Combined analysis of single cell RNA-Seq and ATAC-Seq data reveals 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

We herein 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 in virtual reality, and RNA velocity analysis 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 at three of the four same developmental stages identifying ~125,000 peaks. The number of ATAC-Seq peaks increased with developmental stage in native but not in iPSC-derived retina. We combine single-cell RNA Seq with ATAC-Seq information and demonstrate that approximately halve of the transcription factors that are dynamically

regulated during retinal development act as repressors rather than activators. We provide evidence 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.

Introduction

It has recently become possible to recapitulate retinal development from induced pluripotent stem cells (iPSCs) in human and mice (Eiraku et al., 2011; Nakano et al., 2012). 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 induced 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 (Jin et al., 2011).

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 (f.i. Völkner et al., 2016). 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 (Shekhar et al., 2016). It allows to refine developmental trajectories by identifying cells occupying intermediate states connecting clusters in multidimensional expression space (Saelens et al., 2019) 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") (La Manno et al., 2018). Genes that are defining cellular sub-types can be pinpointed by differential expression analysis between

clusters (Camara et al., 2018), 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 (Trapnell et al., 2014). Recently, scRNA-Seq has been used to compare transcriptome dynamics during native and embryonic stem cells (ESC)- or iPSC-derived retinal development (f.i. Collin et al., 2019; Sridhar et al., 2020; Cowan et al., 2020). This has revealed comparable cellular composition at equivalent ages with, however, some differences in gene expression of particular cell types and as well as structural differences of inner retinal lamination that seems disrupted in advanced organoid stages compared with fetal retina (Sridhar et al., 2020).

Here we report on 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; Yan et al., 2020) collected at three of the four above-mentioned developmental stages, we provide strong evidence for the operation of transcription factor (TF)-based regulatory toggles that lock the transcriptome of distinct constellations of cellular sub-types in both native and iPSC-derived retina likely 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 (Gonzalez-Cordero et al., 2013). 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 (Akimoto et al., 2006) following Eiraku et al. (2011) and Assawachananont et al. (2014) (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

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

We first analyzed all sequencing data jointly (i.e. NaR and 3D-RA) to cover a maximum of intermediate developmental stages and hence generate the most continuous manifold possible. We used Canonical Correlation Analysis (CCA) implemented with Seurat (Butler et al., 2018) to align the NaR and 3D-RA datasets based on the expression profiles of 1,253 selected "most variable" genes (Suppl. Table 2). We projected the corresponding 30-dimensional distances between cells in 2D- and 3D-space using UMAP (McInnes et al., 2018). We assigned all 38,091 cells jointly (i.e. NaR and 3D-RA) to 70 clusters by k-means clustering (Fig. 2A). We defined gene expression signatures for 13 previously recognized retinal cell types using published information (Clark et al., 2019) (Suppl. Fig. 1 and Suppl. Table 3), and regrouped the clusters accordingly in 13 cliques corresponding to neuroepithelium (NE), retinal pigmented epithelium (RPE), early (ERPC), late (LRPC), and neurogenic retinal progenitor cells (NRPC), retinal ganglion cells (RGC), horizontal cells (HC), amacrine cells (AC), photoreceptor precursor cells (PRP), cones (C), rods (R), bipolar cells (BC), and Müller cells (MC) (Fig. 2B). Using previously defined expression signatures specific for the S and G2-M phases of the cell cycle (Tirosh et al., 2016), we defined sub-cliques of actively dividing ERPC, LRPC and NRPC (Suppl. Fig. 2). Clusters 69 (NE) and 60 (ERPC) form a branch that clearly separates from the rest of NE and ERPC. The two clusters share enhanced expression of *Ccnd2* and *Hmga2,* typical of the NE/ciliary marginal zone (CMZ) (Trimarchi et al., 2009) to which they were therefore assigned (Fig. 2B and Suppl. Fig. 1). One cluster (55) did not light up with any of the 14 signatures. Differential expression analysis (see hereafter) revealed that the *Tbr1* transcription factor is exclusively expressed in this cluster hence identifying it as a recently described subgroup of $Tbr1^{+}$ RGC cells located in the inner plexiform layer (Liu et al., 2018). We combined information from (i) the cells' developmental stage (Fig. 2C), (ii) RNA velocity analyses (La Manno et al., 2018) (Fig. 2D), and (iii) the 3D UMAP manifold (Fig. 2E and Suppl. Video: http://www.sig.hec.ulg.ac.be/giga), to deduce developmental trajectories based on our data (Fig. 2F and Suppl. Fig. 3). Briefly, NE generates both RPE and ERPC. ERPC

(detectable during stage 1) mature into LRPC (detectable during stage 2 and 3). Both ERPC and LRPC are actively dividing, and spawn post-mitotic NRPC who sequentially differentiate in RGC (stage 1), HC (stage 1 and 2), AC (stage 2 and 3), and PRP (stage 2 and 3). PRP complete maturation in C, R and BC, while remaining LRPC transform in MC, at stage 4. Our data suggest the occurrence of "re-specification" (Belliveau & Cepko, 1999) of differentiated AC into both HC and PRP. A cellular bridge clearly connects NE with Tbr1^{+} RGC suggesting that this peculiar cell population may directly derive from the neuroepithelium (Fig. 2E). Thus, the joint analysis of serial NaR and 3D-RA scRNA-seq largely recapitulates previously described canonical developmental routes of the mammalian retina and their timing (Reese et al., 2011, Clark et al., 2019), yet reveals evidence for re-specification and the unconventional origin of Tbr1^{+}RGC.

Comparison of NaR and 3D-RA cell fates in UMAP space highlights commonalities and differences in cell developmental trajectories.

We then focused on the comparison between the behavior of NaR and 3D-RA derived cells. Global comparison of the distribution of Nar and 3D-RA cells across the manifold indicates that in vitro neuroretinal differentiation from iPSCs largely recapitulates native development (Fig. 3A-D). This is substantiated by noting that 82% of clusters and 86% of cliques (defined as above) contain at least 10% of the least represented cell origin (NaR vs 3D-RA) (Fig. 3E-F). More granular examination, however, reveals noteworthy differences. The first one is the occurrence of NaR- or 3D-RA specific clusters and cliques: (i) the RPE clique is almost exclusively composed of 3D-RA cells as a result of RPE elimination from NaR by dissection - (which cannot be done with 3D-RA as self-patterning of the NE into presumptive RPE occurs at a molecular but not structural level in 3D-RA; Völkner et al., 2016); (ii) the CMZ is absent in 3D-RA as expected (only recently were culture conditions established for inducing selective CM retinal differentiation in human iPSC-derived RA; Kuwahara et al., 2015) ; (iii) AC cluster 65, which seems to drive re-specification to HC cells, is only present in 3D-RA; (iv) the Tbr1^{+}RGC clique (cluster 55) is largely dominated by 3D-RA cells which can be explained by a lower proportional representation of Tbr1+ cells in NaR compared to 3D-RA at the analyzed stages as confirmed by immunohistochemistry (Fig. 3G). The second difference is the apparent relaxation of pseudo-spatial and pseudo-temporal transcriptome control in 3D-RA

versus NaR. The developmental pathways traversed par NaR cells appear indeed tighter than those of 3D-RA cells, while NaR cells sampled at a specific developmental stage seem to occupy fewer cliques than 3D-RA cells. To quantify the former, we down-sampled cells to equalize NaR and 3D-RA numbers (within developmental stage) and computed the average distance from the *n* closest neighbors, which was indeed highly significantly shorter for NaR than for 3D-RA (Fig. 3H). To quantify the latter, we measured clique entropy within stages, which was indeed significantly lower in NaR than in 3D-RA for all four stages (Fig. 3I). The last noteworthy difference between both systems is the observation that PRP arise earlier in 3D-RA than in NaR and accumulate at the expense of other cell types (particularly RPC), yet partially fail terminal differentiation particularly into BP cells (Fig. 3J).

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 clique against all others, first considering NaR cells only. We identified a total of 7,292 genes with significantly higher expression in a given clique compared to all other merged cliques (log-fold change ≥ 0.25 and p-value ≤ 0.001) hereafter referred to as "clique-specific" 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 (Trapnell et al., 2014) yielded comparable results (Suppl. Fig. 4 and Suppl. Table 5). We then searched for enriched Reactome pathways (Fabregat et al., 2018; Jassal et al., 2020) in the lists of "clique-specific" genes. Two hundred sixty-eight pathways were significantly enriched ($q \leq 0.01$) in at least one clique (Fig. 4B and Suppl. Table 6). These corresponded primarily to: (i) accelerated cell division in ERPC, LRPC and NRPCs, (ii) intense post-transcriptional and translational activity in NE, ERPC, LRPC and NRPCs, (iii) activation of RHO GTPase- and NOTCH-dependent signaling in ERPC, LRPC, NRPCs, RGC and LRPC, NRPCs, respectively, as well as the GPCR-dependent phototransduction cascade in C and R, (iv) activation of mitochondrial citric acid (TCA) cycle and respiratory electron transport in HC, C, R, BC, and MC, of cholesterol synthesis in ERPC and RGC, and of insulin- and glucagon-dependent metabolic integration in RGC and AC, (v) enhanced remodeling of the extracellular matrix in NE, RPE and MC, and GAP junction trafficking in RGC, and (vi) activation of ROBO receptors-dependent axon guidance

in NE, ERPC and LRPC, and of synapse formation in RGC, HC, AC and BC (Fig. 4B). Four hundred and twenty-six genes were more strongly expressed in the CMZ when compared to all other cliques (Suppl. Table 4). Strikingly 14 members of the crystallin gene family ranked amongst the top 24 genes (Suppl. Fig. 5). CMZ-overexpressed genes highlighted 47 Reactome pathways, corresponding largely to the pathways overexpressed in the NE and ERPC to which the CMZ was initially attached (Suppl. Table 6 and Fig. 2). Six hundred and thirty-three genes were more strongly expressed in Tbr1^{+}RGC (3D-RA) when compared to all other cells (NaR + 3D-RA), with *Tbr1* as top ranking differentially expressed gene (Suppl. Table 4 and Suppl. Fig. 6). Twelve Reactome pathways were significantly enriched ($q \leq 0.01$) in the list of Trb1^{+}RGC-specific genes pertaining to: (i) axon guidance, (ii) L1CAM interactions, and (iv) membrane trafficking, which are in accordance with the role of Tbr1+ cells instructing the organization of dendrites of RGCs in the outer plexiform layer of the developing retina (Suppl. Table 6).

At first sight, genes that were differentially expressed between cliques in NaR appeared to recapitulate their in vivo expression profile remarkably well in 3D-RA (Fig. 4A). Yet, to better apprehend the differences between in vivo and in vitro retinal differentiation, we performed differential expression analysis between Nar and 3D-RA within cliques. For each clique, we generated lists of genes that were respectively under- and over-expressed in 3D-RA when compared to NaR ($q \leq 0.01$; Suppl. Table 7). We then searched for biological pathways that were over-represented in the corresponding gene lists using Reactome. This yielded 197 downregulated and 134 upregulated pathways in 3D-RA (Fig. 4B and Suppl. Table 8). Strikingly, both down- and upregulated pathways showed considerable overlap with the pathways shown to be dynamically regulated during in vivo retinal 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 (Clique > 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, Hedeghog, 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).

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

Combined analysis of scRNA-Seq and bulk ATAC-Seq data reveals regulatory toggles locking mutually exclusive transcriptome states operating 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 (Corces et al., 2017) on bulk NaR (E13, P0, P5) and 3D-RA (DD13, DD21, DD25) samples to identify gene-switch components accessible to 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 (Zhang et al., 2008)(Suppl. Table 10). Of these, 93,386 (75.6%) of these 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 (3.8%; stage I: 294, stage II: 82, stage III: 4,327) developmental stage-specific. 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 dynamically regulated TF (Homer; Heinz et al., 2010), amounting to a total of 336 motifs (average number of motifs per TF: 2.3; range: 1 - 14). We used Homer (Heinz et al., 2010), to annotate our catalogue of ATAC-Seq peaks for the corresponding motifs. In total Homer identified 7,128,225 putative binding motifs in 98,181 ATAC-seak peaks assigned to 19,171 genes (Suppl. Table 11). To test whether dynamically regulated TF that were overexpressed in a given clique (defined as above, hence corresponding to a cell type) were indeed activating downstream targets in that clique ("autologous" TF), we searched for an enrichment of the cognate binding motifs in the ATAC-Seq peaks mapping in the vicinity of the genes that were overexpressed in that same clique ("clique-specific" genes as described above). Using a lenient significance threshold (uncorrected $p \leq 0.01$), we found this to be the case in 115/646 (=17.8%) of testable instances (i.e. testing the enrichment of a motif corresponding to an overexpressed TF in ATAC-Seq peaks of all genes overexpressed in the corresponding clique) hence well above background expectations (red bars in Fig. 6E). Two additional striking observations were made. The first is that there were 88 instances where the binding motif for a TF that was overexpressed in a given clique was significantly depleted ($p \leq 0.01$) in the ATAC-Seq peaks of the genes that were overexpressed in that clique (dark blue bars in Fig. 6E). This is expected if the TF acts as an inhibitor rather than an activator (and ATAC-Seq peaks as silencers rather than enhancers). Activator effects were significantly ($p \leq 0.05$) more common than inhibitory effects in MC, PRP, C and R, while the reverse was true in NE, RPE, ERPC, RGC, AC, HC and BC (Fig. 6E). The second noteworthy observation is that enrichment and depletion for specific binding motifs amongst overexpressed genes was

remarkably common even in cliques in which the cognate TF was not significantly overexpressed ("heterologous" TF). Using the same threshold as before (uncorrected $p \leq 0.01$), we observed 448 (=10.2%) significant enrichments (pink bars in Fig. 6E) and 845 (=19.6%) significant depletions (light blue bars in Fig. 6E) for the binding motifs of an "heterologous" TF out of ~13x336=4,368 testable instances. There was a striking correlation between the ratio of enrichment over depletion for binding motifs of autologous and heterologous TF (Fig. 6E).

We reasoned that this unexpected pattern could reflect the operation of "regulatory toggles" in which mutually exclusive subsets of the transcriptome, defining for instance a precursor (P) and differentiated (D) cell state, would be controlled by mutually exclusive sets of activator/inhibitor TFs (Fig. 6D). Such a scenario would indeed account for the observation that genes overexpressed in cell state P would not only be enriched (respectively depleted) in binding motifs for activators (a_p) (respectively inhibitors, i_p) expressed in cell state P, but also enriched (respectively depleted) in bindings motifs for inhibitors (i_d) (respectively activators, a_d) expressed in cell state D (and vice versa). To search for mutually exclusive cell states that would be controlled by such regulatory toggles, we analyzed the correlation between the levels of enrichment/depletion (measured by the corresponding -log(p) value) of the 336 binding motifs in overexpressed genes for all pairs of cliques. We predicted that for cliques corresponding to mutually exclusive states (blue and orange states in Fig. 6D) the correlation would be negative (blue state: enrichment of a_p and i_d, depletion of i_p and a_d; orange state: enrichment of i_p and a_d, depletion of a_p and i_d in Fig. 6D), while for cliques corresponding to the same state (blue or orange states in Fig. 6D) the correlation would be positive (2 x blue: consistent enrichment of a_p and i_d, and depletion of i_p and a_d; 2 x orange: consistent enrichment of i_p and a_d, and depletion of a_p and i_d in Fig. 6D). This analysis revealed three major clique groups which we will refer to hereafter as "constellations": (C1) ERPC, LRPC and NRPC, (C2) HC, RGC and AC, and (C3) PRP, R, C and BC (Fig. 6E). Cliques within constellations were characterized by highly correlated patterns of binding motif enrichment/depletion (red in Fig. 6F). NE and MC appeared more isolated yet closer to C1 for NE, and to C3 for MC (results for RPE were considered with caution as the NaR samples contained few RPE cells; see above). Conversely, the enrichment/depletion patterns of C1 and C3 cliques were strongly negatively correlated (blue in Fig. 6F), indicative of the operation of the above-postulated "regulatory toggles" locking mutually exclusive C1 and C3

transcriptome states. Enrichment/depletion patterns of C2 cliques appeared uncorrelated with those of C3 cliques, and mildly negatively correlated with those of C1 cliques suggestive of possibly less tight C1 to C2 toggles. The pattern of MC was also negatively correlated with those of C1 cliques. Examples of TF, including both activators and inhibitors, that may underpin the postulated regulatory toggles and the constellations that they may hereby differentiate are shown in Table 1 and Fig. 6G. They include TF that were previously known to play key roles in retinal development as well as at least seven TF not yet described in retinal context (Nfyb, Tcf4, Arid3b, Foxo3, Mafg, Zbtb7a and Etv1). All results are provided in Supplemental Fig. 8. The possibility to operate as suppressor has been reported for several of the TF inferred to act as inhibitors in this analysis including f.i. Zbtb7a (Laudes et al., 2004), Tgif1&2 (Satoh et al. (2008), and Tcf3 (Kuwahara et al., 2014), hence supporting the validity of our hypothesis and approach. One TF (Otx2) appears to act as transcriptional activator in PRP (enrichment of binding motifs with $p = 3 \times 10^{-24}$), and as a transcriptional repressor in NRPC (depletion of binding motifs with $p = 9 \times 10^{-6}$) (Supplemental Fig. 8).

We repeated the same analysis on 3D-RA data. We obtained comparable patterns of activator and inhibitory effects across cliques when compared to NaR, with a 68:81 ratio of activator versus inhibitory effects for autologous TF, and a 279:710 ratio for heterologous TF (Fig. 6H). The most striking differences pertained to MC for which 28 activator effects were detected in NaR yet none in 3D-RA. All results are provided in Supplemental Fig. 9. Analyzing the correlations of binding motif enrichment/depletion between cliques revealed an even more pronounced clustering of cliques in three constellations (Fig. 6I). NE appeared as an integral part of C1. RPE (with many more cells in 3D-RA than in NaR) were closer to C1 than to C3 (contrary to NaR). MC appeared more isolated from all other cliques than in NaR with neither stout positive nor negative correlations with any other clique.

To further compare NaR and 3D-RA, we computed the correlations between the vectors of binding motif enrichment/depletion as describe above but for NaR vs 3D-RA clique apirs. This revealed that (i) 3D-RA HC resemble NaR RGC and AC more than NaR HC, (ii) 3D-RA cones resemble NaR rods more than NaR cones, (iii) 3D-RA NE share features with NaR MC, (iv) 3D-RA MC, BC and (to a lesser extend) R and C share features with NaR HC, and (v) 3D-RA MC share features with NaR BC and to a lesser extend R, C and PRP (Fig. 6J). We did not find striking perturbations of enrichment/depletion patterns of specific binding motifs in 3D-RA that may explain these differences, suggesting that these alterations may result from the

combined effect of subtle alterations of multiple toggle components. One example of induction of a C1 activator (Etv5; see Table 1) in MC that may significantly affect the MC transcriptome is shown in Fig. 6J.

Discussion

We herein use scRNA-seq to compare the unfolding of the epigenetic program in *in vivo* versus *in vitro* (from iPS cells) derived murine retina at four match stages of development encompassing the presumed emergence times of the major retinal cell types. Results obtained by combining information from (i) the analysis of four developmental stages (E13 vs DD13, P0 vs DD21, P5 vs DD25 and P9 vs DD29), (ii) 3D UMAP manifolds visualized in virtual reality (http://www.sig.hec.ulg.ac.be/giga), and (iii) RNA velocity analysis, are in good agreement with the previously reported main retinal developmental trajectories (Fig. 2F). More specifically, we confirm the existence of a recently characterized RGC subpopulation that specifically expresses the Tbr1 transcription factor. We provide evidence that this subpopulation may directly derive from NE via a narrow, proprietary developmental bridge. We provide evidence supporting the occurrence of « respecification » of AC into both HC and PRP. It is noteworthy that the evidence for both the unconventional origin of Tbr1+ RGC and respecification of AC cells to HC cells stems from 3D-RA cell populations only. Further work is needed to confirm that these phenomena also occur in NaR. We identify >7,000 genes that are dynamically regulated during *in vivo* retinal differentiation corresponding to tens of biological pathways pertaining to the cell cycle, gene expression, signal transduction, metabolism, cell biology and development (Fig. 4). Dynamically regulated genes include ~300 TF, of which ~100 are at least 1.5x as strongly expressed in one specific retinal cell type when compared to all other ones. The latter include 22 TF not yet described in the field of retinal development which could serve as a starting point for functional investigations of the roles of these genes in retinogenesis and physiology.

By combining scRNA-Seq with ATAC-Seq data generated on bulk samples corresponding to three of the analyzed developmental stages we provide evidence (based on significant underrepresentation of binding motifs in ATAC-Seq peaks close to overexpressed genes) that nearly halve of the TF act as transcriptional suppressors rather than transcriptional activators in the cells in which they are overexpressed (referred to as autologous TF). We make the

unexpected observation that binding motifs corresponding to TF which are not overexpressed in the corresponding cell type (heterologous TF) may nevertheless be very significantly enriched or depleted in ATAC-Seq peaks of overexpressed genes. We interpret this observation as reflecting the operation of "regulatory toggles" that lock cells in distinct stable cell states (Fig. 6D). We provide evidence (from correlation analysis of enrichment/depletion vectors between clique pairs) that retinal development relies on three main toggles, corresponding respectively to ERPC, LRPC and NRPC (C1), RGC, AC and HC (C2), and PRP, C, R and BC (C3) (Fig. 6F). Detailed analysis of the enrichment/depletion patterns indicates how specific TF may further contribute to the stabilization of sub-states within constellations (Table 1). The signals of binding motif depletion reported in Fig. 6 were obtained by quantifying the density of binding motifs in ATAC-Seq peaks between genes that are significantly overexpressed in a given clique versus those that are significantly under-expressed in the same clique. A very similar signal was obtained when contrasting genes that were overexpressed in a given clique with either genes that were not dynamically regulated during retinal development (i.e. not significantly overexpressed in any clique), or with all genes (data not shown). This suggests that the corresponding motifs are "actively" removed from the corresponding genes during evolution by purifying selection. A decrease (over genome-wide background) of the density of binding motifs for TF determining alternate cell states within a tissue would indeed reduce the probability of mis-expression of genes that may confuse cell state. Our "regulatory toggle" model makes a number of testable predictions. As an example, the outcome of Fig. 6F (defining C1, C2 and C3), suggests specific contrasts for differential expression analysis (f.i. ERPC+LRPC+NRPC versus PRP+C+R+BC) in order to better define the corresponding state-specific genes (blue and orange gene sets in Fig. 6D) including activators and inhibitors. The signature of binding motif enrichment and depletion should be even stronger for these genes sets. Also, the function of the TFs that are candidate components of "regulatory toggles" (f.i. Table 1) could be tested using knock-out/down experiments conducted in 3D-RA. Our results suggest that "regulatory toggles" involve sets of multiple, possibly redundant activators/inhibitors (rather than only one activator/inhibitor pair). It might therefore be necessary to perform pooled screens using CRISPR libraries targeting several candidates at once (at high multiplicity of infection) in Perturb-Seq like experiments conducted in 3D-RA (Dixit et al., 2016) in order to induce detectable alterations in cellular behavior.

We show that 3D-RA broadly recapitulate the in vivo developmental program and trajectories. However, developmental trajectories appear less canalized in 3D-RA when compared to NaR, PRP to develop earlier and at the expense of other cell types, and terminal differentiation of BC to be incomplete (Fig. 3). We identify ~3,000 genes that are differentially regulated between 3D-RA and NaR in at least one cell type, and identify the corresponding biological pathways pertaining in particular to the rate of cell division which is reduced in 3D-RA RPCs when compared to NaR, post-transcriptional and translational mechanisms which appear exacerbated in the majority of 3D-RA cliques when compared to NaR, signal transduction via WNT and Notch pathways which are diminished in 3D-RA RPCs when compared to NaR, 3D-RA differentiated cells which appear less functional with less phototransduction cascade activity and decrease synapse formation, and finally apoptosis and stress response which are increased at the latest stages of 3D-RA culture. Strikingly, the perturbed pathways show a highly significant overlap with those that were shown to be dynamically regulated during the development of the native retina. We show that transcription factors that are dynamically regulated during in vivo retinal development are particularly sensitive to the iPSC culture conditions. This is likely to drive the perturbations of the above-mentioned biological pathways. We have examined the status of the above-mentioned retinal "regulatory toggles" in 3D-RA. The three main constellations (C1, C2 and C3) were dominating the 3D-RA landscape even more than the NaR landscape, indicating that some developmental subtleties might be lost during in vitro development (as expected). Monitoring the toggle landscape may become a valuable approach to monitor how closely 3D-RA systems recapitulate native development. Comparison of NaR versus 3D-RA enrichment/depletion vectors revealed hints of a number of interesting differences between the two systems, particularly with regards to 3D-RA RPE which appear closer to C1 than to C3 suggestive of a rather "presumptive 3D-RA RPE" status rather than a fully differentiated RPE status as observed in NaR, 3D-RA HC which respond more like NaR RGC and AC than NaR HC, 3D-RA cones which respond more like NaR rods than cones, and 3D-RA MC which have striking similarities with NaR HC.

Materials and methods

Generation of iPSC-derived retinal aggregates. *Maintenance of iPSCs:* The mouse iPSC-NrlGFP line was obtained from the laboratory for retinal regeneration from the Center for

Bioloy and Development (CDB) RIKEN (Kobe, Japan). These iPSCs were generated from fibroblasts (Homma et al. 2013) of C57BL/6 Nrl-eGFP transgenic mice (Akimoto et al. 2006). The cells were cultivated in 60-mm Petri dishes (0,6 x 10^5 cells total per dish) coated with gelatine 0,1% (G2625, Sigma-Aldrich) in Glasgow's Minimum Essential Medium (GMEM, Thermo Fisher Scientific) / 10% Fetal Bovine Serum (FBS, Biological Industries) sodium pyruvate 1 mM (Sigma-Aldrich) / MEM Non-Essential Amino Acids Solution 0,1 mM (NEAA, Thermo Fisher Scientific) / 2-mercaptoethanol 0,1 mM (2-ME, Wako) / penicilline-streptomycine 100 U/mL (Thermo Fisher Scientific). 1000 U/mL of Leukemia inhibitory factor (Esgro LIF, Merck), 3 µM of CHIR99021 (BioVision) and 1 µM of PD0325901 (Stemgent) were added to the culture medium. These culture conditions for iPSC maintenance are according to (Iwasaki et al. 2006). *Generation of iPSC-derived retinal aggregates :* Differentiation of iPSCs into retinal aggregates was done using the SFEBq (serum-free floating culture of embryoid body-like aggregates with quick re-aggregation) method according to Eiraku et al. (2011) with some modifications following Assawachananont et al. (2014) and Iwasaki et al. (2016). The iPSCs were dissociated after 4-5 days of maintenance using trypsine 0,25% / 1 mM EDTA (Thermo Fisher Scientific) at 37°C during 2 minutes. Embryoid body-like aggregates were formed by adding 5000 cells/dish in a low binding 96 dish-plat (174925 Nunclon™ Sphera™, Thermo Fisher Scientific) in 100 µL of differentiating medium containing 0,1 mM AGN193109 (Toronto Research Chemicals). The differentiating medium is composed of GMEM (Thermo Fisher Scientific), 5% of Knock-out Serum Replacement (KSR, Thermo Fisher Scientific), Sodium Pyruvate 1 mM (Sigma-Aldrich), NEAA 0,1 mM (Thermo Fisher Scientific) and 2-ME 0,1 mM (Wako). Dissociation of iPSCs corresponds to differentiating day 0 (DD0). At DD1, 20 µL of matrigel reduced in growth factors (Corning) was added to obtain a final concentration equal to 2%. The cells were left in this medium till DD8. At DD8, retinal aggregates were picked up and transfered in Petri dishes 60-mm in maturation medium composed of Dulbecco's Modified Eagle's Medium (DMEM)/F-12 with glutamax (Thermo Fisher Scientific) completed with 1% of N2 supplement (Thermo Fisher Scientific) and penicilline-streptomycine 100 U/mL (Thermo Fisher Scientific). Some AR 0,5 µM (from DD13 to DD18, Sigma-Aldrich), 1 mM of L-taurine (from DD13 to DD29, Sigma-Aldrich) and FBS 1% (from DD21 to DD29, Biological Industries) were added to this maturation medium. Taurine and AR promote rod photoreceptors differentiation (Osakada et al. 2008). From DD8 to DD29 cultures were maintained in hyperoxic conditions (37°C, 40% O_2 / 5% CO_2). The development

of retinal aggregates was monitored and GFP expression was confirmed from DD18 using a digital inverted fluorescence microscope EVOS FL (Thermo Fisher Scientific).

Immunofluorescence. Retinal aggregates were fixed for 20 minutes at room temperature in paraformaldehyde (PFA) 4% diluted in phosphate saline (PBS) at pH 7.4. They were then washed in PBS before cryoprotection in sucrose 30% (diluted in PBS) at 4°C overnight. Eyeballs from wild type C57BL/6 mice, used as positive controls, were enucleated and punctured in the center of the cornea before fixation for 1 hour in PFA 4%, at room temperature, then washed in PBS and added in sucrose 30% at 4° overnight. Samples were embedded in Richard-Allan Scientific™ NEG-50™ medium (Thermo Fisher Scientific). Slices of 10 à 15 µm were collected from cryostat on slides Superfrost Ultra Plus (Thermo Fisher Scientific). For immunofluorescence slides were first incubated in blocking solution Blocking One (Nacalai Tesque) for 1 hour at room temperature, then at 4°C overnight with primary antibodies diluted in Dako REAL Antibody Diluent (Agilent). We used the following primary antibodies: rabbit antibodie against proteinkinase Cα diluted 1:500 (anti-PKCα, Sigma-Aldrich AB_477345), rabbit antibodie against recoverine 1:1000 (Merck AB_2253622), rabbit antibodie against calretinine 1:500 (Swant AB_2313763), rabbit antibodie against Pax6 1:100 (BioLegend AB_2313780), mouse antibodie against RET-P1 1:1000 (anti-rhodopsine, Sigma-Aldrich AB_260838), sheep antibodie against Chx10 1:1000 (Exalpha Biologicals AB_2314191). After 24 hours, slides were washed three times for 5 minutes in PBS-Tween 0,05% then incubated with appropriate secondary antibodies in the dark at room (anti-IgG rabbit A488 et A647, anti-IgG mouse A555 et anti-IgG sheep A555 all from Thermo Fisher Scientific) and some 4',6-diamidino-2-phénylindole (DAPI) 1:1000 diluted in Dako REAL Antibody Diluent. After another wash in PBS-Tween, slides were mounted with FluorSave™ Reagent (Merck). Images were taken with confocal microscope Nikon Eclipse T_i.

Single cell RNA Seq. *Dissociation of native retinal tissue and 3D-culture retinal aggregates:* The dissociation of mouse retinas and 3D retinal aggregates was inspired by the protocol of Macosko et al. (2015). Eyeballs of C57BL/6 wild type mice were enucleated at time points E13, P0, P5 and P9. Dissected retinas were placed in Dulbecco's Phosphate Buffered Saline (DPBS, Thermo Fisher Scientific). Optic vesicule (OV)-like structures of the iPSCs derived 3D retinal aggregates were cut at DD13, DD21, DD25 and DD29 and transferred in DPBS as well.

Papaïne 4 U/mL (Worthington Biochemical Corporation) was added to the samples. The solution containing the retinas and the OV-like structures was maintained at 37°C for 45 and 30 minutes, respectively. Ovomucoïde 0,15% (Worthington Biochemical Corporation) was added for papaïne inhibition. Samples were centrifuged in order to eliminate the supernatant and cells were resuspended again in DPBS. The cell suspension was then manipulated for the different steps of single cell RNA sequencing. Cell numbers and proportion of life cells were estimated by Trypan Blue staining and using a Countess II cell counter (ThermoFisher). *scRNA-Seq:* We generated two biological replicates for stages 1 to 3 (NaR and 3D-RA) and one biological replicate for stage 4 (NaR and 3D-RA). We loaded ~15,700 cells for biological replicate 1 (stage 1-4) and ~10,000 cells for biological replicate 2 (stage 1-3) in a Chromium Controller instrument (10X Genomics, CA). Sequencing libraries were generated using Chromium Next GEM Single Cell 3' reagent kits v2.0 following the recommendations of the manufacturer (10X Genomics, CA). Actual sequencing was conducted on an Illumina NextSeq 500 instrument (Illumina, CA). *Bioinformatic analyses:* Demultiplexing, alignment, filtering, barcode counting, UMI counting, and aggregation of multiple runs were conducted using Cell Ranger (10X Genomics, CA). Further filtering, k-means clustering, UMAP projection were conducted using the Seurat software suite (https://satijalab.org/seurat/; Butler et al., 2018). Velocity analysis was performed using the Velocyto R package (La Manno et al., 2018). Single-cell trajectory inference and pseudo-time analysis was conducted with Monocle2 (http://cole-trapnell-lab.github.io/monocle-release/; Trapnell et al., 2014). All corresponding scripts and datasets are freely available from http://big-cloud19.segi.ulg.ac.be/UAG/Georges_A_2020.

ATAC-Seq. *Data generation:* ATAC-seq libraries were constructed on NaR (E13, P0, P5) and 3D-RA (DD13, DD21, DD25) samples with biological replicates following the Omni ATAC protocol (Corces et al., 2017). We used 50,000 cells per reaction taken from the cell suspensions prepared for the scRNA-seq. We tested two different amounts of the Tagment DNA TDE1 enzyme (1 and 2 µl in a 50 µl reaction volume) (Illumina, San Diego, California) per sample. Genomic DNA (gDNA) libraries were also prepared using 50 ng of gDNA isolated from NaR P5 and 3D-RA DD25 cells by following the Nextera DNA Sample Preparation Guide (Illumina). The libraries were purified using the MinElute PCR purification kit (Qiagen, Venlo, Netherlands) followed by 13 and 5 cycles of PCR-amplifications for ATAC-seq and gDNA

libraries, respectively. After validating library size distribution using the QIAxcel capillary electrophoresis (Qiagen), the libraries were further purified using the SPRIselect reagent to remove large DNA molecules (a right-side size selection with 0.55X followed by 1.5X ratios of beads) (Beckman Coulter, Brea, California). On average 10.6 millions of 38-nucleotide paired-end sequences were obtained using a NextSeq 500 sequencer (Illumina). *Data analyses:* Data was analyzed by following the ENCODE Kundaje lab ATAC-seq pipeline (https://www.encodeproject.org/pipelines/ENCPL792NWO/). Sequences were trimmed using Trimmomatic (Bolger et al., 2014) and aligned on the Mus musculus genome assembly mm10 using Bowtie2 (Langmead and Salzberg, 2012). After filtering out low quality, multiple mapped, mitochondrial, and duplicated reads using SAMtools (Li et al., 2009) and the Picard Toolkit (http://broadinstitute.github.io/picard/), fragments with map length ≤146 bp were kept as nucleosome-free fraction. Genomic loci targeted by TDE1 were defined as 38-bp regions centered either 4 (plus strand reads) or 5-bp (negative strand reads) downstream of the read's 5'-end. ATAC-seq peaks were called using the MACS2 software (narrowPeak; q-value ≤ 0.01; Zhang et al., 2008). FRiP scores were calculated as the fraction of TDE1 targeted loci falling into the called peaks. Overlapping peaks across samples were merged and annotated for the occurrence of TF binding motifs of interest (Suppl. Table 10) and the closest gene using Homer (Heinz et al., 2010). TDE1 targeted loci overlapping the merged peaks were extracted and converted to a bedgraph file with a scaling factor to one million reads using BEDTools (Quinlan and Hall, 2010), and further to tdf format to visualize peaks on the Integrative Genomics Viewer (Robinson et al., 2011). The total number of TDE1 targeted loci overlapping the merged peaks were counted using BEDOPS (Neph et al., 2012), normalized for peak lengths and a sequencing depth with per one million scaling factor, standardized and used for hierarchical cluster analysis using R hclust (Murtagh and Legendre, 2014) and gplots (https://CRAN.R-project.org/package=gplots). The detailed analysis pipeline is provided in the ATAC_seq_analysis_pipeline.docx file. The overall mapping rate with Bowtie2 averaged 98.6%, the mapping rate to the mitochondrial genome 4.1%, the duplicate fragment rate 6.0%, the proportion of usable reads after filtration 83.4%, and the FRiP score 34.1%. The FRiP score was significantly lower for E13 samples (reminiscent of the E14.5 samples in Aldiri et al. (2017)), yet not so in the equivalent DD13 samples (Suppl. Table 12).

Downstream analyses. *Percentage of 3D-RA cells in cliques and clusters:* The number and proportion of 3D-RA cells in cliques and clusters were computed using a Perl script (Dev_path_width.pl) and the corresponding graphs generated in R (Perc_3D_RA_in_cl.R). *Width of developmental trajectories in 2D UMPA space:* To test whether the developmental trajectories were more tightly regulated in NaR than in 3D-RA we computed the average distance (computed as the Euclidian distance in 2D-UMAP space, i.e.$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$) between 500 randomly selected NaR and 500 randomly selected 3D-RA cells and their n nearest neighbors (with n ranging from 1 to 50). The number of cells per developmental stage was adjusted between NaR and 3D-RA by down sampling to the number of the least populated source. The corresponding calculations were performed five times. The curves shown in Fig. 3H correspond to the averages across the five replicates. The grey confidence zone in Fig. 3H is bounded by the maxima and minima across the five replicates. The corresponding script was written in Perl (Dev_path_width.pl) and the graph generated in R (Path_width.R). *Within developmental stage clique entropy:* To compare clique diversity within developmental stage between NaR and 3D-RA, we first equalized the number of cells with developmental stage between NaR and 3D-RA by randomly dropping cells from the most populated source. We then sampled two cells within cell source (NaR and 3D-RA) and developmental stage and checked whether they were from the same clique or not. This was repeated 1,000 times yielding a measure of clique diversity akin to (1-entropy). Down-sampling of cells was repeated 100 times. Each data point in Fig. 3I corresponds to (1-Entropy) for one such random sample. The corresponding script was written in Perl (entropy.pl) and the graph generated in R (Entropy.R). *Pathway analyses:* Pathway enrichment analyses were conducted using the on-line Reactome analysis tools (Fabregat et al., 2018; Jassal et al., 2020). Mouse gene identifiers were converted to human counterparts. Pathway analysis results were downloaded as flat files. A total of 392 pathways with enrichment p-value ≤ 0.01 in at least one analysis were kept and manually sorted according to Reactome hierarchy (Man_processed_reactome_output.txt). A pathway is enriched in a list of genes if it contains more components of the pathway than expected by chance (given the number of genes in the list). The overlapping genes ("Found entities") hence define the enrichment. The same pathway can be enriched in two gene lists due to the same, distinct or partially overlapping sets of "found entities". We quantified the degree of overlap

between sets of "found entities" for the 1,313 pathway enrichments using principal component (PC) analysis in a space defined by the presence/absence of 1,335 genes. The distance between sets of "found entities" in a space consisting of the 20 first PCs was projected in 3D space using t-distributed stochastic neighbor embedding (tSNE) implemented with the *Rtsne* R function. 3D tSNE coordinates were converted to hexadecimal RGB code and used to color the sets of "found entities" (corresponding to the enrichment of a pathway in a specific gene list) when generating 2D tSNE graphs (Fig. 4D), or when generating a tile showing the pathways enriched in specific analyses (CLIQUE>OTHER, NaR > 3D-RA or 3D-RA > NaR) and cliques within analysis (NE, RPE, ERPC, LRPC, NRPC, RGC, HC, AC, PRP, C, R, BC or MC) (Fig. 4B). The corresponding scripts were written in Perl (Reactome_analysis.pl) and R (Reactome_analysis.R). *Analysis of dynamically regulated ("CLIQUES>OTHERS") and differentially regulated ("NaR>3D-RA" and "3D-RA>NaR") TF:* TF were extracted from Suppl. Tables 4 ("CLIQUES>OTHERS") and Suppl. Table 7 ("NaR>3D-RA" and "3D-RA>NaR") using a list of mouse TF obtained from the PHANTOM consortium (Kanamori et al., 2004)(mouse_TF.txt), and a file formatted using a Perl script (TF_prep.pl) to generate the tile shown in Fig. 5B in R (TF_tile.R). *Identifying regulatory toggles:* We used Homer (Heinz et al., 2010) to compile the number of occurrences of 336 binding motifs for 151 of 307 dynamically regulated TF in 98,181 ATAC-Seq peaks assigned to 19,170 genes. For each gene, the data were summarized as (i) the total number of occurrences, and (ii) the mean number of occurrences per peak (i.e. density), for each of the 336 binding motifs (Suppl. Table 11 and 336_BM_in_19170_genes.txt file). We then checked - for each of the 336 binding motifs separately - whether the number ("total" in Suppl. Table 11) and density ("mean" in Suppl. Table 2) of motifs differed significantly between genes that were upregulated versus downregulated in every one of the 13 cliques. Differential expression analyses to identify genes that are up- and downregulated in specific cliques were performed with the *Findmarkers* function in Seurat (https://satijalab.org/seurat/). The corresponding results are summarized in a series of files labelled, respectively, "NaR/<CLIQUE_ACRONYM>_markers.txt" for NaR, and "3D_RA/IPS_<CLIQUE_ACRONYM>_markers.txt" for 3D_RA. We used a threshold q-value of 0.05 to declare a gene as significantly up- or down-regulated in a given clique. The statistical significance of the difference in number and density of binding motifs between up- and down-regulated genes was computed using Wilcoxon rank-based test implemented with the

wilcox.test R function in the Comb_scRNA_ATAC_seq.R script. Differences were deemed significant if the p-value was ≤ 0.01. The input files needed for this analysis in R were generated using the RET_UP_prep.pl, RET_DOWN_prep.pl, RET_UP.pl, and RET_DOWN.pl for NaR (respectively IPS_UP_prep.pl, IPS_DOWN_prep.pl, IPS_UP.pl, and IPS_DOWN.pl for 3D_RA) Perl scripts and associated mouse_tf.txt, de_tf_order.txt and Expr_by_cell_type.txt input files. The results of these analyses are summarized in 2 x (total and mean) 336 graphs (similar to Fig. 6 G&K) generated as .pdf files (Suppl_Fig_8.pdf for NaR and Suppl_Fig_9.pdf for 3D_RA) with the Comb_scRNA_ATAC_seq.R script. Significant differences in the number/density of binding motifs were sorted in (i) enrichments versus depletion in ATAC-Seq peaks of overexpressed genes, and (ii) for TF that are or are not enriched in the corresponding clique. The corresponding clique-specific bar graphs were generated with the Comb_scRNA_ATAC_seq.R script (Fig. 6E and 6H). The previous analyses generate for each clique a vector of signed (enrichment (+) versus depletion (-)) log(1/p) values of differences in number/density between up- and downregulated genes for 336 binding motifs for 151 TF. We computed Spearman's correlations between all pairs of vectors using the R cor function, and reported the ensuing results as heatmaps (Fig. 6F and 6I) which were generated using the Comb_scRNA_ATAC_seq.R script. All corresponding scripts and datasets are freely available from http://big-cloud19.segi.ulg.ac.be/UAG/Georges_A_2020.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGEMENTS

We are grateful to Tomoyo Hashiguchi for teaching the 3D-RA culture protocol and to Tomohiro Masuda and Akishi Onishi for their comments and suggestions. We thank the GIGA Genomics platform for their help both with scRNA-seq experiments and early bio-informatic analyses (especially Wouter Coppieters and Benoit Charloteaux). This work was financially supported by the "King Baudouin Foundation", the "Global Ophthalmology Awards Program

(GOAP) from Bayer", and the "Fonds Leon Fredericq". LD and LN are respectively postdoctoral researcher and senior associate researcher of the FRS-FNRS.

DATA AVAILABILITY

All data generated as part of this work are available without restrictions. They have been deposited under accession numbers [*in process*]. All data and analysis pipelines are available at http://big-cloud19.segi.ulg.ac.be/UAG/Georges_A_2020.

REFERENCES

Agaki et al. (2015) ETS-related transcription factors ETV4 and ETV5 are involved in proliferation and induction of differentiation-associated genes in embryonic stem (ES)- cells. J Biol Chem 290: 22460-22473.

Akimoto *et al.* (2006) Targeting of GFP to newborn rods by Nrl promoter and temporal expression profiling of flow-sorted photoreceptors. Proc Natl Acad Sci USA 103: 3890-3895.

Aldiri *et al.* (2015) Brg1 coordinates multiple processes during retinogenesis and is a tumor suppressor in retinoblastoma. Development 142: 4092-4106.

Aldiri *et al.* (2017) The Dynamic Epigenetic Landscape of the Retina During Development, Reprogramming, and Tumorigenesis. Neuron 94:550-568.

Andzelm *et al.* (2015) MEF2D drives photoreceptor development through a genome- wide competition for tissue-specific enhancers. Neuron 86: 247-263.

Apara *et al.* (2017) KLF9 and JNK3 Interact to Suppress Axon Regeneration in the Adult CNS. J Neurosci. 37:9632–9644.

Arbogast *et al.* (2016) Thyroid hormone signaling in the mouse retina. PLoS One 11: e0168003

Assawachananont *et al.* (2014) Transplantation of embryonic and induced pluripotent stem cell-derived 3D retinal sheets into retinal degenerative mice. Stem Cell Rep 2: 662–674.

Balasubramanian *et al.* (2018) Lhx9 is required for the development of retinal nitric oxide-synthesizing amacrine cell subtype. Mol Neurobiol. 55: 2922–2933.

692 Bassett *et al.* (2007) Conditional deletion of activating protein 2 alpha (AP-2α) in the
693 developing retina demonstrates non-cell-autonomous roles for AP-2α in optic cup
694 development. Mol Cell Biol 27:7497–7510.

695 Belliveau & Cepko (1999) Extrinsic and intrinsic factors control the genesis of amacrine cells
696 and cone cells in the rat retina. Development 126: 555-556.

697 Bergen et al. (2019) Generalizing RNA velocity to transient cell states through dynamic
698 modeling. BioRxiv (https://doi.org/10.1101/82093).

699 Bertacchi *et al.* (2019) Mouse Nr2f1 haploinsufficiency unveils new pathological mechanisms
700 of a human optic atrophy syndrome. EMBO Mol Med 11: e10291.

701 Bolger *et al.* (2014). Trimmomatic: a flexible trimmer for Illumina sequence data.
702 Bioinformatics 30: 2114-2120.

703 Boulling *et al.* (2013) Identification of HMX1 target genes: a predictive promoter model
704 approach. Mol Vis 19:1779–1794.

705 Brzezinski *et al*. (2010) Blimp1 controls photoreceptor versus bipolar cell fate choice during
706 retinal development. Development 137: 619–629.

707 Bumsted *et al.* (2007) Expression of homeodomain transcription factor Meis2 in embryonic
708 and postnatal retina. J Comp Neurol 505: 58-72.

709 Burmeister *et al.* (1996) Ocular retardation mouse caused by Chx10 homeobox null allele:
710 impaired retinal progenitor proliferation and bipolar cell differentiation. Nat Genet 12: 376-
711 384.

712 Butler *et al.* (2018) Integrating single-cell transcriptomic data across different conditions,
713 technologies, and species. Nat Biotech 36: 411-420.

714 Camara *et al.* (2018) Methods and challenges in the analysis of single-cell RNA-sequencing
715 data. Curr Opin Syst Biol 7:47-53.

716 Chend *et al.* (2004) Photoreceptor-specific nuclear receptor Nr2e3 functions as a
717 transcriptional activator in rod photoreceptors. Hum Mol Genet 15: 1563-1575.

718 Cherry *et al.* (2011) NeuroD factors regulate cell fate and neurite stratification in the
719 developing retina J Neurosci. 31:7365-79.

Chow *et al.* (2001) Vsx1, a rapidly evolving paired-like homeobox gene expressed in cone bipolar cells. Mech Dev 109:315–322.

Chung *et al.* (2013) Laser Capture Microdissection-Directed Profiling of Glycolytic and mTOR Pathways in Areas of Selectively Ablated Müller Cells in the Murine Retina. Invest Opht & Vis Sci 54: 6578-6584.

Clark *et al.* (2019) Comprehensive analysis of retinal development at single cell resolution identifies NFI factors as essential for mitotic exit and specification of late-born cells. Neuron 102: 1111-1126.

Collin *et al.* (2019) Deconstructing retinal organoids: single cell RNA-Seq reveals the cellular components of human pluripotent stem cell-derived retina. Stem Cells 37: 593-598.

Corces *et al.* (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14:959-962.

Cowan et al. (2020) Cell types of the human retina and its organoids at single-cell resolution : developmental convergence, transcriptomic identity, and disease map. bioRxiv : https://doi.org/10.1101/703348.

Dagnino *et al.* (1997) Expression patterns of the E2F family of transcription factors during mouse nervous system development. Mech Dev 66: 13-25.

De Melo *et al.* (2011) The Spalt family transcription factor Sall3 regulates the development of cone photoreceptors and retinal horizontal interneurons. Development 138: 2325–2336.

De Melo *et al*. (2016) Multiple intrinsic factors act in concert with Lhx2 to direct retinal gliogenesis. Sci Rep 6:32757

Dixit *et al.* (2016) Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell 167: 1853-1866.

Eiraku *et al.* (2011) Self-organizing optic-cup morphogenesis in 3D culture. Nature 472: 51-56.

Eiraku *et al.* (2012) Mouse embryonic stem cell culture for generation of three-dimensional retinal and cortical tissues Nat Prot 7: 69–79.

Elshatory *et al.* (2007) Islet-1 controls differentiation of retinal bipolar and cholinergic amacrine cells. J Neurosci 27: 12707-12720.

748 Fabregat *et al.* (2018) The reactome pathway knowledgebase. Nucl Ac Res. 44: D481-487.

749 Feng *et al.* (2006) Requirement for Bhlhb5 in the specification of amacrine and cone bipolar
750 subtypes in mouse retina. Development 133: 4815–4825.

751 Finkler *et al.* (2007) CAMTAs: Calmodulin-binding transcription activators from plants to
752 human. FEBS Lett 581: 3893-3898.

753 Fujieda et al. (2009) Retinoic acid receptor-related orphan receptor alpha regulates a subset
754 of cone genes during mouse development. J Neurochem 108: 91-101.

755 Furimsky *et al.* (2005) Gli3 Controls Precursor Cell Proliferation and Differentiation in the
756 Developing Mouse Retina. Invest Ophthalmol Vis Sci 46: 578.

757 Furukawa *et al.* (1997) Crx, a Novel Otx-like Homeobox Gene, Shows Photoreceptor-Specific
758 Expression and Regulates Photoreceptor Differentiation. Cell 91: 531-541.

759 Furukawa *et al.* (2000). Rax, Hes1, and Notch1 promote the formation of Müller glia by
760 postnatal retinal progenitor cells. Neuron 26:383–394.

761 Gallina *et al.* (2014). Glucocorticoid receptors in the retina, Müller glia and the formation of
762 Müller glia-derived progenitors. Development 141: 3340–3351.

763 Gan *et al*. (1999) POU domain factor Brn-3b is essential for retinal ganglion cell
764 differentiation and survival but not for initial cell fate specification. Dev Biol 210:469–480.

765 Gonzalez-Cordero *et al.* (2013) Photoreceptor precursors derived from three-dimensional
766 embryonic stem cell cultures integrate and mature within adult degenerate retina. Nat
767 Biotechnol 31: 741- 747.

768 Goodson *et al.* (2018). Prdm13 is required for Ebf3+ amacrine cell formation in the retina.
769 Dev Biol. 434:149-163.

770 Hagstrom *et al.* (2001) A role for the Tubby-like protein 1 in rhodopsin transport. Invest
771 Ophthalmol Vis Sci 42:1955–1962.

772 Hao *et al.* (2012) Transcriptional regulation of rod photoreceptor homeostasis revealed by in
773 vivo NRL targetome analysis. PLoS Genet 8: e1002649.

774 Hauck *et al.* (2003) Proteomic profiling of primary retinal Muller glia cells reveals a shift in
775 expression patterns upon adaptation to in vitro conditions. Glia 44: 251–63.

776 Heine *et al.* (2008) Evidence for an evolutionary conserved role of homothorax/Meis1/2
777 during vertebrate retina development. Development 135: 805-811.

778 Heinz *et al.* (2010) Simple combinations of lineage-determining transcription factors prime
779 cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38:576-589.

780 Hensley *et al*. (2011) Cellular expression of Smarca4 (Brg1)-regulated genes in zebrafish
781 retinas. BMC Dev Biol. 11:45.

782 Hennig *et al.* (2008) Regulation of photoreceptor gene expression by Crx-associated
783 transcription factor network. Brain Res 1192:114–133.

784 Hermans-Borgmeyer *et al*. (2000) Developmental expression of the estrogen receptor-related
785 receptor gamma in the nervous system during mouse embryogenesis. Mech Dev 97: 197–
786 199.

787 Hojo *et al.* (2000) Glial cell fate specification modulated by the bHLH gene Hes5 in mouse

788 retina. Development 127: 2515–2522.

789 Homma *et al. (*2013) Developing rods transplanted into the degenerating retina of Crx-
790 knockout mice exhibit neural activity similar to native photoreceptors. Stem Cells 31:1149–
791 1159.

792 Hu *et al.* (2013) Aryl hydrocarbon receptor deficiency causes dysregulated cellular matrix
793 metabolism and age-related macular degeneration-like pathology. Proc Natl Acad Sci US 110:
794 E4069-E4078.

795 Hu et al. (2019) Dissecting the transcriptome landscape of the human fetal neural retina and
796 retinal pigment epithelium by single-cell RNA-seq analysis. Plos Biology 17: e3000365.

797 Hufnagel *et al.* (2010) Neurog2 controls the leading edge of neurogenesis in the mammalian
798 retina Dev Biol. 340: 490–503.

799 Inoue *et al.* (2002) Math3 and NeuroD regulate amacrine cell fate specification in the
800 retina. Development 129: 831–842.

801 Inoue *et al.* (2010) COUP-TFI and -TFII nuclear receptors are expressed in amacrine cells and
802 play roles in regulating the differentiation of retinal progenitor cells. Exp Eye Res. 90: 49-56.

803 Iwasaki *et al.* (2016) Differentiation/purification protocol for retinal pigment epithelium from

804 mouse induced pluripotent stem cells as a research tool. PLoS One 11: 1–20.

805 Jalali *et al.* (2011) HeyL promotes neuronal differentiation of neural progenitor cells. J
806 Neurosci Res 89: 299–309.

807 Jassal *et al.* (2020) The reactome pathway knowledgebase. Nucleic Acids Res 48: D498-D503.

808 Jelcick *et al.* (2011) Genetic variations strongly influence phenotypic outcome in the mouse
809 retina. PloS One 6: 10.1371.

810 Jia et al. (2009) Retinoid related orphan nuclear receptor RORbeta is an early-acting factor in
811 rod photoreceptor development. Proc Natl Acad Sci US 106: 17534–17539.

812 Jin *et al.* (2010) Early B-Cell Factors Are Required for Specifying Multiple Retinal Cell Types
813 and Subtypes from Postmitotic Precursors. J Neurosci 30:11902–11916.

814 Jin *et al.* (2011) Modeling retinal degeneration using patient-specific induced pluripotent
815 stem cells. PLoS One 6: e17084.

816 Jin *et al.* (2015) Tfap2a and 2b act downstream of Ptf1a to promote amacrine cell
817 differentiation during retinogenesis. Mol Brain 13: 8-28.

818 Jones *et al.* (2007). An intron control region differentially regulates expression of thyroid
819 hormone receptor B2 in the cochlea, pituitary and cone photoreceptors. Mol Endocrinol 21:
820 1108-1119.

821 Jung *et al*. (2015) Transcription factor Prdm8 is required for rod bipolar and type 2 off-cone
822 bipolar cell survival and amacrine subtype identity. Proc Natl Acad Sci USA 112:7115.

823 Kanamori *et al.* (2004). A genome-wide and nonredundant mouse transcription factor
824 database. Biochem Biophys Res Commun 322:787-793.

825 Kim *et al.* (2014) Deletion of Aryl Hydrocarbon Receptor AHR in Mice Leads to Subretinal
826 Accumulation of Microglia and RPE Atrophy. Invest Ophthalmol Vis Sci 55: 6031–6040.

827 Klimova *et al.* (2015) Onecut1 and Onecut2 transcription factors operate downstream of Pax6
828 to regulate horizontal cell development. Dev Biol 402: 48-60.

829 Kuribayashi *et al.* (2014) BMP signaling participates in late phase differentiation of the retina,
830 partly via upregulation of Hey2. Dev Neurobiol 74: 1172-83.

Kuwahara *et al.* (2014) Tcf3 represses Wnt-β-catenin signaling and maintains neural stem cell population during neocortical development. PLoS One 9:e94408.

Kuwahara *et al.* (2015) Generation of a ciliary margin-like stem cell niche from self-organizing human retinal tissue. Nat Commun 6: 6286.

Laboissonniere *et al.* (2019). Molecular signatures of retinal ganglion cells revealed through single cell profiling. Sci Rep 9: 15778.

La Manno *et al.* (2018) RNA velocity of single cells. Nature 560: 494-510.

Langmead & Salzberg (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359 .

Laudes *et al.* (2004) Role of the POZ zinc finger transcription factor FBI-1 in human and murine adipogenesis. J Biol Chem 279:11711-11718.

Lee *et al*. (2004) The distribution of c-myb immunoreactivities in the adult mouse retina**.** Neurosci Lett 366: 297-301.

Li *et al.* (2009) 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

Lin *et al*. (2007). ZacI functions through TGFßII to negatively regulate cell number in the developing retina. Neural Development 2:1-21.

Liu *et al.* (2009) Zeb1 Represses *Mitf* and Regulates Pigment Synthesis, Cell Proliferation and Epithelial Morphology. Invest Ophthalmol Vis Sci 50: 5080-5088.

Liu *et al.* (2018) Tbr1 instructs laminar pattering of retinal ganglion cell dendrites. Nat Neurosci 21: 659-670.

Ma *et al.* (2007) Zac1 promotes a Muller glial cell fate and interferes with retinal ganglion cell di erentiation in Xenopus retina. Dev Dyn 236: 192–202.

Macosko *et al.* (2015) Highly parallel genome wide expression profiling of individual cells using nanoliter droplets. Cell 161: 1202:1214.

Marcucci *et al.* (2016) The ciliary margin zone of the mammalian retina generates ganglion cells. Cell Reports 17: 3153- 3164.

Martersteck *et al.* (2017) Diverse Central Projection Patterns of Retinal Ganglion Cells. Cell Reports 18: 2058-2072

Matar *et al.* (2018) Casz1 controls higher-order nuclear organization in rod photoreceptors. Proc Natl Acad Sci US 115: E7987–E7996.

Matsukawa *et al*. (2018) Mechanisms of RhoA inactivation and CDC42 and Rac1 activation during zebrafish optic nerve regeneration. Neurochem Int 112: 71-80.

McInnes *et al.* (2018) UMAP: Uniform Manifold Approximation and Projection. J. Open Source Software 3: 861.

Mears *et al.* (2001) Nrl is required for rod photoreceptor development. Nat Genet 29: 447-52.

Melo *et al.* (2011). The Spalt family transcription factor Sall3 regulates the development of cone photoreceptors and retinal horizontal interneurons. Development 138: 2325–2336.

Moose *et al.* (2009) Ocular forkhead transcription factors: seeing eye to eye. Int J Dev Biol. 53: 29–36.

Mo *et al.* (2004) Role of the Barhl2 homeobox gene in the specification of glycinergic amacrine cells. Development 131: 1607–1618.

Mollema *et al.* (2011). Nuclear Receptor Rev-erb Alpha (Nr1d1) Functions in Concert with Nr2e3 to Regulate Transcriptional Networks in the Retina. PloS One 6: e17494.

Mowat *et al.* (2010). HIF-1alpha and HIF-2alpha Are Differentially Activated in Distinct Cell Populations in Retinal Ischaemia. PLoS One 5: e11103.

Murtagh et al. (2014). Ward's hierarchical agglomerative clustering method : whihc algorithms implement Ward's criterion ? J Classification 31 : 274-295.

Nakamura *et al.* (2006) Expression of Olig2 in retinal progenitor cells. Neuroreport 17: 345-349.

Nakano *et al.* (2012) Self-formation of optic cups and storable stratified neural retina from human ES cells. Cell Stem Cell 10: 771–785.

884 Nelson *et al.* (2011) Genome-Wide Analysis of Mu¨ller Glial Differentiation Reveals a
885 Requirement for Notch Signaling in Postmitotic Cells to Maintain the Glial Fate. PLoS One 6:
886 e22817.

887 Neph *et al.* (2012) BEDOPS: high performance genomic feature operations. Bioinformatics 28:
888 1919-1920.

889 Oh *et al.* (2007) Transformation of cone precursors to functional rod photoreceptors by bZIP
890 transcription factor NRL. Proc Natl Acad Sci US 104: 1679–1684.

891 Osakada *et al.* (2008) Toward the generation of rod and cone photoreceptors from mouse,
892 monkey and human embryonic stem cells. Nat Biotechnol 26: 215–224.

893 Peng Hu *et al.* (2013) Aryl hydrocarbon receptor deficiency causes dysregulated cellular
894 matrix metabolism and age-related macular degeneration-like pathology. Proc Natl Acad Sci
895 US 110: E4069-E4078.

896 Poché *et al.* (2008). Sox9 is expressed in mouse multipotent retinal progenitor cells and
897 functions in Müller glial cell development. Journal of Comparative Neurology 510: 237–250.

898 Pohl et al (2005). The Fox gene family in Xenopus laevis: FoxI2, FoxM1 and FoxP1 in early
899 development. Int J Dev Biol 49:53-58.

900 Quinlan et al. BEDTools: a flexible suite of utilities for comparing genomic features.
901 Bioinformatics 26: 841-842.

902 Rattner *et al.* (2019) Roles of HIFs and VEGF in angiogenesis in the retina and brain**.** J Clin
903 Invest 129: 3807-3820.

904 Reese *et al.* (2011) Development of the retina and optic pathway. Vision Res 51: 613–632.

905 Reinhardt *et al.* (2015). Sox2, Tlx, Gli3, and Her9 converge on Rx2 to define retinal stem
906 cells *in vivo.* Embo *.* 3: 1572–1588.

907 Rheaume *et al.* (2018) Single cell transcriptome profiling of retinal ganglion cells identifies
908 cellular subtypes. Nat Comm 9: 2759.

909 Robinson et al. (2011). Integrative genomics viewer. Nat Biotechnol 29 : 24-26.

910 Roesch *et al.* (2012). Gene expression changes within Müller glial cells in retinitis pigmentosa.
911 Mol Vis 18: 1197–1214.

912 Saelens *et al.* (2019) A comparison of single cell trajectory inference methods. Nat Biotechnol
913 37: 547–554.

914 Sajgo *et al.* (2017) Molecular codes for cell type specification in Brn3 retinal ganglion cells.
915 Proc Natl Acad Sci US 114 : E3974-E3983.

916 Satoh *et al*. (2008) TGIF, a homeodomain transcription factor, regulates retinal progenitor cell
917 differentiation. Exp Eye Res 87: 571-579.

918 Schuff *et al.* (2007) Foxn3 is required for craniofacial and eye development of Xenopus Laevis.
919 Dev Dyn 236: 226-239.

920 Seritrakul *et al.* (2017) Tet-mediated DNA hydroxymethylation regulates retinal neurogenesis
921 by modulating cell-extrinsic signaling pathways. PLoS Genet 13: e1006987.

922 Shekhar *et al.* (2016) Comprehensive classification of retinal bipolar neurons by single-cell
923 transcriptomics. Cell 166: 1308-1323.

924 Sridhar *et al.* (2020) Single Cell Transcriptomic Comparison of human fetal retina, hPSC-
925 derived retinal organoids, and long-term retinal cultures. Cell Rep 30:1644-1659.

926 Tang *et al.* (2010) COUP-TFs regulate eye development by controlling factors essential for
927 optic vesicle morphogenesis. Development 137:725-734.

928 Tirosh *et al.* (2016) Dissecting the multicellular ecosystem of metatastic melanoma by single-
929 cell RNA-seq. Science 352:189-196.

930 Trapnell *et al.* (2014) The dynamics and regulators of cell fate decisions are revealed by
931 pseudotemporal ordering of single cells. Nat Biotechnol 32:381-386.

932 Trimarchi *et al.* (2008) Individual retinal progenitor cells display extensive heterogeneity of
933 gene expression. PloS One 3: e1588.

934 Trimarchi *et al.* (2009) Identification of genes expressed preferentially in the developping
935 peripheral margin of the optic cup. Dev Dyn 238: 2327.

936 Ueno *et al.* (2015) Transition of differential histone H3 methylation in photoreceptors and
937 other retinal cells during retinal differentiation. Sci Rep 6:29264.

938 Vazquez-Chona *et al.* (2005) Genetic networks controlling retinal injury. Mol Vis 11:958.

939 Völkner *et al.* (2016) Retinal organoid from pluripotent stem cells efficiently recapitulate
940 retinogenesis. Stem Cells Rep 6: 525-538.

941 Wang *et al.* (2013) Gene Expression Changes under Cyclic Mechanical Stretching in Rat Retinal
942 Glial (Müller) Cells. PLoS One 8: e63467.

943 Watabe *et al.* (2011) The role of Zic family zinc finger transcription factors in the proliferation
944 and differentiation of retinal progenitor cells. Biochem Biophys Res Commun 415: 42-47.

945 Wohl *et al.* (2017) Müller glial microRNAs are required for the maintenance of glial
946 homeostasis and retinal architecture. Nat Commun 8: 1603.

947 Wu *et al.* (2013) Onecut1 is essential for horizontal cell genesis and retinal integrity. J Neurosci
948 33: 13053-65.

949 Xi *et al.* (2003) A comprehensive analysis of the expression of cristallins in mouse retina. Mol
950 Vis 9:410-409.

951 Yan *et al.* (1998) NeuroD induces photoreceptor cell overproduction in vivo and de novo
952 generation in vitro. J Neurobiol 36: 485–496.

953 Yan *et al.* (2020) From reads to insight: a hitchhiker's guide to ATAC-Seq data analysis.
954 Genome Biol 21: 22.

955 Yasuda *et al.* (2014) Retinal transcriptome profiling at transcription start sites: a cap analysis
956 of gene expression early after axonal injury. BMC Genomics 15: 982.

957 Zagozewski *et al.* (2014) The role of homeobox genes in retinal development and disease. Dev
958 Biol 393:195-208.

959 Zhang *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137.

960 Zhou *et al.* (1996) Retina-derived POU-domain factor-1: a complex POU-domain gene
961 implicated in the development of retinal ganglion and amacrine cells. J Neurosci 16:2261–
962 2274.

963 Zhou *et al.* (2018) Retinal progenitor cells release extracellular vesicles containing
964 developmental transcription factors, microRNA and membrane proteins. Sci Rep 8: 2823.

965

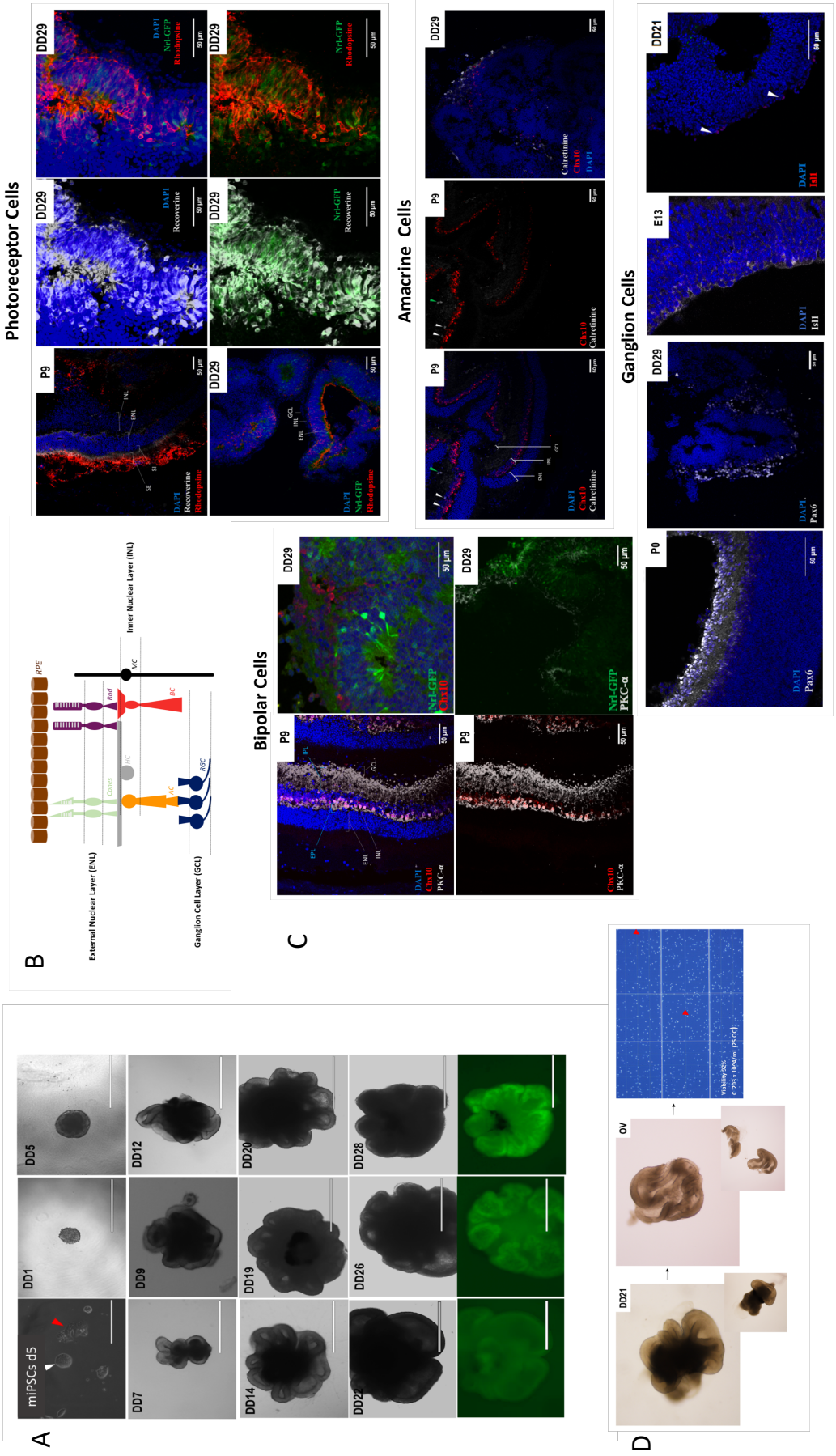
A
miPSCs d5
DD1
DD5
DD7
DD9
DD12
DD14
DD19
DD20
DD22
DD26
DD28
B
RPE
Rod
Cones
HC
MC
BC
AC
RGC
External Nuclear Layer (ENL)
Inner Nuclear Layer (INL)
Ganglion Cell Layer (GCL)
C
Bipolar Cells
P9
DD29
DAPI
Chx10
PKC-α
Nrl-GFP
D
DD21
OV
P0
Pax6
Amacrine Cells
Calretinine
Ganglion Cells
E13
Isl1

Figure 1: (A) In vitro differentiation of miPSC-Nrl*GFP*-derived retinal aggregates (3D-RA). Morphology of iPSCs five days post-thawing. Cells in undifferentiated state are circumferential, "domed shaped" and surrounded by a luminous halo (white arrow). Some unstable colonies tend to differentiate; they have a fibroblastic morphology instead of round shape (red arrow). 3D-RA differentiating retinal aggregates from differentiation DD1 to DD28 obtained following the modified SFEBq protocol (Eiraku *et al.,* 2012): (DD1) Rapid re-aggregation of iPSCs after dissociation and passage in 96-well plate, (DD5) Appearance of retinal neuro-epithelium on the retinal aggregates (light edges), (DD7) Evagination of retinal neuro-epithelium, (DD8-DD20) Growing of evaginating OV like structures, (DD22-DD28) GFP expression under influence of the *Nrl* promotor in the photoreceptor layer of retinal aggregates. Scale: 400 µm (miPSCs d5), 1000 µm (others). **(B) Schematic of anatomy of retinal layers. (C) Expected layered expression of cell-type specific immunochemical markers.** Immuno-histochemical markers of cellular subtypes in native retina (E13, P5 and P9) and Nrl*GFP* iPSCs-derived 3D-RA at DD21 and DD29. Photoreceptor cells: GFP (green), recoverine (white) and rhodopsine (red) are specific for rod photoreceptor cells; DAPI (blue) marks the nucleus of all cells. INL = inner nuclear layer; ENL = external nuclear layer; SI = internal segment; SE = external segment; GCL = ganglion cell layer. Bipolar cells: Chx10 (red) and Pkc-α (white) are specific for bipolar cells; Nrl-GFP (green) and DAPI (blue) as above; INL,ENL,GCL = as above; IPL = inner plexiform layer; EPL = external plexiform layer. Amacrine cells: calretinine (white) is specific for amacrine cells; Chx10 (red) and DAPI (blue) as above; the white arrows show two amacrine cells in the internal nuclear layer, while the green arrow shows two amacrine cells probably delocalized during the experiment; CNE = external nuclear layer; CNI = inner nuclear layer; CCG = ganglionic cell layer. Retinal ganglionic cells: Pax6 (white) and Isl1 (white or red) are specific for retinal ganglion cells; DAPI (blue) as above; the white arrows show Isl1 positive retinal ganglion cells. Scale: 50 µm for all. **(C) Dissection of OV-like structures and dissociation of cells into viable single cell solution.** Manual dissection of 3D-retinal aggregates at stage 2 (DD21). After dissection, the retinal neuro-epithelium/evaginating tissue is isolated from the dark/pigmented inner cell mass. Cells from retinal neuro-epithelium are dissociated into a homogenous solution of single cells (no doublets). 10 µl of the solution is used for cell counting and appraisal of cell viability.

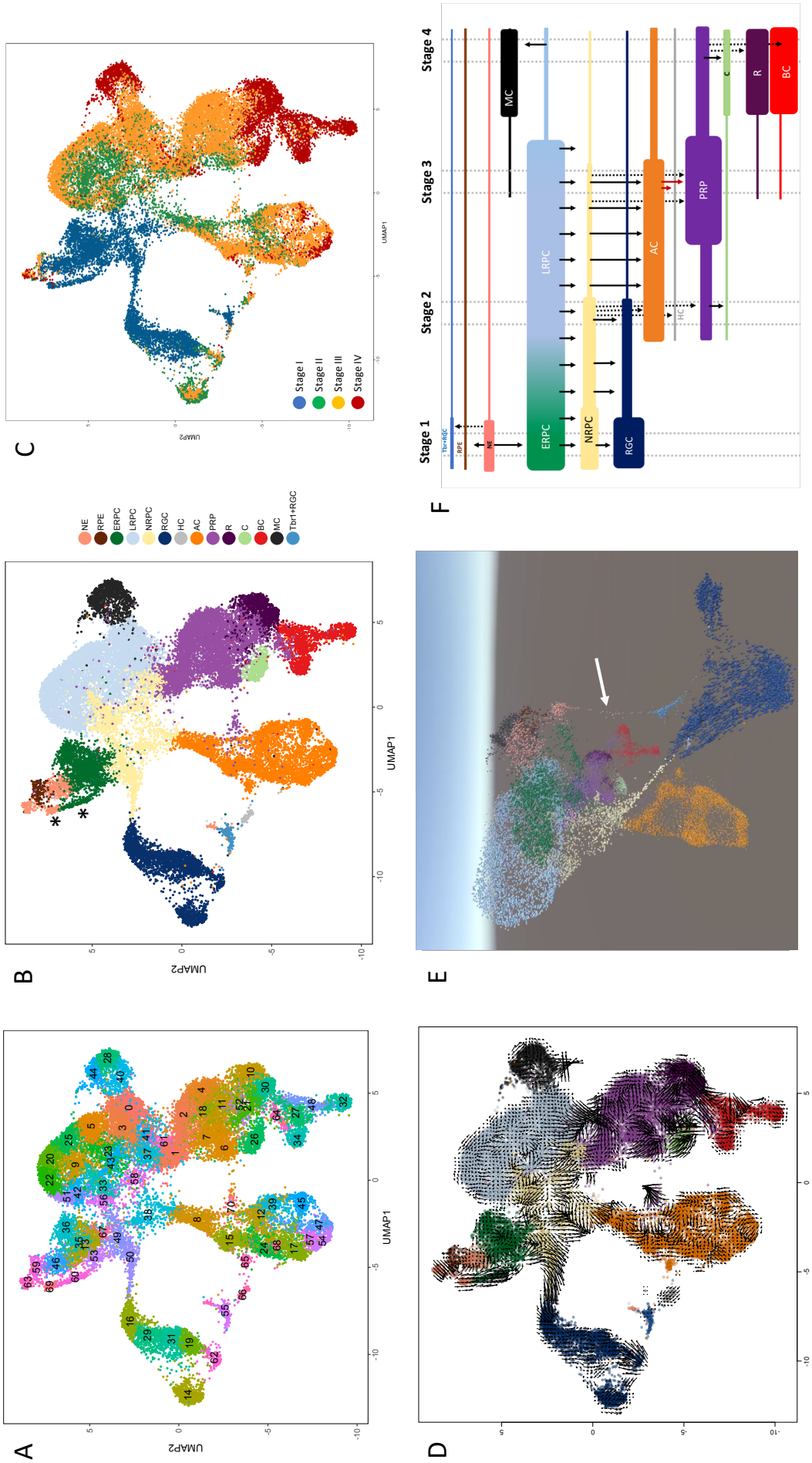
A
UMAP2
UMAP1
B
UMAP2
UMAP1
NE
RPE
ERPC
LRPC
NRPC
RGC
HC
AC
PRP
R
C
BC
MC
Tbr1+RGC
C
UMAP2
UMAP1
Stage I
Stage II
Stage III
Stage IV
D
E
F
Stage 1
Stage 2
Stage 3
Stage 4
Tbr1+RGC
RPE
NE
ERPC
NRPC
RGC
LRPC
MC
AC
HC
PRP
R
BC

Figure 2: (A) 2D UMAP manifold showing NaR and 3D-RA cells jointly and their assignment to 70 clusters by k-means clustering. **(B)** Merging of the clusters in 14 cliques corresponding to neuroepithelium (NE), retinal pigmented epithelium (RPE), early (ERPC), late (LRPC), neurogenic retinal progenitor cells (NRPC), retinal ganglionic cells (RGC), Tbr1 positive retinal ganglionic cells ($Tbr1^{+}$RGC), horizontal cells (HC), amacrine cells (AC), photoreceptor precursor cells (PRP), cones (C), rods (R), bipolar cells (BC), and Müller cells (MC), on the basis of the expression of known marker genes (Suppl. Table 3 and Suppl. Figure 1). Cluster 69 (NE) and 60 (ERPC) (marked by asterisks) correspond to the ciliary marginal zone (CMZ) which forms a branch that clearly separates from the rest of NE and ERPC. **(C)** Cells colored by developmental stage: 1. blue = DD13 + E13, 2. green = DD21 + P0, 3. orange = DD25 + P5, 4. red = DD29 + P9. **(D)** Local RNA velocities (La Manno et al., 2018) embedded in 2D UMAP. **(E**) Preview of virtual reality visit of 3D UMAP manifold (http://www.sig.hec.ulg.ac.be/giga/) showing the cellular bridge directly connecting NE and Trb1+RGC (white arrow). **(F)** Summary of inferred developmental trajectories. Red arrows mark presumed re-specification events.

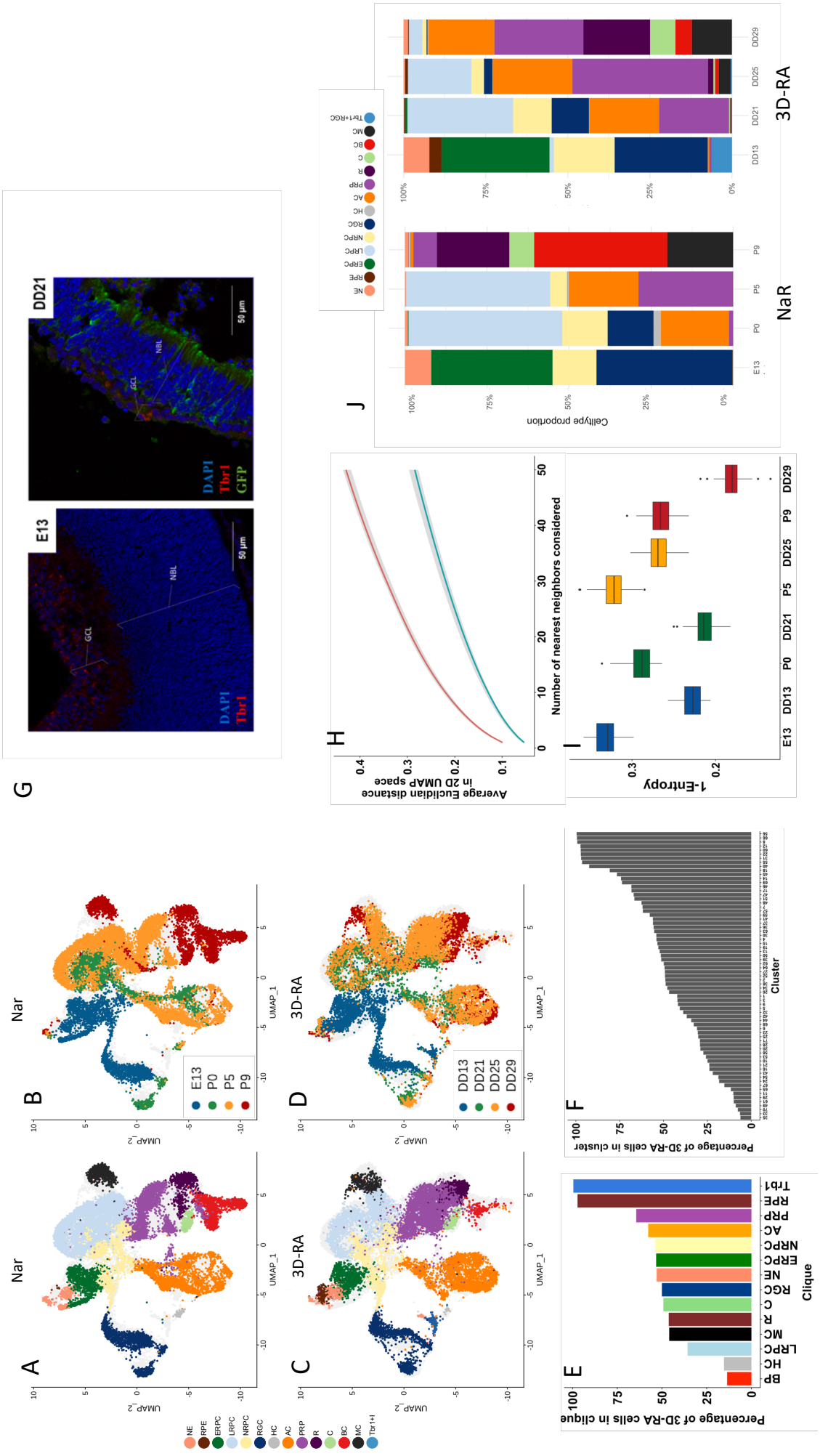
A
Nar
B
C
3D-RA
D
UMAP_1
UMAP_2
NE
RPE
ERPC
LRPC
NRPC
RGC
HC
AC
PRP
R
C
BC
MC
Tbr1+
E13
P0
P5
P9
DD13
DD21
DD25
DD29
E
Percentage of 3D-RA cells in clique
Clique
F
Percentage of 3D-RA cells in cluster
Cluster
G
E13
DD21
DAPI
Tbr1
GFP
GCL
NBL
50 µm
H
Average Euclidian distance in 2D UMAP space
Number of nearest neighbors considered
I
1-Entropy
J
NaR
3D-RA
Celltype proportion

Figure 3: (A-D) Distribution of Nar (A-B) versus 3D-RA (C-D) cells across the UMAP manifold, sorted by clique (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 cliques (E) and 70 clusters (F). 86% of cliques 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 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 than for NaR cells. **(I)** Larger clique diversity (sampling-based measure of entropy) in three out of four developmental stages for 3D-Ra than for NaR. **(J)** Proportions of cliques within developmental stage for NaR (left) and 3D-RA (right).

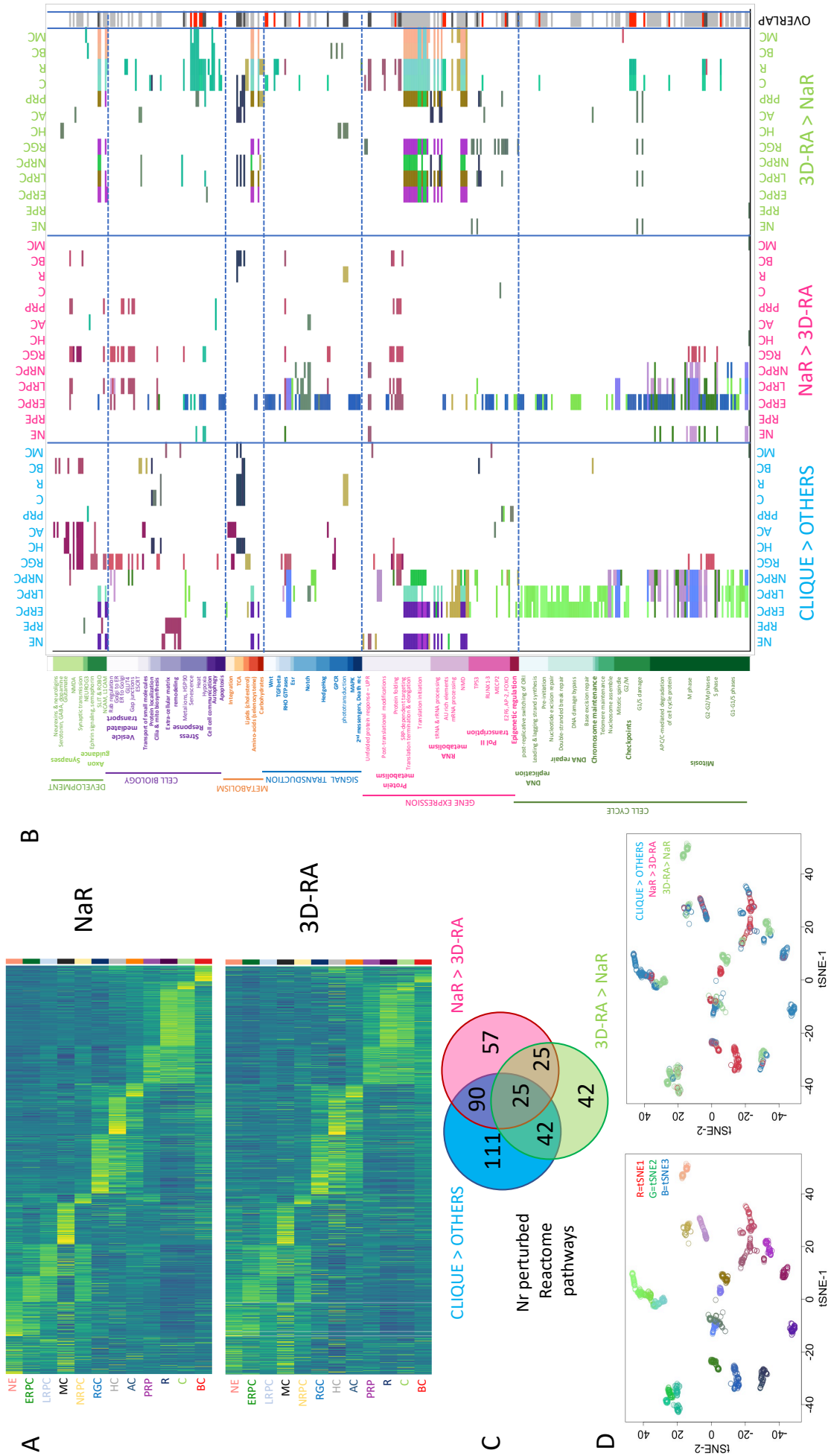
A
NaR
3D-RA
NE
ERPC
LRPC
MC
NRPC
RGC
HC
AC
PRP
R
C
BC
B
CLIQUE > OTHERS
NaR > 3D-RA
3D-RA > NaR
OVERLAP
CELL CYCLE
GENE EXPRESSION
METABOLISM
SIGNAL TRANSDUCTION
CELL BIOLOGY
DEVELOPMENT
C
CLIQUE > OTHERS
NaR > 3D-RA
3D-RA > NaR
57
25
90
25
42
111
42
Nr perturbed Reactome pathways
D
CLIQUE > OTHERS
NaR > 3D-RA
3D-RA> NaR
R=tSNE1
G=tSNE2
B=tSNE3
tSNE-1
tSNE-2

Figure 4: (A) Expression profiles in 12 cliques of 7,292 genes that are dynamically regulated during in vivo retinal development (i.e. significantly overexpressed in at least one clique 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 ("Clique > other": when comparing expression levels between specific cell types (i.e. cliques) and all other cells in NaR only; "NaR > 3D-RA" and "3D-RA> NaR": when comparing expression levels between NaR and 3D-RA cells within cliques). Y-axis: Reactome pathways are colored by "top level" system (cell cycle, gene expression, signal transduction, metabolism, cell biology and development) and sub-level therein. X-axis: Type of contrast for DE and cell type / clique labels are as in the remainder of the manuscript and figure ("Clique > other": turquoise; "NaR > 3D-RA": magenta; "3D-RA> NaR": lime). Tiles mark the pathways that are significantly enriched in the corresponding contrast and cell type/clique. The colors of the tiles reflect similarity in gene content ("found entities") as described in (C). Last column ("Overlap"): White: pathways significant in one contrast only, Black: pathways significant in all three contrasts, Grey: pathways significant in "Clique > other" and ("NaR > 3D-RA" or "3D-RA> NaR"), Red: pathways significant in "NaR > 3D-RA" and "3D-RA> NaR". **(C)** Number of unique and shared Reactome pathways between "Clique > other", "NaR > 3D-RA" and "3D-RA> NaR". All overlaps are highly significant ($p < 10^{-6}$) assuming random sampling from 2,365 reactome pathways. **(D)** (Left) tSNE map (dimensions 1 and 2 of 3) of 1,313 gene sets ("found entities") marking enriched Reactome pathways (each little circle corresponding to a tile in B). Overlapping sets (in terms of gene content) are close to each other in tSNE space (3D). The gene sets are colored in RGB scale where tSNE dimension 1 (tSNE1) determines the intensity of red, tSNE2 the intensity of green, and tSNE3 the intensity of blue. (Right) Same datapoints as for (Left) but colored by contrast of origin as in B and C: "Clique > other": turquoise; "NaR > 3D-RA": magenta; "3D-RA> NaR": lime. One can see that many clusters encompass gene sets corresponding to distinct contrasts, hence highlighting the strong overlap between Reactome pathways that are essential for normal retinal development ("Clique > other"), and those that are perturbed in 3D-RA relative to NaR ("NaR > 3D-RA" and "3D-RA> NaR").

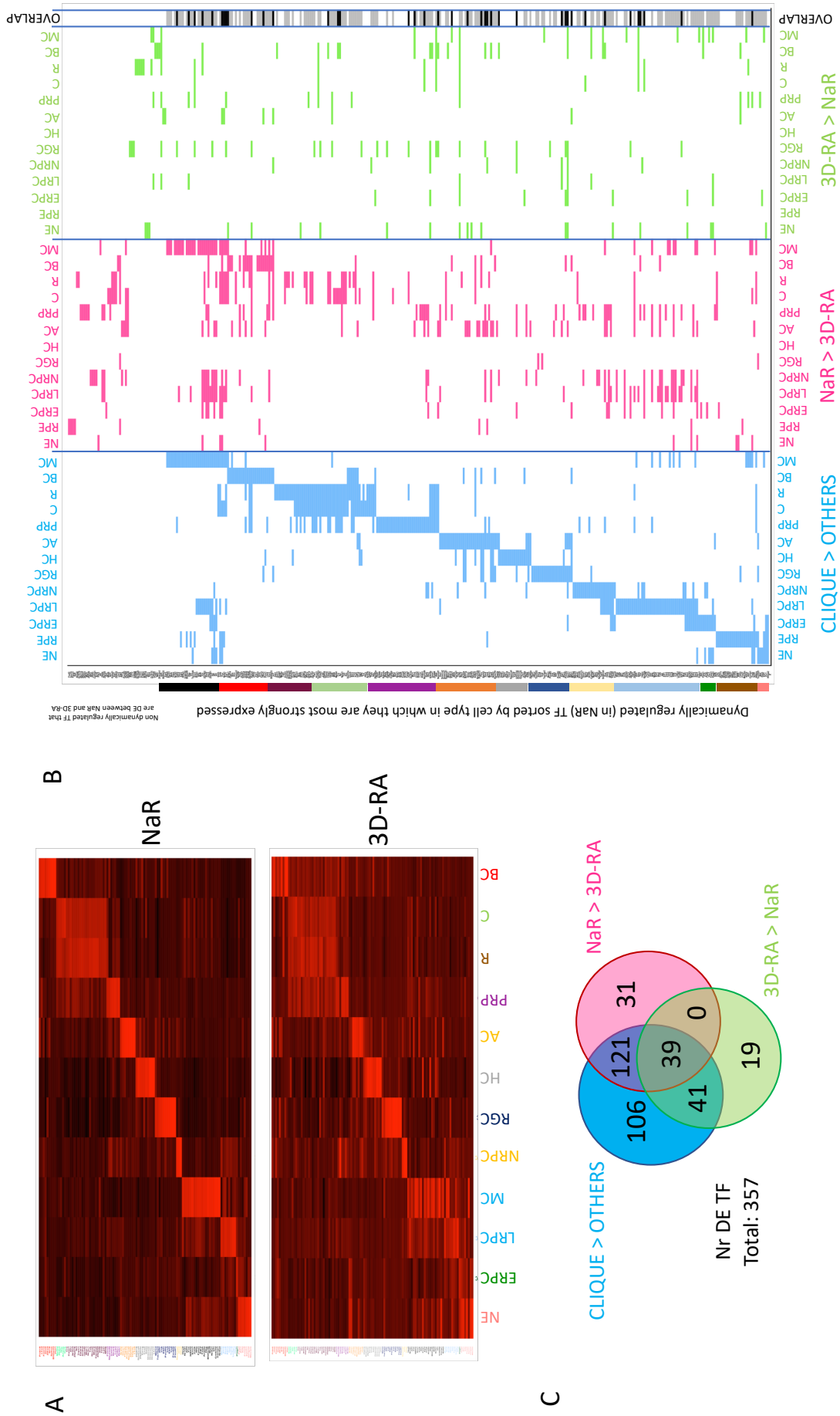
A
NaR
3D-RA
NE
ERPC
LRPC
MC
NRPC
RGC
HC
AC
PRP
R
C
BC
B
Dynamically regulated (in NaR) TF sorted by cell type in which they are most strongly expressed
Non dynamically regulated TF that are DE between NaR and 3D-RA
CLIQUE > OTHERS
NaR > 3D-RA
3D-RA > NaR
OVERLAP
C
NaR > 3D-RA
3D-RA > NaR
CLIQUE > OTHERS
31
0
121
39
19
106
41
Nr DE TF
Total: 357

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 (Kanamori et al., 2004).

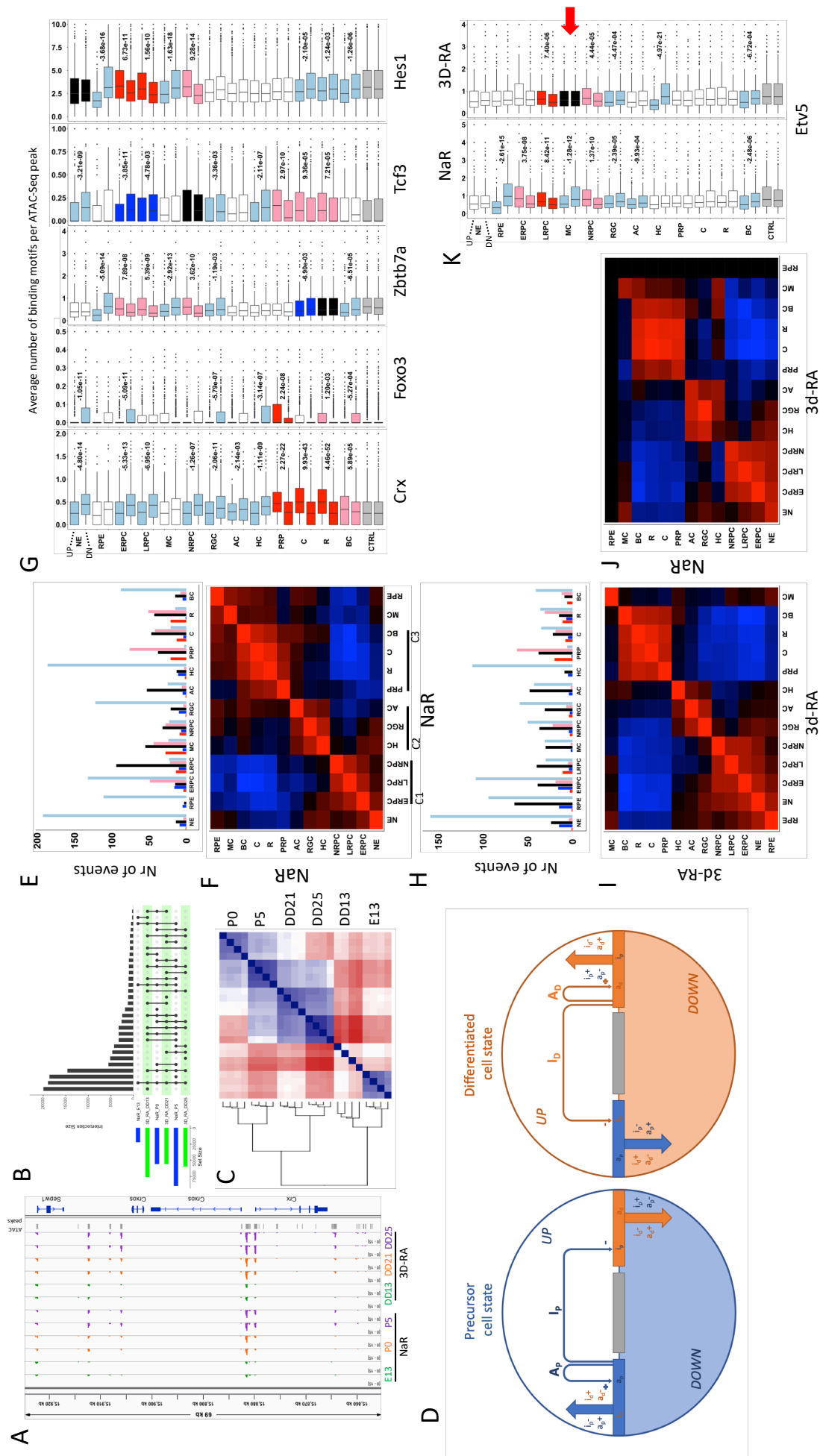
A
B
C
D
E
F
G
H
I
J
K
NaR
3D-RA
P0
P5
DD13
DD21
DD25
E13
Crx
Chrx
Sebw1
peaks
ATAC
Nr of events
Average number of binding motifs per ATAC-Seq peak
Hes1
Tcf3
Zbtb7a
Foxo3
Crx
Etv5
Precursor cell state
Differentiated cell state
UP
DOWN
RPE
MC
BC
R
C
PRP
AC
RGC
HC
NRPC
LRPC
ERPC
NE
CTRL

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. **(D) 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-inhibitor sets of TF. In the example, we show an activator ($\mathbf{A_p}$) – inhibitor ($\mathbf{I_p}$) pair (small blue arrows) turned ON in the precursor state (OFF in the differentiated state), and an activator ($\mathbf{A_d}$) – inhibitor ($\mathbf{I_d}$) pair (small orange arrows) turned ON in the differentiated state (OFF in the precursor state). Genes respond to these activator and inhibitor 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 I_p (labelled as i_p), but (given their expression pattern in differentiated cells) also enriched in binding motifs for $\mathbf{I_d}$ (labelled as i_d) and depleted in binding motifs for $\mathbf{A_d}$ (labelled as a_d) (and vice versa). **(E)** Number of instances of enrichment (red) and depletion (blue) for binding motifs of autologous TF (i.e. overexpressed in the corresponding clique), and enrichment (pink) and depletion (light blue) for binding motifs of heterologous TF (i.e. TF overexpressed in another clique) in ATAC-Seq peaks of genes overexpressed in the corresponding clique, for NaR. Black bars correspond to

TF that are overexpressed in the corresponding clique, yet without binding motif enrichment or depletion in ATAC-Seq peaks of overexpressed (same clique) genes. **(F)** Correlation (red: positive; blue: negative; black: null) between the vectors of enrichment/depletion of 336 binding motifs (corresponding to 157 TF) for all pairs of NaR cliques. **(G)** Examples of TF (from Table 1) that are thought to underpin regulatory toggles. The boxplots illustrate the density of binding motifs for the corresponding TF in genes that are, respectively, over- ("UP") and under-expressed ("DN") in the corresponding clique. The color codes are as in (E), i.e. red: TF over-expressed + enrichment of cognate (autologous) binding motif (=> activator), blue: TF over-expressed + depletion of cognate (autologous) binding motif (=> inhibitor), pink: TF not over-expressed + enrichment of cognate (heterologous) binding motif, light blue: TF not over-expressed + depletion of cognate (heterologous) binding motif, black : TF over-expressed + no enrichment or depletion of cognate binding motif. The p-values of the corresponding enrichment (positive value) or depletion (negative value) are given when $p < 0.01$. Control genes (CTRL in grey) correspond to genes with stable expression across all cliques. **(H)** as in (E) for 3D-RA. **(I)** as in (F) for 3D-RA. **(J)** Correlation (red: positive; blue: negative; black: null) between the vectors of enrichment/depletion of 336 binding motifs (corresponding to 157 TF) for all pairs of NaR vs 3D-RA cliques. RPE were ignored in this analysis (blackened) as the observed differences may results from the low number of RPE cells available for analysis in NaR. **(K)** Possible example of the perturbation of a component of a regulatory toggle in MC. C1 activator Etv5 was found not to be overexpressed in NaR MC yet to be in 3D-RA MC. This may underpin the observation that the genes overexpressed in NaR MC were very significantly depleted in Etv5 binding motifs ($p = 1.3 \times 10^{-12}$), yet that genes overexpressed in 3D-RA MC were not. Color codes and organisation as in (G).

Table 1: Candidate constituents of regulatory toggles operating in the mouse retina (NaR). Expression: Constellation(s) ("Const" defined as in main text) in which the corresponding TF is overexpressed. The activator (A) (respectively inhibitor (I) status) is defined based on the observation of a significant enrichment (respectively depletion) of its binding motif in genes overexpressed in the same clique (autologous TF). Toggle: The sets of constellations (Const A vs Const B) with contrasted enrichment/depletion of binding motifs for the corresponding TF (heterologous TF). Constellations/cliques for which the effects are less pronounced are bracketed. References are provided for TF that have been previously implicated in retinal development.

Transcription factor	Expression		Toggle		References
	Const	Effect	Const A vs Const B		
Hes1	NE-C1	A	C1	C3-MC	Furakawa et al. (2000); Hojo et al. (2000)
Foxp1	NE-C1	I	NE-C1-C2	C3(not BC)	Pohl et al. (2005)
Nr2f1	NE-C1	I	NE-C1-C2	C3	Tang et al. (2010); Bertacchi et al. (2019)
E2f1	C1	A	C1	C2-C3-MC	Dagnino et al. (1997)
Etv5	C1	A	C1	C2-C3-MC	Trimarchi et al. (2008)
Myb	C1	A	C1	NE-C2-BC-MC	Lee et al. (2004)
Nfyb	C1	A	C1	C2-C3-MC	
Tgif1	C1	I	NE-C1	C3	Satoh et al. (2008)
Tgif2	C1	I	NE-C1	C2-C3	Satoh et al. (2008)
Neurog2	C1	I	NE-C1-C2	C3	Hufnagel et al. (2010)
Olig2	C1	I	NE-C1-C2	C3	Nakamura et al. (2006)
Tcf3	C1	I	NE-C1-C2	C3	Trimarchi et al. (2008)
Sox9	C1-(MC)	A	C1	NE-C2-C3	Poche et al. (2008)
Tcf4	C1-(MC)	I	NE-C1-C2	C3	
Barhl2	C2	I	NE-C2	C1-C3(not BC)	Mo et al. (2004)
Meis1	C2	I	NE-C1-C2	C3	Heine et al. (2008)
Neurod2	C2	I	NE-C1-C2-(MC)	C3(not BC)	Inoue et al. (2002)
Nr2f2	C2	I	NE-C1-C2	C3	Tang et al. (2010)
Thra	C2-(BC)	I	NE-C1-C2	C3-MC	Arbogast et al. (2016)
Meis2	C2-C3	A	NE-C1	C3	Bumsted et al. (2007)
Hif1	C2-C3	I	C1	(C2)-C3-MC	Rattner et al. (2019)
Arid3b	C3	A	NE-C1-C2	C3	
Ahr	C3	I	C1	C3-MC	Kim et al. (2014)
Crx	C3	A	NE-C1-C2	C3	Furukawa et al. (1997)
Foxo3	C3	A	NE-C1-C2	C3(not BC)	
Mafg	C3	A	NE-C1-C2	C3	
Neurod1	C3	A	NE-C1-C2	C3	Yan et al. (1998)
Nrl	C3	A	NE-C1-C2	C3	Oh et al. (2007)
Rora	C3	A	NE-C1-C2	C3-(MC)	Fujieda et al. (2009)
Zeb1	C3	A	NE-C1	(C2)-C3	Liu et al. (2009)
Zbtb7a	C3	I	C1	(C2)-C3-MC	
Rorb	(C1)-C3	A	NE-C1-C2	C3-MC	Jia et al. (2009)
Etv1	(NE)-MC	I	C1	C2-C3-MC	