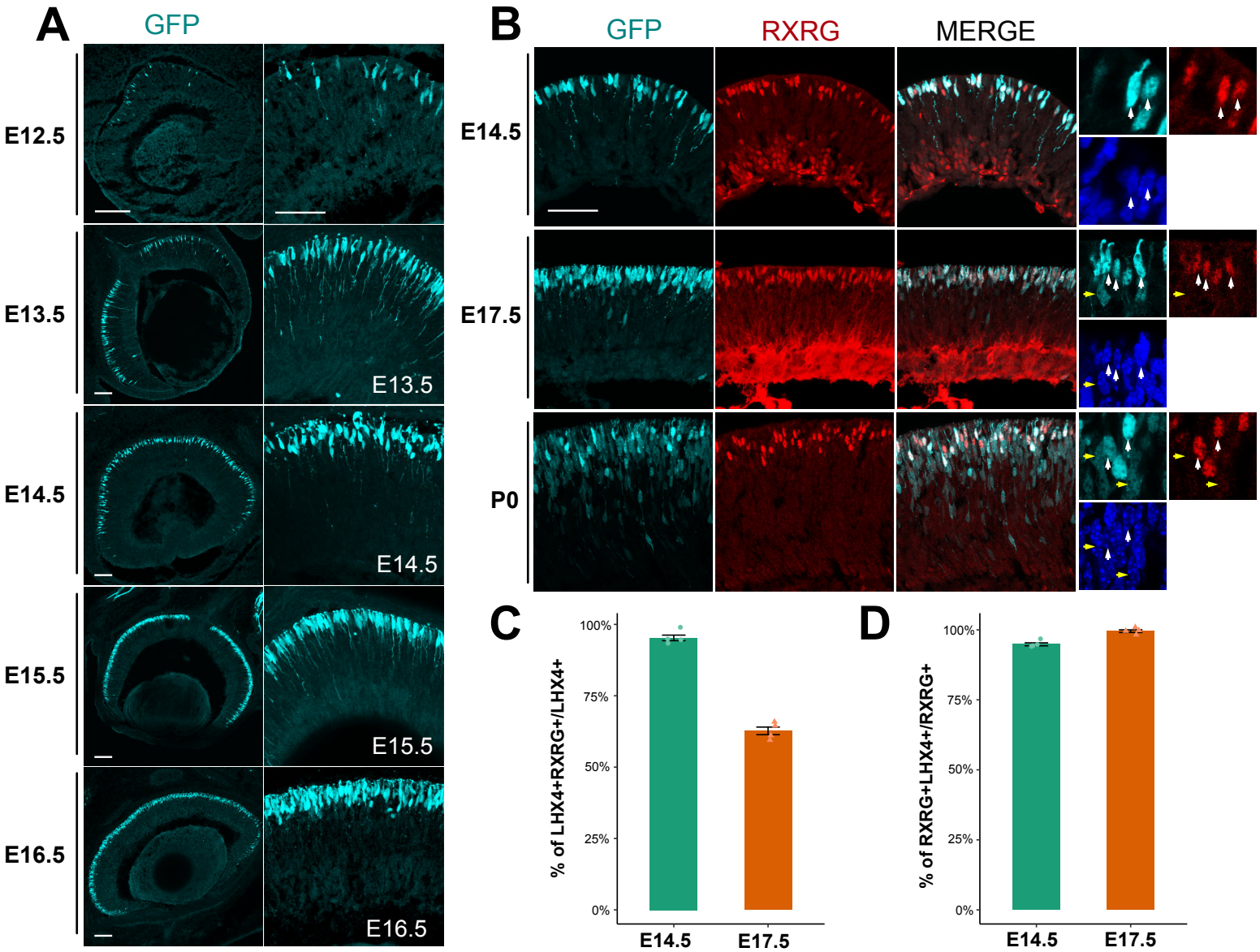


Figure 3 - **LHX4 BAC-EGFP is a reliable reporter of photoreceptors during retinal development.** Large panels are maximum intensity projections Z-stacks and small panels are single planes of the same Z-stacks. Scale bar represents 50 µm

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.

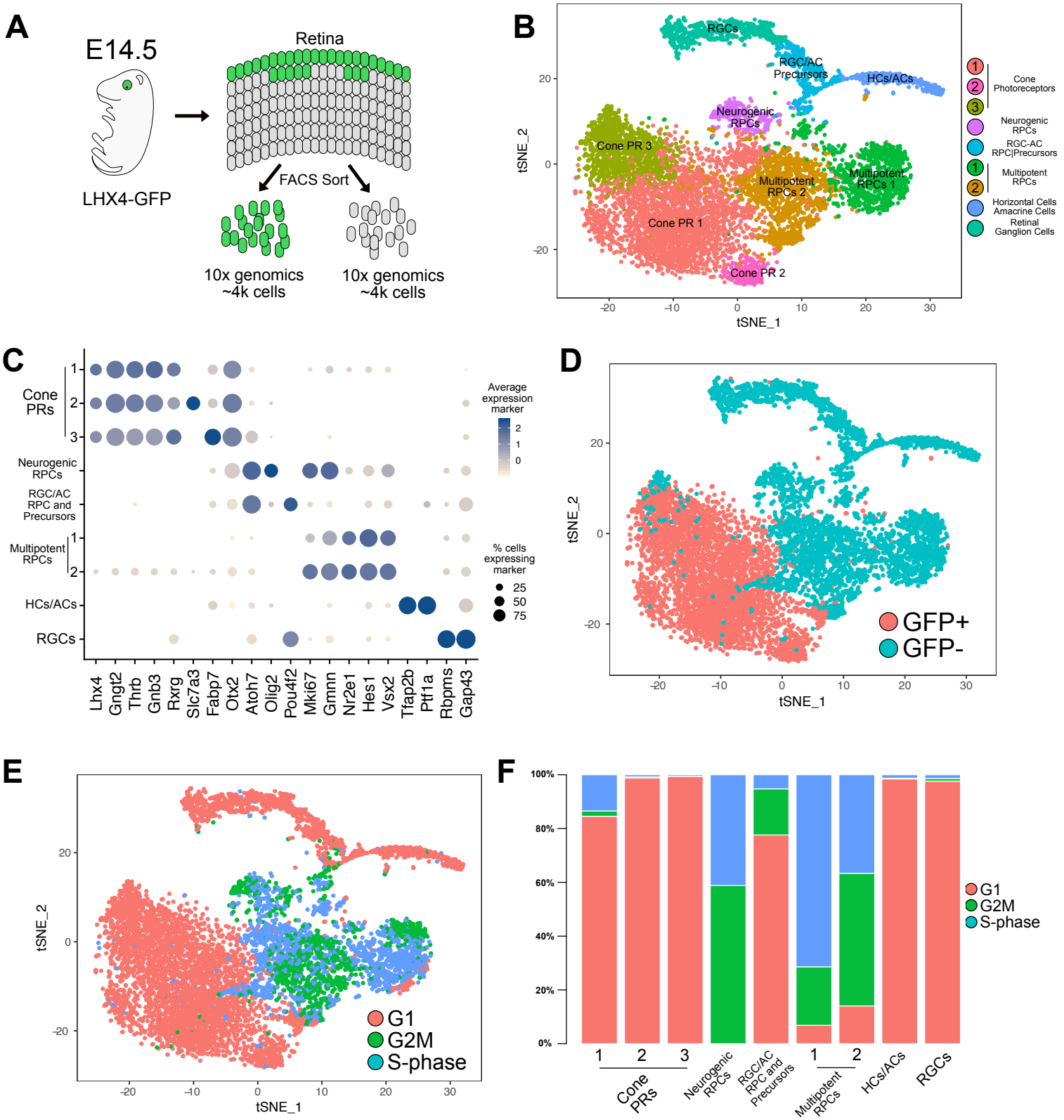


Figure 4 - TSNE - **Single cell sequencing of LHX4-GFP E14.5 developing retina.**

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

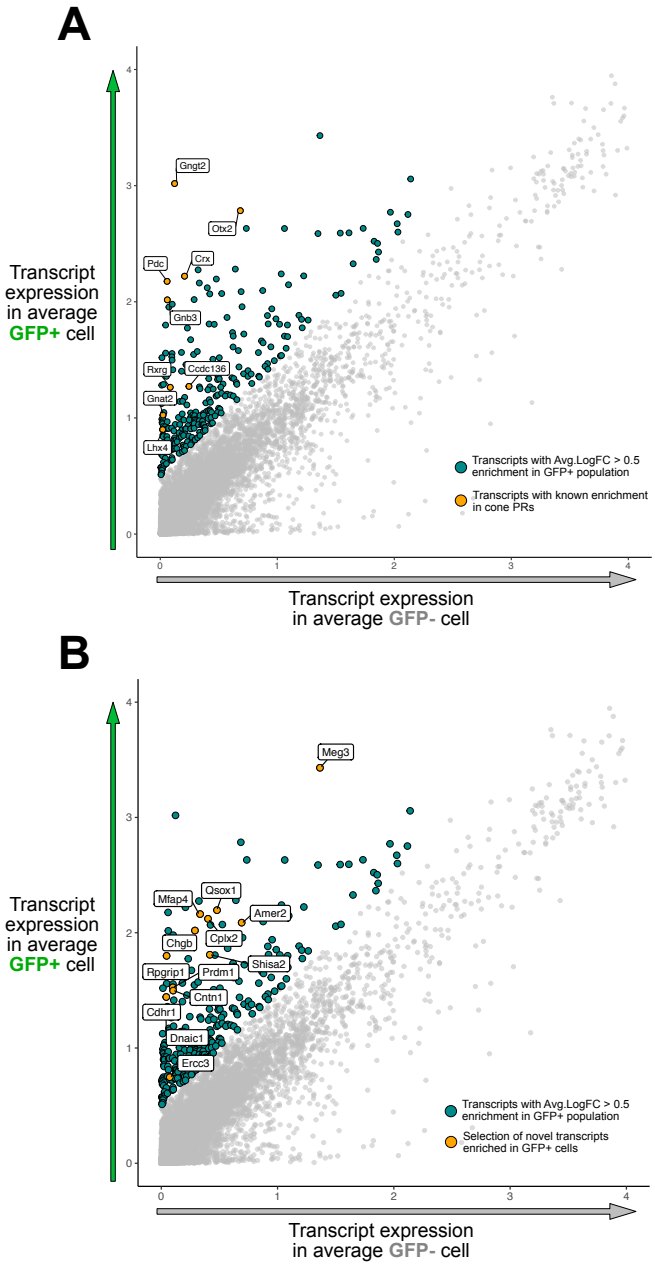
(B) tSNE plot of all GFP+ and GFP- cells collected, analyzed in tandem, and displaying the results of unsupervised cluster analysis with the assigned cell class according to their molecular signature.

(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.



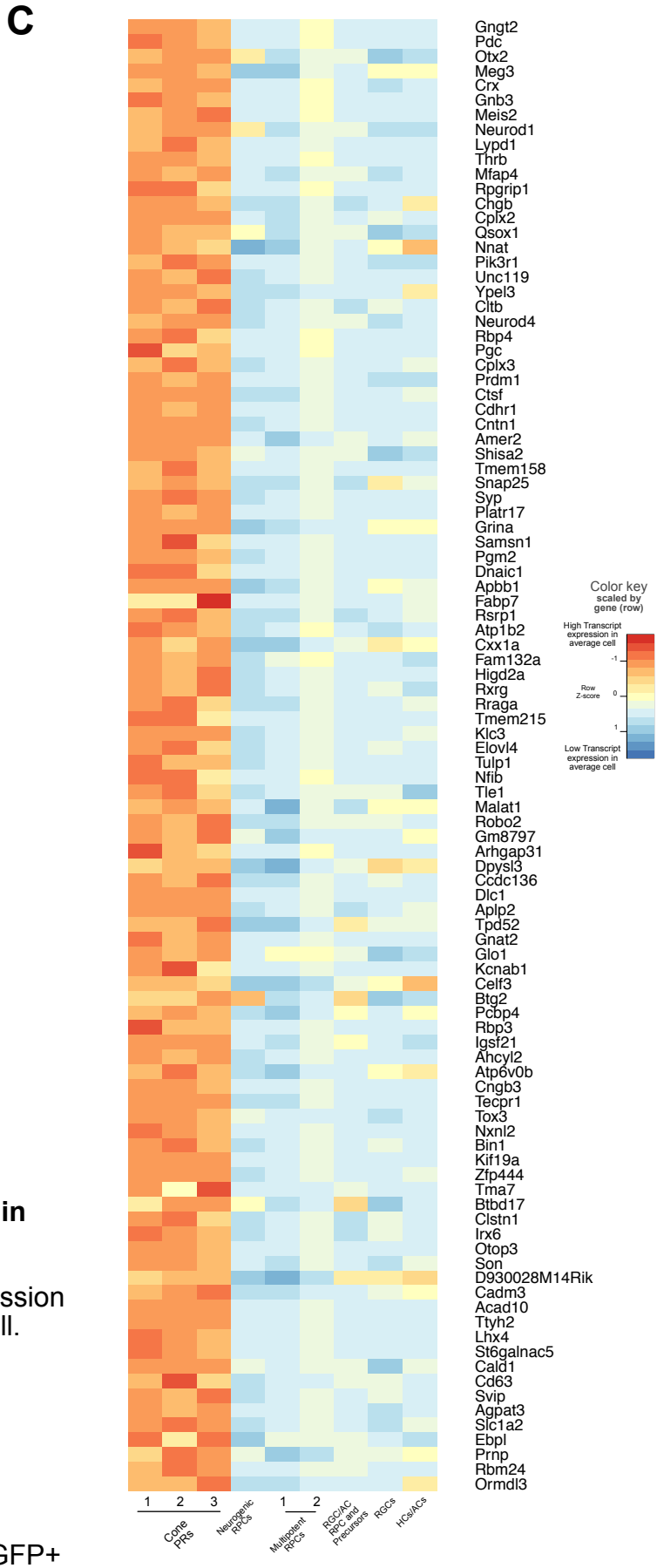


Figure 5 - **Identification of known and novel transcripts enriched in cone photoreceptors in LHX4-GFP+ cells.**

(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.

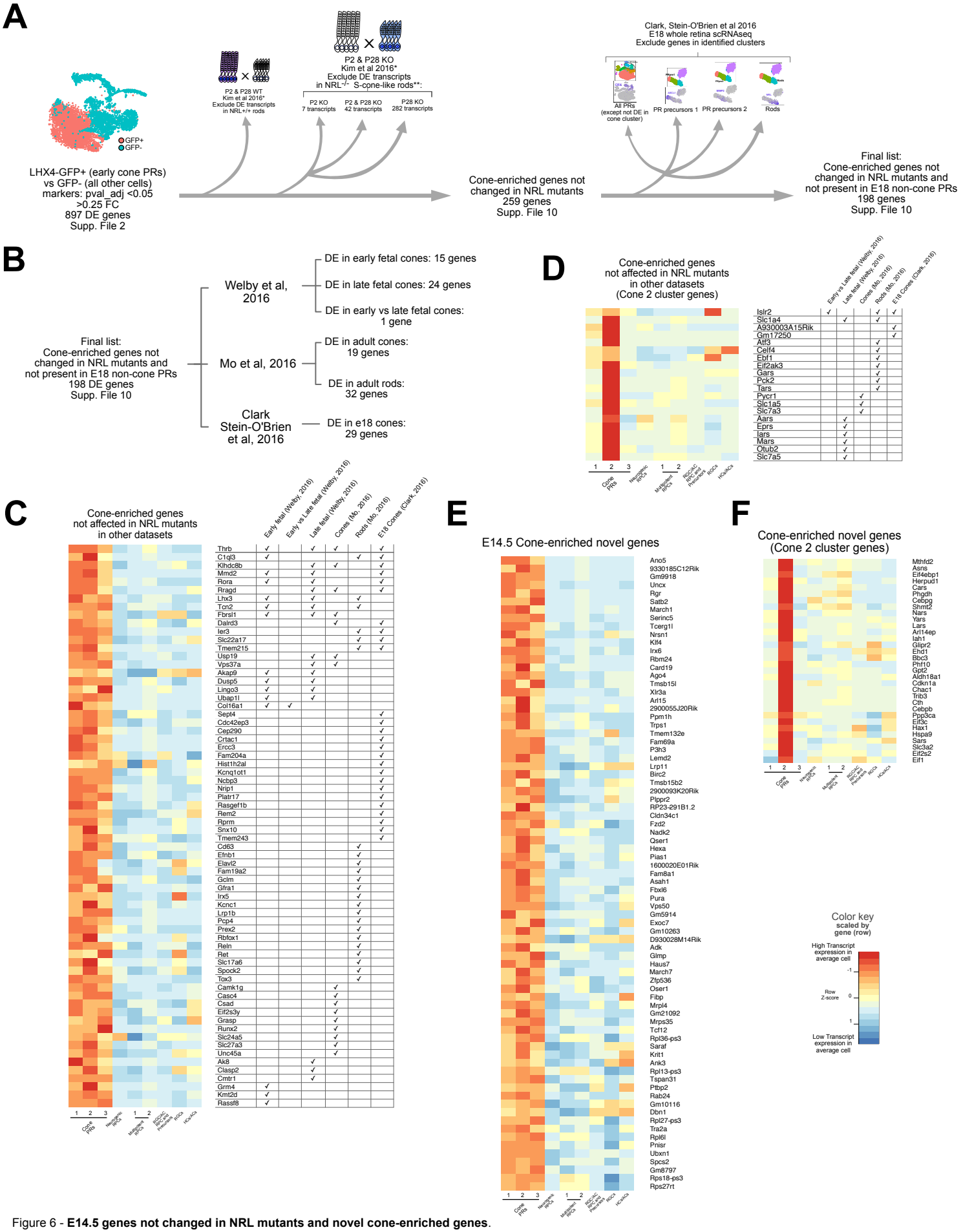


Figure 6 - **E14.5 genes not changed in NRL mutants and novel cone-enriched genes**.

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.