Combined analysis of single cell RNA-Seq and ATAC-Seq data reveals putative regulatory toggles operating in native and iPS-derived retina.

Anouk Georges[1,2#], Haruko Takeda[3#], Arnaud Lavergne[4], Michiko Mandai[5], Fanny Lepiemme[1], Latifa Karim[4], Loic Demeulenaere[3], Michael Schyns[6], Laurent Nguyen[1], Jean-Marie Rakic[2], Masayo Takahashi[7$] & Michel Georges[3$].

[1]GIGA Stem Cells, GIGA Institute, University of Liège, Belgium. [2]Department of Ophthalmology, Faculty of Medicine and CHU University Hospital, University of Liège, Belgium. [3]GIGA Medical Genomics, GIGA Institute, University of Liège, Belgium. [4]GIGA Genomics Platform, GIGA Institute, University of Liège, Belgium. [5]Laboratory for Retinal Regeneration, Center for Developmental Biology, RIKEN, Japan. [6]Digital Business, HEC Management School, University of Liège, Belgium. [7]Senior Visiting Scientist, Laboratory for Retinal Regeneration, Center for Biosystems Dynamics Research, RIKEN, Japan.

[#]Co-first authors. [$]Co-senior author. Correspondence: michel.georges@uliege.be

Abstract

Background: It has recently become possible to recapitulate retinal development from induced pluripotent stem cells, opening new investigative and therapeutic opportunities. Single cell RNA sequencing allows comparison of transcriptome unfolding during in vivo and in vitro development at single cell resolution, which can be integrated with information about accessible regulatory elements identified by ATAC-Seq.

Results: We report the generation and analysis of single-cell RNA-Seq data (> 38,000 cells) from native and iPSC-derived murine retina at four matched developmental stages spanning the emergence of the major retinal cell types. We combine information from temporal sampling, visualization of 3D UMAP manifolds, and RNA velocity to show that iPSC-derived 3D retinal aggregates broadly recapitulate the native developmental trajectories with evidence supporting re-specification from amacrine cells to horizontal and photoreceptor precursor cells, as well as a direct differentiation of Tbr1$^+$ retinal ganglion cells from neuro-epithelium cells. We show relaxation of spatial and temporal transcriptome control, premature emergence and dominance of photoreceptor precursor cells, and susceptibility of dynamically regulated pathways and transcription factors to culture conditions in iPSC-derived retina. We

generate bulk ATAC-Seq data for native and iPSC-derived murine retina identifying ~125,000 peaks. We combine single-cell RNA-Seq with ATAC-Seq information and obtain evidence that approximately halve the transcription factors that are dynamically regulated during retinal development may act as repressors rather than activators. We propose that sets of activators and repressors with cell-type specific expression control "regulatory toggles" that lock cells in distinct transcriptome states underlying differentiation, with subtle but noteworthy differences between native and iPSC-derived retina.

Conclusions: Combined analysis of single-cell RNA-Seq and ATAC-Seq information has refined the comparison of native and iPS-derived retinal development.

Keywords: retina, iPSC, single cell RNA-Seq, ATAC-Seq

Introduction

It has recently become possible to recapitulate retinal development from induced pluripotent stem cells (iPSCs) in human and mice [1,2]. This has opened new avenues to explore the molecular mechanisms underlying developmental competence, commitment and differentiation for each of the major cell types during retinal neurogenesis. It offers hope to improve therapies for retinal degenerative diseases which afflict hundreds of million people in the US and Europe alone where they account for approximately 50% of all cases of blindness. Stem cells derived from patient-specific somatic cells also offer unique opportunities to study and model the effects of gene defects on human retinal development in vitro and to test small molecules or other therapies for the corresponding disorders [3].

How faithfully iPSC-derived 3D retinal aggregates recapitulate specific developmental programs has typically been assessed by monitoring the expression of limited numbers of cell-type specific markers and examining the spatial patterning of the corresponding groups of cells [4]. Interrogating the expression of a handful of marker genes/proteins does not fully inform about the proper temporal and spatial execution of the epigenetic program, nor does it inform about the presence of aberrant cell types. Single cell RNA sequencing (scRNA-Seq) now enables the profiling of samples of the transcriptome (typically between 3% and 15% of mRNAs present in a cell depending on the methodology) of individual cells. This permits the clustering of cells based on the similarity of their transcriptome and the identification of cellular subtypes including some that may not have been recognized before [5]. It allows to refine developmental trajectories by identifying cells occupying intermediate states

connecting clusters in multidimensional expression space [6] and by predicting the developmental orientation taken by individual cells based on measured deviations from the steady-state ratio between spliced and unspliced RNA molecules ("RNA velocity") [7]. Genes that are defining cellular sub-types can be pinpointed by differential expression analysis between clusters [8], while genes that drive the differentiation process may be identified by searching for gene sets that are dynamically regulated across real and/or pseudo-time [9]. Recently, scRNA-Seq has been used to compare transcriptome dynamics during native and embryonic stem cells (ESC)- or iPSC-derived retinal development [10-12]. This has revealed comparable cellular composition at equivalent ages with, however, some differences in gene expression of particular cell types as well as structural differences of inner retinal lamination that seems disrupted in advanced organoid stages compared with fetal retina [11].

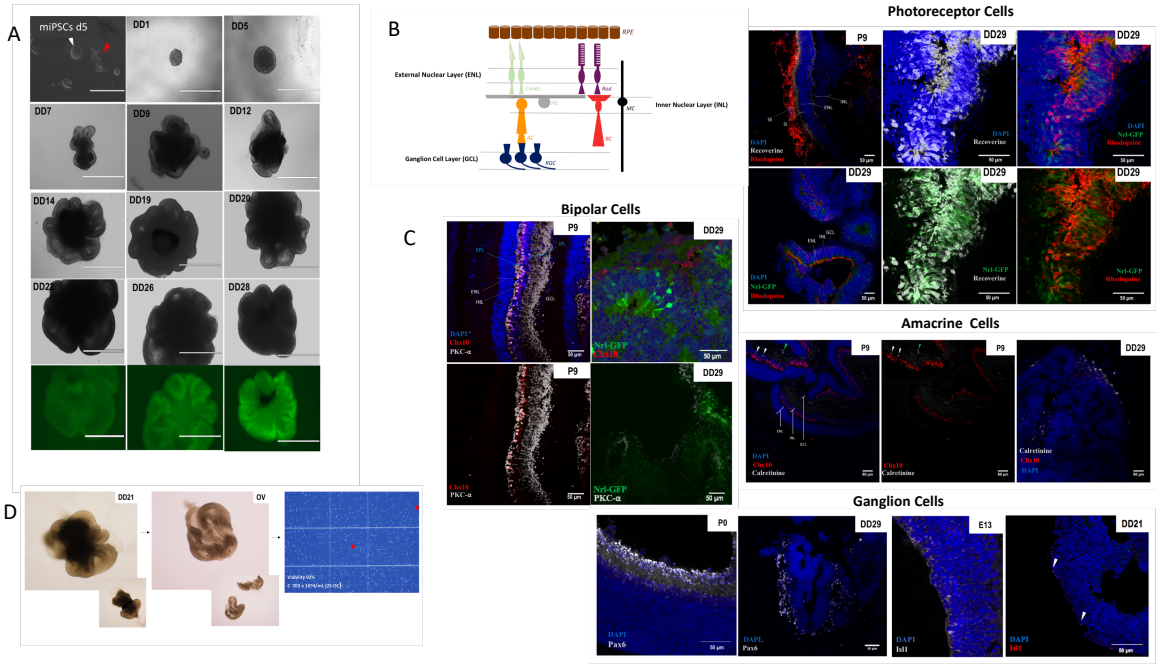
Here we report the generation and use of scRNA-Seq data collected at four matched stages of native and iPSC-derived retinal development in the mouse to study the dynamics of the transcriptome and compare it between the two systems. By integrating scRNA-Seq data with bulk ATAC-seq data (which identifies active gene regulatory elements by virtue of local chromatin openness [13] collected at three of the four above-mentioned developmental stages, we provide evidence for the operation of transcription factor (TF)-based regulatory toggles that lock the transcriptome of distinct cellular sub-types in both native and iPSC-derived retina thereby underpinning the different cellular identities.

Results

Joint analysis of scRNA-Seq data from native retina and iPSC-derived retinal aggregates highlights canonical developmental trajectories while supporting the occurrence of re-specification and the unconventional origin of Tbr1$^+$ RGC.

To improve the comparison of the developmental trajectories of native retina (NaR) and iPSC-derived 3D retinal aggregates (3D-RA), we performed scRNA-Seq of murine NaR and 3D-RA at four matched stages of development: embryonic day (E)13 vs differentiation day (DD)13, postnatal day (P)0 vs DD21, P5 vs DD25 and P9 vs DD29 [14]. NaR were dissected from two to 11 C57Bl6 mice (of both sexes) per stage. Mouse 3D-RA were generated from the iPSC Nrl-GFP line [15] following [1] and [16] (Fig. 1A). Immunolabeling revealed the presence of all major retinal cell types (Fig. 1B&C). Optic vesicle-like structures (OV) were manually dissected from 3D-RA (Fig. 1D). Cells from NaR and OV were dissociated and subjected to droplet-based

97 scRNA-Seq using a 10X Genomics Chromium platform (see Methods). We obtained sequence
98 information for 21,249 cells from NaR and 16,842 cells from 3D-RA, distributed evenly
99 amongst developmental stages. We generated an average of 74,808 reads per cell,
100 corresponding to 5,940 unique molecular identifiers (UMIs) and 2,471 genes per cell (Suppl.
101 Table 1).
102

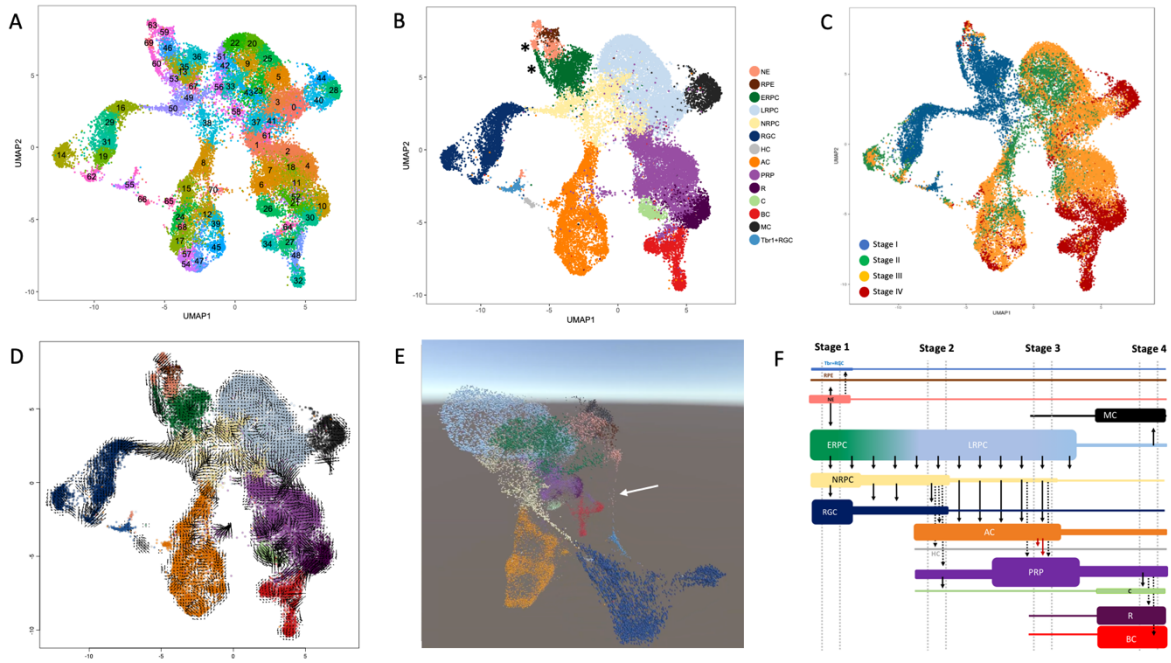


103

104 **Figure 1: (A) In vitro differentiation of miPSC-Nrl*GFP*-derived retinal aggregates (3D-RA).** Morphology of iPSCs
105 five days post-thawing. Cells in undifferentiated state are circumferential, "domed shaped" and surrounded by
106 a luminous halo (white arrow). Some unstable colonies tend to differentiate; they have a fibroblastic
107 morphology instead of round shape (red arrow). 3D-RA differentiating retinal aggregates from differentiation
108 DD1 to DD28 obtained following the modified SFEBq protocol [1]: (DD1) Rapid re-aggregation of iPSCs after
109 dissociation and passage in 96-well plate, (DD5) Appearance of retinal neuro-epithelium on the retinal
110 aggregates (light edges), (DD7) Evagination of retinal neuro-epithelium, (DD8-DD20) Growing of evaginating OV
111 like structures, (DD22-DD28) GFP expression under influence of the *Nrl* promotor in the photoreceptor layer of
112 retinal aggregates. Scale: 400 µm (miPSCs d5), 1000 µm (others). **(B) Schematic of anatomy of retinal layers.**
113 **(C) Expected layered expression of cell-type specific immunochemical markers.** Immuno-histochemical
114 markers of cellular subtypes in native retina (E13, P5 and P9) and Nrl*GFP* iPSCs-derived 3D-RA at DD21 and DD29.
115 Photoreceptor cells: GFP (green), recoverine (white) and rhodopsine (red) are specific for rod photoreceptor
116 cells; DAPI (blue) marks the nucleus of all cells. INL = inner nuclear layer; ENL = external nuclear layer; SI = internal
117 segment; SE = external segment; GCL = ganglion cell layer. Bipolar cells: Chx10 (red) and Pkc-α (white) are specific
118 for bipolar cells; Nrl-GFP (green) and DAPI (blue) as above; INL,ENL,GCL = as above; IPL = inner plexiform layer;
119 EPL = external plexiform layer. Amacrine cells: calretinine (white) is specific for amacrine cells; Chx10 (red) and
120 DAPI (blue) as above; the white arrows show two amacrine cells in the internal nuclear layer, while the green
121 arrow shows two amacrine cells probably delocalized during the experiment; ENL = external nuclear layer; INL =
122 inner nuclear layer; GCL = ganglionic cell layer. Retinal ganglionic cells: Pax6 (white) and Isl1 (white or red) are
123 specific for retinal ganglion cells; DAPI (blue) as above; the white arrows show Isl1 positive retinal ganglion cells.
124 **(C) Dissection of OV-like structures and dissociation of cells into viable single cell suspension.** Manual
125 dissection of 3D-retinal aggregates at stage 2 (DD21). After dissection, the retinal neuro-epithelium/evaginating
126 tissue is isolated from the dark/pigmented inner cell mass. Cells from retinal neuro-epithelium are dissociated
127 into a homogenous suspension of single cells.

128 We first analyzed all sequencing data jointly (i.e. NaR and 3D-RA) to cover a maximum of
129 intermediate developmental stages and hence generate the most continuous manifold
130 possible. We used Canonical Correlation Analysis (CCA) implemented with Seurat [17] to align
131 the NaR and 3D-RA datasets based on the expression profiles of 1,253 selected "most
132 variable" genes (Suppl. Table 2). We projected the corresponding 30-dimensional distances
133 (based on the 30 first CCA) between cells in 2D- and 3D-space using UMAP [18]. We assigned
134 all 38,091 cells jointly (i.e. NaR and 3D-RA) to 70 clusters by k-means clustering (Fig. 2A). We
135 defined gene expression signatures for 13 previously recognized retinal cell types using
136 published information [19] (Suppl. Fig. 1 and Suppl. Table 3), and regrouped the clusters
137 accordingly in 13 cell types corresponding to neuroepithelium (NE), retinal pigmented
138 epithelium (RPE), early (ERPC), late (LRPC), and neurogenic retinal progenitor cells (NRPC),
139 retinal ganglion cells (RGC), horizontal cells (HC), amacrine cells (AC), photoreceptor
140 precursor cells (PRP), cones (C), rods (R), bipolar cells (BC), and Müller cells (MC) (Fig. 2B).
141 Using previously defined expression signatures specific for the S and G2-M phases of the cell
142 cycle [20], we defined cellular subtypes of actively dividing ERPC, LRPC and NRPC (Suppl. Fig.
143 2). Clusters 69 (NE) and 60 (ERPC) form a branch that clearly separates from the rest of NE
144 and ERPC. The two clusters share enhanced expression of *Ccnd2* and *Hmga2,* typical of the
145 NE/ciliary marginal zone (CMZ) [21] to which they were therefore assigned (Fig. 2B and Suppl.
146 Fig. 1). One cluster (55) did not light up with any of the 14 signatures. Differential expression
147 analysis (see hereafter) revealed that the *Tbr1* transcription factor is exclusively expressed in
148 this cluster hence identifying it as a recently described subgroup of Tbr1$^+$ RGC cells located in
149 the inner plexiform layer [22]. We combined information from (i) the cells' developmental
150 stage (Fig. 2C), (ii) RNA velocity analyses [7] (Fig. 2D), and (iii) the 3D UMAP manifold (Fig. 2E
151 and Suppl. Video: http://www.sig.hec.ulg.ac.be/giga), to deduce developmental trajectories
152 based on our data (Fig. 2F and Suppl. Fig. 3). Briefly, NE generates both RPE and ERPC. ERPC
153 (detectable during stage 1) mature into LRPC (detectable during stage 2 and 3). Both ERPC
154 and LRPC are actively dividing, and spawn post-mitotic NRPC who sequentially differentiate
155 in RGC (stage 1), HC (stage 1 and 2), AC (stage 2 and 3), and PRP (stage 2 and 3). PRP complete
156 maturation in C, R and BC, while remaining LRPC transform in MC, at stage 4. Our data suggest
157 the occurrence of "re-specification" [23] of differentiated AC into both HC (cluster 65) and
158 PRP (cluster 70). A cellular bridge clearly connects NE with Tbr1$^+$ RGC (cluster 63) suggesting
159 that this peculiar cell population may directly derive from the neuroepithelium (Fig. 2E). The

160 average number of UMIs in clusters 63, 65 and 70 was within normal range ($p \geq 0.19$)
161 precluding doublets as alternative explanation. Thus, the joint analysis of serial NaR and 3D-
162 RA scRNA-seq largely recapitulates previously described canonical developmental routes of
163 the mammalian retina and their timing [24,25], yet reveals evidence for re-specification and
164 the unconventional origin of Tbr1^{+}RGC.
165



166

167 **Figure 2: (A)** 2D UMAP manifold showing NaR and 3D-RA cells jointly and their assignment to 70 clusters by k-
168 means clustering. **(B)** Merging of the clusters in 14 cliques corresponding to neuroepithelium (NE), retinal
169 pigmented epithelium (RPE), early (ERPC), late (LRPC), neurogenic retinal progenitor cells (NRPC), retinal
170 ganglionic cells (RGC), Tbr1 positive retinal ganglionic cells (Tbr1^{+}RGC), horizontal cells (HC), amacrine cells (AC),
171 photoreceptor precursor cells (PRP), cones (C), rods (R), bipolar cells (BC), and Müller cells (MC), on the basis of
172 the expression of known marker genes (Suppl. Table 3 and Suppl. Figure 1). Cluster 69 (NE) and 60 (ERPC)
173 (marked by asterisks) correspond to the ciliary marginal zone (CMZ) which forms a branch that clearly separates
174 from the rest of NE and ERPC. **(C)** Cells colored by developmental stage: 1. blue = DD13 + E13, 2. green = DD21
175 + P0, 3. orange = DD25 + P5, 4. red = DD29 + P9. **(D)** Local RNA velocities [7] embedded in 2D UMAP. **(E)** Preview
176 of virtual reality visit of 3D UMAP manifold (http://www.sig.hec.ulg.ac.be/giga/) showing the cellular bridge
177 directly connecting NE and Trb1+RGC (white arrow). **(F)** Summary of inferred developmental trajectories. Red
178 arrows mark presumed re-specification events.
179

180 ***Comparison of NaR and 3D-RA cell fates in UMAP space highlights commonalities and***
181 ***differences in developmental trajectories.***
182 We then focused on the comparison between the behavior of NaR and 3D-RA derived cells.
183 Global comparison of the distribution of NaR and 3D-RA cells across the manifold indicates
184 that in vitro neuroretinal differentiation from iPSCs largely recapitulates native development

185 (Fig. 3A-D). This is substantiated by noting that 82% of clusters and 86% of cell types contain
186 at least 10% of the least represented cell origin (NaR or 3D-RA) (Fig. 3E-F). More granular
187 examination, however, reveals noteworthy differences. The first one is the occurrence of
188 NaR- or 3D-RA specific clusters and cell types: (i) the RPE cell type is almost exclusively
189 composed of 3D-RA cells as a result of RPE elimination from NaR by dissection (which cannot
190 be done with 3D-RA as self-patterning of the NE into presumptive RPE occurs at a molecular
191 but not structural level in 3D-RA [4]; (ii) the CMZ is absent in 3D-RA as expected (only recently
192 were culture conditions established for inducing selective CM retinal differentiation in human
193 iPSC-derived RA [25]; (iii) AC cluster 65, which seems to drive re-specification to HC cells, is
194 only present in 3D-RA; (iv) the Tbr1^{+}RGC cluster (55) is largely dominated by 3D-RA cells which
195 can be explained by a lower proportional representation of Tbr1+ cells in NaR compared to
196 3D-RA at the analyzed stages as confirmed by immunohistochemistry (Fig. 3G). The second
197 difference is the apparent relaxation of pseudo-spatial and pseudo-temporal transcriptome
198 control in 3D-RA versus NaR. The developmental pathways traversed par NaR cells appear
199 indeed tighter than those of 3D-RA cells, while NaR cells sampled at a specific developmental
200 stage seem to populate fewer cell types than 3D-RA cells. To quantify the former, we down-
201 sampled cells to equalize NaR and 3D-RA numbers (within developmental stage) and
202 computed the average distance from the *n* closest neighbors, which was indeed highly
203 significantly shorter for NaR than for 3D-RA (Fig. 3H). To quantify the latter, we measured the
204 diversity of cell types within stages (using a measure of entropy), which was indeed
205 significantly lower in NaR than in 3D-RA for all four stages (Fig. 3I). The last noteworthy
206 difference between both systems is the observation that PRP arise earlier in 3D-RA than in
207 NaR and accumulate at the expense of other cell types (particularly RPC), yet partially fail
208 terminal differentiation particularly into BC cells (Fig. 3J).

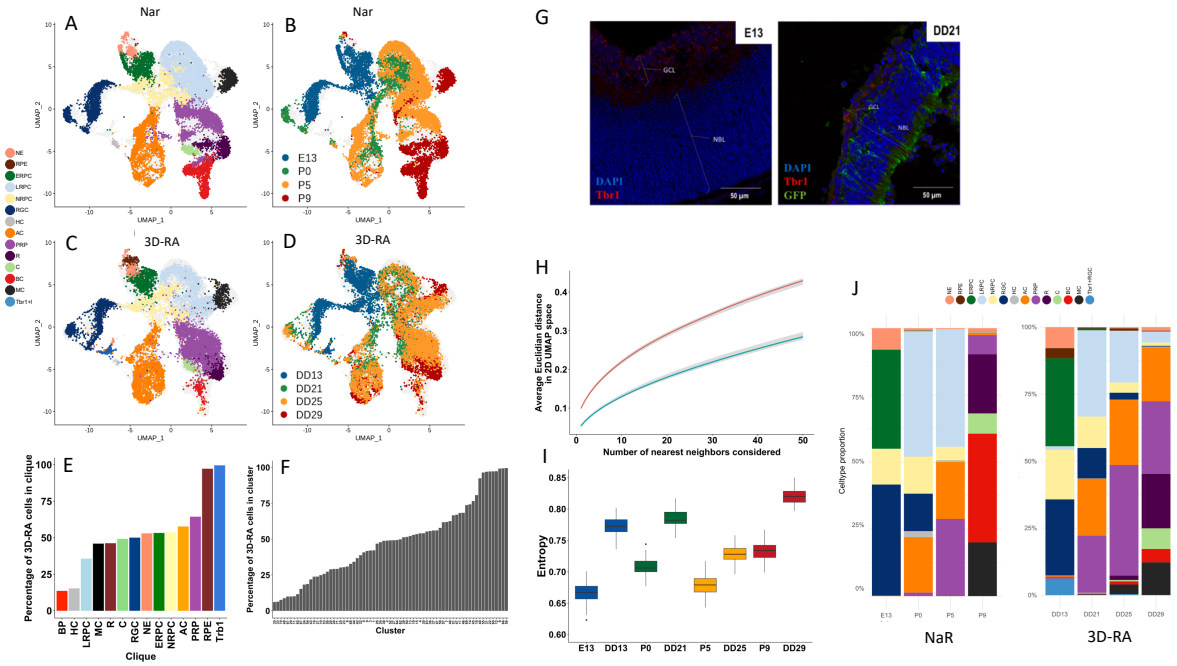


Figure 3: (A-D) Distribution of NaR (A-B) versus 3D-RA (C-D) cells across the UMAP manifold, sorted by cell type (A,C) and developmental stage (B,D). **(E-F)** Proportion of 3D-RA cells (adjusted for number of NaR and 3D-RA cells) in 14 cell types (E) and 70 clusters (F). 86% of cell types and 82% of clusters contain at least 10% of the least represented cell origin (NaR vs 3D-RA). **(G)** Expression of Tbr1 immunochemical marker (red), specific for Tbr1^{+}RGC, in NaR (P0) and in 3D-RA (DD21), showing the lower proportion of Tbr1^{+}RGC cells in NaR (dominated by NBL = Neuroblastic layer) compared to 3D-RA. GCL = ganglion cell layer. **(H)** Larger average distance in 2D UMAP space (Y-axis) from *n* nearest neighbors (X-axis) for 3D-RA (red) than for NaR cells (blue). **(I)** Larger cell type diversity (sampling-based measure of entropy) in the four developmental stages for 3D-RA than for NaR. **(J)** Proportions of cell types within developmental stage for NaR (left) and 3D-RA (right).

3D-RA culture conditions mainly perturb genes and pathways that play key roles in NaR development, including transcription factors.

To identify key retinal differentiation genes, we performed differential expression analysis for each cell type against all others, first considering NaR cells only. We identified a total of 7,292 genes with significantly higher expression in one of the 13 main cell-types (as defined above) compared to all others merged (log-fold change $\geq$ 0.25 and p-value $\leq$ 0.001) hereafter referred to as "cell type-specifying" genes (Fig. 4A and Suppl. Table 4). Searching for genes that were dynamically regulated as a function of pseudo-time along a developmental trajectory inferred with Monocle 2 [9] yielded comparable results (Suppl. Fig. 4 and Suppl. Table 5). We then searched for enriched Reactome pathways [26,27] in the 13 lists of "cell type-specifying" genes. Two hundred sixty-eight pathways were significantly enriched (q $\leq$ 0.01) in at least one cell-type (Fig. 4B and Suppl. Table 6). These corresponded primarily to: (i) accelerated cell division in ERPC, LRPC and NRPCs, (ii) intense post-transcriptional and

234 translational activity in NE, ERPC, LRPC and NRPCs, (iii) activation of RHO GTPase- and NOTCH-
235 dependent signaling in ERPC, LRPC, NRPCs, RGC and LRPC, NRPCs, respectively, as well as the
236 GPCR-dependent phototransduction cascade in C and R, (iv) activation of mitochondrial citric
237 acid (TCA) cycle and respiratory electron transport in HC, C, R, BC, and MC, of cholesterol
238 synthesis in ERPC and RGC, and of insulin- and glucagon-dependent metabolic integration in
239 RGC and AC, (v) enhanced remodeling of the extracellular matrix in NE, RPE and MC, and GAP
240 junction trafficking in RGC, and (vi) activation of ROBO receptors-dependent axon guidance
241 in NE, ERPC and LRPC, and of synapse formation in RGC, HC, AC and BC (Fig. 4B). In addition
242 to the 13 major cell-types, 633 genes were more strongly expressed in Tbr1^{+}RGC (3D-RA)
243 when compared to all other cell-types (NaR + 3D-RA), with *Tbr1* as top ranking differentially
244 expressed gene (Suppl. Table 4 and Suppl. Fig. 5). Twelve Reactome pathways were
245 significantly enriched ($q \leq 0.01$) in the list of Trb1^{+}RGC-specific genes pertaining to: (i) axon
246 guidance, (ii) L1CAM interactions, and (iv) membrane trafficking, which are in accordance
247 with the role of Tbr1+ cells instructing the organization of dendrites of RGCs in the outer
248 plexiform layer of the developing retina (Suppl. Table 6). Moreover, 426 genes were more
249 strongly expressed in the CMZ when compared to all other cell types (Suppl. Table 4).
250 Strikingly 14 members of the family of crystallin genes ranked amongst the top 24 genes
251 (Suppl. Fig. 6). We noted that the average number of UMIs detected in cluster 60 cells was
252 nearly twice the experiment-wide average ($p = 0.01$). We therefore surmise that this cluster
253 contains a large proportion of doublets of physically connected neuroretinal and posterior
254 epithelial lens cells.
255 At first sight, genes that were differentially expressed between cell-types in NaR appeared to
256 recapitulate their in vivo expression profile remarkably well in 3D-RA (Fig. 4A). Yet, to better
257 apprehend the differences between in vivo and in vitro retinal differentiation, we performed
258 differential expression analysis between NaR and 3D-RA within cell-type. For each of the 13
259 major cell types, we generated lists of genes that were respectively under- and over-
260 expressed in 3D-RA when compared to NaR ($q \leq 0.01$; Suppl. Table 7). We then searched for
261 biological pathways that were over-represented in the corresponding gene lists using
262 Reactome. This yielded 197 downregulated and 134 upregulated pathways in 3D-RA (Fig. 4B
263 and Suppl. Table 8). Strikingly, both down- and upregulated pathways showed considerable
264 overlap with the pathways shown to be dynamically regulated during in vivo retinal
265 development (115/197, $p < 10^{-6}$ and 67/134, $p < 10^{-6}$) (Fig. 4C). The differentially expressed

genes that accounted for the overrepresentation of specific pathways ("found entities" in Reactome) were generally very similar during in vivo development of NaR (Cell type > others) and when comparing NaR and 3D-RA (NaR > 3D-RA and 3D-RA > NaR) (Fig. 4D). Thus, the genes and pathways that appear to be the most perturbed by the 3D-RA culture conditions are also the ones that play key roles in NaR development. More specifically, (i) the rate of cell division in NE, ERPC, LRPC and NRPC was reduced in 3D-RA when compared to NaR, (ii) post-transcriptional and translational mechanisms were exacerbated in in 3D-RA ERPC, LRPC, NRPC, RGC, PRP, C, R, BC and MC, when compared to NaR, (iii) signal transduction via WNT, TGF-beta, RHO GTPases, Esr, Notch, Hedgehog, MAPK, and Death receptors was diminished in 3D-RA when compared to NaR, particularly in ERPC and LRPC, while the phototransduction cascade was less active in 3D-RA-derived R than in NaR-derived R, (iv) mitochondrial citric acid (TCA) cycle and respiratory electron transport was increased in 3D-RA LRPC, NRPC, AC, PRP and C (yet increased in BC), cholesterol synthesis increased in 3D-RA C and R, and gluconeogenesis increased in PCP and R, (v) stress response and apoptosis was reduced in 3D-RA ERPC, yet increased in 3D-RA C, R, BC and MC (i.e. at the latest stages of 3D-RA culture), and (vi) vesicle mediated transport and synapse formation was decreased in 3D-RA LRPC, RGC and PRP (Fig. 4B).

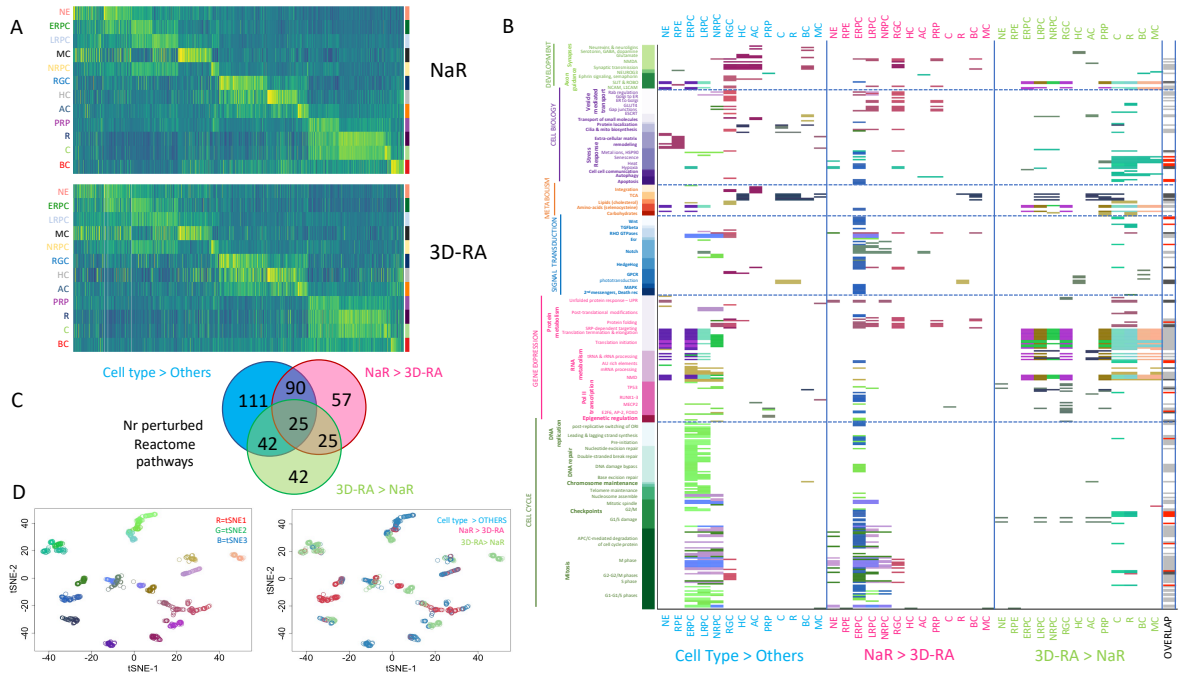


Figure 4: (A) Expression profiles in 12 cell types of 7,292 genes that are dynamically regulated during in vivo retinal development (i.e. significantly overexpressed in at least one cell type when compared to all other ones in NaR) in NaR (upper panel) and 3d-RA (lower panel). **(B)** Reactome pathways that are significantly ($p \leq 0.001$) enriched amongst differentially expressed genes ("Cell type > other": when comparing expression levels between specific cell types and all other cells in NaR only; "NaR > 3D-RA" and "3D-RA> NaR": when comparing

289 expression levels between NaR and 3D-RA cells within cell type). Y-axis: Reactome pathways are colored by "top
290 level" system (cell cycle, gene expression, signal transduction, metabolism, cell biology and development) and
291 sub-level therein. Tiles mark the pathways that are significantly enriched in the corresponding contrast and cell
292 type. The colors of the tiles reflect similarity in gene content as described in (C). Last column ("Overlap"): White:
293 pathways significant in one contrast only, Black: pathways significant in all three contrasts, Grey: pathways
294 significant in "Cell type > other" and ("NaR > 3D-RA" or "3D-RA> NaR"), Red: pathways significant in "NaR > 3D-
295 RA" and "3D-RA> NaR". **(C)** Number of unique and shared Reactome pathways between "Cell type > other",
296 "NaR > 3D-RA" and "3D-RA> NaR". All overlaps are highly significant ($p < 10^{-6}$) assuming random sampling from
297 2,365 Reactome pathways. **(D)** (Left) tSNE map (dimensions 1 and 2 of 3) of 1,313 gene sets ("found entities")
298 marking enriched Reactome pathways (each little circle corresponding to a tile in B). Overlapping sets (in terms
299 of gene content) are close to each other in tSNE space (3D). The gene sets are colored in RGB scale where tSNE
300 dimension 1 (tSNE1) determines the intensity of red, tSNE2 the intensity of green, and tSNE3 the intensity of
301 blue. (Right) Same data points as for (Left) but colored by contrast of origin as in B and C: "Cell type > other":
302 turquoise; "NaR > 3D-RA": magenta; "3D-RA > NaR": lime. One can see that many clusters encompass gene sets
303 corresponding to distinct contrasts, hence highlighting the strong overlap between Reactome pathways that are
304 essential for normal retinal development ("Cell type > other"), and those that are perturbed in 3D-RA relative to
305 NaR ("NaR > 3D-RA" and "3D-RA > NaR").
306

307 The 7,292 genes that were found to be dynamically regulated during in vivo development (i.e.
308 NaR only, cfr above) comprised 307 transcription factors (TF), including 110 that were at least
309 1.5 times more strongly expressed in one cell type (or pair of cell types for C and R) when
310 compared to any of the other (Fig. 5A&B and Suppl. Table 4). The latter comprised 87 factors
311 that were previously reported in the context of retinal development, as well as 22 new ones
312 (NE: Mecom, Klf2, Peg3; LRPC: Lrrfip1; MC: Arid5a, Creb312, Csrnp1, Dbp, Nr4a1, Nr4a3; HC:
313 Zfp618, Zfp804a; AC: Zfp503; PRP: Foxo3, Lcorl; R: Zfp516, Trps1, Ppard, Zc3h3, Mier1, Mier2,
314 Lyrar; BC: St18) (Suppl. Table 9). Contrary to the overall expression profile (i.e. 7,292 genes;
315 Fig. 4A), visual examination of the expression profiles of the 110 most differentially expressed
316 TF indicated considerable loss of cell-type specificity in 3D-RA (Fig. 5A). Indeed, 160 of the
317 307 (52%) differentially expressed TF were significantly ($q < 0.01$) under-expressed in at least
318 one cell type in 3D-RA when compared to NaR, while 80/307 (26%) were significantly ($q <$
319 0.01) over-expressed in at least one cell type (Fig. 5B&C). An additional 31 (non-dynamically
320 regulated) TF were down-regulated in 3D-RA, while 19 were upregulated (Fig. 5B&C). Thus,
321 the expression profile of a remarkably high proportion of TF that are dynamically regulated
322 during normal retinal development in vivo appears perturbed in 3D-RA, and this may in part
323 drive the differences observed between both systems including with regards to Reactome
324 pathways (Fig. 4B).

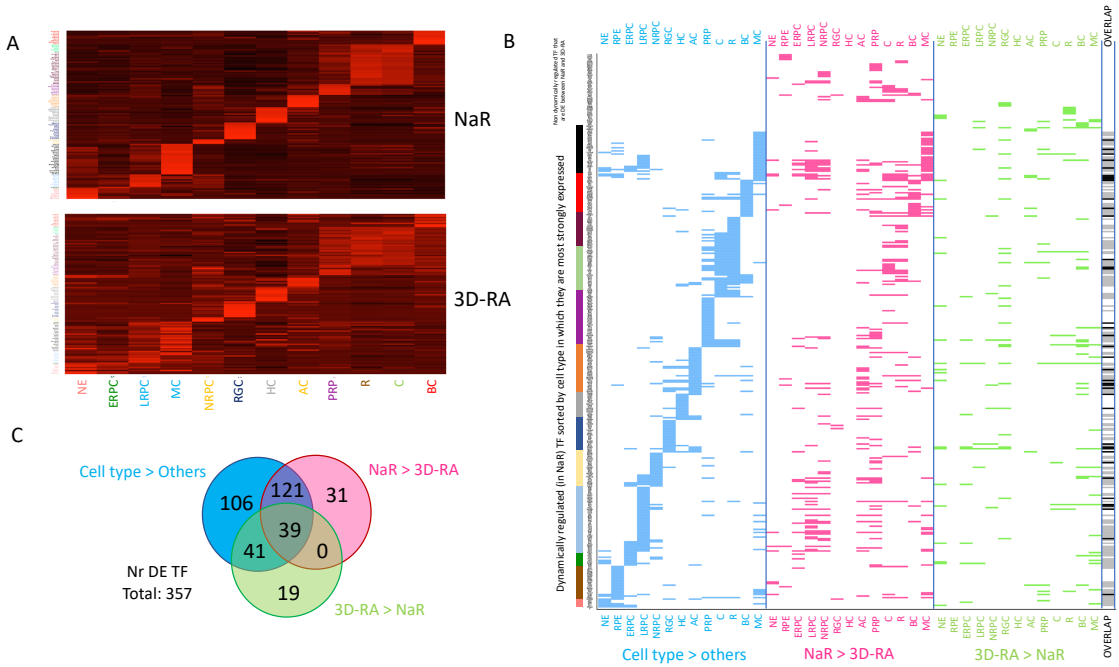


Figure 5: (A) Standardized expression levels of 110 most clique-specific TF across 12 cliques in NaR (upper panel) and 3D-RA (lower panel). **(B)** Transcription factors (TF) that are (i) differentially regulated between cell types (turquoise), (ii) under-expressed in 3D-RA when compared to NaR (magenta), or (iii) over-expressed in 3D-RA when compared to NaR (lime). OVERLAP: TF that are differentially expressed in the three conditions (Clique > others, NaR>3D-RA and 3D-RA> NaR) are marked in black. TF that are differentially expressed during retinal development (Clique > others) and in one of the NaR vs 3D-RA conditions (NaR>3D-RA or 3D-RA> NaR) are marked in grey. Acronyms for cell types are as in the remainder of the manuscript. **(C)** Number of differentially expressed TF in "Clique > others", "NaR > 3D-RA", and "3D-RA>NaR", with corresponding overlaps. The overlaps are highly significant ($p < 10^{-6}$) assuming that TF are sampled randomly from the full collection of ~1,500 TFs [28].

Combined analysis of scRNA-Seq and bulk ATAC-Seq data reveals putative regulatory toggles in NaR and 3D-RA

It is generally assumed that execution of the epigenetic program underlying differentiation is controlled by dynamically regulated TF that activate downstream target genes. To verify this assertion, we performed ATAC-Seq [29] on our bulk NaR (E13, P0, P5) and 3D-RA (DD13, DD21, DD25) samples to identify gene-switch components accessible by TF during retinal development (Fig. 6A). For each sample type, we analyzed two technical replicates of two biological replicates for a total of 24 libraries. We defined a total of 123,482 peaks using MACS2 [30] (Suppl. Table 10). Of these, 93,386 (75.6%) were detected in NaR, 97,333 (78.8%) in 3D-RA. 18,933 (15.3%) were common to all samples, 26,149 (30.0%) NaR-specific, 30,096 (24.4%) 3D-RA-specific, and 4,703 developmental stage-specific (3.8%; stage I: 294, stage II: 82, stage III: 4,327). The number of peaks increased with developmental stage in NaR but not

in 3D-RA (highest number of peaks in DD13) (Fig. 6B). Nevertheless, stage I samples (E13 and DD13) clustered together, while for subsequent stages samples clustered by origin (NaR vs 3D-RA) (Fig. 6C). DNA binding motifs are reported for 151 of the 307 TF found to be dynamically regulated by scRNA-Seq (*vide supra*), amounting to a total of 336 motifs (average number of motifs per TF: 2.3; range: 1 - 14) [31]. We used Homer to annotate our catalogue of ATAC-Seq peaks for the corresponding motifs [31]. In total Homer identified 7,128,225 putative binding motifs in 98,181 ATAC-seak peaks assigned to 19,171 genes (closest transcription start site)(Suppl. Table 11).

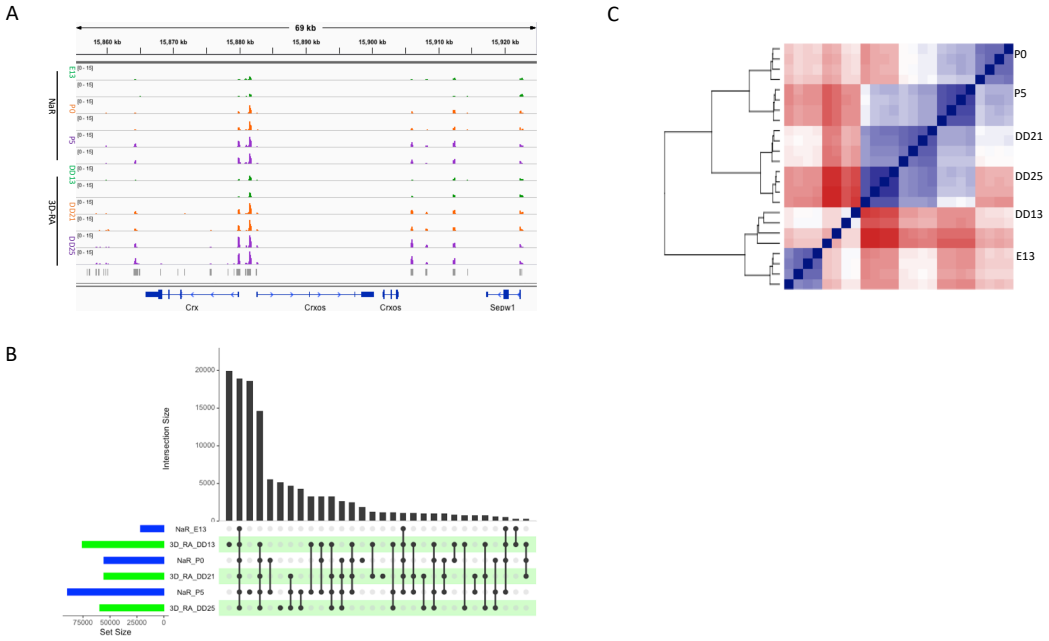


Figure 6: (A) Example of ATAC-Seq results obtained in NaR and 3D-RA samples in the vicinity of the *Crx* gene. The intensity of several peaks is increasing with developmental stage in both NaR and 3D-RA as expected for this TF that is primarily expressed in PRP, C and R. **(B) Numbers and overlap between ATAC-Seq peaks detected in different sample types**. The number of peaks increases with developmental stage in NaR but not in 3D-RA. A large proportion of peaks are either DD13- (16.1%) or P5-specific (15.1%), or shared between all samples (15.3%), or all samples minus E13 (11.9%). **(C) Hierarchical clustering of the samples based on the intensity of 123,482 ATAC-Seq peaks.** Sample types cluster by developmental stage at stage I (E13 and D13), but by origin (NaR vs 3D-RA) at stages II and II.

To test whether dynamically regulated TF that were overexpressed in a given cell type were indeed activating downstream targets in that cell type, we searched for an enrichment of the cognate binding motifs in the ATAC-Seq peaks of genes that were significantly overexpressed in that cell type, relative to ATAC-Seq peaks of genes that were significantly underexpressed in the same cell type. We used average number of binding motifs per ATAC-Seq peak as metric to correct for gene length. We first analyzed NaR. We found 84 instances of binding

374 motif enrichment (q-value < 0.01) involving 37 TF overexpressed in the corresponding cell
375 type (Fig. 7A, Suppl. Table 12&13). Examples of such "activator" TF include Crx (3/3 binding
376 motifs) operating in PRP, R and C, and Etv5 (1/1 binding motif) acting in LRPC (Fig. 7B).
377 Intriguingly, we observed 45 instances of binding motif depletion (q-value < 0.01), involving
378 26 overexpressed TF. Stated otherwise, for 26 TF that were overexpressed in one or more
379 cell types, binding motifs were significantly more abundant in ATAC-Seq peaks of genes that
380 were underexpressed than in ATAC-Seq peaks of genes that were overexpressed in the
381 corresponding cell types (Fig. 7A, Suppl. Table 12&13). One explanation of this finding is that
382 these TF act as repressors rather than as activators. Examples of such candidate "repressor"
383 TF are Arh (1/2 binding motifs) acting in C, and Tgif2 (2/3 binding motifs) acting in LRPC and
384 NRPC (Fig. 7B). Such repressor TF may be components of regulatory toggles that ensure cell
385 type-specific gene expression not only by inducing expression of required genes (via activator
386 TF), but also by precluding expression of undesired genes (via repressor TF) (Fig. 7C). To gain
387 insights in the nature of the genes targeted by these putative repressor TF, we searched for
388 the enrichment of the corresponding binding motifs in genes overexpressed in other cell types
389 (i.e. cell types in which the repressor TF are not overexpressed and in which the target genes
390 of the repressor are therefore not repressed). Fig. 7D shows the corresponding results for all
391 repressor TF expressed in a given cell type jointly (average signed (enrichment +, depletion -)
392 log(1/p) over all repressor TF). For comparison, similar graphs are shown for the combined
393 effect of all activator TF expressed in a given cell type. It appears that the 12 cell types
394 analyzed in NaR (excluding RPE) form three clusters: (i) NE, ERPC, LRPC and NRPC, (ii) RGC, AC
395 and HC, and (iii) PRP, C, R, BC and MC. Assuming that the repressor hypothesis is valid,
396 repressor TF expressed in cluster (i) primarily target genes overexpressed in cluster (iii),
397 repressor TF expressed in cluster (ii) target both cluster (i) and (iii), and repressor TF expressed
398 in cluster (iii) primarily target cluster (i). There was considerable overlap between the TF
399 (activator and repressor) overexpressed in cell types from the same cluster. It is noteworthy,
400 that there was some evidence that the same mechanisms might underpin differentiation of
401 cell types within clusters. As an example, NRPC express activator TF that do not operate in
402 NE and ERPC, while ERPC express repressor TF that target LRPC and NRPC-specific genes (Fig.
403 7D).
404 Six TF overexpressed in more than one cell type, were characterized by the enrichment of
405 binding motifs in overexpressed genes of some cell type(s), yet the depletion of binding motifs

in overexpressed genes of other cell type(s) (Etv1, Fosb, Otx2, Pax6, Tcf3, Sox11) (Suppl. Table 12&13). A good example is Otx2, which is significantly overexpressed in PRP and NRPC, and whose three binding motifs are enriched in genes that are overexpressed in PRP while being enriched in genes that are underexpressed in NRPC (which include the genes that are overexpressed in PRP)(Fig. 7B). One possible explanation of these observations is that the corresponding TF are necessary but not sufficient to induce expression of target genes. As an example, Otx2 may be induced in NRPC that will develop into PRP, but only exert its transcriptional effects after maturation into PRP. Consistent with this hypothesis, NRPC expressing Otx2 tended to cluster in the vicinity of PRP in the UMAP manifold (Suppl. Fig. 7). We refer to these six TF as "conditional activators". Three TF were characterized by the enrichment of one binding motif yet the depletion of another binding motif in the same cell type (Lhx1, Plagl1, Zic1) (Suppl. Table 12&13). These will be referred to as "dual TF" yet were considered with caution.

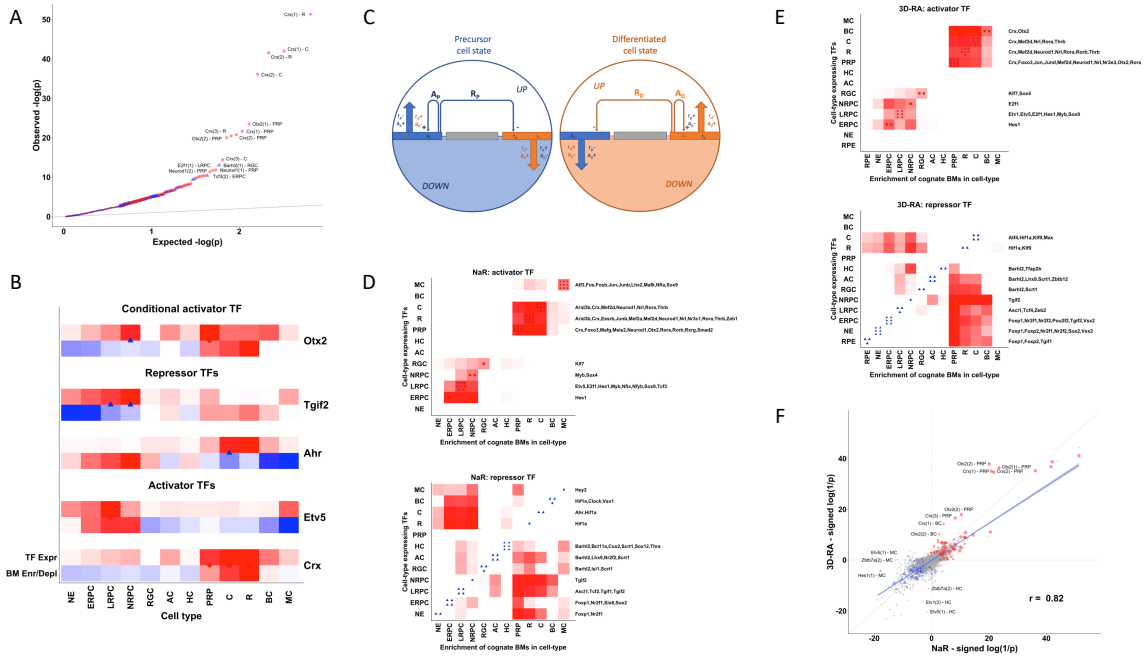


Figure 7: (A) QQ plot of log(1/p) values of the difference in average number of binding motifs in ATAC Seq peaks of overexpressed versus underexpressed genes for TF overexpressed in the corresponding cell type. Enrichments are indicated by red dots, depletions by blue triangles. Symbols are large for instances with q-value < 0.01, and with small symbols for instances with q-value ≥ 0.01. For the most significant effects the name of the TF is given, as well as – between brackets - the number of the binding motif, and the cell type. The line corresponds to expectations assuming that all tests are true null hypotheses. **(B) Examples of activator, repressor and conditional activator TF.** Upper lines ("TF Expr", white-red color code): standardized expression pattern of corresponding TF across 12 cell types. Lower lines ("BM Enr/Depl", blue-white-red color code): standardized enrichment (red) or depletion (blue) of binding site of corresponding TF across 12 cell types. The red circles and blue triangles mark the instances that would be part of (A). They require that the TF be overexpressed in the corresponding cell type, and that there is either a significant enrichment of binding motifs in overexpressed genes (red circles) or a significant depletion (blue triangles). **(C) Components of regulatory**

toggles and principles underlying their detection. Shown are a hypothetical precursor (blue) and derived differentiated (orange) cell. The genes (horizontal rectangles) are subdivided in genes that define the precursor state (blue rectangles), genes that define the derived state (orange rectangles), and genes that do not participate in the differentiation (grey rectangles). Differential expression analysis between precursor cells and differentiated cells reveals (i) the genes that are upregulated in precursor cells (and consequently downregulated in differentiated cells)(large blue arrows), and (ii) the genes that are upregulated in differentiated cells (and consequently downregulated in precursor cells)(large orange arrows). In this work we focused on the genes that were upregulated in specific cliques. It is assumed that the differentiation between cell states is "locked" by a toggle comprising mutually exclusive activator-repressor sets of TF. In the example, we show an activator ($\mathbf{A_p}$) – repressor ($\mathbf{R_p}$) pair (small blue arrows) turned ON in the precursor state (OFF in the differentiated state), and an activator ($\mathbf{A_d}$) – repressor ($\mathbf{R_d}$) pair (small orange arrows) turned ON in the differentiated state (OFF in the precursor state). Genes respond to these activator and repressor TF by means of binding motifs in cis-acting regulatory elements. One can predict that genes activated in the precursor cell state will be enriched in binding motifs for $\mathbf{A_p}$ (labelled as a_p) and depleted in binding motifs for R_p (labelled as r_p), but (given their expression pattern in differentiated cells) also enriched in binding motifs for $\mathbf{R_d}$ (labelled as r_d) and depleted in binding motifs for $\mathbf{A_d}$ (labelled as a_d) (and vice versa). **(D) Combined NaR enrichment profile (across all cell types; X-axis) of binding motifs for all activator (upper graph) and repressor (lower graph) TF expressed in a given cell type (Y-axis).** Standardized (across entire array) sum of signed (+ enrichment, - depletion) log(1/p) values for binding motifs of TF expressed in a given cell type (Y-axis). Positive values are measured by a white-red color code; values ≤ 0 are in white. For each cell type the number of overexpressed activator TF (upper graph, red circles on diagonal) and repressor TF (lower graph, blue triangles on diagonal) are given, and their names provided on the right. **(E)** As in (D) for 3D-RA. **(F) Comparison of the log(1/p) values for enrichment (+)/depletion (-) of TF binding motifs across cell types between NaR (X-axis) and 3D-RA (Y-axis).** Red circles: values for activator TF in the cell type in which they are overexpressed. Blue triangles: values for repressor TF in the cell type in which they are overexpressed. Small grey circles: values for in cell types in which the corresponding TF is not overexpressed. The identity of outliers is labelled as in (A). "r" = correlation coefficient.

We repeated the same analyses for the 13 major cell types detected in 3D-RA (including RPE). We identified 18 activator TF, 24 repressor TF, 3 conditional activator TF, and 2 dual TF (Suppl. Table 12&14). Thirty-eight of these 47 TF overlapped with those found in NaR. Analyzing the combined effect of TF overexpressed in a given cell type largely confirmed the three clusters detected in NaR. RPE appear to be part of cluster (i), while 3D-RA MC could not be classified as no overexpressed TF (whether activator or repressor) could be detected in this cell type (Fig. 7E). The degree of enrichment/depletion of TF binding motifs (measured by the signed (enrichment +, depletion -) log(1/p)) in the different cell types was highly correlated between NaR and 3D-RA ($r= 0.82$, $p < 2.2e-16$)(Fig. 7F). The slope of the regression line was significantly < 1 suggesting that, overall, the TF's activator and repressor effects might be slightly reduced in 3D-RA when compared to NaR. Outliers included Crx and Otx2 which appeared to have more pronounced activator effects in PRP of 3D-RA than of NaR. This may be related to the fact that the 3D-RA culture conditions were designed to "push" PRP development (see also Fig. 3J). Other outliers were Etv1, Etv5, Hes1 and Zbtb7a which pointed towards TF-driven differences in the NaR and 3D-RA transcriptomes of HC and MC. Binding motifs for this squadron of TF were enriched in genes that were underexpressed in

476 MC of NaR but not of 3D-RA, while being enriched in genes that were underexpressed in HC
477 of 3D-RA but not of NaR (Suppl. Fig. 8). We identified the corresponding genes, and amongst
478 those, the ones with the highest density (in ATAC-Seq peaks) of binding motifs for the TF
479 squadron (one SD above the mean), and submitted them to Reactome. The pathways that
480 were enriched pertained mainly to RNA stability and translation and to a lesser the cell cycle
481 (Suppl. Table 15). They coincided remarkably well with the Reactome pathways that were
482 highlighted by the genes that were more strongly expressed in the MC of 3D-RA than of NaR
483 (group of genes labelled in orange in Fig. 4B). Taken together, our results suggested that the
484 differences in the transcriptome of Nar versus 3D-RA MC are at least in part due to perturbed
485 transcriptional regulations by the Etv1, Etv5, Hes1 and Zbtb7a TF squadron. .
486

487 **Discussion**

488 We herein use scRNA-seq to compare the unfolding of the epigenetic program in *in vivo*
489 versus *in vitro* (from iPS cells) derived murine retina at four match stages of development
490 encompassing the presumed emergence times of the major retinal cell types. Results
491 obtained by combining information from (i) the analysis of four developmental stages (E13 vs
492 DD13, P0 vs DD21, P5 vs DD25 and P9 vs DD29), (ii) 3D UMAP manifolds visualized in virtual
493 reality (http://www.sig.hec.ulg.ac.be/giga), and (iii) RNA velocity analysis, are in good
494 agreement with the previously reported main retinal developmental trajectories (Fig. 2F).
495 More specifically, we confirm the existence of a recently characterized RGC subpopulation
496 that specifically expresses the Tbr1 transcription factor. We provide evidence that this
497 subpopulation may directly derive from NE via a narrow, proprietary developmental bridge.
498 We provide evidence supporting the occurrence of « respecification » of AC into both HC and
499 PRP. It is noteworthy that the evidence for both the unconventional origin of Tbr1+ RGC and
500 respecification of AC cells to HC cells stems from 3D-RA cell populations only. Further work
501 is needed to confirm that these phenomena also occur in NaR. We identify >7,000 genes
502 that are dynamically regulated during *in vivo* retinal differentiation corresponding to tens of
503 biological pathways pertaining to the cell cycle, gene expression, signal transduction,
504 metabolism, cell biology and development (Fig. 4). Dynamically regulated genes include ~300
505 TF, of which 110 are at least 1.5x as strongly expressed in one specific retinal cell type when
506 compared to all other ones. The latter include 22 TF not yet described in the field of retinal

507 development which could serve as a starting point for functional investigations of the roles of
508 these genes in retinogenesis and physiology.
509 By combining scRNA-Seq with ATAC-Seq data generated on bulk samples corresponding to
510 three of the analyzed developmental stages, we were able to show – for 75 of 151 analyzed
511 TF - a significant (q-value < 0.01) enrichment (31), depletion (31) or both (13) of the TF's
512 binding motifs in ATAC-Seq peaks of the genes that are overexpressed in the same cell type
513 as the TF (considering NaR and 3D-RA results jointly)(Suppl. Table 12). Enrichment is what is
514 expected for a regular TF that activates the transcription of target genes in the cell type in
515 which it is (over)expressed. We refer to these as "activator" TF. A depletion is against
516 expectations, but is what is expected if the TF acts as a repressor in the cell type in which it is
517 (over)expressed. Although we have no formal proof of their repressing activity, we
518 tentatively refer to these as "repressor" TF. We propose that combinations of activator and
519 repressor TF constitute regulatory toggles that help ensure cell type-specific gene expression
520 and hence cellular identity (Fig. 7C). Our results suggest that the hypothesized regulatory
521 toggles involve multiple activators and repressors. This may confer robustness to the system,
522 and enable differentiation of multiple cell types. It also indicates that perturbing TF one at a
523 time, whether by overexpression or knock-out/down, may not be effective to dissect such
524 multifactorial toggles. It may be necessary to perform pooled screens using CRISPR libraries
525 targeting several candidates at once (at high multiplicity of infection) in Perturb-Seq like
526 experiments conducted in 3D-RA [32] in order to induce detectable alterations in cellular
527 behavior. For 10 TF, overexpression was accompanied by binding motif enrichment in some
528 cell type(s), and depletion in other(s). These may include "conditional activator" TF, meaning
529 that the presence of the TF is necessary but not sufficient to exert its effect on transcription.
530 This could for instance reflect the need for post-translational modification of the TF [33], or
531 for cooperation with other TF or cofactors [34]. Finally, for three TF, enrichment was
532 observed for one binding motif and depletion for another binding motif, yet in the same cell
533 type. We refer to these as "dual" TF. Although we consider these results cautiously, context-
534 dependent TF effects are well documented [35].
535 We assessed enrichment versus depletion of binding motifs by comparing genes that were
536 overexpressed with those that were underexpressed in a given cell type. Alternatively, one
537 can compare the density of binding motif in ATAC-Seq peaks of up- and down-regulated genes
538 (i.e. over- and underexpressed genes) with that in ATAC-Seq peaks of genes that are stably

539 expressed across cell types, i.e. genes that are neither significantly over- nor underexpressed
540 in any given cell type. When doing this for the 84 detected instances of binding motif
541 enrichment and the 45 instances of binding motif depletion, we observed that the binding
542 motif density in stably expressed genes was intermediate between that of up- and
543 downregulated genes in 73% (enrichments) and 71% (depletions) of cases, respectively.
544 Although we recognize the ascertainment bias, these figures are at least compatible with the
545 notion that the density of binding motifs for TF is increased by positive selection in target
546 genes, while being decreased by negative selection in non-target genes.
547 We show that 3D-RA broadly recapitulate the in vivo developmental program and
548 trajectories. However, developmental trajectories appear less canalized in 3D-RA when
549 compared to NaR, PRP to develop earlier and at the expense of other cell types, and terminal
550 differentiation of BC to be incomplete (Fig. 3). We identify ~3,000 genes that are differentially
551 regulated between 3D-RA and NaR in at least one cell type, and identify the corresponding
552 biological pathways pertaining in particular to the rate of cell division which is reduced in 3D-
553 RA RPCs when compared to NaR, post-transcriptional and translational mechanisms which
554 appear exacerbated in the majority of 3D-RA cliques when compared to NaR, signal
555 transduction via WNT and Notch pathways which are diminished in 3D-RA RPCs when
556 compared to NaR, 3D-RA differentiated cells which appear less functional with less
557 phototransduction cascade activity and decrease synapse formation, and finally apoptosis
558 and stress response which are increased at the latest stages of 3D-RA culture. Strikingly, the
559 perturbed pathways show a highly significant overlap with those that were shown to be
560 dynamically regulated during the development of the native retina. We show that
561 transcription factors that are dynamically regulated during in vivo retinal development are
562 particularly sensitive to the iPSC culture conditions. This is likely to drive the perturbations
563 of the above-mentioned biological pathways.
564 We show how scRNA-Seq and bulk ATAC-Seq can be combined to gain preliminary insights
565 into what may underpin differences observed between the NaR and 3D-RA transcriptomes.
566 As an example, the comparison between MC from NaR and 3D-RA revealed 418 genes that
567 were more strongly expressed in NaR when compared to 3D-RA, and 424 that were more
568 strongly expressed in 3D-RA (Suppl. Table 7). The first list of genes (NaR>3D-RA) was not
569 enriched for any Reactome pathway, but the second (3D-RA) highlighted 53 of these (Suppl.
570 Table 8), corresponding to five subsets of genes (five colors in MC column (3D-RA>NaR) in Fig.

571 4B). One of these subsets (orange set, rich in genes encoding ribosomal proteins) highlighted
572 pathways related to RNA stability (NMD), RNA translation, selenocystein metabolism and
573 signaling by ROBO receptors. Combined scRNA-Seq and ATAC-Seq data indicated that the
574 transcriptome of MC is depleted in genes with high density of binding motifs for Etv1, Etv5,
575 Hes1 and Zbtb7a in their ATAC-Seq peaks in NaR but not so in 3D-RA (Fig. 7F and Suppl. Fig.
576 8). These four TF are expressed at relatively high and comparable levels in both NaR and 3D-
577 RA. We identified the corresponding genes and subjected them to Reactome analysis. They
578 identified – to a large extend – the same pathways as the orange gene subset defined above.
579 Thus, the differences between NaR and 3D-RA MC with regards to the corresponding
580 pathways are most likely driven by perturbations of the Etv1, Etv5, Hes1 and Zbtb7a TF
581 squadron. Examination of their transcriptional and binding motif enrichment/depletion
582 profile across cell types (Suppl. Fig. 8) suggests that Etv1, Etv5 and Hes1 operate as
583 "conditional TF" (as defined above) in NaR: despite still being present in MC they do not have
584 the activator effect in these cells as seen in retinal progenitor cells (hence the observed
585 depletion in the NaR MC transcriptome). In 3D-RA MC, they may still have "residual" activator
586 activity which would explain why the depletion is not seen. Zbtb7a is clearly distinct from the
587 other three: it most strongly expressed in PRP, C and R, yet its binding motifs are primarily
588 found in genes expressed in retinal progenitor cells. How and why genes rich in Zbtb7a motifs
589 would be underexpressed in NaR but not 3D-RA MC remains unclear. Yet these examples
590 show how studying the regulatory toggle landscape may become a valuable approach to
591 monitor how closely organoids recapitulate native development.
592

593 **Materials and methods**

594 ***Generation of iPSC-derived retinal aggregates.*** *Maintenance of iPSCs:* The mouse iPSC-
595 NrlGFP line was obtained from the laboratory for retinal regeneration from the Center for
596 Bioloy and Development (CDB) RIKEN (Kobe, Japan). These iPSCs were generated from
597 fibroblasts [36] of C57BL/6 Nrl-eGFP transgenic mice [15]. The cells were cultivated in 60-mm
598 Petri dishes (0,6 x 10^5 cells total per dish) coated with gelatine 0,1% (G2625, Sigma-Aldrich)
599 in Glasgow's Minimum Essential Medium (GMEM, Thermo Fisher Scientific) / 10% Fetal
600 Bovine Serum (FBS, Biological Industries) sodium pyruvate 1 mM (Sigma-Aldrich) / MEM Non-
601 Essential Amino Acids Solution 0,1 mM (NEAA, Thermo Fisher Scientific) / 2-mercaptoethanol
602 0,1 mM (2-ME, Wako) / penicilline-streptomycine 100 U/mL (Thermo Fisher Scientific). 1000

603 U/mL of Leukemia inhibitory factor (Esgro LIF, Merck), 3 µM of CHIR99021 (BioVision) and 1
604 µM of PD0325901 (Stemgent) were added to the culture medium. These culture conditions
605 for iPSC maintenance are according to [37]. *Generation of iPSC-derived retinal aggregates :*
606 Differentiation of iPSCs into retinal aggregates was done using the SFEBq (serum-free floating
607 culture of embryoid body-like aggregates with quick re-aggregation) method according to [1]
608 with some modifications following [37] and [38]. The iPSCs were dissociated after 4-5 days of
609 maintenance using trypsine 0,25% / 1 mM EDTA (Thermo Fisher Scientific) at 37°C during 2
610 minutes. Embryoid body-like aggregates were formed by adding 5000 cells/dish in a low
611 binding 96 dish-plat (174925 Nunclon™ Sphera™, Thermo Fisher Scientific) in 100 µL of
612 differentiating medium containing 0,1 mM AGN193109 (Toronto Research Chemicals). The
613 differentiating medium is composed of GMEM (Thermo Fisher Scientific), 5% of Knock-out
614 Serum Replacement (KSR, Thermo Fisher Scientific), Sodium Pyruvate 1 mM (Sigma-Aldrich),
615 NEAA 0,1 mM (Thermo Fisher Scientific) and 2-ME 0,1 mM (Wako). Dissociation of iPSCs
616 corresponds to differentiating day 0 (DD0). At DD1, 20 µL of matrigel reduced in growth
617 factors (Corning) was added to obtain a final concentration equal to 2%. The cells were left in
618 this medium till DD8. At DD8, retinal aggregates were picked up and transfered in Petri dishes
619 60-mm in maturation medium composed of Dulbecco's Modified Eagle's Medium (DMEM)/F-
620 12 with glutamax (Thermo Fisher Scientific) completed with 1% of N2 supplement (Thermo
621 Fisher Scientific) and penicilline-streptomycine 100 U/mL (Thermo Fisher Scientific). Some
622 AR 0,5 µM (from DD13 to DD18, Sigma-Aldrich), 1 mM of L-taurine (from DD13 to DD29,
623 Sigma-Aldrich) and FBS 1% (from DD21 to DD29, Biological Industries) were added to this
624 maturation medium. Taurine and AR promote rod photoreceptors differentiation [39]. From
625 DD8 to DD29 cultures were maintained in hyperoxic conditions (37°C, 40% O_2 / 5% CO_2). The
626 development of retinal aggregates was monitored and GFP expression was confirmed from
627 DD18 using a digital inverted fluorescence microscope EVOS FL (Thermo Fisher Scientific).
628
629 ***Immunofluorescence.*** Retinal aggregates were fixed for 20 minutes at room temperature in
630 paraformaldehyde (PFA) 4% diluted in phosphate saline (PBS) at pH 7.4. They were then
631 washed in PBS before cryoprotection in sucrose 30% (diluted in PBS) at 4°C overnight.
632 Eyeballs from wild type C57BL/6 mice, used as positive controls, were enucleated and
633 punctured in the center of the cornea before fixation for 1 hour in PFA 4%, at room
634 temperature, then washed in PBS and added in sucrose 30% at 4° overnight. Samples were

635 embedded in Richard-Allan Scientific™ NEG-50™ medium (Thermo Fisher Scientific). Slices of
636 10 à 15 µm were collected from cryostat on slides Superfrost Ultra Plus (Thermo Fisher
637 Scientific). For immunofluorescence slides were first incubated in blocking solution Blocking
638 One (Nacalai Tesque) for 1 hour at room temperature, then at 4°C overnight with primary
639 antibodies diluted in Dako REAL Antibody Diluent (Agilent). We used the following primary
640 antibodies: rabbit antibody against protein kinase Cα diluted 1:500 (anti-PKCα, Sigma-Aldrich
641 AB_477345), rabbit antibody against recoverine 1:1000 (Merck AB_2253622), rabbit antibody
642 against calretinine 1:500 (Swant AB_2313763), rabbit antibody against Pax6 1:100 (BioLegend
643 AB_2313780), mouse antibody against RET-P1 1:1000 (anti-rhodopsine, Sigma-Aldrich
644 AB_260838), sheep antibody against Chx10 1:1000 (Exalpha Biologicals AB_2314191). After
645 24 hours, slides were washed three times for 5 minutes in PBS-Tween 0,05% then incubated
646 with appropriate secondary antibodies in the dark at room (anti-IgG rabbit A488 et A647, anti-
647 IgG mouse A555 et anti-IgG sheep A555 all from Thermo Fisher Scientific) and some 4',6-
648 diamidino-2-phénylindole (DAPI) 1:1000 diluted in Dako REAL Antibody Diluent. After another
649 wash in PBS-Tween, slides were mounted with FluorSave™ Reagent (Merck). Images were
650 taken with confocal microscope Nikon Eclipse T_i.
651
652 ***Single cell RNA Seq.*** *Dissociation of native retinal tissue and 3D-culture retinal aggregates:*
653 The dissociation of mouse retinas and 3D retinal aggregates was inspired by the protocol of
654 Macosko et al. [40]. Eyeballs of C57BL/6 wild type mice were enucleated at time points E13,
655 P0, P5 and P9. Dissected retinas were placed in Dulbecco's Phosphate Buffered Saline (DPBS,
656 Thermo Fisher Scientific). Optic vesicule (OV)-like structures of the iPSCs derived 3D retinal
657 aggregates were cut at DD13, DD21, DD25 and DD29 and transferred in DPBS as well. Papaïne
658 4 U/mL (Worthington Biochemical Corporation) was added to the samples. The solution
659 containing the retinas and the OV-like structures was maintained at 37°C for 45 and 30
660 minutes, respectively. Ovomucoïde 0,15% (Worthington Biochemical Corporation) was
661 added for papaïne inhibition. Samples were centrifuged in order to eliminate the supernatant
662 and cells were resuspended again in DPBS. The cell suspension was then manipulated for the
663 different steps of single cell RNA sequencing. Cell numbers and proportion of life cells were
664 estimated by Trypan Blue staining and using a Countess II cell counter (ThermoFisher). *scRNA-*
665 *Seq:* We generated two biological replicates for stages 1 to 3 (NaR and 3D-RA) and one
666 biological replicate for stage 4 (NaR and 3D-RA). We loaded ~15,700 cells for biological

667 replicate 1 (stage 1-4) and ~10,000 cells for biological replicate 2 (stage 1-3) in a Chromium
668 Controller instrument (10X Genomics, CA). Sequencing libraries were generated using
669 Chromium Single Cell 3' reagent kits v2.0 following the recommendations of the
670 manufacturer (10X Genomics, CA). Actual sequencing was conducted on an Illumina NextSeq
671 500 instrument (Illumina, CA). *Bioinformatic analyses:* Demultiplexing, alignment, filtering,
672 barcode counting, UMI counting, and aggregation of multiple runs were conducted using Cell
673 Ranger v2.1.1 (10X Genomics, CA). Further filtering, k-means clustering, UMAP projection
674 were conducted using the Seurat software suite (https://satijalab.org/seurat/)[17]. Velocity
675 analysis was performed using the Velocyto R package [7]. Single- cell trajectory inference and
676 pseudo-time analysis was conducted with Monocle2 (http://cole-trapnell-
677 lab.github.io/monocle-release/)[9].
678
679 ***ATAC-Seq.*** *Data generation:* ATAC-seq libraries were constructed on NaR (E13, P0, P5) and
680 3D-RA (DD13, DD21, DD25) samples with biological replicates following the Omni ATAC
681 protocol [29]. We used 50,000 cells per reaction taken from the cell suspensions prepared for
682 the scRNA-seq. We tested two different amounts of the Tagment DNA TDE1 enzyme (1 and 2
683 µl in a 50 µl reaction volume) (Illumina, San Diego, California) per sample. Genomic DNA
684 (gDNA) libraries were also prepared using 50 ng of gDNA isolated from NaR P5 and 3D-RA
685 DD25 cells by following the Nextera DNA Sample Preparation Guide (Illumina). The libraries
686 were purified using the MinElute PCR purification kit (Qiagen, Venlo, Netherlands) followed
687 by 13 and 5 cycles of PCR-amplifications for ATAC-seq and gDNA libraries, respectively. After
688 validating library size distribution using the QIAxcel capillary electrophoresis (Qiagen), the
689 libraries were further purified using the SPRIselect reagent to remove large DNA molecules (a
690 right-side size selection with 0.55X followed by 1.5X ratios of beads) (Beckman Coulter, Brea,
691 California). On average 10.6 millions of 38-nucleotide paired-end sequences were obtained
692 using a NextSeq 500 sequencer (Illumina). *Data analyses:* Data was analyzed by following the
693 ENCODE Kundaje lab ATAC-seq pipeline
694 (https://www.encodeproject.org/pipelines/ENCPL792NWO/). Sequences were trimmed
695 using Trimmomatic [41] and aligned on the Mus musculus genome assembly mm10 using
696 Bowtie2 [42]. After filtering out low quality, multiple mapped, mitochondrial, and duplicated
697 reads using SAMtools [43] and the Picard Toolkit (http://broadinstitute.github.io/picard/),
698 fragments with map length ≤146 bp were kept as nucleosome-free fraction. Genomic loci

699 targeted by TDE1 were defined as 38-bp regions centered either 4 (plus strand reads) or 5-bp
700 (negative strand reads) downstream of the read's 5'-end. ATAC-seq peaks were called using
701 the MACS2 software (narrowPeak; q-value ≤ 0.01) [30]. FRiP scores were calculated as the
702 fraction of TDE1 targeted loci falling into the called peaks. Overlapping peaks across samples
703 were merged and annotated for the occurrence of TF binding motifs of interest (Suppl. Table
704 10) and the closest gene using Homer [31]. TDE1 targeted loci overlapping the merged peaks
705 were extracted and converted to a bedgraph file with a scaling factor to one million reads
706 using BEDTools [44], and further to tdf format to visualize peaks on the Integrative Genomics
707 Viewer [45]. The total number of TDE1 targeted loci overlapping the merged peaks were
708 counted using BEDOPS [46], normalized for peak lengths and a sequencing depth with per
709 one million scaling factor, standardized and used for hierarchical cluster analysis using R
710 hclust [47] and gplots (https://CRAN.R-project.org/package=gplots). The detailed analysis
711 pipeline is provided in the ATAC_seq_analysis_pipeline.docx file. The overall mapping rate
712 with Bowtie2 averaged 98.6%, the mapping rate to the mitochondrial genome 4.1%, the
713 duplicate fragment rate 6.0%, the proportion of usable reads after filtration 83.4%, and the
714 FRiP score 34.1%. The FRiP score was significantly lower for E13 samples (reminiscent of the
715 E14.5 samples in [48]), yet not so in the equivalent DD13 samples (Suppl. Table 16).
716
717 ***Downstream analyses***. *Percentage of 3D-RA cells in cliques and clusters:* The number and
718 proportion of 3D-RA cells in cliques and clusters were computed using a Perl script
719 (Dev_path_width.pl) and the corresponding graphs generated in R (Perc_3D_RA_in_cl.R).
720 *Width of developmental trajectories in 2D UMPA space:* To test whether the developmental
721 trajectories were more tightly regulated in NaR than in 3D-RA we computed the average
722 distance (computed as the Euclidian distance in 2D-UMAP space,
723 i.e.$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$) between 500 randomly selected NaR and 500 randomly
724 selected 3D-RA cells and their n nearest neighbors (with n ranging from 1 to 50). The number
725 of cells per developmental stage was adjusted between NaR and 3D-RA by down sampling to
726 the number of the least populated source. The corresponding calculations were performed
727 five times. The curves shown in Fig. 3H correspond to the averages across the five replicates.
728 The grey confidence zone in Fig. 3H is bounded by the maxima and minima across the five
729 replicates. The corresponding script was written in Perl (Dev_path_width.pl) and the graph

730 generated in R (Path_width.R). *Within developmental stage clique entropy:* To compare
731 clique diversity within developmental stage between NaR and 3D-RA, we first equalized the
732 number of cells with developmental stage between NaR and 3D-RA by randomly dropping
733 cells from the most populated source. We then sampled two cells within cell source (NaR and
734 3D-RA) and developmental stage and checked whether they were from the same clique or
735 not. This was repeated 1,000 times yielding a measure of clique diversity akin to (1-entropy).
736 Down-sampling of cells was repeated 100 times. Each data point in Fig. 3I corresponds to (1-
737 Entropy) for one such random sample. The corresponding script was written in Perl
738 (entropy.pl) and the graph generated in R (Entropy.R). *Pathway analyses:* Pathway
739 enrichment analyses were conducted using the on-line Reactome analysis tools [26,27].
740 Mouse gene identifiers were converted to human counterparts. Pathway analysis results
741 were downloaded as flat files. A total of 392 pathways with enrichment p-value ≤ 0.01 in at
742 least one analysis were kept and manually sorted according to Reactome hierarchy
743 (Man_processed_reactome_output.txt). A pathway is enriched in a list of genes if it contains
744 more components of the pathway than expected by chance (given the number of genes in
745 the list). The overlapping genes ("Found entities") hence define the enrichment. The same
746 pathway can be enriched in two gene lists due to the same, distinct or partially overlapping
747 sets of "found entities". We quantified the degree of overlap between sets of "found
748 entities" for the 1,313 pathway enrichments using principal component (PC) analysis in a
749 space defined by the presence/absence of 1,335 genes. The distance between sets of "found
750 entities" in a space consisting of the 20 first PCs was projected in 3D space using t-distributed
751 stochastic neighbor embedding (tSNE) implemented with the *Rtsne* R function. 3D tSNE
752 coordinates were converted to hexadecimal RGB code and used to color the sets of "found
753 entities" (corresponding to the enrichment of a pathway in a specific gene list) when
754 generating 2D tSNE graphs (Fig. 4D), or when generating a tile showing the pathways enriched
755 in specific analyses (CLIQUE>OTHER, NaR > 3D-RA or 3D-RA > NaR) and cliques within analysis
756 (NE, RPE, ERPC, LRPC, NRPC, RGC, HC, AC, PRP, C, R, BC or MC) (Fig. 4B). The corresponding
757 scripts were written in Perl (Reactome_analysis.pl) and R (Reactome_analysis.R). *Analysis of*
758 *dynamically regulated ("CLIQUES>OTHERS") and differentially regulated ("NaR>3D-RA" and*
759 *"3D-RA>NaR") TF:* TF were extracted from Suppl. Tables 4 ("CLIQUES>OTHERS") and Suppl.
760 Table 7 ("NaR>3D-RA" and "3D-RA>NaR") using a list of mouse TF obtained from the
761 PHANTOM consortium [28] (mouse_TF.txt), and a file formatted using a Perl script

762 (TF_prep.pl) to generate the tile shown in Fig. 5B in R (TF_tile.R). *Identifying regulatory*
763 *toggles:* We used Homer [31] to compile the number of occurrences of 336 binding motifs
764 for 151 of 307 dynamically regulated TF in 98,181 ATAC-Seq peaks assigned to 19,170 genes.
765 For each gene, the data were summarized as (i) the total number of occurrences, and (ii) the
766 mean number of occurrences per peak (i.e. density), for each of the 336 binding motifs (Suppl.
767 Table 11). We then checked - for each of the 336 binding motifs separately - whether the
768 number ("total" in Suppl. Table 11) and density ("mean" in Suppl. Table 11) of motifs differed
769 significantly between genes that were upregulated versus downregulated in every one of the
770 13 cliques. Differential expression analyses to identify genes that are up- and downregulated
771 in specific cliques were performed with the *Findmarkers* function in Seurat
772 (https://satijalab.org/seurat/). The corresponding results are summarized in a series of files
773 labelled, respectively, "NaR/RET_<CELL_TYPE>_markers.txt" for NaR, and
774 "3D_RA/IPS_<CELL_TYPE>_markers.txt" for 3D_RA. We used a threshold q-value of 0.05 to
775 declare a gene as significantly up- or down-regulated in a given clique. The statistical
776 significance of the difference in number and density of binding motifs between up- and down-
777 regulated genes was computed using Wilcoxon rank-based test implemented with the
778 wilcox.test R function. Differences were deemed significant if the q-value (computed with
779 the qvalue R function) was ≤ 0.01. Corresponding results are provided as Suppl. Table 13 for
780 NaR and Suppl. Table 14 for 3D-RA. The graphs for figure 7 were generated using the
781 Comb_scRNA_ATAC_seq R script. All corresponding scripts and datasets are freely available
782 from http://web.giga.ulg.ac.be/pubdata/UAG/Georges_A_2020.
783

784 **AUTHOR CONTRIBUTIONS**

785 Conceived and designed the experiments: AG, MT, MM, HT, MG. Performed the experiments:
786 AG, HT, FL, LK. Analyzed the data: AG, AL, HT, MS, LD, MG. Contributed
787 reagents/materials/analysis tools/supervision: AG, AL, LN, JMR, LD, MS, MT, MG. Wrote the
788 paper: AG, HT, MG.
789

790 **ACKNOWLEDGEMENTS**

791 We are grateful to Tomoyo Hashiguchi for teaching the 3D-RA culture protocol and to
792 Tomohiro Masuda and Akishi Onishi for their comments and suggestions. We thank the GIGA
793 Genomics platform for their help both with scRNA-seq experiments and early bio-informatic

794 analyses (especially Wouter Coppieters and Benoit Charloteaux). This work was financially
795 supported by the "King Baudouin Foundation", the "Global Ophthalmology Awards Program
796 (GOAP) from Bayer", and the "Fonds Leon Fredericq". LD and LN are respectively postdoctoral
797 researcher and senior associate researcher of the FRS-FNRS.
798

799 **DATA AVAILABILITY**

800 All data generated as part of this work are available without restrictions. They have been
801 deposited under accession numbers E-MTAB-9440 and E-MTAB-9395. All data and analysis
802 pipelines are available at http://web.giga.ulg.ac.be/pubdata/UAG/Georges_A_2020. The
803 authors declare that they have no competing interests.
804

805 **REFERENCES**

806 1. Eiraku M, Sasai Y. Mouse embryonic stem cell culture for generation of three-dimensional
807 retinal and cortical tissues. Nat Prot. 2012;7: 69–79.
808 2. Nakano T, Ando S, Takata N, Kawada M, Muguruma K, Sekiguchi K, Saito K, Yonemura
809 S, Eiraku M, Sasai Y. Self-formation of optic cups and storable stratified neural retina
810 from human ES cells. Cell Stem Cell. 2012;10: 771–785.
811 3. Jin ZB, Okamoto S, Osakada F, Homma K, Assawachananont J, Hirami Y, Iwata T, Takahashi
812 M. Modeling retinal degeneration using patient-specific induced pluripotent stem cells.
813 PLoS One. 2011;6:e17084.
814 4. Völkner M, Zschätzsch M, Rostovskaya M, Overall RW, Busskamp V, Anastassiadis K, O
815 Karl M. Retinal organoid from pluripotent stem cells efficiently recapitulate retinogenesis.
816 Stem Cells Rep. 2016;6: 525-538.
817 5. Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, Adiconis X, Levin
818 JZ, Nemesh J, Goldman M, McCarroll SA, Cepko CL, Regev A, Sanes JR. Comprehensive
819 classification of retinal bipolar neurons by single-cell transcriptomics. Cell. 2016;166:
820 1308-1323.
821 6. Saelens W, Cannoodt R, Todorov H & Saeys Y. A comparison of single cell trajectory
822 inference methods. Nat Biotechnol. 2019;37: 547–554.
823 7. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K,
824 Kastriti ME, Lonnerberg P, Furlan A, Fan J, Borm LE, Liu Z, Van Bruggen D, Guo J, He X,

825 Barker R, Sundstrom E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S,
826 Kharchenko P. RNA velocity of single cells. Nature. 2018;560: 494-510.

827 8. Camara PG. Methods and challenges in the analysis of single-cell RNA-sequencing data.
828 Curr Opin Syst Biol. 2018; 7:47-53.

829 9. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ,
830 Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by
831 pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32:381-386.

832 10. Collin J, Queen R, Zerti D, Dorgau B, Hussain R, Coxhead J, Cockell S, Lako M.
833 Deconstructing retinal organoids: single cell RNA-Seq reveals the cellular components of
834 human pluripotent stem cell-derived retina. Stem Cells. 2019;37: 593-598.

835 11. Sridhar A, Hoshino A, Finkbeiner CR, Chitsazan A, Dai L, Haugan AK, Eschenbacher KM,
836 Jackson DL, Trapnell C, Bermingham-McDonogh O, Glass I, Reh TA. Single Cell
837 Transcriptomic Comparison of human fetal retina, hPSC-derived retinal organoids, and
838 long-term retinal cultures. Cell Rep. 2020;30:1644-1659.

839 12. Cowan CS, Renner M, GrossScherf B, Goldblum D, Munz M, Krol J, Szikra T, Papasaikas
840 P, Cuttat R, Waldt A, Diggelmann R, Patino-Alvarez CP, Gerber-Hollbach N, Schuierer
841 S, Hou Y, Srdanovic A, Balogh M, Panero R, Hasler PW, Kusnyerik A, Szabo A, Stadler
842 MB, Orgül S, Hierlemann A, Scholl HPN, Roma G, Nigsch F, Roska B. Cell types of the
843 human retina and its organoids at single-cell resolution: developmental convergence,
844 transcriptomic identity, and disease map. 2020;bioRxiv :
845 https://doi.org/10.1101/703348.

846 13. Yan F, Powell DR, Curtis D & Wong NC. From reads to insight: a hitchhiker's guide to ATAC-
847 Seq data analysis. Genome Biol. 2020;21: 22.

848 14. Gonzalez-Cordero A, West EL, Pearson RA, Duran Y, Carvalho LS, Chu CJ, Naeem A,
849 Blackford SJ, Georgiadis A, Lakowski J, Hubank M, Smith AJ, Bainbridge JWB, Sowden JC,
850 Ali RR. Photoreceptor precursors derived from three-dimensional embryonic stem cell
851 cultures integrate and mature within adult degenerate retina. Nat Biotechnol. 2013;31:
852 741- 747.

853 15. Akimoto M, Cheng H, Zhu D, Brzezinski JA, Khanna R, Filippova E, C T Oh E, Jing Y, Linares
854 JL, Brooks M, Zareparsi S, Mears AJ, Hero A, Glaser T, Swaroop A. Targeting of GFP to

855 newborn rods by Nrl promoter and temporal expression profiling of flow-sorted
856 photoreceptors. Proc Natl Acad Sci USA. 2006;103: 3890-3895.
857 16. Assawachananont J, Mandai M, Okamoto S, Yamada C, Eiraku M, Yonemura S, Sasai Y,
858 Takahashi M. Transplantation of embryonic and induced pluripotent stem cell-derived
859 3D retinal sheets into retinal degenerative mice. Stem Cell Rep. 2014;2: 662–674.
860 17. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic
861 data across different conditions, technologies, and species. Nat Biotech. 2018;36: 411-
862 420.
863 18. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection. J.
864 Open Source Software. 2018;3: 861.
865 19. Clark BS, Stein-O'-Brien GL, Shiau F, Cannon GH, Davis-Marcisak E, Sherman T, Santiago
866 CP, Hoang TV, Rajaii F, James-Esposito RE, Gronostajski RM, Fertig EJ, Goff LA, Blackshaw
867 S. Comprehensive analysis of retinal development at single cell resolution identifies NFI
868 factors as essential for mitotic exit and specification of late-born cells. Neuron. 2019;102:
869 1111-1126.
870 20. Tiroshi I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, Rotem A, C
871 Rodman, Lian C, Murphy G, Fallahi-Sichani M,Dutton-Regester K, Lin JR, Cohen O, Shah P,
872 Lu D, Genshaft A, Hughes TK, Ziegler CGK, Kazer SW, Gaillard A, Kolb KE, Villani AC,
873 Johannessen CM, Andreev AY, Van Allen EM, Bertagnolli M, Sorger PK, Sullivan RJ,Flaherty
874 RT,Frederick DT, Jané-Valbuena J, Yoon CH, Rozenblatt-Rosen O, Shalek AK, Regev A,
875 Garraway LA. Dissecting the multicellular ecosystem of metatastic melanoma by single-
876 cell RNA-seq. Science. 2016;352:189-196.
877 21. Trimarchi JM, Cho SH, Cepko CL. Identification of genes expressed preferentially in the
878 developping peripheral margin of the optic cup. Dev Dyn. 2009;238: 2327.
879 22. Liu J, Reggiani JDS, Laboulaye ME, Pandey S, Chen B, Rubenstein JLR, Krishnaswamy A,
880 Sanes JR. Tbr1 instructs laminar pattering of retinal ganglion cell dendrites. Nat Neurosci.
881 2018;21: 659-670.
882 23. Belliveau MJ, Cepko CL. Extrinsic and intrinsic factors control the genesis of amacrine cells
883 and cone cells in the rat retina. Development. 1999; 126: 555-556.
884 24. Reese B. Development of the retina and optic pathway. Vision Res. 2011;51: 613–632.

885 25. Kuwahara A, Ozone C, Nakano T, Saito K, Eiraku M, Sasai Y. Generation of a ciliary margin-
886 like stem cell niche from self-organizing human retinal tissue. Nat Commun. 2015;6:6286.
887 26. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B,
888 Korninger F, May B, Milacic M, Roca CD, Rothfels K, Sevilla C, Shamovsky V, Shorser S,
889 Varusai T, Viteri G, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P. The reactome
890 pathway knowledgebase. Nucl Ac Res. 2018 ;44: D481-487.
891 27. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J,
892 Gillespie M, Haw R, Loney F, May B, Milacic M, Rothfels K, Sevilla C, Shamovsky V, Shorser
893 S, Varusai T, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P. The reactome pathway
894 knowledgebase. Nucleic Acids Res. 2020;48: D498-D503.
895 28. Kanamori M, Konno H, Osato N, Kawai J, Hayashizaki Y, Suzuki H. A genome-wide and
896 nonredundant mouse transcription factor database. Biochem Biophys Res Commun.
897 2004;322:787-793.
898 29. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S,
899 Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC,
900 Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ & Chang
901 HY. An improved ATAC-seq protocol reduces background and enables interrogation of
902 frozen tissues. Nat Methods. 2017;14:959-962.
903 30. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM,
904 Brown M, Li W & Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol.
905 2008;9:R137.
906 31. Heinz S, Benner C, span N, Bertolino E, Lin YC, Laslo P, cheng JX, Murre C, Singh H, Glass
907 CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory
908 elements required for macrophage and B cell identities. Mol Cell. 2010; 38:576-589.
909 32. Dixit A, Parnas O, Li B, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T,
910 Raychowdhury R,Adamson B, Norman TM, Lander ES, Weissman JS, Friedman N, Regev A.
911 Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled
912 genetic screens. Cell. 2016;167: 1853-1866.
913 33. Filtz TM, Vogel WK, Leid M. Regulation of transcription factor activity by interconnected,
914 post-translational modifications. Trends Pharmacol Sci. 2014;35: 76-85.

915 34. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR,
916 Weirauch MT. The human transcription factors. Cell. 2018;172: 650-665.
917 35. Ernst J, Melnikov A, Zhang X, Wang L, Rogov P, Mikkelsen TS, Kellis M. Genome-scale high-
918 resolution mapping of activating and repressive nucleotides in regulatory regions. Nat
919 Biotechnol. 2016;34:1180-1190.
920 36. Homma K, Okamoto S, Mandai M, Gotoh N, Rajasimha HK, Chang YS, Chen S, Cogliati T,
921 Swaroop A, Takahashi M. Developing rods transplanted into the degenerating retina of
922 Crx-knockout mice exhibit neural activity similar to native photoreceptors. Stem Cells.
923 2013;31:1149–1159.
924 37. Iwasaki Y, Sugita S, Mandai M, Yonemura S, Onishi A, Ito SI, Mochizuki M, Ohno-Matsui K,
925 Takahashi M. Differentiation/purification protocol for retinal pigment epithelium from
926 mouse induced pluripotent stem cells as a research tool. PLoS One. 2016;11: 1–20.
927 38. Eiraku M, Takata N, Ishibashi H, Kawada M, Sakakura E, Okuda S, Sekiguchi K, Adachi T &
928 Sasai Y. Self-organizing optic-cup morphogenesis in 3D culture. Nature. 2011;472: 51-56.
929 39. Osakada F, Ikeda H, Mandai M, Wataya T, Watanabe K, Yoshimura N, Akaike A, Sasai Y &
930 Takahashi M. Toward the generation of rod and cone photoreceptors from mouse,
931 monkey and human embryonic stem cells. Nat Biotechnol. 2008;26: 215–224.
932 40. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR,
933 Kamitaki N, Martersteck, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll
934 SA. Highly parallel genome wide expression profiling of individual cells using nanoliter
935 droplets. Cell. 2015;161: 1202:1214.
936 41. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
937 data. Bioinformatics. 2014; 30: 2114-2120.
938 42. Langmead B & Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
939 2012;9: 357-359.
940 43. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R.
941 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format
942 and SAMtools. Bioinformatics. 2009;25: 2078-2079.
943 44. Quinlan AR & Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
944 features. Bioinformatics. 2010;26: 841-842.

945 45. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G & Mesirov JP.
946 Integrative genomics viewer. Nat Biotechnol. 2011;29:24-26.
947 46. Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano
948 MT, Vierstra J, Thomas S, Sandstrom R, Humbert R, Stamatoyannopoulos JA. BEDOPS: high
949 performance genomic feature operations. Bioinformatics. 2012;28: 1919-1920.
950 47. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: whihc
951 algorithms implement Ward's criterion ? J Classification. 2014;31:274-295.
952 48. Aldiri I, Ajioka I, Xu B, Zhang J, Chen X, Benavente C, Finkelstein D, Johnson D, Akiyama J,
953 Pennacchio LA, Dyer MA. Brg1 coordinates multiple processes during retinogenesis and is
954 a tumor suppressor in retinoblastoma. Development. 2015;142: 4092-4106.
955