Diversification of multipotential postmitotic mouse retinal ganglion cell precursors into discrete types

Karthik Shekhar[1,2,3,4*], Irene E. Whitney[4,5], Salwan Butrus[1], Yi-Rong Peng[4,6], and Joshua R. Sanes[4*]

Author Affiliations:

1. Department of Chemical and Biomolecular Engineering; Helen Wills Neuroscience Institute; Center for Computational Biology; California Institute for Quantitative Biosciences, QB3, University of California, Berkeley, CA, USA 94720

2. Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA USA 94720

3. Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

4. Center for Brain Science and Department of Molecular and Cellular Biology, Harvard University, Cambridge MA, 02138

5. Present address: Honeycomb Biotechnologies 940 Winter Street Waltham MA 02451

6. Department of Ophthalmology, Stein Eye Institute, UCLA David Geffen School of Medicine, Los Angeles, CA 90095, USA

* Correspondence
Karthik Shekhar (kshekhar@berkeley.edu)
Joshua R Sanes (sanesj@mcb.harvard.edu)

Abstract (150 words)

The genesis of broad neuronal classes from multipotential neural progenitor cells has been extensively studied, but less is known about the diversification of a single neuronal class into multiple types. We used single-cell RNA-seq to study how newly-born (postmitotic) mouse retinal ganglion cell (RGC) precursors diversify into ~45 discrete types. Computational analysis provides evidence that RGC type identity is not specified at mitotic exit, but acquired by gradual, asynchronous fate restriction of postmitotic multipotential precursors. Some types are not identifiable until a week after they are generated. Immature RGCs may be specified to project ipsilaterally or contralaterally to the rest of the brain before their type identity has been determined. Optimal transport inference identifies groups of RGC precursors with largely non-overlapping fates, distinguished by selectively expressed transcription factors that could act as fate determinants. Our study provides a framework for investigating the molecular diversification of discrete types within a neuronal class.

Keywords: development, diversification, retina, retinal ganglion cell, single-cell RNA-seq, optimal transport

Introduction

A central question in developmental neurobiology is how the brain's diverse neuronal types arise from multipotential progenitors (Lodato & Arlotta, 2015; McConnell, 1991; Wamsley & Fishell, 2017). The vertebrate retina has been a valuable model for addressing this question: it is about as complicated as any other region of the brain, but has several features that facilitate mechanistic analysis (Dowling, 2012). The retina contains five classes of neurons – photoreceptors that sense light, three interneuronal classes (horizontal, bipolar and amacrine cells) that process visual information, and retinal ganglion cells (RGCs) that pass the information to the rest of the brain through the optic nerve (**Fig. 1a**) (Masland, 2012). These classes can be divided into numerous types, ~130 in mouse and chick, each of which has characteristic morphological, physiological and molecular properties, and plays distinct roles in information processing (Baden et al., 2016; Franke et al., 2017; Goetz et al., 2021; Macosko et al., 2015; Rheaume et al., 2018; Shekhar et al., 2016; Shekhar & Sanes, 2021; Tran et al., 2019; Yamagata, Yan, & Sanes, 2021; Yan et al., 2020). Remarkably, nearly all types are distributed across the entire retina (Kay, Chu, & Sanes, 2012; Keeley, Eglen, & Reese, 2020; Rockhill, Euler, & Masland, 2000), so morphogen gradients, which play a critical role in other parts of the central nervous system (e.g. (Sagner & Briscoe, 2019)), cannot provide an explanation for retinal neuronal diversification (Marquardt & Gruss, 2002).

Seminal studies have provided deep insights into how retinal classes arise(Bassett & Wallace, 2012; Cepko, 2014). First, lineage tracing in rodents and frogs showed that single retinal progenitor cells (RPCs) can give rise to neurons of all classes as well as glia, and are therefore multipotential (Holt, Bertsch, Ellis, & Harris, 1988; Turner & Cepko, 1987; Turner, Snyder, & Cepko, 1990; Wetts & Fraser, 1988). Second, the competence of multipotential RPCs to generate cells of particular classes changes over time, accounting for their sequential (but overlapping) birth windows (Cepko, 2014; Livesey & Cepko, 2001). Such segregation of birth windows is a hallmark of many neuronal systems (Holguera & Desplan, 2018), and is believed to arise from the differential temporal regulation of gene expression in RPCs (Blackshaw et al., 2004; Brown, Patel, Brzezinski,

& Glaser, 2001; S. Chen et al., 1997; Clark et al., 2019; Trimarchi, Stadler, & Cepko, 2008). Third, competence is probabilistic rather than deterministic, with stochastic factors accounting for variations in the distribution of cell classes generated by individual RPCs (Boije, MacDonald, & Harris, 2014; Gomes et al., 2011; Johnston & Desplan, 2010).

In contrast to these well-established tenets of neuronal class generation, we know far less about how immature postmitotic neurons (which we call neuronal precursors here) committed to a specific class identity diversify into distinct types. We address this issue here, focusing on RGCs. All RGCs are similar in many respects: for example, they all elaborate dendrites that receive input from amacrine and bipolar interneurons, send axons through the optic nerve, and use glutamate as a neurotransmitter (Sanes & Masland, 2015). However, they differ in molecular, morphological and physiological details, which have led to their division into ~45 distinct types in mice (Baden et al., 2016; Bae et al., 2018; Goetz et al., 2021; Rheaume et al., 2018; Tran et al., 2019). Most of these types appear to be feature detectors that collectively transmit a diverse set of highly processed images of the visual world to the rest of the brain (Baden, Euler, & Berens, 2020; Sanes & Masland, 2015). Several genes have been implicated in maturation of a few mouse RGC types (Clark et al., 2019; Kiyama et al., 2019; Liu et al., 2018; Lo Giudice, Leleu, La Manno, & Fabre, 2019; Lyu & Mu, 2021; Peng et al., 2017; Sajgo et al., 2017), but a comprehensive investigation of RGC diversification has been lacking.

To gain insight into how and when adult RGC types emerge, we used high-throughput single-cell RNA-seq (scRNA-seq) to profile RGC precursors during embryonic and postnatal life in mice. We find that the number and distinctiveness of molecularly defined groups of precursors increases with developmental age, implying that types arise by a gradual process rather than from ~45 committed precursor types. Using statistical inference approaches, we identify fate associations among immature RGCs as transcriptomically distinct types emerge. These analyses suggest a model in which types arise from multipotential precursors by a process of restriction that we term fate decoupling. The decoupling is gradual and asynchronous, resulting in different types emerging at different times. We also use markers of RGCs that project to contralateral or

ipsilateral retinorecipient areas to subdivide each type by its projection pattern, leading to the conclusion that laterality may be specified prior to type identity is fixed. Together, our results provide both a model of RGC diversification and a computational framework that can be applied generally to analyze the diversification of closely related neuronal types within a class.

Results

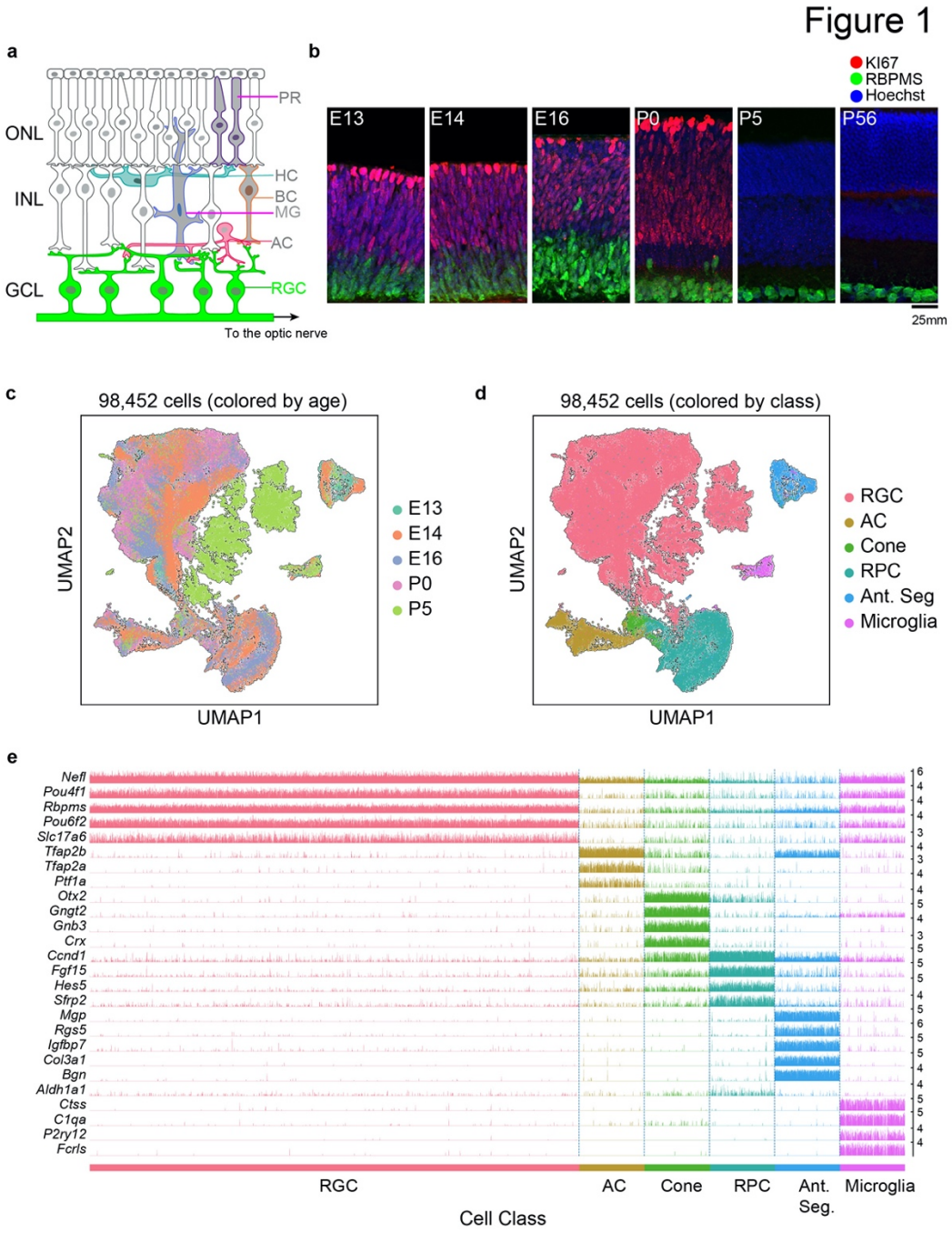
Transcriptomic atlas of developing mouse RGCs

Mouse RGCs are born between approximately embryonic days (E) 11 and 17 with new-born RGCs exiting the mitotic cycle near the apical margin, then migrating basally to form the ganglion cell layer (Drager, 1985; Marcucci, Soares, & Mason, 2019; Voinescu, Kay, & Sanes, 2009) (**Fig. 1b**). Reported birthdates differ among publications, and are complicated by naturally occurring cell death and the central-peripheral developmental gradient, but a detailed analysis concludes that >95% of RGCs in adult mouse retina are born after E12.8 and >85% before E16 (Farah & Easter, 2005). Shortly after they are born, RGCs extend axons through the optic nerve, with some reaching retinorecipient areas by E15 (Godement, Salaun, & Imbert, 1984; Osterhout et al., 2011) and forming diverse projection patterns (Martersteck et al., 2017). During early postnatal life, they extend dendrites apically into the inner plexiform layer of the retina, receiving synapses from amacrine cells by postnatal day (P)4 and bipolar cells a few days later (Kim, Zhang, Meister, & Sanes, 2010; Lefebvre, Sanes, & Kay, 2015). Light responses are detected in RGCs by P10 but image-forming vision does not begin until eye-opening, around P14 (Hooks & Chen, 2020).

To determine when and how RGCs diversify, we used droplet-based scRNA-seq (Macosko et al., 2015; Zheng et al., 2017) to profile them at 5 stages: E13 and E14 (during the period of peak RGC genesis), E16 (by which time RGCs axons are reaching target retinorecipient areas), P0 (as dendrite elaboration begins), and P5 (shortly after RGCs begin to receive synapses). As RGCs comprise ≤1% of retinal cells (Jeon, Strettoi, & Masland, 1998), we enriched them with antibodies to two RGC-selective cell-surface

166 markers, Thy1/CD90 (Barres, Silverstein, Corey, & Chun, 1988) and L1cam
167 (Demyanenko & Maness, 2003) (**Supplementary Fig. 1a**).
168
169 We obtained a total of 98,452 single-cell transcriptomes with acceptable quality metrics
170 (**Methods**). Of these, we identified 75,115 (76%) as RGCs based on their expression of
171 canonical RGC markers including *Rbpms* (an RNA-binding protein) and *Slc17a6* (the
172 vesicular glutamate transporter VGLUT2) (**Figs. 1c-e, Supplementary Figs. 1b-c**). Non-
173 RGCs included amacrine cells (*Tfap2a*+*Tfap2b*+), cone photoreceptors (*Otx2*+*Crx*+),
174 microglia (*P2ry12*+*C1qa*+), anterior segment cells (*Mgp*+*Bgn*+), and retinal progenitor
175 cells (RPCs). Anterior segment cells were found only in E13 and E14 samples because
176 whole eyes were dissociated at these stages. RPCs formed a continuum, containing both
177 "primary" RPCs expressing cell-cycle related genes (e.g. *Mki67, Ccnd5, Birc5*) and
178 previously described RPC regulators (e.g. *Sfrp2*, *Vsx2* and *Fgf15*), and "neurogenic"
179 RPCs expressing proneural transcription factors (e.g. *Hes6*, *Ascl1*, *Neurog2*)(Clark et al.,
180 2019). Importantly, these markers were not expressed in cells annotated as RGCs
181 (**Supplementary Fig. 1d**). These stringent criteria ensured that our dataset comprised
182 postmitotic committed RGCs, allowing us to focus on their diversification and maturation.
183
184 Overall, we recovered ~5,900 to 18,500 RGCs at each of the five time points. Of the two
185 surface markers used for enriching RGCs, Thy1 was effective at later stages as shown
186 previously (Kay et al., 2011; Rheaume et al., 2018; Tran et al., 2019), whereas L1cam
187 expression was more selective at E13 and E14 (**Supplementary Figs. 1b,c**). To evaluate
188 the effectiveness of our enrichment strategy at early stages, we compared our data with
189 two recent studies in which developing retinal cells were profiled using scRNA-seq without
190 any enrichment (Clark et al., 2019; Lo Giudice et al., 2019). A joint analysis of these
191 datasets at embryonic time points showed consistency in the transcriptional signatures of
192 major cell groups without discernible biases (**Supplementary Figs. 1e-g**). However, our
193 enrichment protocols increased the fractional yield of RGCs by >3X at E14 and E16 and
194 by >100X at P0 (**Supplementary Fig. 1h**), which enabled us to resolve heterogeneity
195 within this class at immature stages. For the analysis that follows, we compared precursor

196 RGCs (E13-P5) to a previously described dataset of 35,699 mature RGCs at P56(Tran
197 et al., 2019).



198

199 **Figure 1, Transcriptomic profiling of single postmitotic RGCs during embryonic and**
200 **postnatal development in mice.**
201 a. Sketch of a section of the mouse retina showing major cell classes - photoreceptors (PRs; rods
202 and cones), horizontal cells (HCs), bipolar cells (BCs), amacrine cells (ACs), Müller glia (MGs)
203 and retinal ganglion cells (RGCs). PRs reside in the outer nuclear layer (ONL), while BCs, HCs

and most ACs reside in the inner nuclear layer (INL). RGCs and some ACs reside in the ganglion cell layer (GCL). Axons of RGCs project to higher visual areas via the optic nerve.
b. Retinal section of the indicated ages labeled for the cell-cycle marker *MKI67* (red) and the RGC marker *RBPMS* (green); nuclei are counterstained by the Hoeschst dye (blue). Micrographs are orientated as the schematic in panel a.
c. Visualization of transcriptional diversity of 98,452 cells using Uniform Manifold Approximation and Projection (UMAP), a nonlinear dimensionality reduction algorithm that assigns proximal x-y coordinates to cells (dots) with similar transcriptional profiles(Becht et al., 2019).
d. Same as c, with cells colored by cell class, assigned based on transcriptional signatures displayed in panel e. RPC, retinal progenitor cells; Ant. Seg., anterior segment cells.
e. Tracksplot showing expression patterns of cell-class specific marker genes (rows) across single cells (columns). Cells are grouped by class as in d. For each class, we randomly sampled 20% of total cells covering all immature time points (E13, E14, E16, P0, P5). For each gene, the scale on the y-axis (right) corresponds to normalized, log-transformed transcript counts detected in each cell.

Immature RGCs diversify postmitotically

One can envision two extreme models of RGC diversification. In one, RGC type would be specified at or before mitotic exit, with each type arising from a distinct set of committed precursors. At the other extreme, all precursor RGCs would be identical when they exit mitosis, and gradually acquire distinct identities as they mature (**Fig. 2a**). Intermediate models could involve multiple groups of precursor RGCs, each biased towards a distinct set of terminal types.

To distinguish among these alternatives, we analyzed the transcriptomic diversity of RGCs at each developmental stage using the same dimensionality reduction and graph clustering approaches devised for analysis of adult RGCs (Tran et al., 2019) (see **Methods**). This analysis led to three main results.

First, RGCs were already heterogeneous soon after mitotic exit. There were 10 transcriptionally defined precursor clusters at E13 (**Fig. 2b**), before or at the peak time of RGC birth. The number of discrete clusters increased only slightly by E14 (from 10 to 12; **Fig. 2c**), arguing against a model in which the number of precursor types extrapolated back to one. No single cluster dominated the frequency distribution at either time, as would be expected if a totipotent precursor RGC were to exist shortly after terminal mitosis.

Second, the number of transcriptionally defined clusters increased gradually, between E13 and adulthood, reaching 45 only after P5 (**Figs. 2b-g**). Several arguments indicate that this increase is biologically significant rather than being an artifact of the data or computational analysis. (1) We used the same clustering procedure at all ages. (2) The qualitative trends were robust against variations in clustering parameters. (3) All embryonic clusters contained cells isolated with both cell markers, L1cam and *Thy1* (**Supplementary Figs. 2a-c**), indicating that lower cluster numbers at early stages did not result from biased collection methods. (4) The increase in the number of effective molecular types was robust as demonstrated by three diversity indices - Rao, Simpson and Shannon – all of which buffer against artificial inflation of diversity due to small clusters (**Fig. 2h**, **Supplementary Fig. 2d; see Methods**). (5) There was no systematic dependence of the number of clusters on the number of cells. For example, we identified 12 clusters from 17,100 cells at E14 and 38 clusters from 17,386 cells at P5.

Third, the transcriptomic variation became increasingly discrete with age. We quantified this increase in inter-cluster separation by calculating (1) the average cross-validation error of a multi-class classifier, and (2) the ratio of mean cluster diameter to mean inter-cluster distance in the low dimensional embedding (**Methods**). Both metrics decrease in numerical value as the clusters are more-well defined. From these trends, we conclude that the boundaries between RGC clusters become sharper as development proceeds (**Figs. 2i,j)**.

Taken together, our results show that transcriptomic clusters of RGCs increase in number and distinctiveness over time, making it unlikely that RGC type identity is fully specified at the progenitor stage.

Figure 2

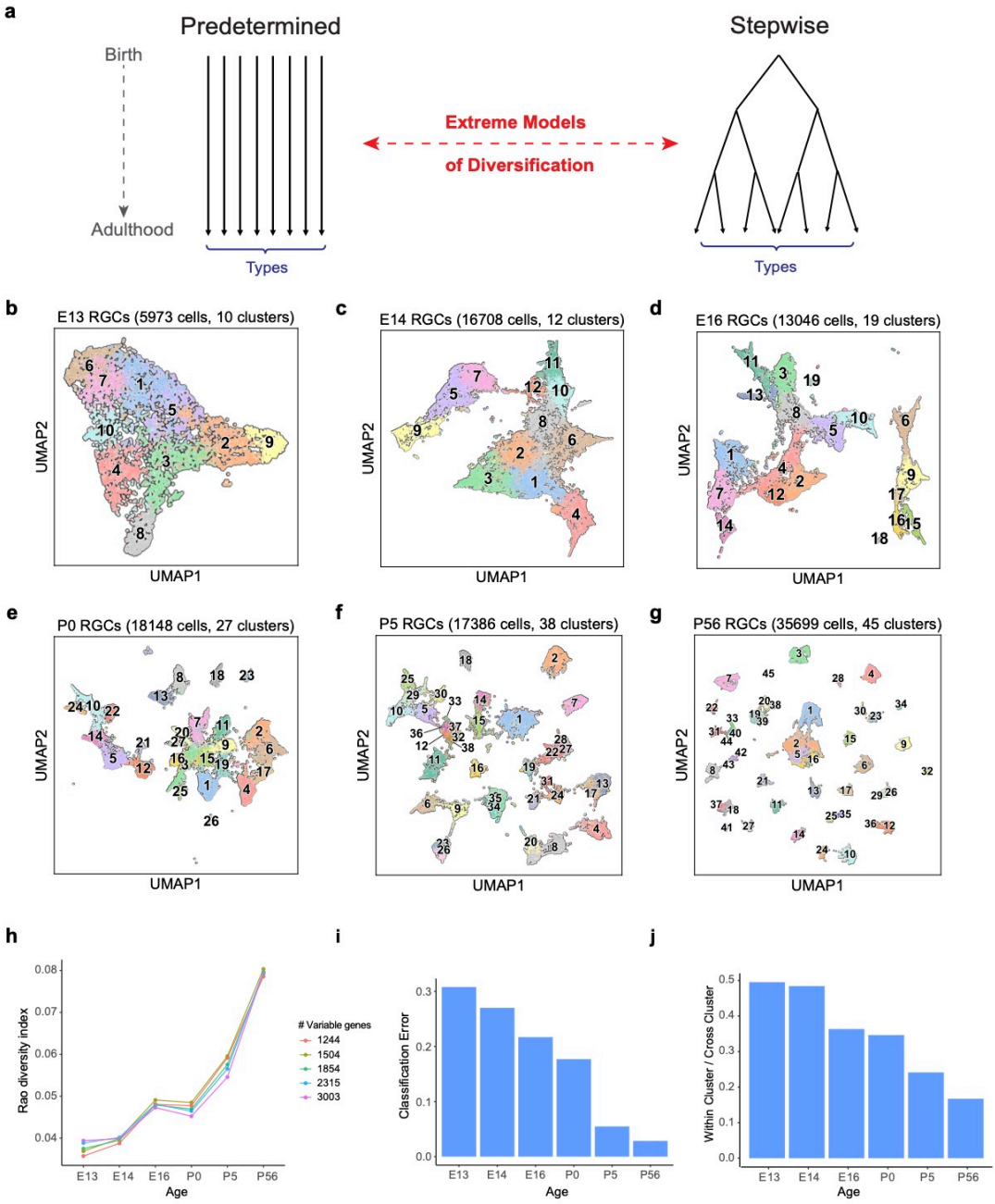


Figure 2. The number and discreteness of transcriptomic clusters of RGCs increases with age.

a. Extreme models of RGC diversification. In one scenario (*left*) immature RGCs commit to one of the terminal types by the time of birth (i.e. mitotic exit) or shortly after. Alternatively (*right)*, initially identical postmitotic RGC precursors acquire distinct molecular identities in a gradual process of restriction.

b-g. Visualization of transcriptomic diversity of immature RGCs at E13 (b), E14 (c), E16 (d), P0 (e), P5 (f) and P56 (g) using UMAP. Cells are colored by their cluster identity, determined independently using dimensionality reduction and graph clustering (**Methods**). Clusters are numbered based on decreasing size at each age. Data for adults (P56) are replotted from (Tran et al., 2019). In that study 45 transcriptomic types were identified via unsupervised approaches, one of which was mapped to 2 known functional types by supervised approaches. We do not distinguish them in this study.
h. Transcriptional diversity of RGCs as measured by the Rao diversity index (y-axis) increases with age (x-axis). The trend is insensitive to the number of genes used to compute inter-cluster distance (colors). See **Methods** for details underlying the calculation.
i. Transcriptomic distinctions between RGC clusters become sharper with age as shown by decreasing average per-cluster error of a multiclass-classifier with age. Gradient boosted decision trees(T. Chen & Guestrin, 2016) were trained on a subset of the data, and applied on held out samples to determine the test error.
j. RGC clusters also become better separated in the UMAP embedding, as shown by decreasing values of the average relative cluster diameter with age.

Temporal relationships among immature RGC clusters

We next investigated the temporal relationships among precursor RGC clusters identified at different ages. We again consider two extreme models. In a "specified" model, each terminal type arises from a single cluster at every preceding developmental stage (**Fig. 3a, left**). In this model, distinct transcriptomic states among precursor RGCs correspond to distinct groups of fates. At the other extreme, distinct clusters would share similar sets of fates (**Fig. 3a, right**). In an intermediate model, fates of precursor clusters would exhibit partial overlap.

As a first step in discriminating among these scenarios, we used transcriptome-wide correspondence among clusters as a proxy for fate association. We identified mappings between clusters across each pair of consecutive developmental stages (E13-E14, E14-E16, E16-P0, P0-P5, and P5-P56) using gradient boosted trees (T. Chen & Guestrin, 2016), a supervised classification approach (**Methods**). In each case, a classifier trained on transcriptional clusters at the older stage was used to assign older cluster labels to cells at the younger stage (e.g. E16 labels assigned to E14 RGCs). Patterns expected for the extreme models are schematized as "confusion matrices"(Stehman, 1997) in the lower panels of **Fig. 3a.**

Correspondence fell between the two extremes (**Figs. 3d-h and Supplementary Figs. 3a-d**). We quantified the extent of correspondence using two metrics: Normalized

Conditional Entropy (NCE) and the Adjusted Rand Index (ARI) (**Methods**). Both NCE and ARI are restricted to the range (0,1), with lower values of NCE and higher values of ARI consistent with a specified mode of diversification. Both metrics exhibited an increased degree of specificity with age (**Figs. 3b,c**). Since NCE and ARI provide a single measure of specificity for the entire datasets being compared, we also computed a "local metric", the Occupancy Fraction, which quantifies mapping specificity for each cluster (**Methods**). Results based on this metric were consistent with increased specificity of correspondence with age (**Supplementary Fig. 3e**). Overall, this analysis of transcriptomic correspondence suggests that poorly specified relationships among transcriptomic clusters at early stages are gradually refined to yield increasingly specific associations at later stages.

Figure 3

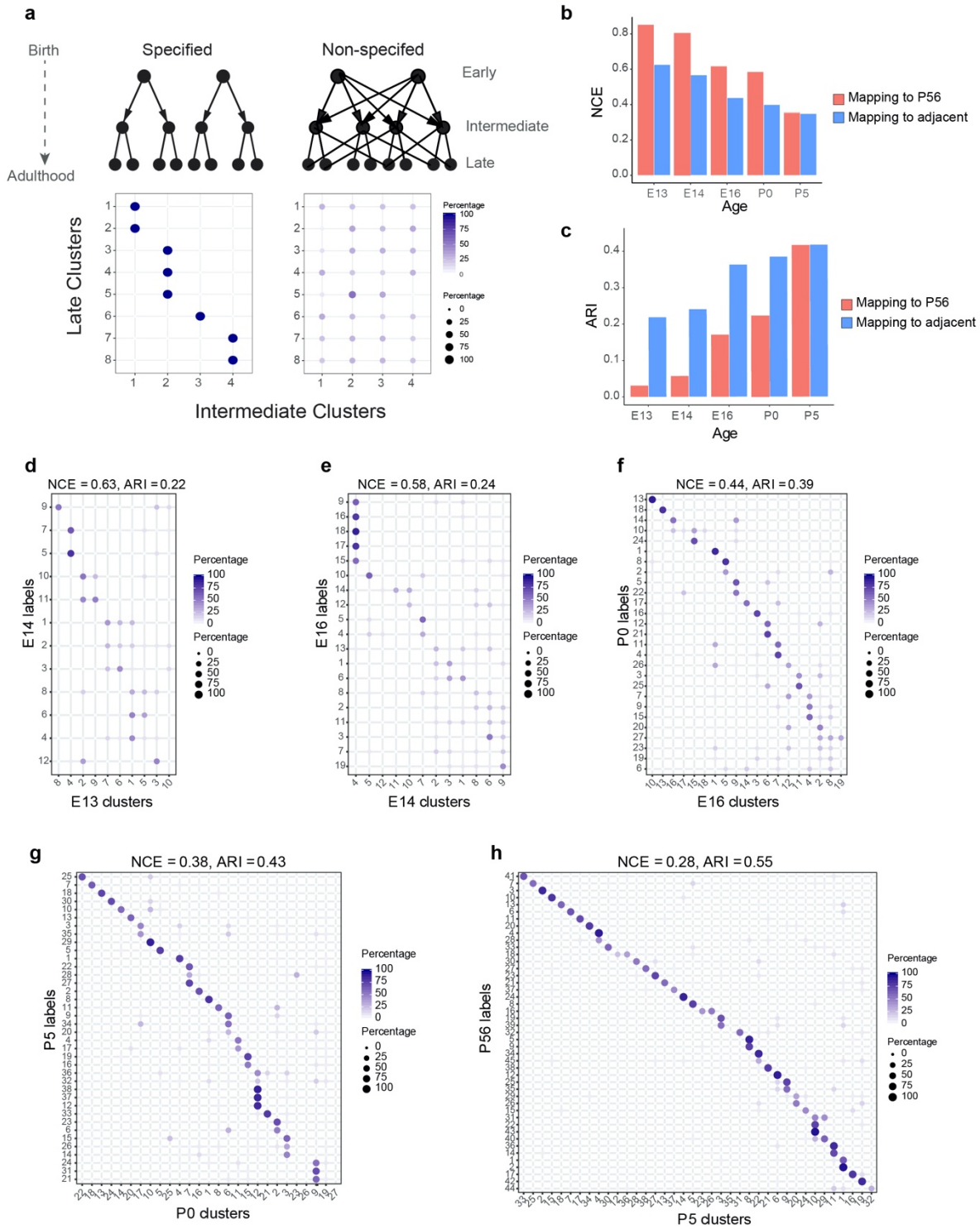


Figure 3. Incompletely specified temporal relationships among RGC clusters.
a. *Top:* Specified (*left*) and non-specified (*right*) modes of diversification. Nodes denote transcriptomic clusters of immature RGCs, and arrows denote fate relationships. *Bottom:*

Confusion matrices depicting transcriptomic correspondence between late and early clusters expected for the two modes. Circles and colors indicate the percentage of a given late cluster (row) assigned to a corresponding early cluster (column) by transcriptome-based classifier trained on early clusters. The number of late and early clusters have been set to eight and four for illustration purposes.
b. Barplot showing values of the normalized conditional entropy (NCE) for each age calculated using the transcriptional cluster IDs and the Xgboost-assigned cluster IDs corresponding to the next age or to P56 (E.g. for E13, the NCE was calculated across E13 RGCs by comparing their transcriptional cluster ID with assigned E14 cluster IDs based on a classifier trained on the E14 data). Lower values indicate specific mappings.
c. Same as b, but plotting values of the adjusted Rand Index (ARI), where larger values correspond to higher specificity.
d-h. Confusion matrices (representation as in a), showing transcriptomic correspondence between consecutive ages: E14-E13 (d), E16-E14 (e), P0-E16 (f), P5-P0 (g), P56-P5 (h). In each case, the classifier was trained on the late time point and applied to the early time point. Rows sum to 100%.

Immature RGCs are multipotential

The analysis presented so far relied on comparing clusters between ages and was therefore unable to link individual precursors to specific terminal fates. At one extreme, individual precursor clusters might contain several groups of cells, each committed to a distinct, small number of fates. Alternatively, individual cells might be as multipotential as the clusters in which they reside (**Fig. 4a**).

Cluster-based classification frameworks do not afford a straightforward way to explore variations in patterns of fate associations within clusters. We therefore turned to Waddington-Optimal Transport (WOT), a computational method rooted in optimal transport theory (Kantorovich, 1942; Monge, 1781) that utilizes scRNA-seq measurements at multiple stages, to infer developmental relationships (Schiebinger et al., 2019). Briefly, WOT computes a "transport matrix" Π between each pair of consecutive ages with elements Π_{ij} encoding fate associations between a single RGC *i* at the younger age and RGC *j* at the older age (see **Methods**). Thus, WOT identifies fate associations between individual cells without invoking clustering. We conducted extensive computational tests to assess the numerical stability of associations reported by WOT (**Supplementary Fig. 4**). We also determined that when collapsed to the level of clusters, the WOT inferred transport maps strikingly mirrored the confusion matrices obtained from multi-class classification (**Supplementary Fig. 5**).

Based on the success of these tests, we applied WOT to compute the "terminal fate" for each precursor RGC. We leveraged the fact that in WOT, fate associations between RGCs at non-consecutive ages (e.g. E16 and P56) can be estimated in a principled way by multiplying the intermediate transport matrices. This yielded a fate vector $\vec{f}$ for each of the 75,115 immature RGCs, whose *kth* element f_k represents the predicted probability of commitment to adult type $k \in (1, 2, \ldots, 45)$ (**Methods**). A fully committed precursor would have all but one element of $\vec{f}$ equal to zero, whereas a partially committed precursor would have multiple non-zero elements in $\vec{f}$. Since the elements of $\vec{f}$ are interpreted as probabilities, they are normalized such that $\sum_k f_k = 1$.

We quantified the commitment of each precursor by computing its "potential" $P = \frac{1}{\sum_k f_k^2}$, which is defined analogously to the "inverse participation ratio" in physics (Fyodorov & Mirlin, 1992). In our case, the value of *P* for a given RGC ranges continuously between 1 and 45, with lower values implying a commitment to specific fates, and higher values reflecting indeterminacy. Importantly, this measure of commitment does not rely on arbitrary thresholding of the f_k values to assign precursors to types.

Five results emerged from this analysis.

- Nearly all prenatal RGCs (i.e. on or before P0) were multipotential rather than committed to a single terminal fate, with individual potentials distributed across a range of values (**Fig. 4b**).

- Multipotentiality was a general feature of immature RGCs, being present in cells of all clusters at E13, E14 and E16 (**Figs. 4c-f**).

- At early stages the average value of *P* varied among transcriptomic clusters, reflecting asynchronous specification (**Fig. 4c**). The tempo of commitment is further explored in the next section.

- Although they were multipotential, no precursor RGC was totipotential (i.e. completely unspecified, corresponding to *P*=45). At E13 the average value of P was 11.6 ± 4.9 which was 4-fold lower than the maximum possible value of 45, and no precursor had *P* > 30.

- Finally, inferred multipotentiality decreased gradually during development, and some persisted postnatally, (average $P = 3.4 \pm 2.1$ at P0, and 1.6 ± 0.9 at P5; **Figs. 4g,h**).

From these results, we conclude that early postmitotic RGCs are multipotential but not totipotential, and that type identity is specified gradually via progressive restriction.

Figure 4

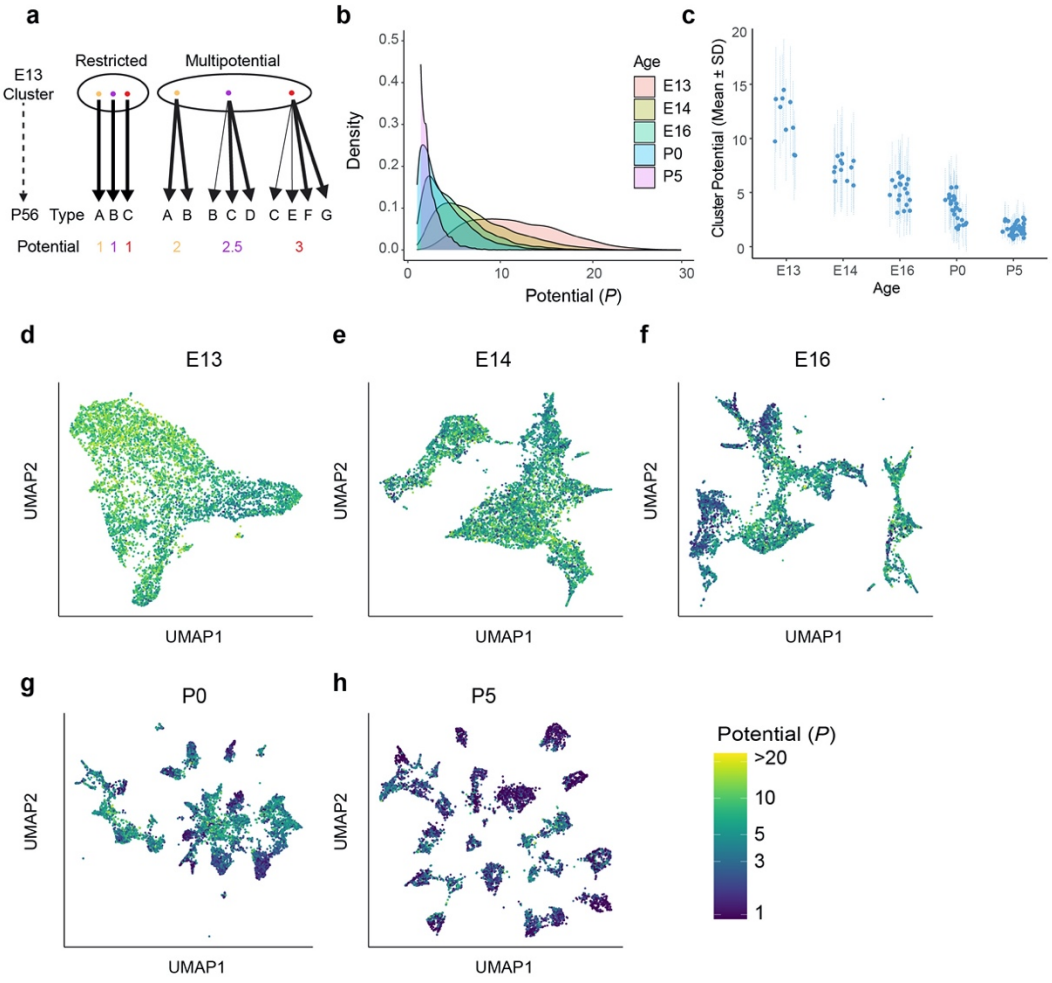


Figure 4. Multipotential fate associations between immature RGCs and terminal types inferred via Optimal Transport

a. Extreme models of diversification at single-cell resolution. Multipotential fate associations in a transcriptionally defined cluster (ellipse) could arise from a mixture of unipotential RGCs (*left*) or from multipotential RGCs (*right*).

b. Distributions of potential *P* across immature RGCs by age showing that restriction increases with age.
c. Inter- and intra-cluster variation of potential by age. At each age, variation in the potential values are shown for each transcriptomically defined cluster at that age. Dots denote the average potential and dotted lines depict the standard deviation for cells within each cluster.
d-h. UMAP projections of E13 (d), E14 (e), E16 (f), P0 (g) and P5 (h) RGCs as in **Fig. 2**, but with individual cells colored by their inferred potential. Potential of all RGCs at P56 =1. The colorbar on the lower right is common to all panels, and values are thresholded at *P* = 20.

Asynchronous specification of mouse RGC types via fate decoupling

As a first step in understanding the progressive restriction of RGC fate, we analyzed the extent to which pairs of mature types were likely to have arisen from the same set of immature precursors. To this end we computed a "fate coupling" value $C(l,m;age)$ for each pair of terminal RGC types (*l* and *m*), defined as the Pearson correlation coefficient between the values of f_l and f_m across all precursors at a given age (**Methods**). f_l and f_m are fate probabilities corresponding to types *l* and *m* as defined in the previous section. Values of $C(l,m;age)$ in our data ranged from -0.11 to 0.95. Higher values of $C(l,m;age)$ indicate strong fate coupling between types *l* and *m*, implying the existence of common postmitotic precursors, whereas low $C(l,m;age)$ values suggest that types *l* and *m* arose from largely nonoverlapping sets of precursors. We visualized the pattern of fate couplings as network graphs, where the nodes represent types and the edge weights represent values of $C(l,m;age)$. The arrangement of nodes was determined at E13 using a force directed layout algorithm (Fruchterman & Reingold, 1991), with pairwise distances being inversely proportional to values of $C(l,m;E13)$, the fate coupling values at E13 (**Fig. 5a**). To visualize the temporal evolution of these fate couplings, we retained the same layout of nodes while updating edge weights according to $C(l,m;age)$ (**Figs. 5b-e)**.

Types that were coupled in fate at the earliest time point gradually decoupled as development proceed. For example, at E13, 118/990 pairs (12%) were strongly coupled (threshold of $C(l,m;age)$>0.2 as determined by randomization tests; see **Methods**), while at P5, only 8/990 (<1%) passed this criterion (**Figs. 5a,e**). Lowering this threshold for coupling to 0.05 increased the number of strongly coupled pairs at P5 to only 2% (20/990).

Different pairs of types decoupled at different rates (**Fig. 5f**). As they decoupled, RGC precursors became increasingly restricted to a single type (i.e. $f_k \gg f_{l \neq k}$ for a precursor favoring type *k*). This corresponded to a "localization" of precursors in transcriptomic space, and is a proxy for specification (see **Methods**). We modeled the extent of localization vs. age via a logistic function (**Fig. 5g and Supplementary Fig. 6d**), and used this to calculate a specification time for each type (τ_{sp}) (see **Methods** for details). Based on this analysis, 7/45 types are specified postnatally. The average τ_{sp} for RGCs was E17.8, but individual RGC types exhibited a wide range from E13.9 to P5.2 (**Fig. 5h**). The inferred specification time was not correlated to adult frequency (**Fig. 5i**).

We illustrate this range by considering three pairs of RGC types in **Supplementary Fig. 6**. C12 and C22 (numbered as in Tran et al., 2019; see **Fig. 2g**) exhibit low fate coupling at all ages profiled (**Supplementary Fig. 6a**), indicative of separate precursor populations. In contrast, C19 and C20 decouple only at P0, implying the existence of a common precursor throughout embryogenesis (**Supplementary Fig. 6b**). C21 and C34 display an intermediate pattern, decoupling around E16 (**Supplementary Fig. 6c**). Taken together, these results suggest that RGC types emerge by asynchronous fate decoupling of multipotential precursors.

Figure 5

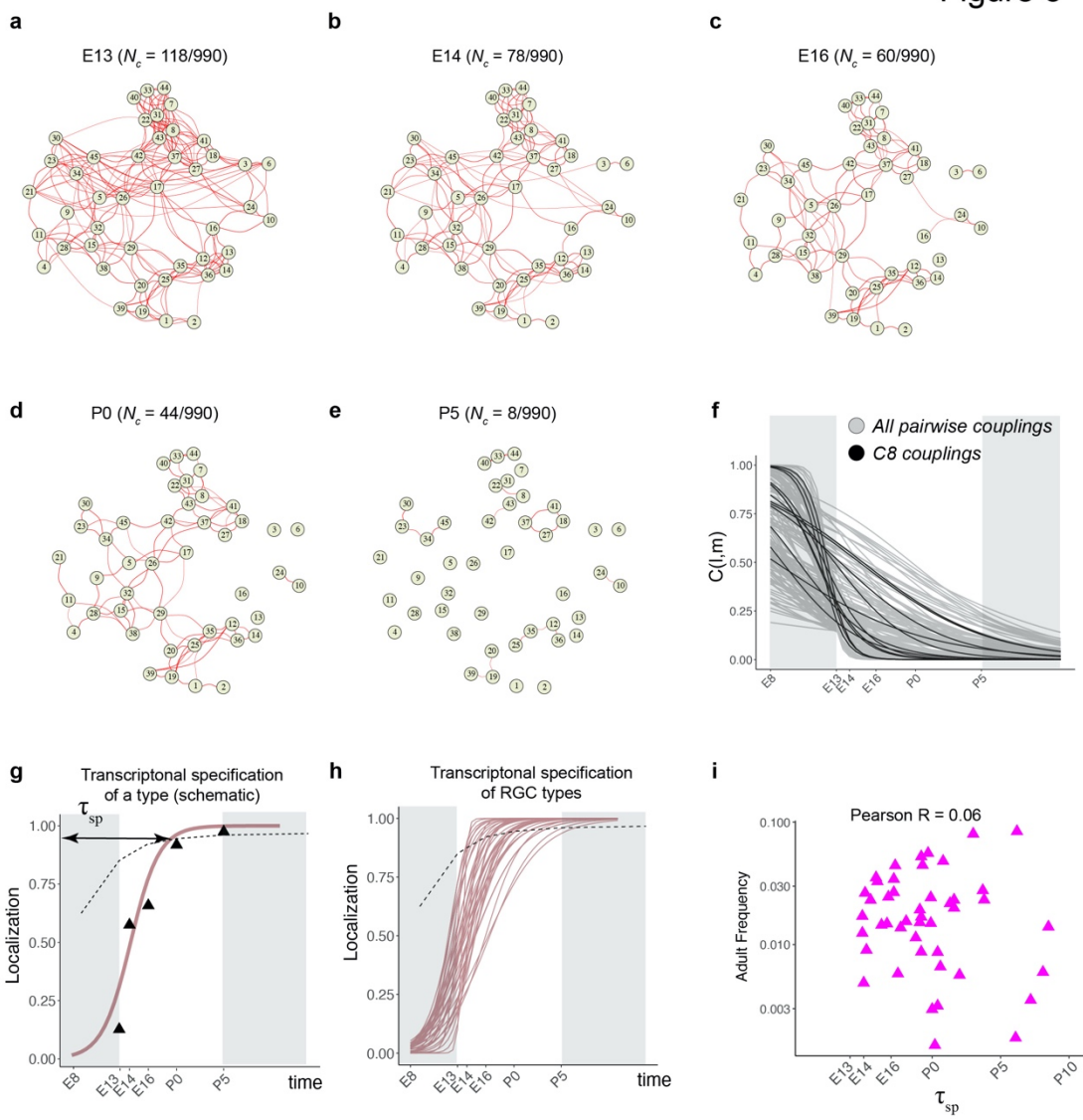


Figure 5. Fate decoupling of RGC types

a. Force-directed layout visualization of fate couplings at E13, with nodes representing RGC types (numbered as in Tran, 2019) and the thickness of edges representing values of *C(l,m;E13)*. Edges with *C(l,m; E13)* < 0.2 are not shown. Number of edges with *C(l,m; E13)* > 0.2 are indicated on top.

b-e. Visualization of fate couplings at E14 (B), E16 (C), P0 (D) and P5 (E). The positions of the nodes are maintained as in panel a, but the edges are redrawn based on values of *C(l,m;age)* at each age. As in panel a, we only show edges *C(l,m; age)* > 0.2.

f. The decay of pairwise fate couplings (y-axis) with age (x-axis). Each line corresponds shows the temporal decay of *C(l,m)* for RGC pair l and m estimated via a logistic model (**Methods**). For each pair, couplings at each age were fit to a model $C(l,m;age) = 1/(1+e^{\beta_0+\beta_1*age})$ with β_0, β_1 representing fitted parameters. The fitting was performed using data for ages E13, E14, E16, P0 and P5. The shaded portions correspond to the periods E8-E13 and P5- represent extrapolations

of the model. Black lines highlight the decay of all non-zero pairwise couplings for RGC type *C8* as an example.

g. Schematic showing logistic modeling to estimate specification time τ_{sp} for a particular type. The y-axis is a measure of the extent to which precursors biased towards the type are present in a single transcriptomically defined cluster (i.e. localization, see **Methods** for details). Localization is defined as a numerical value in the range (0, 1) with higher values consistent with increasing specification. Individual triangles represent the localization values computed using WOT inferred fate couplings at each age, while the curve represents the fit using the logistic model. Dotted line shows the minimum threshold a type to be specified at each age. Its curved shape arises due to the increase in the number of clusters with age.

h. Localization curves (as in panel g) for the 38 RGC types showing the range of inferred specification times. 7 low frequency types have been excluded from display (see **Supplementary Fig. 6d**).

i. Scatter plot showing poor correlation between adult frequency of a type (from (Tran et al., 2019)) and its predicated specification time (calculated from h).

Fate decoupled groups of RGC types defined by transcription factors

Because fate coupling is a metric of inferred overlap of developmental history, it is likely that tightly coupled types share common precursors. This relationship implies that tightly coupled types might also be specified by common transcriptional programs. As a step towards identifying candidate fate determinants, we identified 8 transcription factors (TFs) that are expressed by distinct groups of mature RGC types (**Fig.6a, Supplementary Fig. 7a**). Three of these are well-characterized RGC-selective TFs: *Foxp2*, expressed by 5 F-RGC types (Rousso et al., 2016); *Tbr1*, expressed by 5 T-RGC types (Liu et al., 2018); and *Eomes* (also known as *Tbr2*), expressed by 7 types (C.-A. Mao et al., 2020; Tran et al., 2019). The 7 Eomes/Tbr2 types include the melanopsin expressing intrinsically photosensitive (ip) RGC types (Berson, Dunn, & Takao, 2002). The remaining five were *Neurod2*, *Irx3*, *Mafb*, *Tfap2d,* and *Bnc2* that label 8, 5, 4, 6, and 3 types respectively. *Eomes* types also co-expressed *Tbx20* and *Dmrbt1* while *Neurod2* types also co-express *Satb2*. Together, 40/45 mature types expressed at least one of these TFs in a manner that was, with few exceptions, mutually exclusive. In many cases, the fate proximity of types that shared TF expression was obvious (**Fig. 6a).**

We refer to these TF-based groups as fate-restricted RGC subclasses – an intermediate taxonomic level between class and type based on inferred fate relationships. Consistent with their definition, the pairwise fate coupling among types from different subclasses was significantly lower than among types from the same subclass (**Fig. 6b**). Thus, precursor

RGC states associated with any two subclasses are more distinct than those associated with any two types. This is evident by the significant separation at E13 and negligible overlap at P5 for precursors favoring the *Eomes, Mafb* and *Neurod2* subclasses respectively, as shown in **Fig. 6c-e**.

We also asked whether the TF-based subclasses differed in inferred transcriptomic specification time τ_{sp}, as defined in **Fig. 5g**. As shown in **Figs. 6f-h,** and **Supplementary Figs. 7b-e**, four subclasses were specified within a narrow interval (E16.8-E17.2), but three others differed substantially. The average specification time for the *Eomes* group was E14.6 ($p < 0.0001$, Student's t-test, compared to the mean for all types), while that for the *Mafb* and *Neurod2* groups were E16.9 ($p < 0.001$) and E18.5 ($p < 0.0001$), respectively. The early specification of the *Eomes* group is consistent with birthdating studies showing the average earlier birthdate of ipRGCs compared to all RGCs (McNeill et al., 2011).

In summary, our results suggest the existence of fate-restricted RGC subclasses that arise from distinct sets of precursors and diversify into individual types. This method of defining RGC groups, which relies on inferred proximity of precursors in transcriptomic space, is distinct from previous definitions of RGC subclass based on shared patterns of adult morphology, physiology or gene expression (see **Discussion**). Accordingly, the fate couplings at E13 were only weakly correlated with transcriptomic proximity in the adult retina (**Supplementary Fig. 7f**). Further, while TF-based groups align with some previously defined subclasses (e.g. ipRGCs or *Tbr1*+ RGCs), they do not map to others subclasses such as alpha-RGCs (4 types) or T5-RGCs (9 types) (**Supplementary Fig. 7g**).

Figure 6

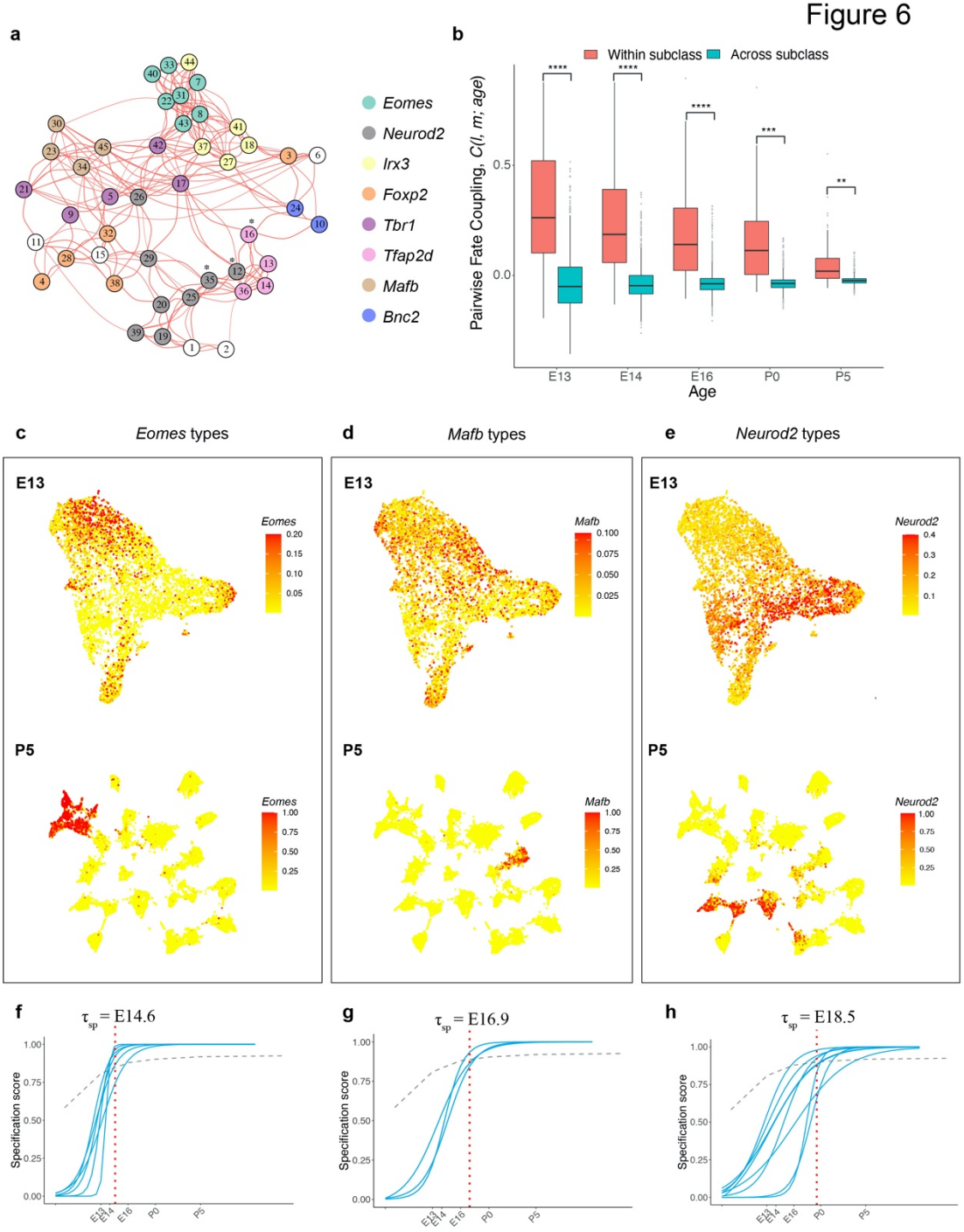


Figure 6. Temporal dynamics of RGC subsets expressing specific TFs

a. E13 network graph of fate couplings from **Fig. 5a**, with RGC types colored based on their selective expression of TFs at P56. Asterisks denote 3/45 types that express more than 1 TF (also see **Supplementary Fig. 7a**).

b. Box-and-whisker plots showing that pairwise fate couplings are higher between types within the same TF subclass than between types in different TF subclasses at all immature ages. Black horizontal line, median; bars, interquartile range; vertical lines, 1st and 99th percentile; dots,

outliers. Stars indicate significant p-values based on a two-sided t-test (****, $p < 10^{-7}$; ***, $p < 10^{-5}$; **, $p < 10^{-2}$).
c. *Eomes*+ types. *Top*: UMAP representation of E13 RGCs with cells colored based on their cumulative fate association towards the 7 *Eomes*+ types. *Bottom*: UMAP representation of P5 RGCs with cells colored based on their cumulative fate association towards the 7 *Eomes*+ types. The value corresponding to the color of each cell (colorbar, right) can be interpreted as the probability of commitment towards the corresponding subclass.
d. Same as c for *Mafb*+ types
e. Same as c for *Neurod2*+ types
f-h. Localization curves (as in **Fig. 5g**) for *Eomes*+ types (f), *Mafb*+ types (g) and *Neurod2*+ types (h). The mean inferred specification time τ_{sp} for each group is indicated on top of each panel.

Transcriptomic profiles of ipsilateral and contralateral RGCs

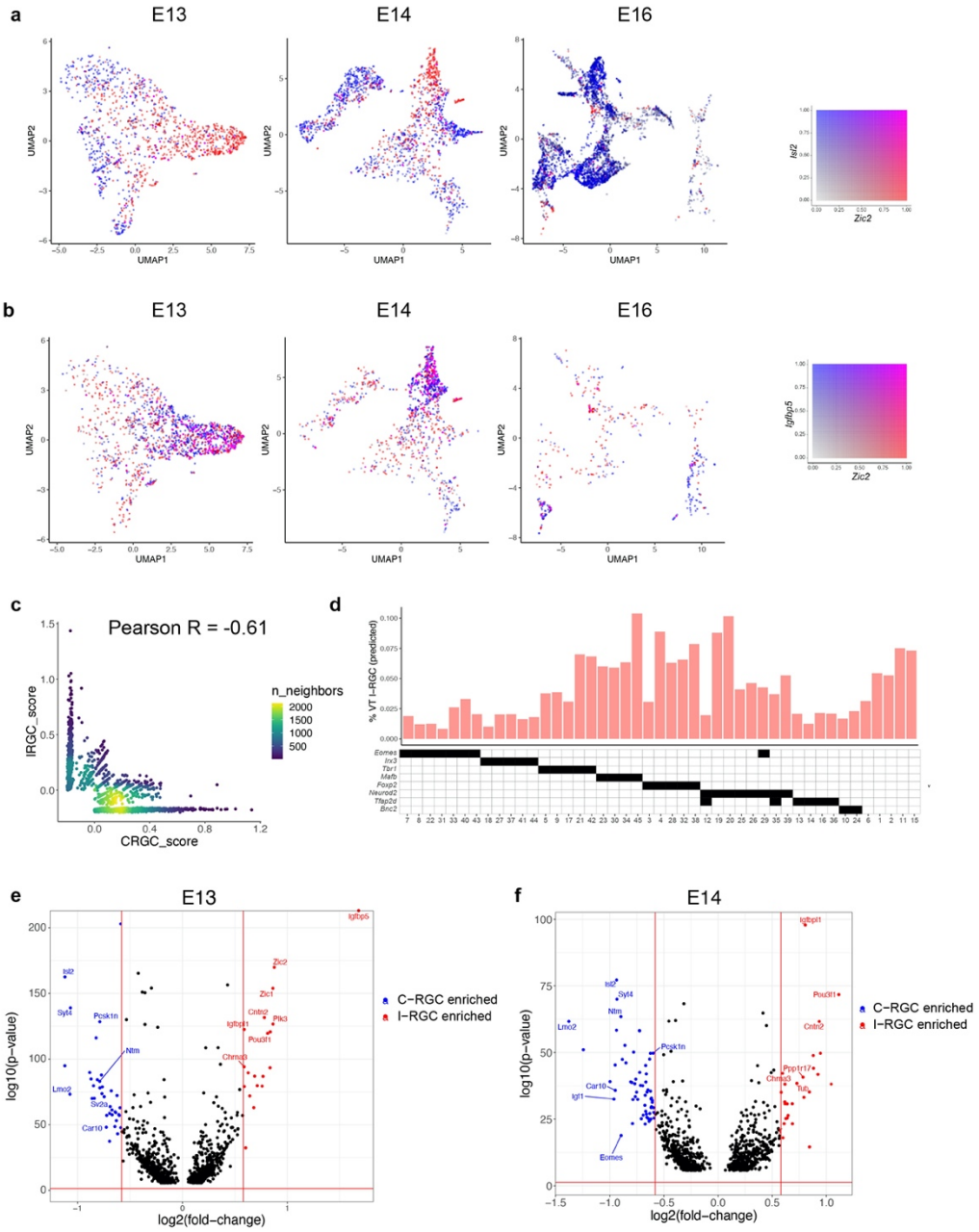
Finally, we considered the origin of two RGC groups defined by their projections: those with axons that remain ipsilateral at the optic chiasm (I-RGCs) and those that cross the midline to innervate contralateral brain structures (C-RGCs). The proportion of I-RGCs varies among vertebrates, in rough correspondence to the extent of binocular vision, ranging from none in most lower vertebrates to ~50% in primates. In mice, 3-5% of RGC axons remain ipsilateral, with most I-RGCs residing in the ventrotemporal (VT) retinal crescent (Mason & Slavi, 2020). While some I-RGCs have been observed to project from the dorsocentral retina during embryonic stages, these are rapidly eliminated so-called "transient" I-RGCs (Soares & Mason, 2015). Thus, in adulthood, C-RGCs are present throughout the retina while "permanent" I-RGCs are confined to the VT crescent.

The zinc-finger transcription factor *Zic2* is expressed in a subset of postmitotic RGCs in VT retina, and is both necessary and sufficient for establishing their ipsilateral identity (Herrera et al., 2003); transient dorsolateral I-RGCs do not express *Zic2* (Pak, Hindges, Lim, Pfaff, & O'Leary, 2004). *Isl2* marks a subset of C-RGCs throughout the retina and appears to specify a contralateral identity in part by repressing Zic2 (Pak et al., 2004). These two transcription factors were expressed in a mutually exclusive fashion in RGC precursors between E13 and E16 (**Fig. 7a**); *Zic*2 was down-regulated at later ages (**Supplementary Fig. 8a**). Furthermore, *Zic2* expression at E13 correlated with *Igfbp5* and *Zic1,* and anti-correlated with *Igf1* and *Fgf12,* consistent with recent reports (Wang, Marcucci, Cerullo, & Mason, 2016) (**Fig. 7b** and **Supplementary Figs. 8b,c**). We scored each cell at E13 based on its expression of ipsilateral genes (**Methods)**, confirming that

587 expression of ipsilateral and contralateral gene signatures were anticorrelated (**Fig. 7c**).
588 Together, these observations support the idea that at E13 *Zic2*+ cells represent I-RGCs,
589 and *Isl2*+ cells represent some but not all C-RGCs.
590
591 Using WOT, we then identified the descendants of presumptive I-RGCs at later ages. We
592 found that I-RGCs comprised 4.3% of adult RGCs, consistent with the range of 3-5%
593 noted above. We queried these cells to identify genes that distinguished putative I-RGCs
594 and C-RGCs throughout the developmental time course. At a fold-change of ≥1.5 we
595 found 59 DE genes at E13 and 89 at E14 (**Figs. 7e,f**). In addition to *Zic2, Igfbp5, Isl2* and
596 *Igf1*, which had been used to define I-RGCs and C-RGCs at E13, they included *Igfbpl1,*
597 *Pou3f1* and *Cntn2* enriched in I-RGCs, and *Lmo2*, *Pcsk1n* and *Syt4* enriched in C-RGCs.
598 The number of genes DE between I- and C-RGCs decreased after E14, with 20, 9 and 0
599 significant genes at E16, P0 and P5, respectively (**Supplementary Figs. 8d,e**),
600 presumably reflecting the downregulation of axon guidance programs once
601 retinorecipient targets have been reached (see **Discussion**).
602
603 We also asked which RGC types included I-RGCs. At E13, putative I-RGCs were highly
604 enriched in 2 of 10 clusters, comprising 38-40% of clusters 2 and 9, 9-14% of clusters 3
605 and 5, and <2% of the other 6 clusters (**Fig. 7d**). In adults, RGCs expressing *Tbr1*, *Mafb,*
606 *Foxp2 and Neurod2* contained 3-4X more I-RGCs than RGCs expressing *Eomes*, *Irx3* or
607 *Tfap2d*. These results are consistent with previous observations that I-RGCs are
608 morphologically and physiologically heterogenous but not uniformly distributed across
609 types(Hong, Kim, & Sanes, 2011; Johnson et al., 2021). Lastly, the WOT-predicted
610 relationship between E13 precursor RGC clusters and I-RGC-rich or -poor terminal types
611 were consistent with these patterns. The top six I-RGC-rich types (C4, C15, C19, C20,
612 C38, C45) derived 14% and 4% of their relative fate association from E13 clusters 2 and
613 9, while the top six I-RGC-poor types (C8, C14, C18, C22, C31 and C41) derived only 3.8
614 and 0.2 % of their relative fate association from E13 clusters 2 and 9. Thus, E13 clusters
615 2 and 9 are preferred precursors of adult types that are relatively rich in I-RGCs.
616
617

618

Figure 7



619

620 **Figure 7. Transcriptomic identification of ipsilaterally projecting RGCs**

621 a. *Zic2*, an I-RGC marker and *Isl2*, a C-RGC marker, are expressed in a mutually exclusive pattern
622 at E13 (left), E14 (middle) and E16 (right). *Zic2* is undetectable after E16 (**Supplementary Fig.**
623 **8a**). Cells are colored based on a bivariate color scale representing co-expression of two markers
624 (colorbar, right).

625 b. *Zic2* and *Igfbp5,* two I-RGC markers, are co-expressed at E13 (left) and E14 (middle).
626 Representation as in panel a.

c. Scatter-plot of gene signatures used to identify I-RGCs (y-axis) and C-RGCs (x-axis) at E13 are negatively correlated (Pearson R = -0.61). Each dot corresponds to a cell, the color represents the number of cells located at a particular (x,y) location (see colorbar, right).
d. Barplot showing % of putative I-RGCs (y-axis) within each of the 45 adult RGC types, estimated by computing the descendants of E13 I-RGCs using WOT. RGC types are arranged along the x-axis based on their membership of TF-groups shown in **Fig. 6a** (annotation matrix, bottom).
e. Volcano plot showing DE genes (MAST test, $p < 10^{-6}$) between predicted I-RGCs and C-RGCs at E13. The x- and the y-axes show the fold-change and the p-value in log2- and log10- units, respectively. Dots represent genes, with red and blue dots highlighting I- and C-RGC enriched genes respectively at fold-change > 1.5 and Bonferroni corrected p-value < $5x10^{-5}$. The two vertical bars correspond to a fold-change of 1.5 in either direction. Select I-RGC and C-RGC enriched genes are labeled.
f. Same as panel e, for E14.

Discussion

The staggering diversity of its neurons underlies the computational power of the nervous system. Accordingly, a major quest in developmental neurobiology is to understand the mechanisms that diversify progenitors. A generally accepted way to deal with this diversity is to divide neurons into classes, and then subdivide classes into subclasses and subclasses into types (Zeng & Sanes, 2017). While much has been learned about how neural progenitors give rise to distinct neuronal classes, little is known about how classes diversify into subclasses and types.

Here, we used mouse RGCs to address this issue. We recently generated a molecular atlas that divided RGCs into ~45 distinct types based on their patterns of gene expression (Tran et al., 2019). We used this atlas here as a foundation to ask how these types are specified during development. We conclude that the earliest precursor RGCs are multipotential and exhibit continuous variation in transcriptomic identity, then diversify into definitive types by a gradual process of fate restriction. Interestingly, these features resemble those that have been discovered to control the generation of retinal cell classes from cycling progenitors (RPCs): multipotentiality, progressive restriction of fate, and stochastic rather than deterministic fate choice (see **Introduction**). We suggest that, at least in this case, similar strategies are used to generate cell classes from mitotically active progenitors and cell types from postmitotic precursors.

663 **Classes, subclasses and types**

664 Definitions of neuronal class, subclass and type have been contentious (Yuste et al.,
665 2020). In general, classes share general features of structure, function, molecular
666 architecture and location, whereas types comprise the smallest groups within classes that
667 can be qualitatively distinguished from other groups based on these and other criteria.
668 Subclasses lie in-between. For RGCs, class identity has been clear for a century, but
669 inventories of subclasses and types have emerged only over the last few decades, as
670 high-throughput methods have been implemented for quantifying structural (primarily
671 dendritic morphology), functional (responses to an array of visual stimuli) and molecular
672 properties (gene and transgene expression) of large numbers of RGCs. Fortunately, to
673 the extent that they have been compared, there is excellent concordance among types
674 defined by molecular, structural and physiological criteria (Bae et al., 2018; Goetz et al.,
675 2021; Tran et al., 2019) (see www.rgctypes.org). Moreover, RGCs of a single type exhibit
676 a regular spacing, called a mosaic arrangement, in that they tend to avoid other members
677 of the same cell type whereas their association with members of other types is random
678 (Kay et al., 2012; Keeley et al., 2020; Rockhill et al., 2000). The molecular basis of this
679 property is poorly understood, but it provides an additional criterion for defining a type.
680 Thus, while no two RGCs are identical, and variation may be continuous in some other
681 structures (Cembrowski & Spruston, 2019), there is strong reason to believe that RGC
682 types are discrete.
683
684 **The adult RGC atlas**

685 Developmental trajectories of cell types cannot be better than the adult types at which
686 they are aimed. We have two reasons to believe that our adult RGC atlas (Tran et al.,
687 2019) is accurate and complete.
688
689 First, the atlas is based on detailed analysis of 35,699 cells, and is therefore powered to
690 detect types occurring at ~0.1% frequency (>40 cells per type;
691 https://satijalab.org/howmanycells/). Results were stable over a variety of parameters
692 (Tran et al., 2019). Moreover, in the course of studies on responses of RGCs to injury,

693 we recently profiled an additional ~120,000 cells (A. Jacobi, N. Tran, W Yan and J.R.S,
694 in preparation), without identifying additional types.
695
696 Second, RGCs have now been classified by functional and structural properties, based
697 on physiological responses to visual stimuli (Baden et al., 2016; Goetz et al., 2021) and
698 serial section electron microscopy (Bae et al., 2018). The numbers of types defined in
699 these ways (47 in (Bae et al., 2018), 42 in (Goetz et al., 2021) and > 32 in (Baden et al.,
700 2016)) match well to the 45-46 defined molecularly (Tran et al., 2019).
701
702 **Multipotentiality of precursor RGCs**
703 The multipotentiality of dividing progenitor cells can be demonstrated by indelibly labeling
704 a progenitor and then examining its progeny at a later stage. For mammals, this was
705 initially done by infecting single cells with a recombinant retrovirus encoding a reporter
706 gene that could be detected following multiple cell divisions (Price, Turner, & Cepko,
707 1987; Sanes, Rubenstein, & Nicolas, 1986; Turner & Cepko, 1987). More recently, it has
708 become possible to greatly increase throughput by tracking scars or barcodes introduced
709 by CRISPR/Cas9 (Baron & van Oudenaarden, 2019; Espinosa-Medina, Garcia-Marques,
710 Cepko, & Lee, 2019; McKenna et al., 2016). In sharp contrast, conclusively
711 demonstrating that a single postmitotic cell is multipotential would require following a cell
712 from an unspecified to a specified state, then turning back time, watching it again, and
713 asking if it acquired the same mature identity. Since this is impossible, we used
714 computational inference to draw tentative conclusions about the extent to which newly
715 postmitotic RGCs are committed to mature into a particular type.
716
717 Our analysis proceeded in three steps. First, to ask whether RGCs were committed to a
718 particular fate before or shortly after they were born, we assessed transcriptomic
719 heterogeneity at a time when a large fraction was newly postmitotic (E13 and E14). We
720 found that heterogeneity was present but limited: 10 transcriptomic clusters were
721 distinguishable at E13 and 12 at E14. Thus, some heterogeneity is present in precursor
722 RGCs, but far less than would be required to specify type identity before or immediately
723 after their birth.

Second, we used a supervised classification approach to ask whether precursor RGC clusters mature into mutually exclusive sets of adult types. This model would imply an orderly, step-wise restriction of cell fates. However, our results indicate substantial overlap in the types derived from cells in different immature clusters. This result argues against a deterministic model of diversification, and suggests that precursor RGCs are incompletely committed to a specific type for a substantial period after they are generated.

Third, we used optimal transport inference (WOT) to ask whether the multipotentiality observed at the level of groups was also a property of individual cells. WOT utilizes time course scRNA-seq snapshots to infer fate associations between individual cells sampled at different time points, without reference to the clusters in which they reside(Schiebinger et al., 2019). While being consistent with supervised classification results at the cluster level, WOT indicated that the majority of individual RGCs were multipotential at E13 and E14. Of equal importance, immature RGCs were not totipotential: the average predicted potential (P) was 11.6 at E13, or ~25% of the maximum possible value of 45, and no RGCs had P >30. We conclude that single multipotential immature RGCs are biased in favor of particular groups of adult RGC types.

Progressive restriction of RGC fate

Further analysis provided insight into the structure of multipotentiality among RGCs. The adult RGC types associated with a precursor RGC were not a randomly chosen subset; rather some were more likely to arise from a common precursor state (“fate coupled”) than others. This suggests a model in which RGC types arise via a progressive decoupling of fates within multipotential precursors. Decoupling is asynchronously, with different types emerging at different times. By modeling the temporal kinetics of fate decoupling, we were able to estimate a tentative specification time for each type – that is, the time at which precursors become irrevocably committed to a particular fate. Analysis of transcriptomic changes that occur during this process, and the effects of visual experience on maturation, will be presented elsewhere (K.S., I.E.W., S.B. and J.R.S., in preparation).

Fate-restricted RGC subclasses

For RGCs, class identity has been clear for a century, and type identity has been solidified during over the last few decades, but criteria for defining subclasses remain unclear. Tentative classifications have used molecular, physiological and morphological criteria (Sanes & Masland, 2015; Tran et al., 2019); In general, these criteria correlate imperfectly with each other, the main exception being that ON and OFF RGCs (responding preferentially to increases and decreases in illumination, respectively) have dendrites that arborize in the inner and outer portion of the inner plexiform layer (Famiglietti & Kolb, 1976).

The pattern of fate couplings between RGC types at E13-14 provides an alternative way to define RGC subclasses – groups of RGC types that arise from the restriction of a common transcriptionally defined precursor state. We identified transcription factors selectively expressed within these subclasses. Our rationale was that among them would be fate determinants, an idea that could be tested by conventional genetic manipulations. Support for this idea is that there is already strong evidence that one such factor is a fate determinant in mouse: *Eomes* is selectively expressed by ipRGCs (and a few other types), and *Eomes* mutants fail to form ipRGCs although their retinas are normal in many respects (C. A. Mao et al., 2014). This encourages the hope that some of the other transcription factors in this set are also fate determinants. It will also be interesting to determine whether members of fate-restricted subclasses share structural or functional properties.

Laterality

The transcription factors Isl2 and Zic2 selective markers of embryonic RGCs that project contralaterally or ipsilaterally, respectively, and are critical determinants of this choice (Herrera et al., 2003; Pak et al., 2004). We found that their expression was largely nonoverlapping in RGCs at E13, that they were co-expressed with previously reported markers of contralaterally and ipsilaterally projecting RGCs, respectively. Because few RGC axons reach the optic chiasm before E14, our results are consistent with genetic

786 evidence that this differential expression is a cause rather than a consequence of the
787 divergent choices the axons make at the chiasm. Among many genes co-expressed with
788 *Isl2* or *Zic2* may be others that play roles in this choice.
789
790 Zic2 is downregulated later in embryogenesis, so we used WOT to infer the fates of
791 putative I-RGCs. We found that they give rise to many types, consistent with previous
792 results (Hong et al., 2011; Johnson et al., 2021). Surprisingly, however, there were few if
793 any genes differentially expressed between the putative mature I- and C-RGCs.
794 Assuming that WOT results are valid – an assertion that can be tested directly in the
795 future – this result suggests a model in which newborn RGCs are doubly specified – by
796 laterality and by type – but that once axonal choice has been made the laterality program
797 is shut down.
798
799 **Beyond the retina**
800 Generation of neuronal classes has been analyzed in many parts of the vertebrate
801 nervous system but we are aware of few reports on how classes diversify into types. A
802 recent study addressed this issue for primary sensory neurons and reached the
803 conclusion that newborn neurons in dorsal root ganglia are transcriptionally unspecialized
804 and become type-restricted as development proceeds (Sharma et al., 2020). Similarly,
805 both excitatory neuronal subclasses appear to diversify postmitotically in the mouse
806 cerebral cortex (Di Bella et al., 2021; Lodato & Arlotta, 2015), and there is suggestive
807 evidence that the same is true for interneuronal subclasses (Wamsley & Fishell, 2017).
808 In all of these cases, it is attractive to speculate that diversification may occur by a process
809 of fate decoupling in subpopulations of distinct multipotential precursors, akin to that
810 documented here for RGCs. Our study provides a computational framework for
811 investigating this issue.
812
813 **Acknowledgments**
814 Funding for this study was provided by National Institute of Health grants R37NS029169,
815 R01EY022073 (JRS), NSF GRP DGE1752814 (SB), and R00EY028625 (KS), and

startup funding from UC Berkeley (KS, SB). The authors would like to thank Joshua Hahn for valuable feedback.

Author Contributions

K.S., I.E.W and J.R.S. designed the study. I.E.W designed the experimental methods and contributed to the initial analysis of the data. K.S. designed the computational methods. I.E.W and Y.P performed experiments. K.S. and S.B performed computational analyses. K.S. and J.R.S analyzed the data and wrote the paper with contributions from other authors.

Competing interests

The authors declare no competing interests.

Correspondence and requests for materials should be addressed to K.S. or J.R.S.

Data availability

All scRNA-seq data collected in this study are submitted to the Gene Expression Omnibus (GEO) under GSE185671.

Code availability

The scripts (written in R) generated for this study are shared at https://github.com/shekharlab/mouseRGCdev.

Materials and Methods

Mice

All animal experiments were approved by the Institutional Animal Care and Use Committees (IACUC) at Harvard University. Mice were maintained in pathogen-free facilities under standard housing conditions with continuous access to food and water. Animals used in this study include both males and females. A meta-analysis (not shown) did not show any systematic sex-related effects in either differentially expressed genes or cell-type proportions. For scRNA-seq and histology, we used C57Bl/6J (JAX #000664). Embryonic and early post-natal C57Bl/6J mice were acquired either from Jackson Laboratories (JAX) from time-mated female mice or time-mated in-house. For timed-matings, a male was placed with a female overnight and removed the following morning (with the corresponding time recorded as day E0.5).

Cell preparation

RGCs were enriched from dissociated retinal cells as previously described with minor modifications (Tran et al., 2019). All solutions were prepared using Ames' Medium with L-glutamine and sodium bicarbonate (equilibriated with 95% O_2/5% CO_2), and all spin steps were done at 450g for 8 minutes. Retinas were dissected out in their entirety immediately after enucleation and digested in ∼80U of papain at 37°C, with the exception of some E13 and E14 eyes which were digested whole, followed by manual trituration in ovomucoid solution. Clumps were removed using a 40μm cell strainer and the cell suspension was spun down and re-suspended in Ames + 4% BSA at a concentration of 10 million cells per 100μl. Cells from E13, E14, E16, and P0 were incubated for 15 minutes at room temperature with antibodies to Thy1 (also known as CD90) and L1CAM pre-conjugated to the fluorophores APC (ThermoFisher Scientific#17-0902-82) and PE (Miltenyi Biotec 130-102-243), respectively. Cells were washed with 6ml of Ames + 4% BSA, spun down and resuspended at a concentration of ∼7 million cells/ml, and calcein blue was added to label metabolically active cells.

Viable Thy1 or L1CAM positive cells were sorted using a MoFlo Astrios sorter into ∼100μl of AMES + 4% BSA. Sorted cells were spun down a final time and resuspended in PBS + 0.1% BSA at a concentration of 500-2000 cells/μl. P5 RGCs were enriched using only CD90, with either magnetic-activated cell sorting (MACS) using large cell columns and CD90 pre-conjugated to microbeads (#130-042-202 and #130-049-101, Miltenyi Biotec), fluorescence activated cell sorting (FACS) with CD90 pre-conjugated to PE/Cy7 (ThermoFisher Scientific #25-0902-81), or both.

Droplet based single-cell (sc) RNA-seq

Statement on replicates: We profiled immature RGCs using scRNA-seq at five developmental time points: E13, E14, E16, P0 and P5. At each age, data was collected from four replicate experiments. Experiments at E13, E14, E16 and P0 involved two biological replicates (distinct mice). Each of these biological replicates were further subdivided into two equal pools and the cells were subjected to two different enrichment methods (anti-Thy1 and anti-L1cam). Thus, each of these time points consisted of four replicate experiments. RGC enrichment at P5 exclusively utilized anti-Thy1, but four biological replicate experiments were performed. One of these was profiled using 10X and three of these were profiled using Drop-seq.

Drop-seq: A subset of P5 RGC dataset was collected using Drop-seq (Macosko et al., 2015), performed largely as described previously (Shekhar et al., 2016). Briefly, cells were diluted to an estimated final droplet occupancy of 0.05, and co-encapsulated in droplets with barcoded beads, which were diluted to an estimated final droplet occupancy of 0.06. The beads were purchased from ChemGenes Corporation, Wilmington MA (# Macosko201110). Individual droplet aliquots of 2 ml of aqueous volume (1 ml each of cells and beads) were broken by perfluorooctanol, following which beads were harvested, and hybridized RNA was reverse transcribed. Populations of 2,000 beads (∼100 cells) were separately amplified for 14 cycles of PCR (primers, chemistry, and cycle conditions identical to those previously described) and pairs of PCR products were co-purified by the addition of 0.6x AMPure XP beads (Agencourt).Fifteen experimental replicates were sequenced in total from 5 biological replicates using an Illumina NextSeq 500. Read 1 was 20bp; read 2 (paired-end) was 60bp.

10X Genomics: Single cell libraries were prepared using the Single-cell gene expression 3' kit on the Chromium platform (10X Genomics, Pleasanton, CA) following the manufacturer's protocol. As our datasets were collected over a long period of time, we used a combination of v1 (a single channel of P5 RGCs), v2 (E13, E14, E16, P0). Briefly, single cells were partitioned into Gel beads in EMulsion (GEMs) in the 10X Chromium instrument followed by cell lysis and barcoded reverse transcription of RNA, amplification, enzymatic fragmentation, 5' adaptor attachment and sample indexing. On average, approximately 8,000-12,000 single cells were loaded on each channel and approximately 3,000-7,000 cells were recovered. Libraries were sequenced on the Illumina HiSeq 2500 platforms at the Broad institute (Paired end reads: Read 1, 26 bases, Read 2, 98 bases).

Statement on power analysis: An important question in all single-cell experiments is that of the number of cells to profile. A widely used approach is the power analysis tool published by the Satija lab (https://satijalab.org/howmanycells/). Fortunately, in this study we were also guided by our previous study of adult RGCs, where we had knowledge of the frequency distribution of adult RGC types, with the rarest type being approximately 0.2% (Tran et al., 2019). In that study, we also found that when classification is performed in a supervised fashion based on an existing atlas, approximately ∼8000 RGCs were sufficient to recover the accurate relative frequency distribution of 45 RGC types. We therefore aimed to profile ∼ 8000 cell at each time point as our analysis involved mapping immature RGCs to the adult atlas. With the exception of E13, all time points contain 1.5-2X more cells than this target value.

Histology

Tissue Fixation: Adult (P56) mice were intracardially perfused with 2-5ml of PBS followed by 15ml of 4% PFA, followed by additional fixation of eyes for 15 minutes in 4% PFA, with the exception of XXXX. P0 and P5 mice were not perfused, rather eyes were fixed in 4% PFA for 30 minutes. At E13, 14, and 16 embryos were fixed whole for 30 minutes in 4% PFA, following which eyes were removed. Following fixation eyes from all time points were transferred to PBS and stored at 4°C until subsequent use.

Sectioning: Cross sections for immunohistochemistry (IHC) were generated using a Leica CM1850 cryostat. For some early developmental time points eyes were kept whole for IHC, otherwise retinas were either (1) dissected out in their entirety from eyes, or (2) the cornea, iris and lens was removed leaving the sclera and retina intact. Tissues were sunk in 30% sucrose overnight at 4°C, embedded in tissue freezing medium, and cryo-sectioned into 25mm slices. Slides with tissue sections were air

dried for ∼3 hours and stored at -80°C until staining.

Immunohistochemistry (IHC): All IHC solutions were made up in PBS + 0.3% Triton-X, and all incubation steps were carried out in a humidified chamber. Following a 1 hour protein block in 5% Normal Donkey Serum at room temperature, slides were incubated overnight at 4°C with primary antibodies, washed twice for 5 minutes each in PBS, incubated for 2 hours at room temperature with secondary antibodies conjugated to various fluorophores (1:1000, Jackson Immunological Research) and Hoechst (1:10000, Life Technologies), and washed again twice for 5 minutes each in PBS before coverslipping with Fluoro-Gel (#17985, Electron Microscopy Sciences). Primary antibodies used include: guinea pig anti-RBPMS (1:1000, #1832-RBPMS, PhosphoSolutions), rabbit anti-KI67 (1:250, #MA5-14520, ThermoFisher Scientific), rat anti-L1CAM (1:10, #130-102-243, Miltenyi Biotec), rabbit anti-TBX20 (1:500, #A04704, Boster Bio), chicken anti-EOMES (1:500, #AB15894, Millipore Sigma), rabbit anti-NEUROD2 (1:500, #ab104430, Abcam), guinea pig anti-PRDM8 (1:2000, kind gift from Sarah E. Ross; Ross et al., 2012), mouse anti-POU4F3 (1:500, #MAB1585, Millipore Sigma), goat anti-VSX2 (1:200, #sc-21690, Santa Cruz Biotechnology), rabbit anti-SOX9 (1:1000, #AB5535, Millipore Sigma), and chicken anti-GFP (1:1000, #ab13970, Abcam).

Imaging

All images were acquired using an Olympus Fluoview 1000 scanning laser confocal microscope, with a 20Xx oil immersion objective and 2x optical zoom. Optical slices were taken at 1μm steps. Fiji (Schindelin et al., 2012) was used to pseudocolor each channel and generate a maximum projection from image stacks. Brightness and contrast was adjusted in Adobe Photoshop.

Alignment and quantification of gene expression in single cells

All single-cell libraries were aligned to the UCSC mm10 transcriptomic reference (*M. musculus*) and gene expression matrices were quantified using standard protocols described previously. For the single-cell libraries generated using the 10X platform (E13, E14, E16, P0 and P5), these steps were performed using cellranger v2.1.0 (10X Genomics). For the single-cell libraries generated using Drop-seq (P5), we used Drop-seq tools (v1.12; Macosko et al., 2015), following procedures described earlier (Shekhar et al., 2016). Alignment and quantification was done for each sample library separately to generate a genes × cells expression matrix of transcript counts. These matrices were column-concatenated for further analysis.

We retained cells that expressed at least 700 genes, resulting in 98,452 cells. We also removed genes expressed in fewer than 10 cells. The resulting M genes × N cells matrix of UMI counts C_{mn} was normalized along each column (cell) to sum to $8,340$, the median of the column sums resulting in a normalized matrix X_{mn}. This was followed by the transformation $\tilde{X}_{mn} \leftarrow \log(X_{mn} + 1)$.

Overview of batch correction, dimensionality reduction and clustering

The following procedure was adopted to perform batch correction, dimensionality reduction and clustering throughout the paper. The procedure was first applied on the entire dataset to separate RGCs from other cell classes, and then to RGCs at each age to identify transcriptomically distinct groups.

1. *Identification of highly variable genes (HVGs)*: We used the Gamma-Poisson framework described previously to identify HVGs (Pandey et al., 2018). Briefly, we compute for each gene the mean (μ_m) and the coefficient of variation (CV_m) for the UMI counts C_{mn},

$$\begin{aligned} \mu_m &= \frac{1}{N}\sum_n C_{mn} \\ \sigma_m^2 &= \frac{1}{N}\sum_n (C_{mn} - \mu_m)^2 \\ CV_m &= \frac{\sigma_m}{\mu_m} \end{aligned} \tag{1}$$

For a given μ_m, the Gamma-Poisson model predicts a "null" coefficient of variation (CV_m^{null}) arising from a combination of Poisson "shot" noise and large variations in library size, assumed to be due to technical reasons,

$$CV_m^{null} = \frac{1}{\mu_m} + \frac{1}{\alpha} \tag{2}$$

Here, α is the shape parameter of a Gamma-distribution fit to the distribution of normalized library sizes T_n (using the R package `MASS`),

$$T_n = \frac{\sum_m C_{mn}}{\sum_{m,n} C_{mn}} \tag{3}$$

In practice, CV_m^{null} serves as a tight lower-bound for empirically observed values of CV_m across the full range of μ_m. This enables us to compute for each gene m, a deviation score $d_m = \log \frac{CV_m}{CV_m^{null}}$, quantifying the extent to which its observed coefficient of variation exceeds the predicted null model. Highly-variable genes (HVGs) are selected if they satisfy $d_m > \text{Mean}(d_m) + 0.8 * \text{std}(d_m)$.

2. *Batch correction and Dimensionality Reduction*: We subsetted the rows of the expression matrix X_{mn} to the HVGs identified in Step 1. As our data comprised cells sampled at different developmental ages as well as multiple biological replicates within each age, we used Liger, a non-negative matrix factorization technique, to embed the data in a reduced dimensional latent space of shared factors (Welch et al., 2019). Liger computes a factorized representation for each matrix that separate "shared" and "dataset-specific" gene expression modules (factors). We use Liger's normalized H factor loadings for cells to build a nearest neighbor graph and define clusters.

   As in any matrix factorization technique, Liger requires the user to choose k, the dimensionality of the latent space. To find k, we use a Random Matrix Theory approach (outlined in Peng et al., 2019). Briefly, k is estimated as as the number of eigenvalues of the sample gene-gene correlation matrix that exceed the 99^{th} percentile of the distribution of the largest eigenvalue of a random Hermitian matrix of the same dimensions. This is given by the Tracy Widom distribution (Tracy and Widom, 1993). For these calculations, we used the R package `RMTstat`.

3. *Clustering and 2D Visualization*: To cluster cells based on transcriptomic similarity, we first built a nearest-neighbor graph on the cells based on their normalized H factor coordinates computed using Liger. The number of nearest neighbors was chosen to be 30. The edges were weighted based on the Jaccard overlap metric, as described previously (Shekhar et al., 2016). The graph was clustered using the Louvain method (Blondel et al., 2008). The normalized

H factor coordinates were also used as input to project cells on to a nonlinear 2D space using the Uniform Manifold Approximation and Projection algorithm (UMAP; Becht et al., 2019). Graph-construction, clustering and the UMAP projection were performed using the R packages `FNN`, `igraph` and `umap`, respectively.

Separation of major cell classes

We began by clustering the full dataset combining all ages using the procedure outlined above. For applying Liger, each age was regarded as a separate "batch" but replicates within an age were not distinguished at this stage.

Approximately 40 clusters were distinguished at this stage. For each cluster, we computed a signature expression vector by averaging the expression levels of HVGs across the cells. These expression vectors were used to perform hierarchical clustering, which enabled us to visualize the transcriptional interrelationships between these clusters as a dendrogram. Hierarchical clustering was implemented in the R package `hclust` (correlation distance, average linkage method). Major clades in the resulting dendrogram corresponded to well known retinal classes and this was confirmed by their expression of known class-specific markers. Each cluster was therefore assigned to its class, which included retinal ganglion cells (RGCs; *Rbpms, Slc17a6, Sncg, Nefl*), microglia (*P2ry12, C1qa-c, Tmem119*), photoreceptors (*Otx2, Gngt2, Gnb3*), amacrine cells (*Tfap2a, Tfap2b, Onecut2*), anterior segment cells (*Mgp, Col3a1,Igfbp7*), cycling progenitors (*Ccnd1, Fgf15, Hes5*) and neurogenic progenitors (*Hes6, Ascl1, Neurog2*). Deeper annotation (e.g. of RGC type) was not done at this stage.

No other cellular classes were identified. 3 clusters comprising fewer than 1.2% of the cells expressed markers of more than one class. These were flagged as doublets and removed from further analyses.

Clustering RGCs at each time point

RGCs identified in step "Separation of Major Cell Classes" were segregated by age, and each group was separately analyzed following the clustering pipeline outlined previously. When implementing Liger, each biological replicate was regarded as a separate batch. The nominal clusters identified by the Louvain algorithm were refined as follows:

1. *Removing contaminants*: Clusters were flagged for further examination if they did not exclusively express RGC-specific markers (e.g. *Rbpms, Slc17a6, Sncg, Nefl*). These clusters were small (typically $<$ 1-2% of cells) and in all cases expressed non-RGC markers (e.g. *P2ry12* or *Tfap2b*). These cells, which likely reflect trace contaminants, were discarded from further analysis.

2. *Merging proximal clusters*: Transcriptomic relationships between nominated clusters were visualized on a dendrogram computed using hierarchical clustering, as noted above above. Neighboring clusters on the dendrogram, which were leaves in a terminal branch, were assessed for differential expression using the MAST differential expression (DE) test (Finak et al., 2016). A gene g was regarded as significantly DE between clusters $\mathcal{C}_a$ and $\mathcal{C}_b$ if it satisfied $|logFC_g(\mathcal{C}_a, \mathcal{C}_b)| > 0.5$ and MAST p value was less than 10^{-5} (FDR corrected), where:

$$logFC_g(\mathcal{C}_a, \mathcal{C}_b) = \ln\left(\frac{|\mathcal{C}_b|}{|\mathcal{C}_a|}\frac{\sum_{n\in\mathcal{C}_a} X_{gn}}{\sum_{n\in\mathcal{C}_b} X_{gn}}\right)$$

is defined to be the log-fold change in expression. Clusters that showed fewer than 10 significant DE genes were merged.

In this manner, we identified 10 RGC clusters at E13, 12 at E14, 19 at E16, 27 at P0, and 38 at P5. Using MAST, we identified DE genes that distinguished each cluster against the rest at any given age.

Joint analysis with whole-retina scRNA-seq datasets

We compared our data with two scRNA-seq studies that profiled the whole retina during development:

1. Clark et al., Neuron, 2019: Count matrices and cell/gene level annotations were downloaded from the author's public repository https://github.com/gofflab/developing_mouse_retina_scRNASeq. This dataset contains whole retinal cells sampled at 10 time points (E11, E12, E14, E16, E18, P0, P2, P5, P8, P14) with four of these (E14, E16, P0, P5) common with our study. We excluded P5 from our analysis as only $N = 11$ RGCs were identified by the authors at this time point.

2. Giudice et al., Development, 2019: Count matrix corresponding to E15.5 retinal cells was kindly provided by the authors.

For consistency with our filtering parameters, we extracted cells based on a cutoff of 700 genes/cell from each of the above datasets. For the Clark et al. dataset, this selected 17,827 cells at E14, 1,674 cells at E16 and 8,343 cells at P0 respectively (N=27,844 cells). In these samples, RGCs comprised 19%, 28% and 0.45%. For the Giudice et al. dataset, this selected 5218 cells, of which 23% were RGCs.

These were combined with the retinal cells profiled in this study at corresponding time points (25,685 cells at E14; 21,274 cells at E16; and 23,251 cells at P0). Together this resulted in a 14,350 genes $\times$ 103,272 cells expression matrix that was analyzed following steps outlined previously. In the alignment step, cells from each combination of age and study was considered as a separate "batch".

We visualized the transcriptional heterogeneity of the full dataset using UMAP, and used the expression of canonical markers to confirm the co-clustering of cell-classes in **Figure S2** (*Rbpms* for RGCs, *Tfap2b* for ACs, *Fgf15* for RPCs and *Gngt2* for RPCs).

Diversity Indices

We quantified the molecular diversity of RGCs based on clusters at each stage using 3 measures of population diversity - The Rao index (**Figure 2**), the Shannon index, and the Simpson index (**Figure S2**). For N clusters with relative frequencies $p_1, p_2, \ldots, p_N$, these indices are defined as follows,

- Let d_{ij} be a distance measure between clusters i and j ($0 \leq d_{ij} \leq 1$). The *Rao index* is defined as,

$$R = \sum_{i \neq j} d_{ij} p_i p_j$$

  We used varying number of genes ($\approx$ 1200-3000) to calculate d_{ij}. The computed Rao index was insensitive to these variations.

- The *Shannon index* is defined as,

$$H = -\sum_i p_i \log p_i$$

- The *Simpson index* is defined as

$$S = \sum_i p_i^2$$

While the Rao and Shannon indices increase with diversity, the Simpson index decreases with diversity.

Analysis of cluster distinctiveness

We quantified the mutual separation of clusters at each age using two approaches:

Multi-class classification: We trained a multi-class classifier (R package `xgboost`) at each age on 50% of the cells using their cluster IDs. The remaining 50% of the cells were used to test the learned classifier and estimate a classification error per cluster, which were averaged at each age. As clusters become better separated, the average classification error decreases.

Relative positions in PCA: At each age, the top 20 PCA coordinates were first standardized by z-scoring. For each cluster C at a given age, we computed two quantities:

- r_C, the median of euclidean distances of each cell from the cluster centroid in the standardized PCA coordinates.
- d_C, the median of euclidean distances of each cell from the centroid of the nearest external cluster.

For a cluster C, a low of value r_C/d_C indicates a higher degree of separatedness. Averaging this metric across all the clusters at a given age quantifies the degree to which clusters are separated in the UMAP representation.

Relating clusters across ages using XGBoost

Analysis Overview: To distinguish between "specified" and "non-specified" modes of diversification (**Figure 3**), we first used a supervised classification approach to associate immature RGC clusters at young ages (tests) to cluster IDs determined at a older ages (references). We used XGBoost, a decision-tree based ensemble learning algorithm (Chen and Guestrin, 2016), to train multi-class classifiers on reference clusters, and used these to assign labels to individual test RGCs.

Two kinds of references were used: (1) Classifiers trained on the adult (P56) clusters were used to assign immature RGCs at each of the five developmental ages (five separate analyses) to adult labels. (2) Classifiers trained on E14, E16, P0, and P5 clusters were used to assign E13, E14, E16, and P0 RGCs to labels corresponding to the previous age, respectively (four separate analyses). The correspondence between classifier assigned labels and cluster IDs of test RGCs were visualized using confusion matrices (e.g. **Figure 3D-H**) and quantified using two metrics - the Adjusted Rand Index (ARI) and Normalized Conditional Entropy (NCE) metrics, described below.
Classification Overview: To describe our classification analysis, we introduce some notation to facilitate a description in general terms. Let $\mathcal{A}^R$ and $\mathcal{A}^T$ denote the reference and the test atlases for

the purpose of supervised classification. The number of cells (i.e. RGCs) contained in the reference and test atlases are denoted $|\mathcal{A}^R|$ and $|\mathcal{A}^T|$, respectively. $\mathcal{A}^R$ and $\mathcal{A}^T$ could be any pair of ages as described above. Without loss of generality, let us assume that $\mathcal{A}^R$ contains r transcriptomic clusters denoted denoted $\{\mathcal{C}_1^R, \mathcal{C}_2^R, \ldots, \mathcal{C}_r^R\}$. Similarly, the $\mathcal{A}^T$ is assumed to contain t transcriptomically defined clusters denoted $\{\mathcal{C}_1^T, \mathcal{C}_2^T, \ldots, \mathcal{C}_t^T\}$.

Each cell in our dataset is the member of a particular atlas, and is assigned to a single cluster within the atlas based on its transcriptome. The transcriptome of each cell is a vector (denoted using lowercase boldface symbols, e.g. $\mathbf{u}$ or $\mathbf{v}$) with number of elements equal to the number of HVGs (the features used for classification). Let cluster($\mathbf{u}$) denote the transcriptionally assigned of cell $\mathbf{u}$. For example, the following statement,

$$\mathbf{u} \in \mathcal{A}^T, \text{cluster}(\mathbf{u}) = \mathcal{C}_k^T$$

translates to "Cell $\mathbf{u}$ in atlas $\mathcal{A}^T$ is a member of cluster $\mathcal{C}_k^T$." Our goal is to *assign* each cell $\mathbf{u} \in \mathcal{A}^T$, a second ID cluster$'(\mathbf{u})$ based on its transcriptomic correspondence to the reference atlas $\mathcal{A}^R$. We accomplish this via an XGBoost classifier trained on $\mathcal{A}^R$ and applied it to every cell in $\mathcal{A}^T$, allowing us to infer transcriptomic correspondences between the two sets of clusters. The main steps are as follows,

- The expression matrices in $\mathcal{A}^R$ and $\mathcal{A}^T$ are z-scored along each feature. The initial set of features are chosen as the common HVGs in the two atlases. Parameters are adjusted to select the common top ∼2000-3000 HVGs.

- Classifiers Class_0^R and Class_0^T are trained on $\mathcal{A}^R$ and $\mathcal{A}^T$ independently. For training, we randomly sample 60% of cells in each cluster up to a maximum of 300 cells. The remaining "held-out" cells are used for validation. We ran the training routine for `xgboost` with the following parameter specification *,

```
xgb_params <- list("objective" = "multi:softprob",
                   "eval_metric" = "mlogloss",
                   "num_class" = nClusters,
                   "eta" = 0.2,"max_depth"=6, subsample = 0.6)
```

- When applied to a cell vector $\mathbf{u}$, the classifier Class_0^R (or Class_0^T) returns a vector of $\mathbf{p} = (p_1, p_2, \ldots)$ of length r (or t) with entries representing probability values of predicted cluster memberships in the corresponding atlas. We use these values to compute the "softmax" assignment of $\mathbf{u}$, so that cluster$'(\mathbf{u}) = \arg\max_i p_i$.

- Post training, Class_0^R and Class_0^T are evaluated on the respective validation sets. Using the predicted cluster assignments of the "held out" cells, we compute for each cluster in $\mathcal{A}^R$ and $\mathcal{A}^T$ the error-rate, defined as the fraction of held-out cells that were misclassified. If the error-rate for any cluster was higher than 10%, the classifier is retrained by artificially upsampling cells from the high error-rate clusters. In the final classifiers, the cluster-specific error rates were typically 1-4%, and in no case exceeded 10%.

*https://xgboost.readthedocs.io/en/latest/parameter.html

- The top 500 discriminatory features (genes) are identified based on average information gain (using the function `xgb.importance`) for each of Class_0^R and Class_0^T. These gene sets are denoted as as G^R and G^T, respectively.

- The common features $\mathsf{G} = \mathsf{G}^R \cap \mathsf{G}^T$ are used to train a third classifier Class^R on the reference atlas $\mathcal{A}^R$. This ensures that inferred transcriptiomic correspondences are based on "core" gene expression programs that underlie cell type identity rather than maturation-associated genes.

- Finally, Class^R is applied to each cell $\mathbf{u} \in \mathcal{A}^T$ to generate predicted labels cluster$'(\mathbf{u})$. Global transcriptional correspondence were visualized using confusion matrices between cluster IDs cluster$(\mathbf{u}) \in \{\mathcal{C}_1^T, \mathcal{C}_2^T, \ldots, \mathcal{C}_t^T\}$ and reference assignments cluster$'(\mathbf{u}) \in \{\mathcal{C}_1^R, \mathcal{C}_2^R, \ldots, \mathcal{C}_r^R\}$, and their correspondence was quantified using metrics described below.

Quantifying cluster correspondence using global and local metrics

Let N_{ij} denote the number of cells in $\mathcal{A}^T$ that are part of transcriptomic cluster $\mathcal{C}_j^T$, and are assigned by $\mathcal{C}_\mathcal{R}$ to reference cluster $\mathcal{C}_i^R$. Thus,

$$N_{ij} = \#\{\text{cluster}'(\mathbf{u}) = \mathcal{C}_i^R,\ \text{cluster}(\mathbf{u}) = \mathcal{C}_j^T\ \forall\ \mathbf{u} \in \mathcal{A}^T\} \tag{4}$$

N_{ij} defines a contingency table, whose marginal sums are defined as,

$$\begin{aligned} a_i &= \sum_{j=1}^{t} N_{ij} \\ b_i &= \sum_{i=1}^{r} N_{ij} \end{aligned} \tag{5}$$

Let $N = \sum_{ij} N_{ij} = |\mathcal{A}^T|$, the number of cells in $\mathcal{A}^T$. Then the **Adjusted Rand index (ARI)** corresponding to the assignments can be evaluated using the following equation (Hubert and Arabie, 1985),

$$ARI = \frac{\sum_{ij} \binom{N_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{N}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{N}{2}} \tag{6}$$

The ARI ranges from 0 and 1, with extremes corresponding to random association and 1:1 correspondences between $\mathcal{A}^T$ and $\mathcal{A}^R$, respectively [†].

As an alternative, we also used the **Normalized Conditional Entropy (NCE)**, an information-theoretic measure (Cover and Thomas, 1991). The NCE quantifies the extent to which knowledge of the value of cluster$'(\mathbf{u})$ reduces the uncertainty (measured in information bits) about the value of cluster$(\mathbf{u})$ for $\mathbf{u} \in \mathcal{A}^T$.

We introduce probability weights q_{ij} and the corresponding marginals $q_{i,\cdot}$ and $q_{\cdot,j}$ as follows,

[†]The ARI can technically also take on negative values for certain scenarios, but these are not observed in our data

$$\begin{aligned} q_{ij} &= \frac{N_{ij}}{N} \\ q_{i,.} &= \frac{a_i}{N} \\ q_{.,j} &= \frac{b_i}{N} \end{aligned} \tag{7}$$

The Conditional Entropy (CE) is then given by the expression (Cover and Thomas, 1991),

$$H(\text{cluster}(\mathbf{u}) \mid \text{cluster}'(\mathbf{u})) = -\sum_{ij} q_{ij} \log \frac{q_{ij}}{q_{i,.}}$$

Note that CE is asymmetric, i.e. $H(\text{cluster}(\mathbf{u}) \mid \text{cluster}'(\mathbf{u})) \neq H(\text{cluster}'(\mathbf{u}) \mid \text{cluster}(\mathbf{u}))$. One notes that $H = 0$ if for each cluster $i \in 1, \ldots, r$, $q_{ij} = \delta_{i,k_i}$, where δ_{lm} is the Kronecker delta defined as,

$$\begin{aligned} \delta_{ij} &= 1, \qquad \text{if } \; i = j \\ &= 0, \qquad \text{if } \; i \neq j \end{aligned} \tag{8}$$

Finally, NCE is defined as,

$$NCE = \frac{H(\text{cluster}(\mathbf{u}) \mid \text{cluster}'(\mathbf{u}))}{H(\text{cluster}(\mathbf{u}))} \tag{9}$$

where $H(\text{cluster}(\mathbf{u})) = -\sum_j q_{.,j} \log q_{.,j}$ is the Shannon entropy (Cover and Thomas, 1991). Due to the normalization in equation (9), NCE values range from 0 to 1, with extremes corresponding to fully specific mapping or random association, respectively between $\mathcal{A}^T$ and $\mathcal{A}^R$. ARI (equation (6)) and NCE (equation (9)) are inversely related. Unlike ARI, however, NCE is able to detect specificity in both many:1 and 1:1 mappings. ARI returns a value lower than 1 for specific mappings if the number of clusters in $\mathcal{A}^T$ and $\mathcal{A}^R$ are not equal

ARI and NCE quantify global correspondences between $\mathcal{A}^T$ and $\mathcal{A}^R$. We also computed a local metric, the **Occupancy Fraction (OF)** that quantified whether individual reference labels $\mathcal{C}_i^R$ were distributed in a "localized" or "diffuse" manner between test clusters $\{\mathcal{C}_1^T, \mathcal{C}_2^T, \ldots, \mathcal{C}_t^T\}$

$$OF(\mathcal{C}_i^R) = \frac{1}{t}\left[\frac{1}{\sum_j \left(\frac{q_{ij}}{q_{i,.}}\right)^2}\right] \tag{10}$$

Note that the term $\frac{q_{ij}}{q_{i,.}}$ is simply the fraction of the total test cells belonging to test cluster $\mathcal{C}_j^T$ that are assigned to reference cluster $\mathcal{C}_i^R$ by the classifier. Defined this way, the term in the square brackets computes an occupation number that ranges from 1 to t and can be interpreted as the number of test clusters that are specifically associated with $\mathcal{C}_i^R$. Division by t, the number of test clusters, therefore converts this number into a fraction.

Waddington Optimal Transport (Waddington-OT)

To identify fate relationships among maturing RGCs used Waddington-OT (Schiebinger et al., 2019), a recently developed framework that is rooted in Optimal Transport theory (Villani et al., 2008). Waddington-OT does not rely on clustering, and therefore is able to identify ancestor-descendant

relationships between any pair of temporally separated RGCs in our data.

At its heart, Waddington-OT models cellular transcriptomes $\{\mathbf{u}\}$ measured at a given age t as a probability distribution in gene expression space $\mathbb{P}_t(\mathbf{u})$ [‡]. This probability distribution evolves with time, as cells differentiate and mature. Different temporal measurements collected at times $\ldots, t_{i-1}, t_i, t_{i+1}, \ldots$ represent temporal snapshots of the corresponding cell distributions $\ldots, \mathbb{P}_{t_{i-1}}, \mathbb{P}_{t_i}, \mathbb{P}_{t_{i+1}}, \ldots$. Unfortunately, as each cell can only be measured once, the measurement at different times are from different cells. Therefore, for a particular cell $\mathbf{u}$ at time t_i, it is not clear which cell(s) at time t_{i-1} is likely to be its ancestor(s) and which cell(s) at time t_{i+1} are likely to be descendant(s). It is this problem at Waddington-OT addresses.

Overview of the inference procedure: Briefly, for a given pair of consecutive transcriptomic snapshots $\mathbb{P}_{t_i}(\mathbf{u})$ and $\mathbb{P}_{t_i+1}(\mathbf{v})$, we wish to estimate the joint distribution $\Pi_{t_i,t_{i+1}}(\mathbf{u},\mathbf{v})$, representing the probability that a cell having an expression vector $\mathbf{u}$ at time t_i transitions to a cell with an expression vector $\mathbf{v}$ at time t_{i+1}. $\Pi_{t_i,t_{i+1}}(\mathbf{u},\mathbf{v})$ is also called the temporal coupling, which, owing to the destructive nature of scRNA-seq assays, is not directly observable. Under the assumption that cells move short distances in transcriptomic space when $\Delta t_i = t_{i+1} - t_i$ is "reasonably close", Waddington-OT estimates $\Pi_{t_i,t_{i+1}}(\mathbf{u},\mathbf{v})$ as the solution to the following convex optimization problem,

$$\hat{\Pi}_{t_i,t_{i+i}} = \underset{\Pi}{\operatorname{argmin}} \sum_{\mathbf{u}\in\mathcal{A}^{t_i}} \sum_{\mathbf{v}\in\mathcal{A}^{t_{i+1}}} c(\mathbf{u},\mathbf{v})\Pi(\mathbf{u},\mathbf{v}) - \epsilon \int\int \Pi(\mathbf{u},\mathbf{v}) \log \Pi(\mathbf{u},\mathbf{v}) d\mathbf{u} d\mathbf{v}$$
$$+ \lambda_1 \mathrm{KL}\left[\sum_{\mathbf{u}\in\mathcal{A}^{t_i}} \Pi(\mathbf{u},\mathbf{v}) \| d\hat{\mathbb{P}}_{t_{i+1}}(\mathbf{v})\right] + \lambda_2 \mathrm{KL}\left[\sum_{\mathbf{v}\in\mathcal{A}^{t_{i+1}}} \Pi(\mathbf{u},\mathbf{v}) \| d\hat{\mathbb{Q}}_{t_i}(\mathbf{u})\right]$$

In the above equation,

- $\hat{\mathbb{P}}_{t_i}(\mathbf{v})$ is an empirical distribution constructed from $\mathcal{A}^{t_i}$, which denotes the scRNA-seq atlas at t_i,

  $$\hat{\mathbb{P}}_{t_i}(\mathbf{v}) = \frac{1}{|\mathcal{A}^{t_i}|} \sum_{\mathbf{x}_i\in\mathcal{A}^{t_i}} \delta(\mathbf{v} - \mathbf{x}_i)$$

  where $\delta(\mathbf{v}-\mathbf{x})$ denotes the Dirac delta function, a probability distribution placing all its mass at the location $\mathbf{x}$.

- $\hat{\mathbb{Q}}_{t_i}(\mathbf{u})$ is the cell distribution at t_i rescaled by the relative growth rate to account for cell division/death.

  $$\hat{\mathbb{Q}}_{t_i}(\mathbf{u}) = \hat{\mathbb{P}}_{t_i}(\mathbf{u}) \frac{g(\mathbf{u})^{t_{i+1}-t_i}}{\int g(\mathbf{u})^{t_{i+1}-t_i} d\hat{\mathbb{P}}_{t_i}}$$

  Here $g(\mathbf{u})$ represents the relative growth rate of cell $\mathbf{u}$ in the time interval (t_i, t_{i+1}) and is estimated within the framework of unbalanced optimal transport (Chizat et al., 2008). For more details, we refer the reader to the supplementary information of Schiebinger et al., 2019.

- $c(\mathbf{u},\mathbf{v})$ is a cost function defined as the euclidean distance $\|\mathbf{u}-\mathbf{v}\|^2$. The first term of the objective function minimizes the cost function weighted by the temporal couplings, which may

[‡]Note that $\mathbf{u}$ may represnt the original gene expression space or a reduced dimensional embedding estimated via Principal Component Analysis or Diffusion Maps.

be interpreted as the transport distance between the distributions $\hat{\mathbb{P}}_{t_i}$ and $\hat{\mathbb{P}}_{t_{i+1}}$ (also known as the Wasserstein distance).

- The second term on the RHS represents entropic regularization, and ϵ is the corresponding strength. Classic OT identifies "deterministic" couplings in that one cell at t_i is transported to a single cell at t_{i+1}. Introduction of the entropic regularization term makes this problem non-deterministic, capturing the notion that there may exist immature cells whose fate is not completely determined. Our inferences of multipotentialy is directly a consequence of adding this entropic regularization term. Additionally, entropic regularization also makes the problem strongly convex, which is computationally beneficial.

- The third and the fourth terms are features of unbalanced optimal transport, where equality constraints on the marginals (a consequence of mass conservation) are relaxed. λ_1 and λ_2 are corresponding Lagrange multipliers.

We note that values of the hyperparameters ϵ, λ_1 and λ_2 are held fixed for all pairwise transport map calculations (E13, E14), (E14, E16), ... etc.

Application to RGC diversification and long-range couplings: We apply Waddington-OT to each pair of consecutive ages t_i and t_{i+1} to estimate the transport map $\hat{\Pi}_{t_i,t_{i+i}}$. Transport maps connecting non-consecutive time points t_i and t_{i+k} are estimated through a simple matrix multiplication of intermediate transport maps,

$$\hat{\Pi}_{t_i,t_{i+k}} = \hat{\Pi}_{t_i,t_{i+1}} \hat{\Pi}_{t_{i+1},t_{i+2}} \ldots \hat{\Pi}_{t_{i+k-1},t_{i+k}}$$

The transport matrices $\hat{\Pi}_{t_i,t_j}$ encode fate relationships between cells at t_i and cells at at a later time t_j ($t_i < t_j$). These relationships can be analyzed at the level of clusters at t_j to associate each cell $\mathbf{u} \in \mathcal{A}^{t_i}$ with transcriptomically defined cluster. This is particularly useful in estimating the terminal identity of immature RGCs.

Operationally we compute for each cell $\mathbf{u} \in \mathcal{A}^{t_i}$ a **"cell fate vector"** $f_{t_j}(\beta; \mathbf{u}, t_i)$, $(\beta = 1, 2, \ldots)$ encoding the probabilities that $\mathbf{u}$ is associated with cluster $\mathcal{C}_\beta^{t_j}$ at time t_j,

$$f_{t_j}(\beta; \mathbf{u}, t_i) = \frac{\sum_{\mathbf{v} \in \mathcal{C}_\beta^{t_j}} \hat{\Pi}_{t_i,t_j}(\mathbf{u}, \mathbf{v})}{\sum_\beta \sum_{\mathbf{v} \in \mathcal{C}_\beta^{t_j}} \hat{\Pi}_{t_i,t_j}(\mathbf{u}, \mathbf{v})} \tag{11}$$

It is easy to verify that.

$$\sum_\beta f_{t_j}(\beta; \mathbf{u}, t_i) = 1 \qquad \forall\, \mathbf{u} \in \mathcal{A}^{t_i}$$

The cell fate vector $f_{t_j}(\beta; \mathbf{u}, t_i)$ encodes probabilistic associations between the cell $\mathbf{u}$ and terminal clusters at $t_j > t_i$ indexed by β. The **"cluster ancestry vector"** at an earlier time t_i of a cluster $C_\beta^{t_j}$ at time $t_j > t_i$, denoted $\Gamma_{t_i}(\mathbf{u}; C_\beta^{t_j})$, is defined as follows,

$$\Gamma_{t_i}(\mathbf{u}; C_\beta^{t_j}) = \frac{\sum_{\mathbf{v} \in \mathcal{C}_\beta^{t_j}} \hat{\Pi}_{t_i,t_j}(\mathbf{u}, \mathbf{v})}{\sum_{\mathbf{u} \in \mathcal{A}^{t_i}} \sum_{\mathbf{v} \in \mathcal{C}_\beta^{t_j}} \hat{\Pi}_{t_i,t_j}(\mathbf{u}, \mathbf{v})} \qquad (t_i < t_j) \tag{12}$$

In a similar vein, the **"cluster descendant vector"** at a later time t_o of a cluster $C_\beta^{t_j}$ at time $t_j < t_o$, denoted $\Gamma_{t_o}(\mathbf{u}; C_\beta^{t_j})$, is defined as,

$$\Gamma_{t_o}(\mathbf{u}; C_\beta^{t_j}) = \frac{\sum_{\mathbf{v}\in C_\beta^{t_j}} \hat{\Pi}_{t_j,t_o}(\mathbf{v},\mathbf{w})}{\sum_{\mathbf{w}\in\mathcal{A}^{t_o}} \sum_{\mathbf{v}\in C_\beta^{t_j}} \hat{\Pi}_{t_j,t_o}(\mathbf{v},\mathbf{w})} \qquad (t_j < t_o) \tag{13}$$

Equations (12) and (13) can be used to compute the putative ancestral or descendent cells associated with a cluster $C_\beta^{t_j}$ at time t_j.

Implementation details of WOT: RGC vectors from all ages were combined, median normalized and log-transformed. 1761 HVGs were identified using the Gamma-Poisson model, and Waddington-OT was run on this matrix as follows,

```
wot optimal_transport --matrix RGC_mat.mtx --cell_days cell_day.txt
                      --growth_iters 3 --epsilon 0.005 --out tmaps/RGC
```

Cell days were specified in `cell_day.txt` as 0, 1, 3, 6, 11 and 20 for E13, E14, E16, P0, P5 and P56 respectively. We computed trajectories and fates for each age using the following command illustrated for P0

```
wot trajectory --tmap tmaps/RGC --cell_set cell_sets.gmt --day 6
              -out tmaps/traj_RGC_P0.txt
```

Fates were computed as,

```
wot fates --tmap tmaps/RGC --cell_set cell_sets.gmt --day 6
              -out tmaps/fate_RGC_P0.txt
```

The above process was repeated for each age.

Multipotentiality of precursors: For each cell at ages $E13 - P5$, we used equation (11) to compute a terminal fate association $f_{P56}(\beta; \mathbf{u}, t)$ $(t \in \{E13, E14, E16, P0, P5\})$ quantifying the probability that it is a precursor of type $\beta \in 1, \ldots, 45$. Note that $f_{P56}(\beta; \mathbf{u}, t)$ is denoted as f_β for brevity in the main text. We define,

$$P(\mathbf{u}; t) = \frac{1}{\sum_\beta f_{P56}(\beta; \mathbf{u}, t)^2} \tag{14}$$

as the potential of precursor $\mathbf{u}$ at age t. Values of P range between 1 and 45, with lower values indicating restriction of fate and higher values suggesting multipotentiality.

Network analysis of fate couplings

We define,

$$C(\alpha, \beta; t) = \frac{\frac{1}{|\mathcal{A}^t|}\sum_{\mathbf{u}\in\mathcal{A}^t}\left(f_{P56}(\alpha;\mathbf{u},t) - \overline{f_{P56}(\alpha;t)}\right)\left(f_{P56}(\beta;\mathbf{u},t) - \overline{f_{P56}(\beta;t)}\right)}{\sqrt{\frac{1}{|\mathcal{A}^t|}\sum_{\mathbf{u}\in\mathcal{A}^t}\left(f_{P56}(\alpha;\mathbf{u},t) - \overline{f_{P56}(\alpha;t)}\right)^2}\sqrt{\frac{1}{|\mathcal{A}^t|}\sum_{\mathbf{v}\in\mathcal{A}^t}\left(f_{P56}(\beta;\mathbf{v},t) - \overline{f_{P56}(\beta;t)}\right)^2}} \tag{15}$$

as the fate coupling between RGC types α and β at age t. Clearly, $C(\alpha, \beta; t)$ is simply the Pearson correlation coefficient between $f_{P56}(\alpha; \mathbf{u}, t)$ and $f_{P56}(\beta; \mathbf{u}, t)$, the probabilities that a cell $\mathbf{u} \in \mathcal{A}^t$ is a precursor of α or β precursor. In equation (15),

$$\overline{f_{P56}(\alpha; t)} = \frac{1}{|\mathcal{A}^t|} \sum_{\mathbf{u} \in \mathcal{A}^t} f_{P56}(\alpha; \mathbf{u}, t) \tag{16}$$

is the mean probability that a cell at age t is a precursor of type α. We computed $C(\alpha, \beta; t)$ across all 990 pairs of RGC types at each immature age $t \in \{E13, E14, E16, P0, P5\}$. The values $C(\alpha, \beta; E13)$ were used as edge weights to visualize the fate coupling network of RGC types using the force-directed layout method (Fruchterman and Reingold, 1991) as implemented in the R package `igraph`. The node layout were computed using $C(\alpha, \beta; E13)$ values. For other ages, the node layout at E13 was retained but the edges were replotted based on $C(\alpha, \beta; t)$ values at the corresponding age.

We computed a null distribution of $C(\alpha, \beta; t)$ by randomizing the values of $f_{P56}(\alpha; \mathbf{u}, t)$ within each cell $\mathbf{u}$ across types. Null values of $C(\alpha, \beta; t)$ rarely exceeded 0.1 and never exceeded 0.2, so only edges with larger weights were visualized in Figure 5.

Decay of pairwise couplings

For each pair of RGC types α and β, we fitted a logistic equation to model the decay of pairwise couplings as,

$$C(\alpha, \beta; t) = \frac{1}{1 + \exp(\beta_0 + \beta_1 t)} \tag{17}$$

The values of t corresponding to E13, E14, E16, P0 and P5 were $t = 0, 1, 3, 6, 11$, with $C(\alpha, \beta; t)$ computed using equation (15). We also assumed that $C(\alpha, \beta; t) = 0$ at $t = 36$, corresponding to P30. Thus, six data-points were used to estimate two parameters for each of the 180 pairs of RGC types that had non-zero values of $C(\alpha, \beta; t)$. The `nls` function from the R package `stats` was used to estimate β_0 and β_1. The results are plotted in Fig. 5f.

Logistic modeling of specification and calculation of τ_{sp}

We hypothesized that the specification of a type β corresponds to the localization of its precursors in transcriptomic space. The extent of localization for a RGC type β at across the time course was calculated as follows. At each age t, we identified the set of precursor RGCs $\text{Prec}(\beta; t)$ showing the highest fate probability corresponding to type β,

$$\text{Prec}(\beta; t) = \{\mathbf{u} \in \mathcal{A}^t \quad | \quad f_{P56}(\beta; \mathbf{u}, t) > f_{P56}(\alpha \neq \beta; \mathbf{u}, t)\} \tag{18}$$

Next, we calculated how the precursors of β were distributed across clusters at time t. We computed the occupancy fractions of precursor cells for type β across all cluster $\mathcal{C}_k, k = 1, 2, \ldots, N(t)$ at a particular time t ($N(t)$ is the number of transcriptomically defined clusters at time t),

$$p_k(\beta; t) = \frac{\#\{\text{cluster}(\mathbf{u}) = \mathcal{C}_k \ \ \forall\, \mathbf{u} \in \text{Prec}(\beta; t)\}}{\#\{\mathbf{u} \in \text{Prec}(\beta; t)\}} \tag{19}$$

The localization score for each type β at a given time t was defined as,

$$\text{Localization}(\beta; t) = 1 - \frac{\sum_{k=1}^{N(t)} \frac{1}{p_k(\beta; t)^2}}{\sum_{k=1}^{N(t)} 1} \tag{20}$$

where the index k ranges over the number of clusters at time t. As defined Localization$(\beta; t)$ is restricted to be between 0 to 1, with higher values representing a greater specification. We used a logistic model to approximate the localization of each type as,

$$\text{Localization}(\beta; t) = \frac{\exp(\gamma_0 + \gamma_1 t)}{1 + \exp(\gamma_0 + \gamma_1 t)} \tag{21}$$

As in the previous section the `nls` function was used to estimate the logistic parameters γ_0 and γ_1. We consider a type β as specific if it crosses the line $y(t) = 0.95\left(1 - 1/\sum_{k=1}^{N(t)} 1\right)$. Thus the specification time for a type β is defined as,

$$\tau_{sp}(\beta) = \arg\min_t \text{Localization}(\beta; t) \geq y(t) \tag{22}$$

Note that as defined, τ_{sp} can be any time point in the interval (E13, P30) corresponding to $t \in (0, 36)$.

Inference of laterality in RGC types

To identify putative ipsilateral and contralateral specified-RGC precursors at E13, we scored each precursor RGC based on their expression of bonafide ipsilateral genes (*Zic2, Zic1* and *Igfbp5*) and bonafide contralateral genes (*Isl2, Fgf12, Igf1*) as in (Wang et al., 2016). We refer to these as I-RGC and C-RGC scores. Putative I-RGCs were those cells that expressed the I-RGC score at 1.5 standard deviations higher than the mean across all cells, and those that express the C-RGC score at 1.5 standard deviations lower than the mean across all cells. C-RGCs were defined analogously. Many cells did not express either of these marker sets as shown in Fig. 7c. These are likely to be RGCs that have not declared their laterality, or C-RGCs that are not defined by the expression of *Isl2, Fgf12*, and *Igf1*.

WOT was then used to compute the descendants of E13 I-RGCs at all subsequent ages through P56 using the `wot fates` command introduced above. These descendants were used for two purposes. First, we assessed the proportion of putative I-RGCs across types as in Fig. 7d. We also performed a differential gene expression test between putative I-RGCs and the remaining RGCs at all ages, as shown in Figs. 7e,f and Supplementary Figs. 8d,e.

References

Baden, T., Berens, P., Franke, K., Roman Roson, M., Bethge, M., & Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. *Nature, 529*(7586), 345-350. doi:10.1038/nature16468

Baden, T., Euler, T., & Berens, P. (2020). Understanding the retinal basis of vision across species. *Nat Rev Neurosci, 21*(1), 5-20. doi:10.1038/s41583-019-0242-1

Bae, J. A., Mu, S., Kim, J. S., Turner, N. L., Tartavull, I., Kemnitz, N., . . . Eyewirers. (2018). Digital Museum of Retinal Ganglion Cells with Dense Anatomy and Physiology. *Cell, 173*(5), 1293-1306 e1219. doi:10.1016/j.cell.2018.04.040

Baron, C. S., & van Oudenaarden, A. (2019). Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat Rev Mol Cell Biol, 20*(12), 753-765. doi:10.1038/s41580-019-0186-3

Barres, B. A., Silverstein, B. E., Corey, D. P., & Chun, L. L. (1988). Immunological, morphological, and electrophysiological variation among retinal ganglion cells purified by panning. *Neuron, 1*(9), 791-803. doi:10.1016/0896-6273(88)90127-4

Bassett, E. A., & Wallace, V. A. (2012). Cell fate determination in the vertebrate retina. *Trends Neurosci, 35*(9), 565-573. doi:10.1016/j.tins.2012.05.004

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W., Ng, L. G., . . . Newell, E. W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature biotechnology, 37*(1), 38-44.

Berson, D. M., Dunn, F. A., & Takao, M. (2002). Phototransduction by retinal ganglion cells that set the circadian clock. *Science, 295*(5557), 1070-1073. doi:10.1126/science.1067262

Blackshaw, S., Harpavat, S., Trimarchi, J., Cai, L., Huang, H., Kuo, W. P., . . . Cepko, C. L. (2004). Genomic analysis of mouse retinal development. *PLoS Biol, 2*(9), E247. doi:10.1371/journal.pbio.0020247

Boije, H., MacDonald, R. B., & Harris, W. A. (2014). Reconciling competence and transcriptional hierarchies with stochasticity in retinal lineages. *Curr Opin Neurobiol, 27*, 68-74. doi:10.1016/j.conb.2014.02.014

Brown, N. L., Patel, S., Brzezinski, J., & Glaser, T. (2001). Math5 is required for retinal ganglion cell and optic nerve formation. *Development, 128*(13), 2497-2508. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11493566

884 Cembrowski, M. S., & Spruston, N. (2019). Heterogeneity within classical cell types is the rule:
885 lessons from hippocampal pyramidal neurons. *Nat Rev Neurosci, 20*(4), 193-204.
886 doi:10.1038/s41583-019-0125-5

887 Cepko, C. (2014). Intrinsically different retinal progenitor cells produce specific types of progeny.
888 *Nat Rev Neurosci, 15*(9), 615-627. doi:10.1038/nrn3767

889 Chen, S., Wang, Q. L., Nie, Z., Sun, H., Lennon, G., Copeland, N. G., . . . Zack, D. J. (1997). Crx,
890 a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor
891 cell-specific genes. *Neuron, 19*(5), 1017-1030. doi:10.1016/s0896-6273(00)80394-3

892 Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system.* Paper presented at
893 the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery
894 and data mining.

895 Clark, B. S., Stein-O'Brien, G. L., Shiau, F., Cannon, G. H., Davis-Marcisak, E., Sherman, T., . . .
896 Blackshaw, S. (2019). Single-Cell RNA-Seq Analysis of Retinal Development Identifies
897 NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. *Neuron, 102*(6),
898 1111-1126 e1115. doi:10.1016/j.neuron.2019.04.010

899 Demyanenko, G. P., & Maness, P. F. (2003). The L1 cell adhesion molecule is essential for
900 topographic mapping of retinal axons. *J Neurosci, 23*(2), 530-538. Retrieved from
901 https://www.ncbi.nlm.nih.gov/pubmed/12533613

902 Di Bella, D. J., Habibi, E., Stickels, R. R., Scalia, G., Brown, J., Yadollahpour, P., . . . Arlotta, P.
903 (2021). Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature,*
904 *595*(7868), 554-559. doi:10.1038/s41586-021-03670-5

905 Dowling, J. E. (2012). *The Retina: An Approachable Part of the Brain* (2nd ed.): Harvard University
906 Press.

907 Drager, U. C. (1985). Birth dates of retinal ganglion cells giving rise to the crossed and uncrossed
908 optic projections in the mouse. *Proc R Soc Lond B Biol Sci, 224*(1234), 57-77.
909 doi:10.1098/rspb.1985.0021

910 Espinosa-Medina, I., Garcia-Marques, J., Cepko, C., & Lee, T. (2019). High-throughput dense
911 reconstruction of cell lineages. *Open Biol, 9*(12), 190229. doi:10.1098/rsob.190229

912 Famiglietti, E. V., Jr., & Kolb, H. (1976). Structural basis for ON-and OFF-center responses in
913 retinal ganglion cells. *Science, 194*(4261), 193-195. doi:10.1126/science.959847

914 Farah, M. H., & Easter, S. S., Jr. (2005). Cell birth and death in the mouse retinal ganglion cell
915 layer. *J Comp Neurol, 489*(1), 120-134. doi:10.1002/cne.20615

Franke, K., Berens, P., Schubert, T., Bethge, M., Euler, T., & Baden, T. (2017). Inhibition decorrelates visual feature representations in the inner retina. *Nature, 542*(7642), 439-444. doi:10.1038/nature21394

Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and experience, 21*(11), 1129-1164.

Fyodorov, Y. V., & Mirlin, A. D. (1992). Analytical derivation of the scaling law for the inverse participation ratio in quasi-one-dimensional disordered systems. *Physical review letters, 69*(7), 1093.

Godement, P., Salaun, J., & Imbert, M. (1984). Prenatal and postnatal development of retinogeniculate and retinocollicular projections in the mouse. *J Comp Neurol, 230*(4), 552-575. doi:10.1002/cne.902300406

Goetz, J., Jessen, Z. F., Jacobi, A., Mani, A., Cooler, S., Greer, D., . . . Sanes, J. (2021). Unified classification of mouse retinal ganglion cells using function, morphology, and gene expression. *bioRxiv*.

Gomes, F. L., Zhang, G., Carbonell, F., Correa, J. A., Harris, W. A., Simons, B. D., & Cayouette, M. (2011). Reconstruction of rat retinal progenitor cell lineages in vitro reveals a surprising degree of stochasticity in cell fate decisions. *Development, 138*(2), 227-235. doi:10.1242/dev.059683

Herrera, E., Brown, L., Aruga, J., Rachel, R. A., Dolen, G., Mikoshiba, K., . . . Mason, C. A. (2003). Zic2 patterns binocular vision by specifying the uncrossed retinal projection. *Cell, 114*(5), 545-557. doi:10.1016/s0092-8674(03)00684-6

Holguera, I., & Desplan, C. (2018). Neuronal specification in space and time. *Science, 362*(6411), 176-180. doi:10.1126/science.aas9435

Holt, C. E., Bertsch, T. W., Ellis, H. M., & Harris, W. A. (1988). Cellular determination in the Xenopus retina is independent of lineage and birth date. *Neuron, 1*(1), 15-26. doi:10.1016/0896-6273(88)90205-x

Hong, Y. K., Kim, I. J., & Sanes, J. R. (2011). Stereotyped axonal arbors of retinal ganglion cell subsets in the mouse superior colliculus. *J Comp Neurol, 519*(9), 1691-1711. doi:10.1002/cne.22595

Hooks, B. M., & Chen, C. (2020). Circuitry Underlying Experience-Dependent Plasticity in the Mouse Visual System. *Neuron, 107*(5), 986-987. doi:10.1016/j.neuron.2020.08.004

Jeon, C. J., Strettoi, E., & Masland, R. H. (1998). The major cell populations of the mouse retina. *J Neurosci, 18*(21), 8936-8946. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9786999

950 Johnson, K. P., Fitzpatrick, M. J., Zhao, L., Wang, B., McCracken, S., Williams, P. R., &
951 Kerschensteiner, D. (2021). Cell-type-specific binocular vision guides predation in mice.
952 *Neuron, 109*(9), 1527-1539 e1524. doi:10.1016/j.neuron.2021.03.010

953 Johnston, R. J., Jr., & Desplan, C. (2010). Stochastic mechanisms of cell fate specification that
954 yield random or robust outcomes. *Annu Rev Cell Dev Biol, 26*, 689-719.
955 doi:10.1146/annurev-cellbio-100109-104113

956 Kantorovich, L. V. (1942). *On the translocation of masses.* Paper presented at the Dokl. Akad.
957 Nauk. USSR (NS).

958 Kay, J. N., Chu, M. W., & Sanes, J. R. (2012). MEGF10 and MEGF11 mediate homotypic
959 interactions required for mosaic spacing of retinal neurons. *Nature, 483*(7390), 465-469.
960 doi:10.1038/nature10877

961 Kay, J. N., De la Huerta, I., Kim, I. J., Zhang, Y., Yamagata, M., Chu, M. W., . . . Sanes, J. R.
962 (2011). Retinal ganglion cells with distinct directional preferences differ in molecular
963 identity, structure, and central projections. *J Neurosci, 31*(21), 7753-7762.
964 doi:10.1523/JNEUROSCI.0907-11.2011

965 Keeley, P. W., Eglen, S. J., & Reese, B. E. (2020). From random to regular: Variation in the
966 patterning of retinal mosaics. *J Comp Neurol, 528*(13), 2135-2160. doi:10.1002/cne.24880

967 Kim, I. J., Zhang, Y., Meister, M., & Sanes, J. R. (2010). Laminar restriction of retinal ganglion cell
968 dendrites and axons: subtype-specific developmental patterns revealed with transgenic
969 markers. *J Neurosci, 30*(4), 1452-1462. doi:10.1523/JNEUROSCI.4779-09.2010

970 Kiyama, T., Long, Y., Chen, C. K., Whitaker, C. M., Shay, A., Wu, H., . . . Mao, C. A. (2019).
971 Essential Roles of Tbr1 in the Formation and Maintenance of the Orientation-Selective J-
972 RGCs and a Group of OFF-Sustained RGCs in Mouse. *Cell Rep, 27*(3), 900-915 e905.
973 doi:10.1016/j.celrep.2019.03.077

974 Lefebvre, J. L., Sanes, J. R., & Kay, J. N. (2015). Development of dendritic form and function.
975 *Annu Rev Cell Dev Biol, 31*, 741-777. doi:10.1146/annurev-cellbio-100913-013020

976 Liu, J., Reggiani, J. D. S., Laboulaye, M. A., Pandey, S., Chen, B., Rubenstein, J. L. R., . . . Sanes,
977 J. R. (2018). Tbr1 instructs laminar patterning of retinal ganglion cell dendrites. *Nat*
978 *Neurosci, 21*(5), 659-670. doi:10.1038/s41593-018-0127-z

979 Livesey, F. J., & Cepko, C. L. (2001). Vertebrate neural cell-fate determination: lessons from the
980 retina. *Nat Rev Neurosci, 2*(2), 109-118. doi:10.1038/35053522

981 Lo Giudice, Q., Leleu, M., La Manno, G., & Fabre, P. J. (2019). Single-cell transcriptional logic of
982 cell-fate specification and axon guidance in early-born retinal neurons. *Development,*
983 *146*(17). doi:10.1242/dev.178103

Lodato, S., & Arlotta, P. (2015). Generating neuronal diversity in the mammalian cerebral cortex. *Annu Rev Cell Dev Biol, 31*, 699-720. doi:10.1146/annurev-cellbio-100814-125353

Lyu, J., & Mu, X. (2021). Genetic control of retinal ganglion cell genesis. *Cell Mol Life Sci, 78*(9), 4417-4433. doi:10.1007/s00018-021-03814-w

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., . . . McCarroll, S. A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell, 161*(5), 1202-1214. doi:10.1016/j.cell.2015.05.002

Mao, C.-A., Chen, C.-K., Kiyama, T., Weber, N., Whitaker, C. M., Pan, P., . . . Massey, S. C. (2020). Tbr2-expressing retinal ganglion cells are ipRGCs. *bioRxiv*.

Mao, C. A., Li, H., Zhang, Z., Kiyama, T., Panda, S., Hattar, S., . . . Wang, S. W. (2014). T-box transcription regulator Tbr2 is essential for the formation and maintenance of Opn4/melanopsin-expressing intrinsically photosensitive retinal ganglion cells. *J Neurosci, 34*(39), 13083-13095. doi:10.1523/JNEUROSCI.1027-14.2014

Marcucci, F., Soares, C. A., & Mason, C. (2019). Distinct timing of neurogenesis of ipsilateral and contralateral retinal ganglion cells. *J Comp Neurol, 527*(1), 212-224. doi:10.1002/cne.24467

Marquardt, T., & Gruss, P. (2002). Generating neuronal diversity in the retina: one for nearly all. *Trends Neurosci, 25*(1), 32-38. doi:10.1016/s0166-2236(00)02028-2

Martersteck, E. M., Hirokawa, K. E., Evarts, M., Bernard, A., Duan, X., Li, Y., . . . Harris, J. A. (2017). Diverse Central Projection Patterns of Retinal Ganglion Cells. *Cell Rep, 18*(8), 2058-2072. doi:10.1016/j.celrep.2017.01.075

Masland, R. H. (2012). The neuronal organization of the retina. *Neuron, 76*(2), 266-280. doi:10.1016/j.neuron.2012.10.002

Mason, C., & Slavi, N. (2020). Retinal Ganglion Cell Axon Wiring Establishing the Binocular Circuit. *Annu Rev Vis Sci, 6*, 215-236. doi:10.1146/annurev-vision-091517-034306

McConnell, S. K. (1991). The generation of neuronal diversity in the central nervous system. *Annu Rev Neurosci, 14*, 269-300. doi:10.1146/annurev.ne.14.030191.001413

McKenna, A., Findlay, G. M., Gagnon, J. A., Horwitz, M. S., Schier, A. F., & Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science, 353*(6298), aaf7907. doi:10.1126/science.aaf7907

McNeill, D. S., Sheely, C. J., Ecker, J. L., Badea, T. C., Morhardt, D., Guido, W., & Hattar, S. (2011). Development of melanopsin-based irradiance detecting circuitry. *Neural Dev, 6*, 8. doi:10.1186/1749-8104-6-8

1017 Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie*
1018 *Royale des Sciences de Paris*.

1019 Osterhout, J. A., Josten, N., Yamada, J., Pan, F., Wu, S. W., Nguyen, P. L., . . . Huberman, A. D.
1020 (2011). Cadherin-6 mediates axon-target matching in a non-image-forming visual circuit.
1021 *Neuron, 71*(4), 632-639. doi:10.1016/j.neuron.2011.07.006

1022 Pak, W., Hindges, R., Lim, Y. S., Pfaff, S. L., & O'Leary, D. D. (2004). Magnitude of binocular
1023 vision controlled by islet-2 repression of a genetic program that specifies laterality of retinal
1024 axon pathfinding. *Cell, 119*(4), 567-578. doi:10.1016/j.cell.2004.10.026

1025 Peng, Y. R., Tran, N. M., Krishnaswamy, A., Kostadinov, D., Martersteck, E. M., & Sanes, J. R.
1026 (2017). Satb1 Regulates Contactin 5 to Pattern Dendrites of a Mammalian Retinal
1027 Ganglion Cell. *Neuron, 95*(4), 869-883 e866. doi:10.1016/j.neuron.2017.07.019

1028 Price, J., Turner, D., & Cepko, C. (1987). Lineage analysis in the vertebrate nervous system by
1029 retrovirus-mediated gene transfer. *Proc Natl Acad Sci U S A, 84*(1), 156-160.
1030 doi:10.1073/pnas.84.1.156

1031 Rheaume, B. A., Jereen, A., Bolisetty, M., Sajid, M. S., Yang, Y., Renna, K., . . . Trakhtenberg, E.
1032 F. (2018). Single cell transcriptome profiling of retinal ganglion cells identifies cellular
1033 subtypes. *Nat Commun, 9*(1), 2759. doi:10.1038/s41467-018-05134-3

1034 Rockhill, R. L., Euler, T., & Masland, R. H. (2000). Spatial order within but not between types of
1035 retinal neurons. *Proc Natl Acad Sci U S A, 97*(5), 2303-2307.
1036 doi:10.1073/pnas.030413497

1037 Rousso, D. L., Qiao, M., Kagan, R. D., Yamagata, M., Palmiter, R. D., & Sanes, J. R. (2016). Two
1038 Pairs of ON and OFF Retinal Ganglion Cells Are Defined by Intersectional Patterns of
1039 Transcription Factor Expression. *Cell Rep, 15*(9), 1930-1944.
1040 doi:10.1016/j.celrep.2016.04.069

1041 Sagner, A., & Briscoe, J. (2019). Establishing neuronal diversity in the spinal cord: a time and a
1042 place. *Development, 146*(22). doi:10.1242/dev.182154

1043 Sajgo, S., Ghinia, M. G., Brooks, M., Kretschmer, F., Chuang, K., Hiriyanna, S., . . . Badea, T. C.
1044 (2017). Molecular codes for cell type specification in Brn3 retinal ganglion cells. *Proc Natl*
1045 *Acad Sci U S A, 114*(20), E3974-E3983. doi:10.1073/pnas.1618551114

1046 Sanes, J. R., & Masland, R. H. (2015). The types of retinal ganglion cells: current status and
1047 implications for neuronal classification. *Annu Rev Neurosci, 38*, 221-246.
1048 doi:10.1146/annurev-neuro-071714-034120

1049 Sanes, J. R., Rubenstein, J. L., & Nicolas, J. F. (1986). Use of a recombinant retrovirus to study
1050 post-implantation cell lineage in mouse embryos. *EMBO J, 5*(12), 3133-3142. Retrieved
1051 from https://www.ncbi.nlm.nih.gov/pubmed/3102226

1052 Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., . . . Lander, E.
1053 S. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies
1054 Developmental Trajectories in Reprogramming. *Cell, 176*(6), 1517.
1055 doi:10.1016/j.cell.2019.02.026

1056 Sharma, N., Flaherty, K., Lezgiyeva, K., Wagner, D. E., Klein, A. M., & Ginty, D. D. (2020). The
1057 emergence of transcriptional identity in somatosensory neurons. *Nature, 577*(7790), 392-
1058 398. doi:10.1038/s41586-019-1900-1

1059 Shekhar, K., Lapan, S. W., Whitney, I. E., Tran, N. M., Macosko, E. Z., Kowalczyk, M., . . . Sanes,
1060 J. R. (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell
1061 Transcriptomics. *Cell, 166*(5), 1308-1323 e1330. doi:10.1016/j.cell.2016.07.054

1062 Shekhar, K., & Sanes, J. R. (2021). Generating and Using Transcriptomically Based Retinal Cell
1063 Atlases. *Annu Rev Vis Sci*. doi:10.1146/annurev-vision-032621-075200

1064 Soares, C. A., & Mason, C. A. (2015). Transient ipsilateral retinal ganglion cell projections to the
1065 brain: Extent, targeting, and disappearance. *Dev Neurobiol, 75*(12), 1385-1401.
1066 doi:10.1002/dneu.22291

1067 Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy.
1068 *Remote sensing of Environment, 62*(1), 77-89.

1069 Tran, N. M., Shekhar, K., Whitney, I. E., Jacobi, A., Benhar, I., Hong, G., . . . Sanes, J. R. (2019).
1070 Single-Cell Profiles of Retinal Ganglion Cells Differing in Resilience to Injury Reveal
1071 Neuroprotective Genes. *Neuron, 104*(6), 1039-1055 e1012.
1072 doi:10.1016/j.neuron.2019.11.006

1073 Trimarchi, J. M., Stadler, M. B., & Cepko, C. L. (2008). Individual retinal progenitor cells display
1074 extensive heterogeneity of gene expression. *PLoS One, 3*(2), e1588.
1075 doi:10.1371/journal.pone.0001588

1076 Turner, D. L., & Cepko, C. L. (1987). A common progenitor for neurons and glia persists in rat
1077 retina late in development. *Nature, 328*(6126), 131-136. doi:10.1038/328131a0

1078 Turner, D. L., Snyder, E. Y., & Cepko, C. L. (1990). Lineage-independent determination of cell
1079 type in the embryonic mouse retina. *Neuron, 4*(6), 833-845. doi:10.1016/0896-
1080 6273(90)90136-4

Voinescu, P. E., Kay, J. N., & Sanes, J. R. (2009). Birthdays of retinal amacrine cell subtypes are systematically related to their molecular identity and soma position. *J Comp Neurol, 517*(5), 737-750. doi:10.1002/cne.22200

Wamsley, B., & Fishell, G. (2017). Genetic and activity-dependent mechanisms underlying interneuron diversity. *Nat Rev Neurosci, 18*(5), 299-309. doi:10.1038/nrn.2017.30

Wang, Q., Marcucci, F., Cerullo, I., & Mason, C. (2016). Ipsilateral and Contralateral Retinal Ganglion Cells Express Distinct Genes during Decussation at the Optic Chiasm. *eNeuro, 3*(6). doi:10.1523/ENEURO.0169-16.2016

Wetts, R., & Fraser, S. E. (1988). Multipotent precursors can give rise to all major cell types of the frog retina. *Science, 239*(4844), 1142-1145. doi:10.1126/science.2449732

Yamagata, M., Yan, W., & Sanes, J. R. (2021). A cell atlas of the chick retina based on single-cell transcriptomics. *Elife, 10*. doi:10.7554/eLife.63907

Yan, W., Laboulaye, M. A., Tran, N. M., Whitney, I. E., Benhar, I., & Sanes, J. R. (2020). Mouse Retinal Cell Atlas: Molecular Identification of over Sixty Amacrine Cell Types. *J Neurosci, 40*(27), 5177-5195. doi:10.1523/JNEUROSCI.0471-20.2020

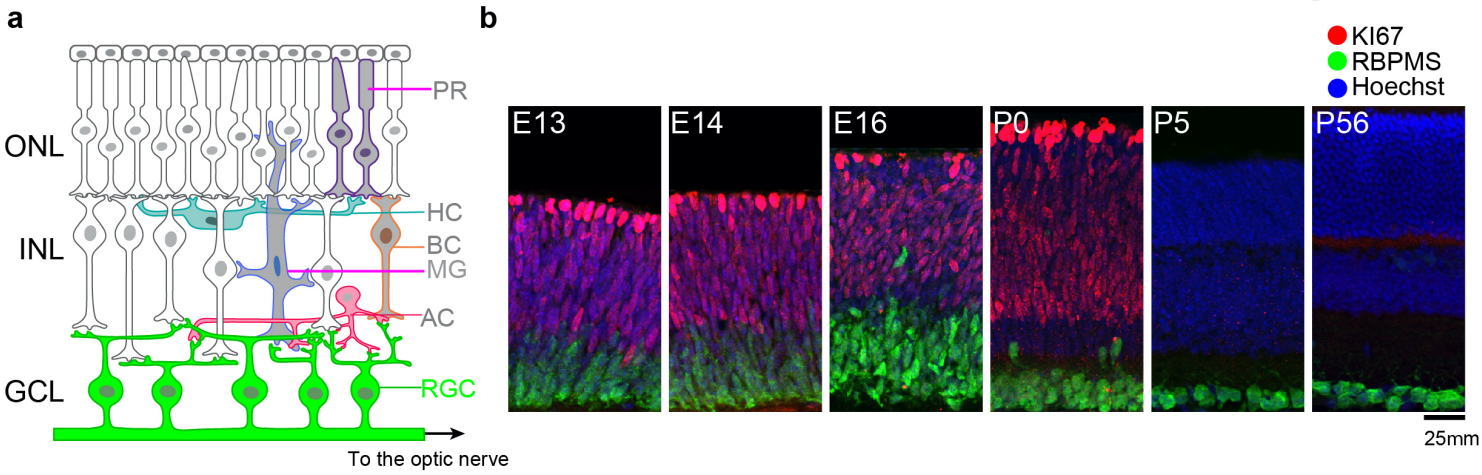
Yuste, R., Hawrylycz, M., Aalling, N., Aguilar-Valles, A., Arendt, D., Armananzas, R., . . . Lein, E. (2020). A community-based transcriptomics classification and nomenclature of neocortical cell types. *Nat Neurosci, 23*(12), 1456-1468. doi:10.1038/s41593-020-0685-8

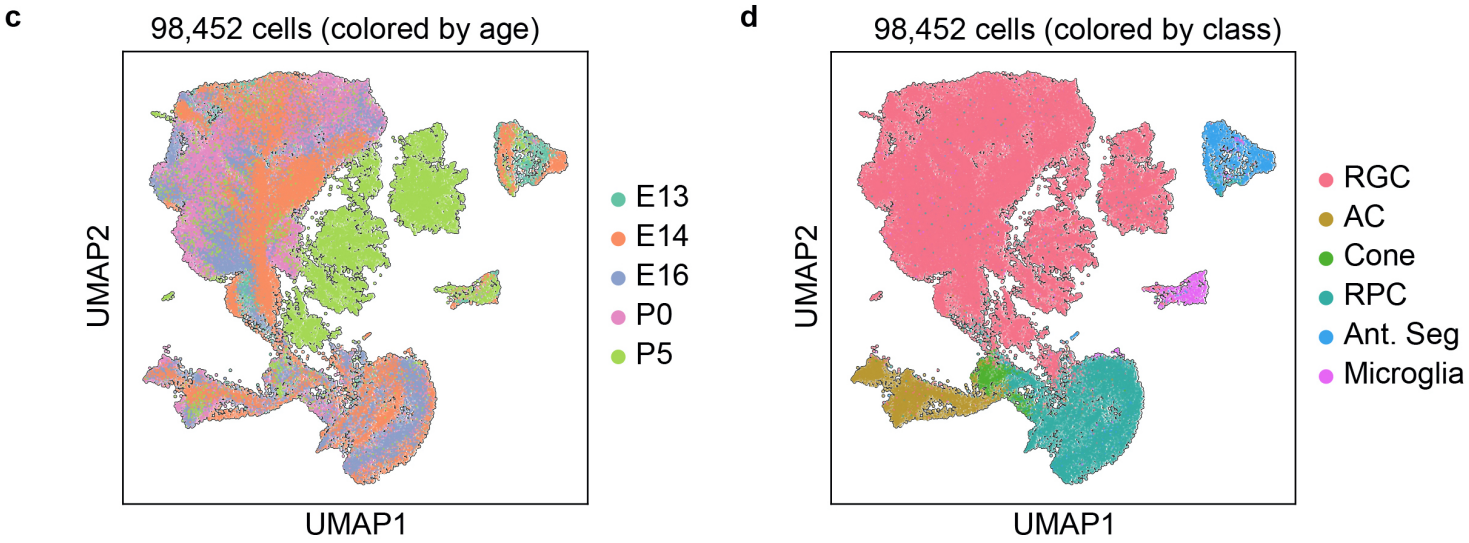
Zeng, H., & Sanes, J. R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat Rev Neurosci, 18*(9), 530-546. doi:10.1038/nrn.2017.85

Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., . . . Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat Commun, 8*, 14049. doi:10.1038/ncomms14049

MAIN FIGURES AND LEGENDS

Figure 1





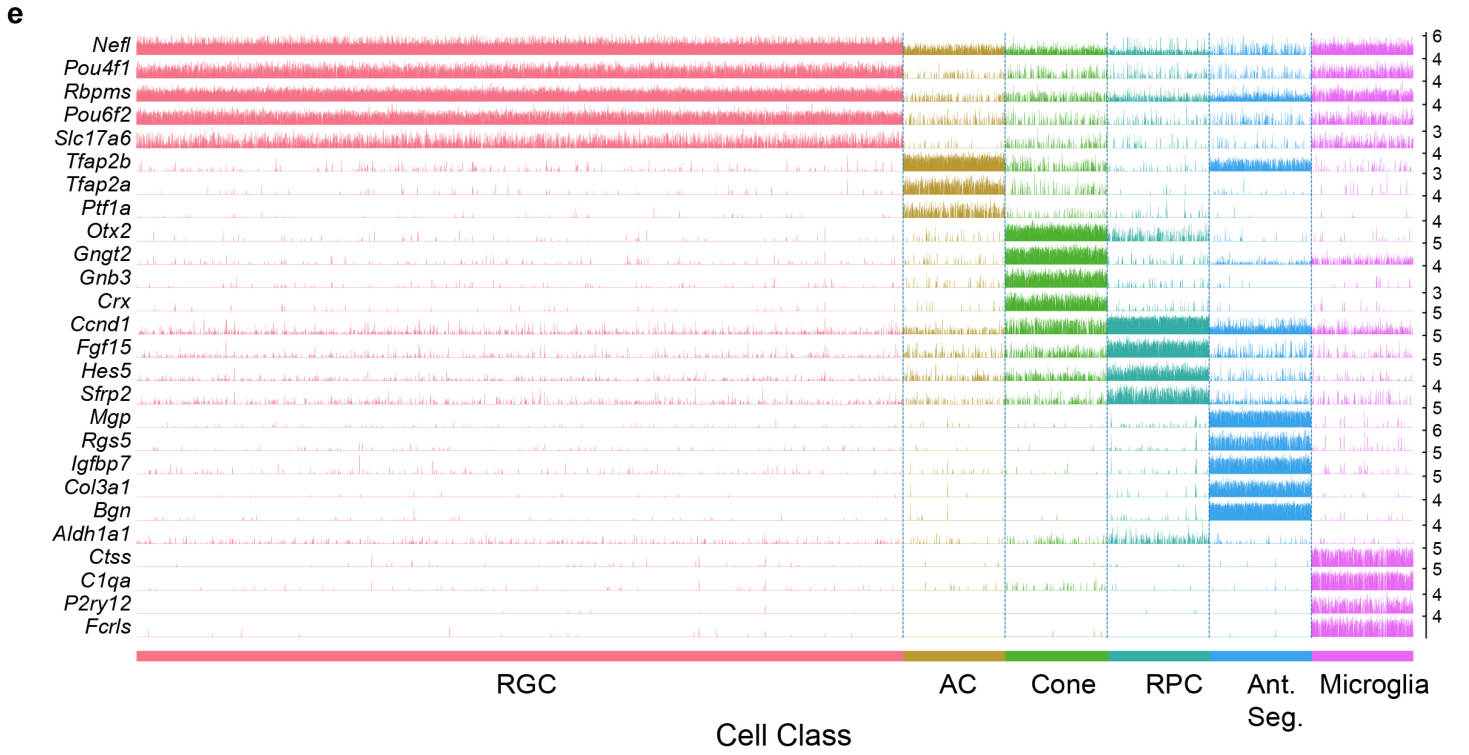


Figure 1, Transcriptomic profiling of single postmitotic RGCs during embryonic and postnatal development in mice.

a. Sketch of a section of the mouse retina showing major cell classes - photoreceptors (PRs; rods and cones), horizontal cells (HCs), bipolar cells (BCs), amacrine cells (ACs), Müller glia (MGs) and retinal ganglion cells (RGCs). PRs reside in the outer nuclear layer (ONL), while BCs, HCs and most ACs reside in the inner nuclear layer (INL). RGCs and some ACs reside in the ganglion cell layer (GCL). Axons of RGCs project to higher visual areas via the optic nerve.

b. Retinal section of the indicated ages labeled for the cell-cycle marker *MKI67* (red) and the RGC marker *RBPMS* (green); nuclei are counterstained by the Hoeschst dye (blue). Micrographs are orientated as the schematic in panel a.

c. Visualization of transcriptional diversity of 98,452 cells using Uniform Manifold Approximation and Projection (UMAP), a nonlinear dimensionality reduction algorithm that assigns proximal x-y coordinates to cells (dots) with similar transcriptional profiles[61].

d. Same as c, with cells colored by cell class, assigned based on transcriptional signatures displayed in panel e. RPC, retinal progenitor cells; Ant. Seg., anterior segment cells.

e. Tracksplot showing expression patterns of cell-class specific marker genes (rows) across single cells (columns). Cells are grouped by class as in d. For each class, we randomly sampled 20% of total cells covering all immature time points (E13, E14, E16, P0, P5). For each gene, the scale on the y-axis (right) corresponds to normalized, log-transformed transcript counts detected in each cell.

Figure 2

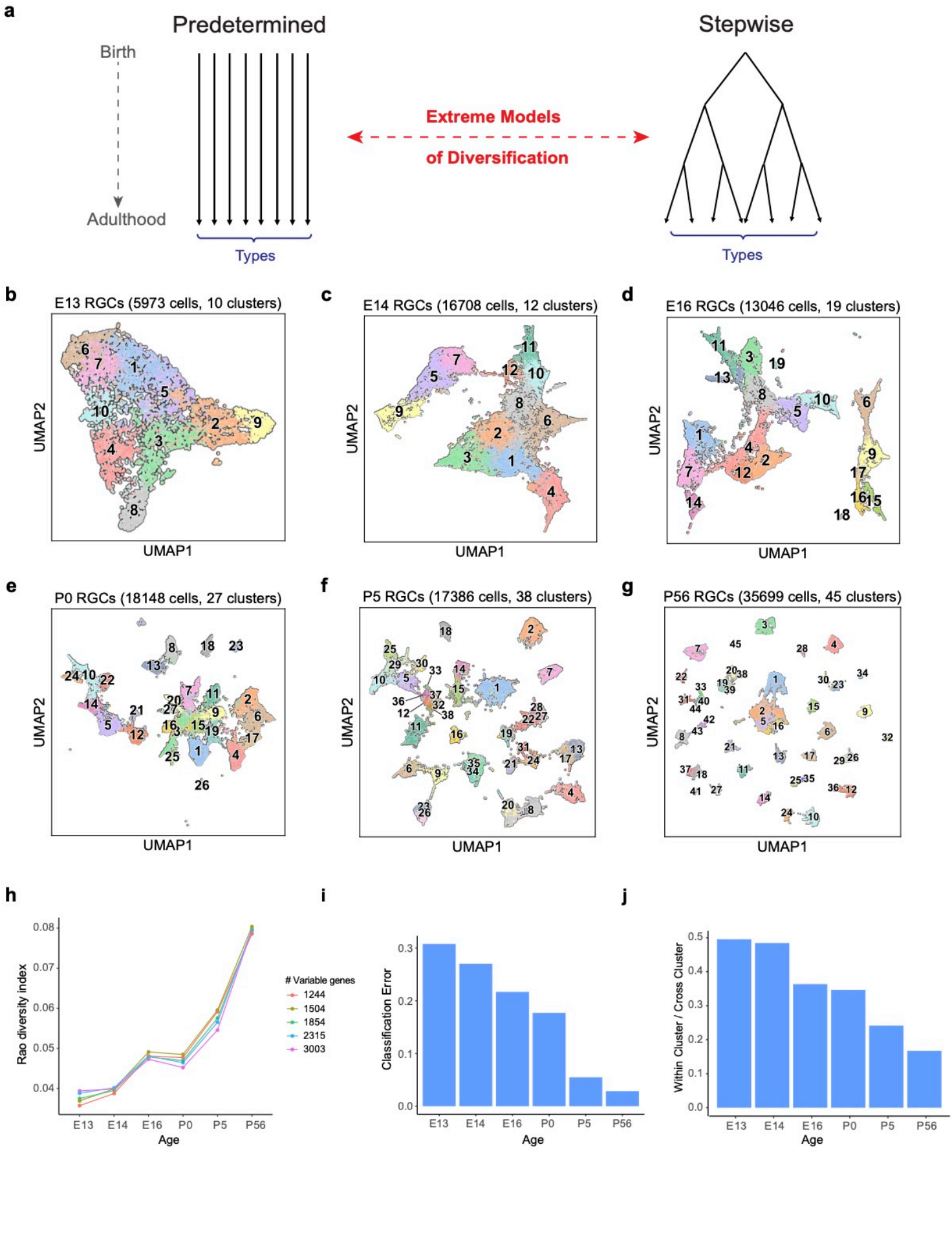


Figure 2. The number and discreteness of transcriptomic clusters of RGCs increases with age.

a. Extreme models of RGC diversification. In one scenario (*left*) immature RGCs commit to one of the terminal types by the time of birth (i.e. mitotic exit) or shortly after. Alternatively (*right)*, initially identical postmitotic RGC precursors acquire distinct molecular identities in a gradual process of restriction.

b-g. Visualization of transcriptomic diversity of immature RGCs at E13 (b), E14 (c), E16 (d), P0 (e), P5 (f) and P56 (g) using UMAP. Cells are colored by their cluster identity, determined independently using dimensionality reduction and graph clustering (**Methods**). Clusters are numbered based on decreasing size at each age. Data for adults (P56) are replotted from ref. 10. In that study 45 transcriptomic types were identified via unsupervised approaches, one of which was mapped to 2 known functional types by supervised approaches. We do not distinguish them in this study.

h. Transcriptional diversity of RGCs as measured by the Rao diversity index (y-axis) increases with age (x-axis). The trend is insensitive to the number of genes used to compute inter-cluster distance (colors). See **Methods** for details underlying the calculation.

i. Transcriptomic distinctions between RGC clusters become sharper with age as shown by decreasing average per-cluster error of a multiclass-classifier with age. Gradient boosted decision trees[62] were trained on a subset of the data, and applied on held out samples to determine the test error.

j. RGC clusters also become better separated in the UMAP embedding, as shown by decreasing values of the average relative cluster diameter with age.

Figure 3

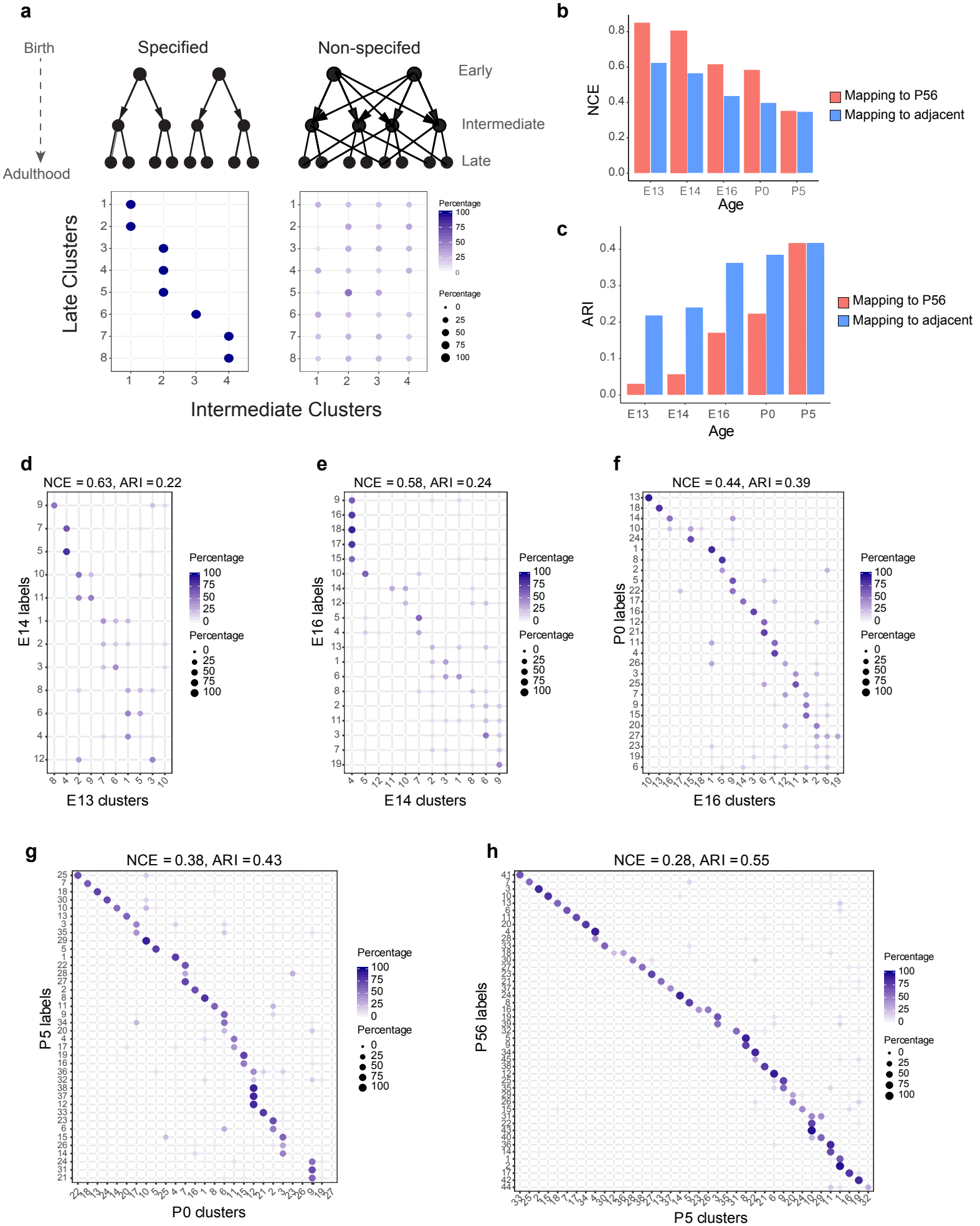


Figure 3. Incompletely specified temporal relationships among RGC clusters.
a. *Top:* Specified (*left*) and non-specified (*right*) modes of diversification. Nodes denote transcriptomic clusters of immature RGCs, and arrows denote fate relationships. *Bottom:* Confusion matrices depicting transcriptomic correspondence between late and early clusters expected for the two modes. Circles and colors indicate the percentage of a given late cluster (row) assigned to a corresponding early cluster (column) by transcriptome-based classifier trained on early clusters. The number of late and early clusters have been set to eight and four for illustration purposes.
b. Barplot showing values of the normalized conditional entropy (NCE) for each age calculated using the transcriptional cluster IDs and the Xgboost-assigned cluster IDs corresponding to the next age or to P56 (E.g. for E13, the NCE was calculated across E13 RGCs by comparing their transcriptional cluster ID with assigned E14 cluster IDs based on a classifier trained on the E14 data). Lower values indicate specific mappings.
c. Same as b, but plotting values of the adjusted Rand Index (ARI), where larger values correspond to higher specificity.
d-h. Confusion matrices (representation as in a), showing transcriptomic correspondence between consecutive ages: E14-E13 (d), E16-E14 (e), P0-E16 (f), P5-P0 (g), P56-P5 (h). In each case, the classifier was trained on the late time point and applied to the early time point. Rows sum to 100%.

Figure 4

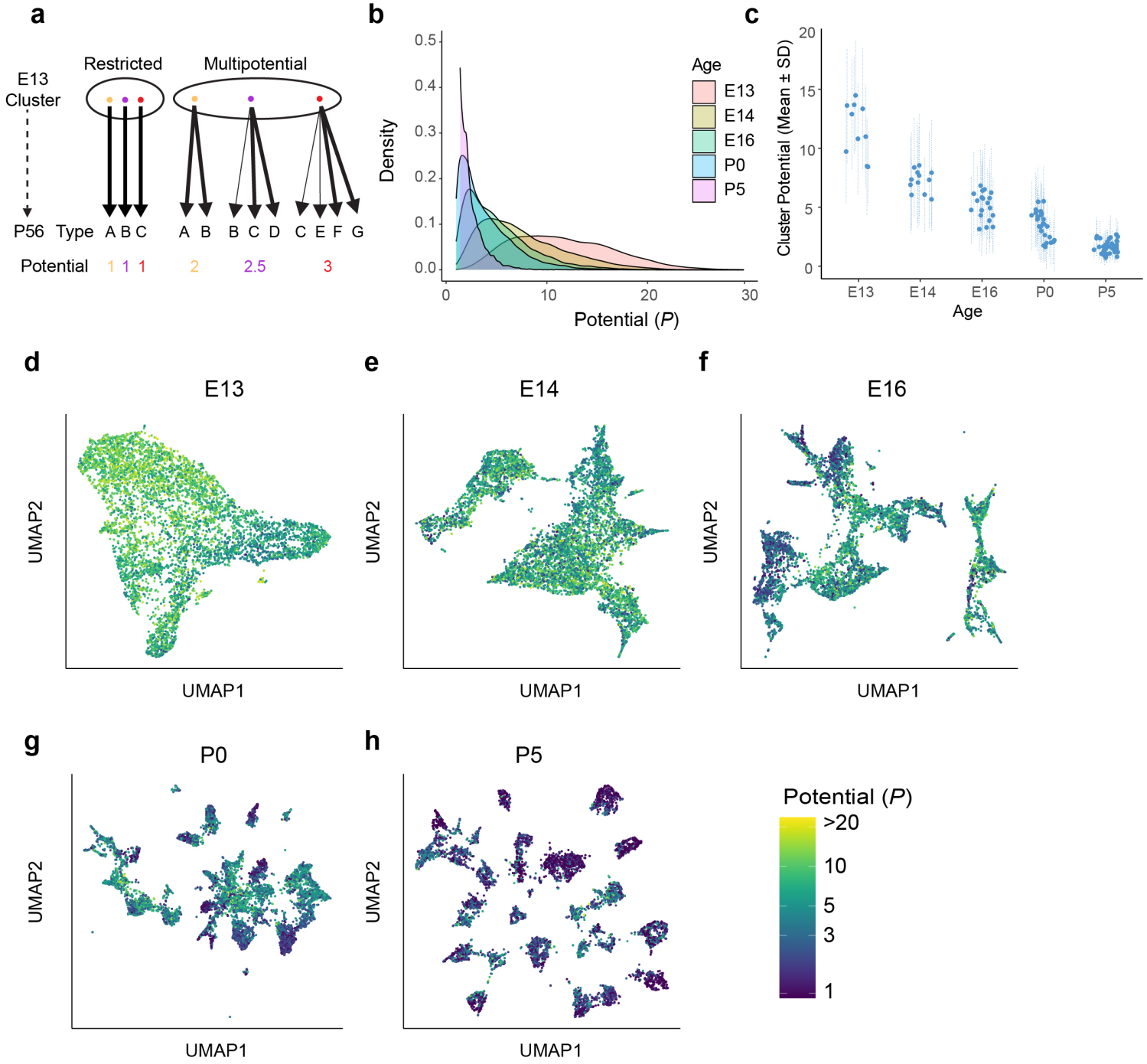


Figure 4. Multipotential fate associations between immature RGCs and terminal types inferred via Optimal Transport

a. Extreme models of diversification at single-cell resolution. Multipotential fate associations in a transcriptionally defined cluster (ellipse) could arise from a mixture of unipotential RGCs (*left*) or from multipotential RGCs (*right*).

b. Distributions of potential *P* across immature RGCs by age showing that restriction increases with age.

c. Inter- and intra-cluster variation of potential by age. At each age, variation in the potential values are shown for each transcriptomically defined cluster at that age. Dots denote the average potential and dotted lines depict the standard deviation for cells within each cluster.

d-h. UMAP projections of E13 (d), E14 (e), E16 (f), P0 (g) and P5 (h) RGCs as in **Fig. 2**, but with individual cells colored by their inferred potential. Potential of all RGCs at P56 =1. The colorbar on the lower right is common to all panels, and values are thresholded at $P = 20$.

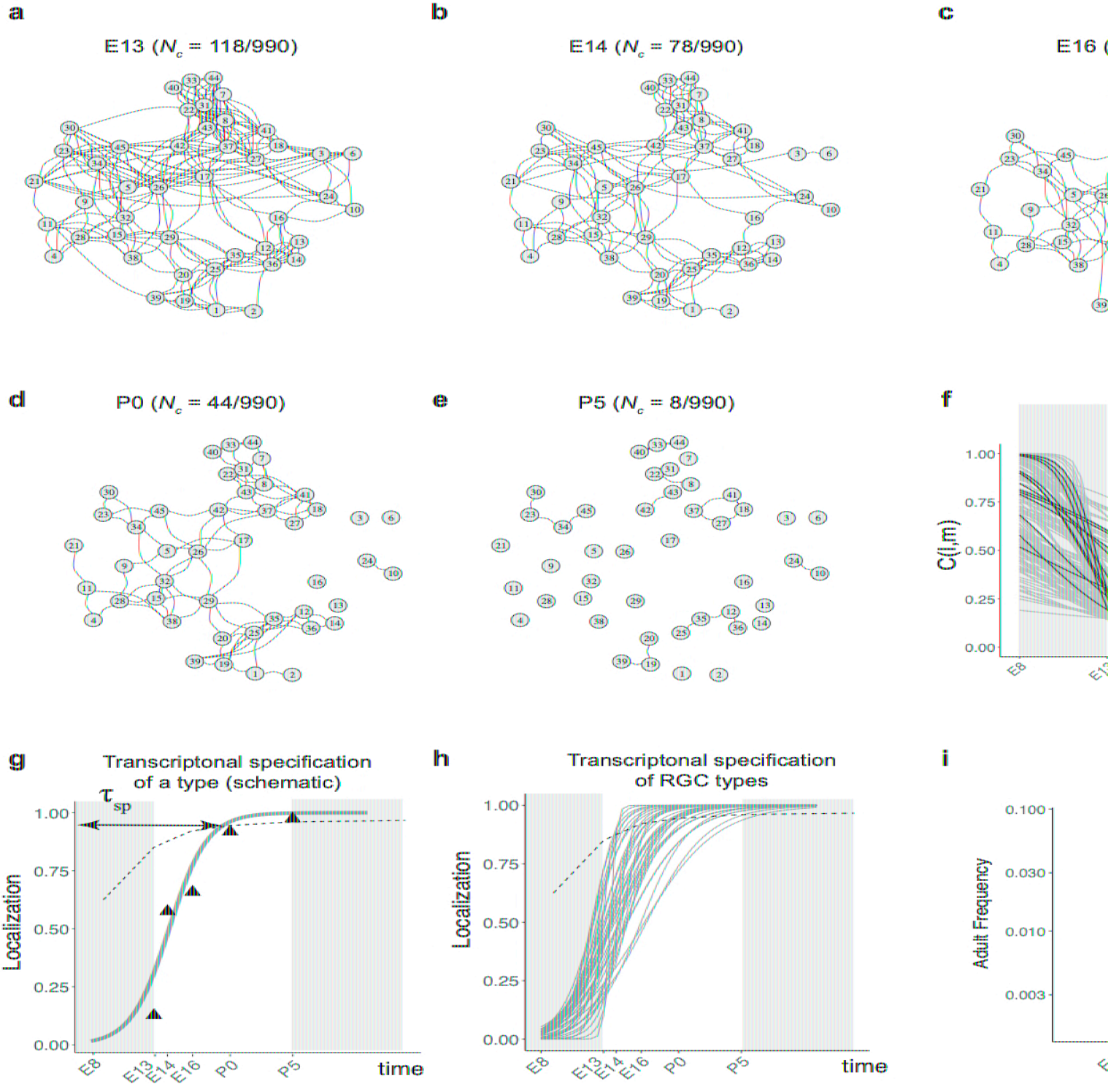


Figure 5. Fate decoupling of RGC types

a. Force-directed layout visualization of fate couplings at E13, with nodes representing RGC types (numbered as in Tran, 2019) and the thickness of edges representing values of *C(l,m;E13)*. Edges with *C(l,m; E13)* < 0.2 are not shown. Number of edges with *C(l,m; E13)* > 0.2 are indicated on top.

b-e. Visualization of fate couplings at E14 (B), E16 (C), P0 (D) and P5 (E). The positions of the nodes are maintained as in panel a, but the edges are redrawn based on values of *C(l,m;age)* at each age. As in panel a, we only show edges *C(l,m; age)* > 0.2.

f. The decay of pairwise fate couplings (y-axis) with age (x-axis). Each line corresponds shows the temporal decay of *C(l,m)* for RGC pair l and m estimated via a logistic model (**Methods**). For each pair, couplings at each age were fit to a model $C(l,m;age) = 1/(1 + e^{\beta_0+\beta_1*age})$ with β_0, β_1 representing fitted parameters. The fitting was performed using data for ages E13, E14, E16, P0 and P5. The shaded portions correspond to the periods E8-E13 and P5- represent extrapolations of the model. Black lines highlight the decay of all non-zero pairwise couplings for RGC type *C8* as an example.

g. Schematic showing logistic modeling to estimate specification time τ_{sp} for a particular type. The y-axis is a measure of the extent to which precursors biased towards the type are present in a single transcriptomically defined cluster (i.e. localization, see **Methods** for details). Localization is defined as a numerical value in the range (0, 1) with higher values consistent with increasing specification. Individual triangles represent the localization values computed using WOT inferred fate couplings at each age, while the curve represents the fit using the logistic model. Dotted line shows the minimum threshold a type to be specified at each age. Its curved shape arises due to the increase in the number of clusters with age.

h. Localization curves (as in panel g) for the 38 RGC types showing the range of inferred specification times. 7 low frequency types have been excluded from display (see **Supplementary Fig. 6d**).

i. Scatter plot showing poor correlation between adult frequency of a type (from ref. 9) and its predicated specification time (calculated from H).

Figure 6

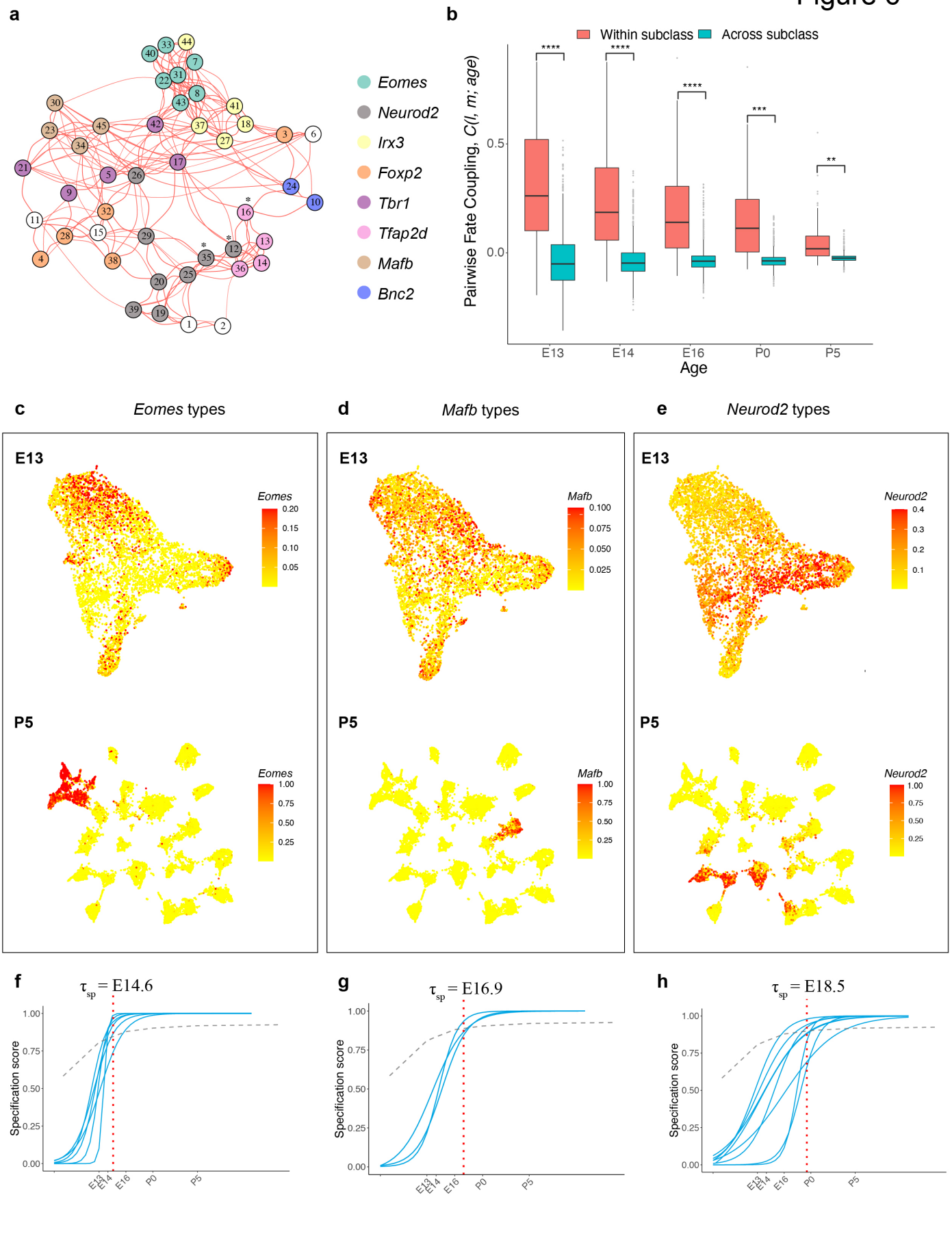


Figure 6. Temporal dynamics of RGC subsets expressing specific TFs

a. E13 network graph of fate couplings from **Fig. 5a**, with RGC types colored based on their selective expression of TFs at P56. Asterisks denote 3/45 types that express more than 1 TF (also see **Supplementary Fig. 7a**).

b. Box-and-whisker plots showing that pairwise fate couplings are higher between types within the same TF subclass than between types in different TF subclasses at all immature ages. Black horizontal line, median; bars, interquartile range; vertical lines, 1st and 99th percentile; dots, outliers. Stars indicate significant p-values based on a two-sided t-test (****, $p < 10^{-7}$; ***, $p < 10^{-5}$; **, $p < 10^{-2}$).

c. *Eomes*+ types. *Top*: UMAP representation of E13 RGCs with cells colored based on their cumulative fate association towards the 7 *Eomes*+ types. *Bottom*: UMAP representation of P5 RGCs with cells colored based on their cumulative fate association towards the 7 *Eomes*+ types. The value corresponding to the color of each cell (colorbar, right) can be interpreted as the probability of commitment towards the corresponding subclass.

d. Same as c for *Mafb*+ types

e. Same as c for *Neurod2*+ types

f-h. Localization curves (as in **Fig. 5g**) for *Eomes*+ types (f), *Mafb*+ types (g) and *Neurod2*+ types (h). The mean inferred specification time τ_{sp} for each group is indicated on top of each panel.

Figure 7

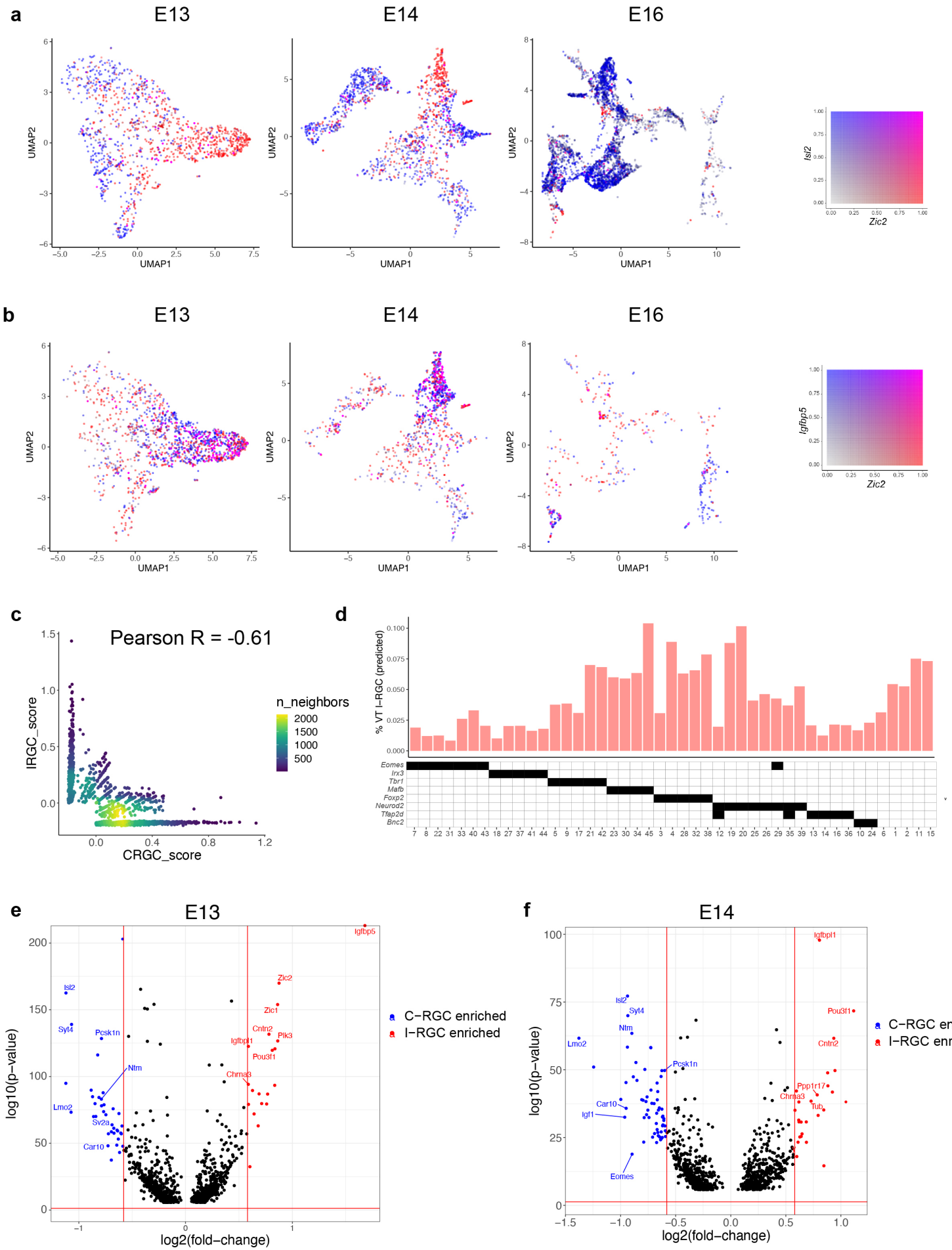


Figure 7. Transcriptomic identification of ipsilaterally projecting RGCs

a. *Zic2*, an I-RGC marker and *Isl2*, a C-RGC marker, are expressed in a mutually exclusive pattern at E13 (left), E14 (middle) and E16 (right). *Zic2* is undetectable after E16 (**Supplementary Fig. 8a**). Cells are colored based on a bivariate color scale representing co-expression of two markers (colorbar, right).

b. *Zic2* and *Igfbp5,* two I-RGC markers, are co-expressed at E13 (left) and E14 (middle). Representation as in panel a.

c. Scatter-plot of gene signatures used to identify I-RGCs (y-axis) and C-RGCs (x-axis) at E13 are negatively correlated (Pearson R = -0.61). Each dot corresponds to a cell, the color represents the number of cells located at a particular (x,y) location (see colorbar, right).

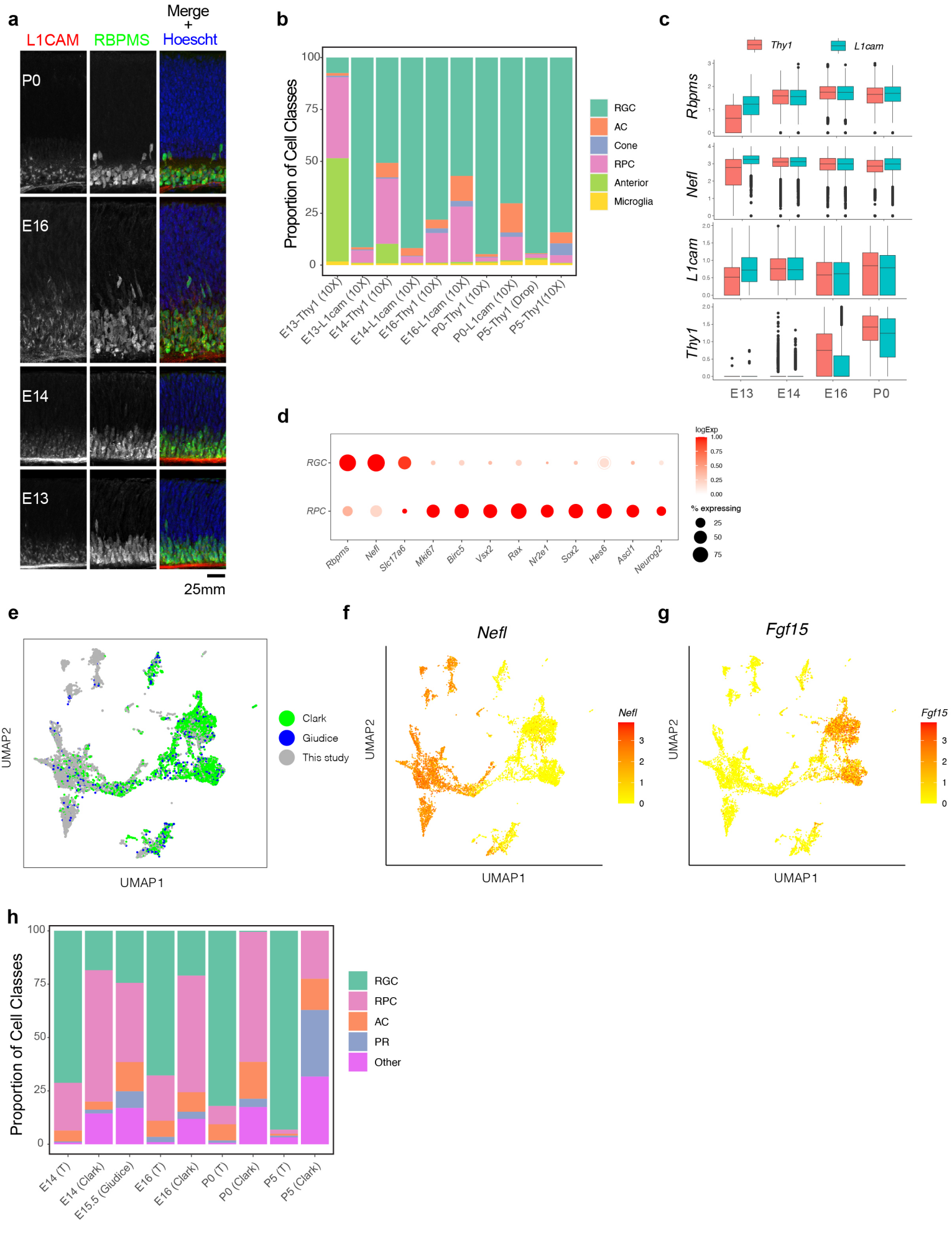
d. Barplot showing % of putative I-RGCs (y-axis) within each of the 45 adult RGC types, estimated by computing the descendants of E13 I-RGCs using WOT. RGC types are arranged along the x-axis based on their membership of TF-groups shown in **Fig. 6a** (annotation matrix, bottom).

e. Volcano plot showing DE genes (MAST test, $p < 10^{-6}$) between predicted I-RGCs and C-RGCs at E13. The x- and the y-axes show the fold-change and the p-value in log2- and log10- units, respectively. Dots represent genes, with red and blue dots highlighting I- and C-RGC enriched genes respectively at fold-change > 1.5 and Bonferroni corrected p-value < $5x10^{-5}$. The two vertical bars correspond to a fold-change of 1.5 in either direction. Select I-RGC and C-RGC enriched genes are labeled.

f. Same as panel e, for E14.

SUPPLEMENTARY FIGURES AND LEGENDS

Supplementary Fig. 1



Supplementary Figure 1. Separation of major transcriptomic groups and assessment of immature RGCs at E13-P5.

a. Retinal sections stained with *L1CAM* (red), which was used to enrich for RGCs at early stages, and the pan-RGC marker *RBPMS,* (green) at E13, E14, E16 and P0. Nuclei are counterstained by the Hoeschst dye (blue).

b. Relative proportions (y-axis) of major cell classes shown in **Fig. 1c-e** at each combination of age and enrichment method. Both anti-Thy1 and anti-L1cam were used to enrich RGCs at E13, E14, E16 and P0, but only anti-Thy1 was used at P5, because L1cam becomes localized to axons postnatally. AC, Amacrine Cells; RPC, retinal progenitor cells.

c. Box and whisker plots show gene expression levels of key markers by RGCs as a function of age and enrichment method. Markers shown shown are two pan-RGC markers, *Rbpms, Nefl*, and the two cell-surface proteins used for enrichment, *Thy1* and *L1cam*. Note that *Thy1* expression is poor at E13, consistent with low RGC yield in anti-Thy1 enriched cells (panel B). Black horizontal line, median; bars, interquartile range; vertical lines, range; dots, outliers.

d. Dotplot showing genes (columns) that are selectively expressed in RGCs and RPCs. The size of each circle is proportional to the percentage of cells expressing the gene, and the color depicts the average log-normalized expression.

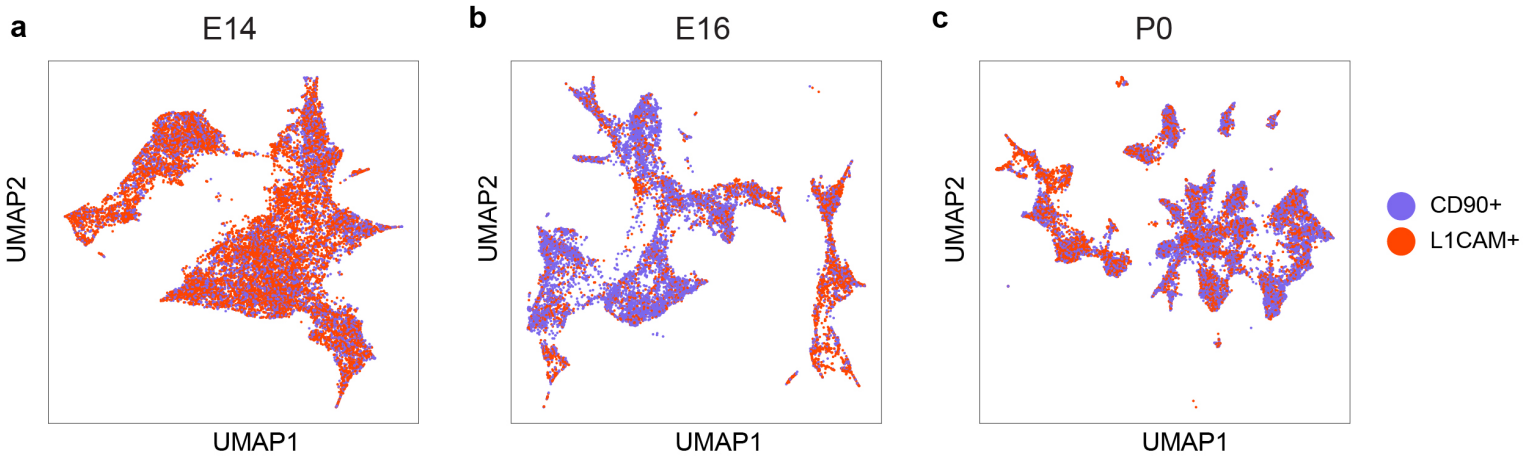
e. Co-embedding analysis of E14, E16 and P0 data collected in this study with whole retina single-cell transcriptomes in independent studies: E14, E16 and P0 data from[31] and E15.5 data from[43]. Cells (points) are visualized in UMAP and colored by study of origin.

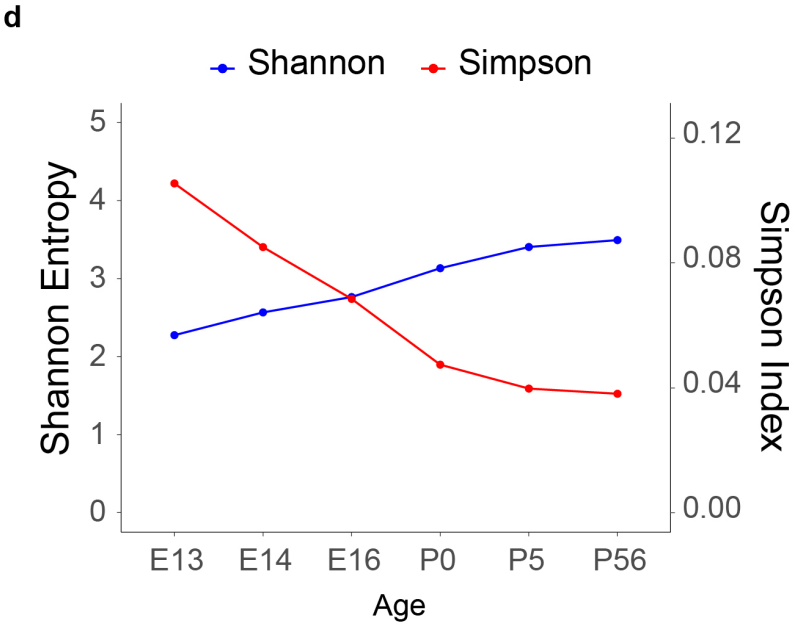
f. Same as e, with cells colored by the expression level of *Nefl,* an RGC marker. This shows the higher enrichment of RGCs in our study compared to[31] and[43].

g. Same as d, with cells colored by expression level of *Fgf15*, an RPC marker.

h. Relative proportions of major cell classes across different datasets analyzed in panel e separated by age. T, this study.

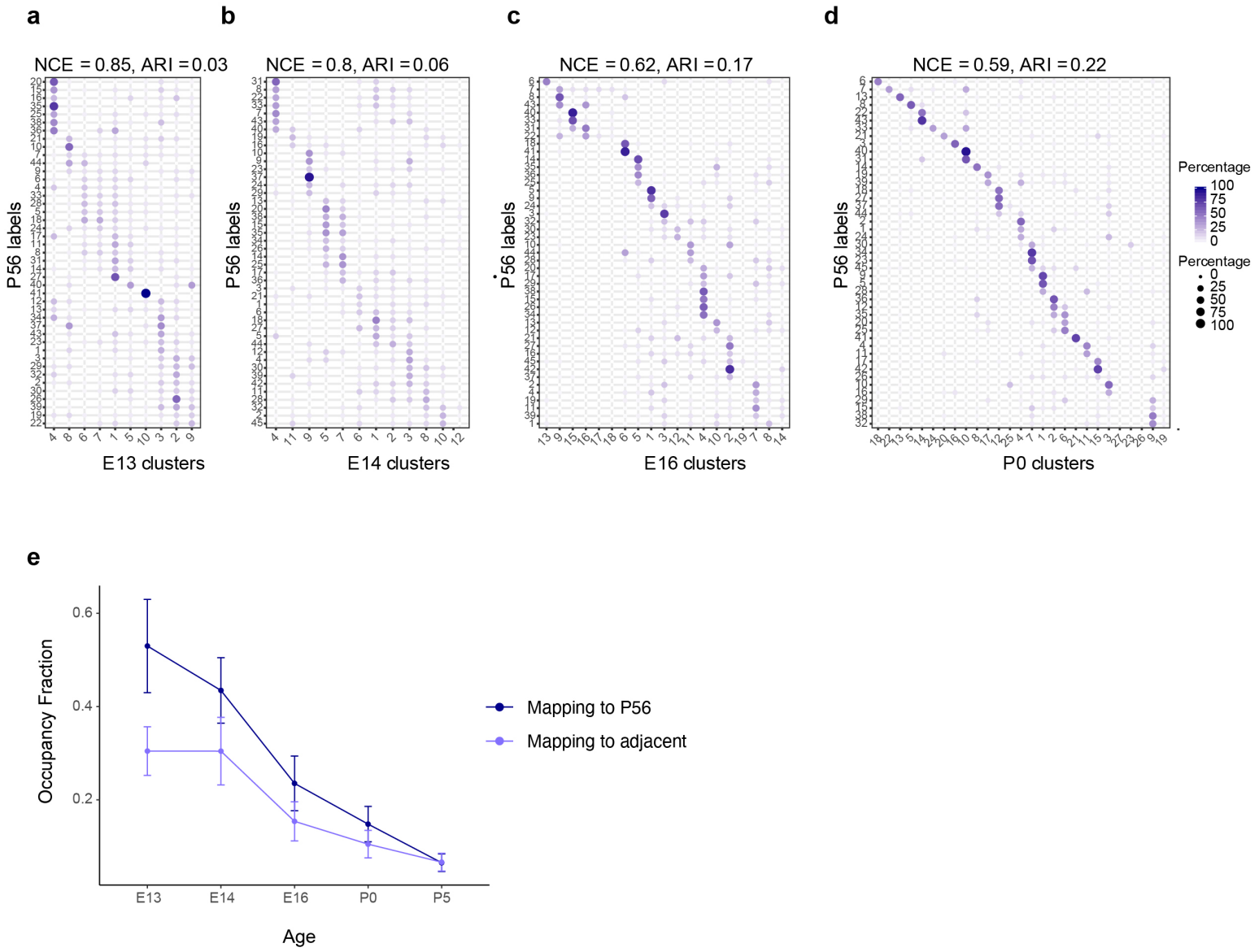
Supplementary Fig. 2





Supplementary Figure 2. Transcriptomic diversity of immature RGCs by age.
a-c. UMAP embedding for RGCs at E14 (a, same as **Fig. 2c)**, E16 (b, same as **Fig. 2d**) and P0 (c, same as **Fig. 2e**) with cells colored by enrichment method showing comparable transcriptomic diversity of immature RGCs enriched by L1cam or Thy1.
d. Simpson and Shannon diversity indices (see **Methods**) associated with clustering decrease and increase with age respectively, consistent with increasing transcriptomic diversity.

Supplementary Fig. 3



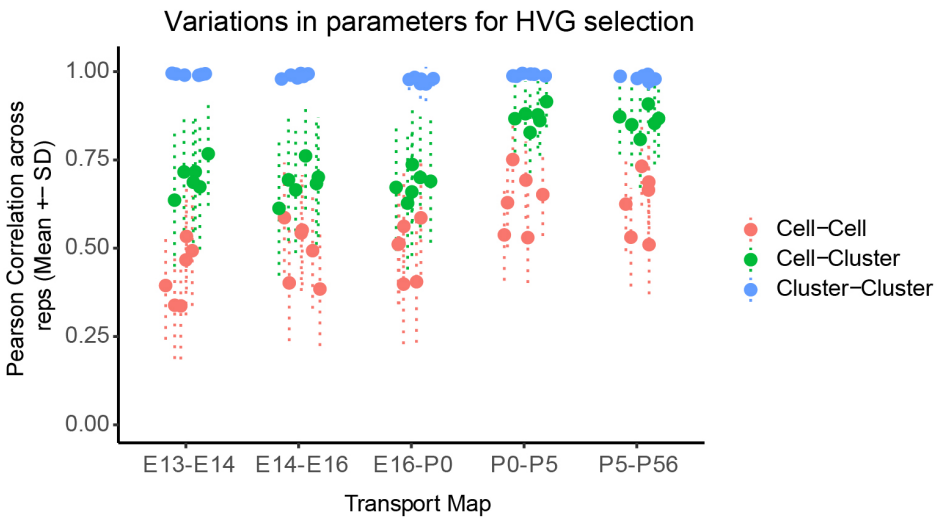
Supplementary Figure 3. Temporal correspondences between transcriptomic clusters evaluated using supervised classification.

a-d. Confusion matrices, showing transcriptomic correspondence between adult RGC types at P56 (rows) and immature RGC clusters (columns) at ages E13 (a), E14 (b), E16 (c) and P0 (d). In each case, immature RGCs were assigned adult labels using an Xgboost classifier trained on adult RGCs. The P56 to P5 mapping is shown in **Fig. 2h**.

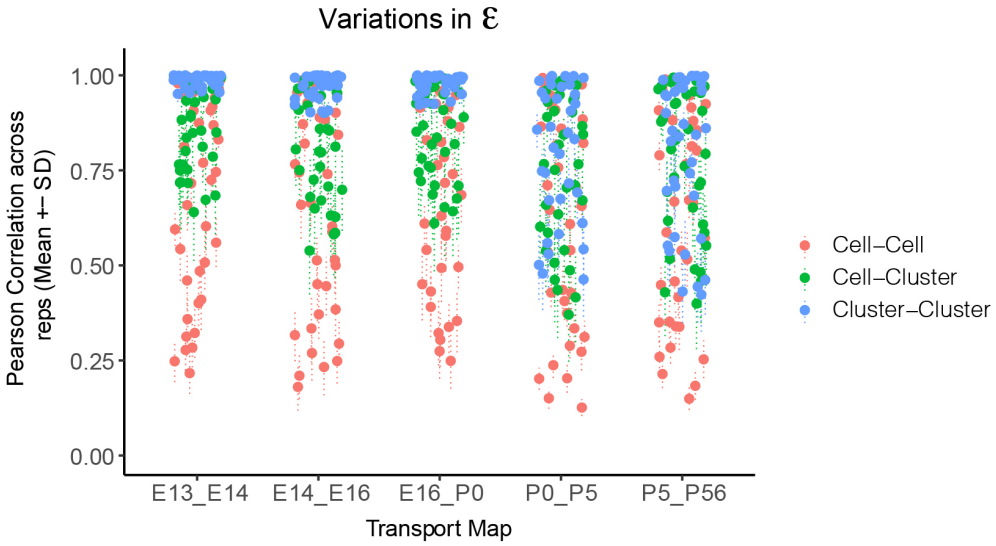
e. Line plots showing occupancy fraction (OF) of mapping of an early cluster to cluster IDs at later ages. OF values quantifies the specificity of mapping of an early cluster to late clusters, with lower values denoting higher specificity. The average occupancy fraction across clusters decreases steadily with age consistent with the decrease and increase in NCE and ARI respectively (**Figs. 3b,c**). Error bars indicate standard deviation computed across clusters. Also, as expected, the occupancy fraction values are lower for mapping to adjacent time points than to P56.

Supplementary Fig. 4

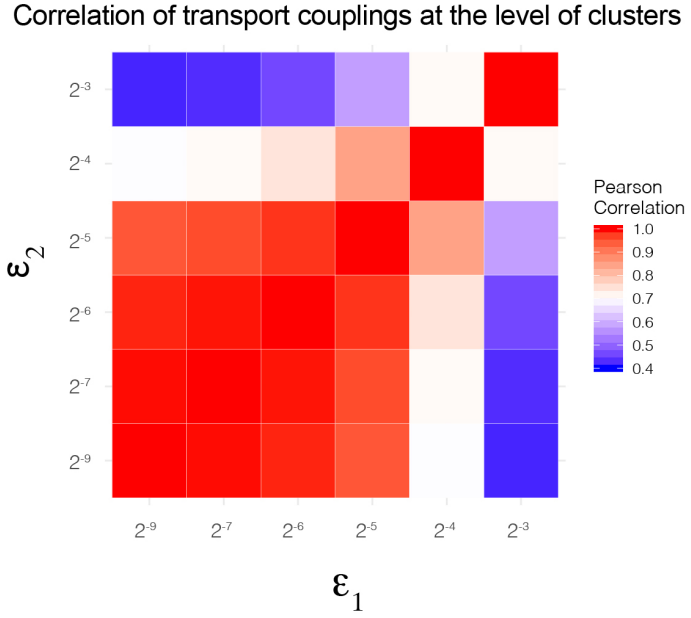
a



b



c



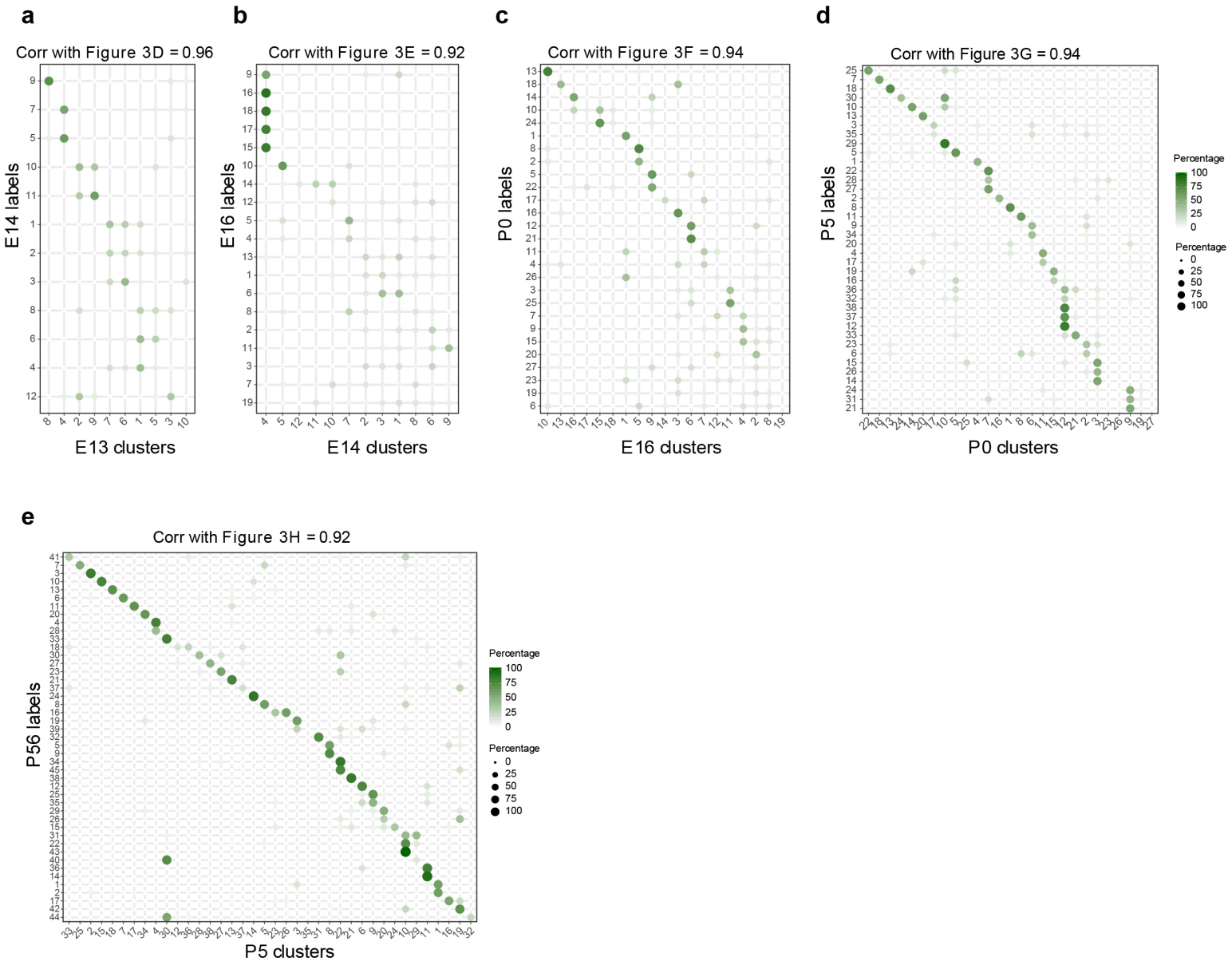
Supplementary Figure 4. Variations in WOT inferred temporal couplings and tests across variations in hyperparameters.

a. Variations in WOT inferred temporal couplings (Π_{ij}) at the level of cells and clusters to changes in the set of highly variable genes (HVGs) used for computing transport maps. Four sets of features were tested corresponding to the top 800, 1100, 1400 and 1800 HVGs based on our previously described Poisson-Gamma model (Pandey et al., 2018). Using these sets, we inferred four corresponding transport maps at each of the 5 age pairs E13-E14, E14-E16, E16-P0, P0-P5 and P5-P56. The entropic regularization hyperparameter ϵ (see panels b, c) was held constant at a value 2^{-7} in these tests. At each age pair, we computed the Pearson correlation coefficient (PCC) between estimated temporal couplings for every older cell (column of the transport map Π) across each pairwise combination of the four transport maps, towards a total of 6 combinations. These are indicated as red dots and lines (mean $\pm$ SD). We then grouped (summed) the rows of the transport map by transcriptomic cluster at the younger age, such that each element of the new matrix indicates cell (column)-cluster (row) couplings. The PCC of these couplings were computed for each older cell (column) within each pairwise combination of the four transport maps and are indicated as green dots and lines (mean $\pm$ SD). Finally, we grouped (summed) both the rows and columns of the transport map by transcriptomic cluster at either age to obtain a matrix of cluster-cluster couplings. The PCC values of these couplings were computed for each older cluster within each pairwise combination of the four transport maps and are indicated as blue dots and lines (mean $\pm$ SD). We find that the cell-cell couplings increase in robustness at later ages, but the cell-cluster and cluster-cluster couplings are quite robust (correlation > 0.6).

b. Variations in WOT inferred temporal couplings at the level of cells and clusters as in panel A, but to changes in the entropic regularization ϵ. Six values were used - (2^{-8} , 2^{-7} , 2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}) with increasing values corresponding to more transport maps with decreasingly localized (or increasingly distributed) couplings. At each age pair, 6 transport maps are computed and PCC values for cell-cell, cell-cluster and cluster-cluster couplings are computed as in panel A for each of 15 transport map pairs. Here too, the cluster-cluster and cell-cluster couplings show higher stability, although at later stages higher values of ϵ exhibit loss of stability even at the cluster-cluster level (see panel c).

c. Heatmap showing cluster-cluster PCC values for P5-P56 transport maps inferred using different values of the entropic regularization parameter, epsilon (rows and columns). Loss of stability occurs at higher values of the entropic regularization, consistent with panel B. Based on this we used epsilon = 2^{-7} to calculate results shown in **Fig. 4**.

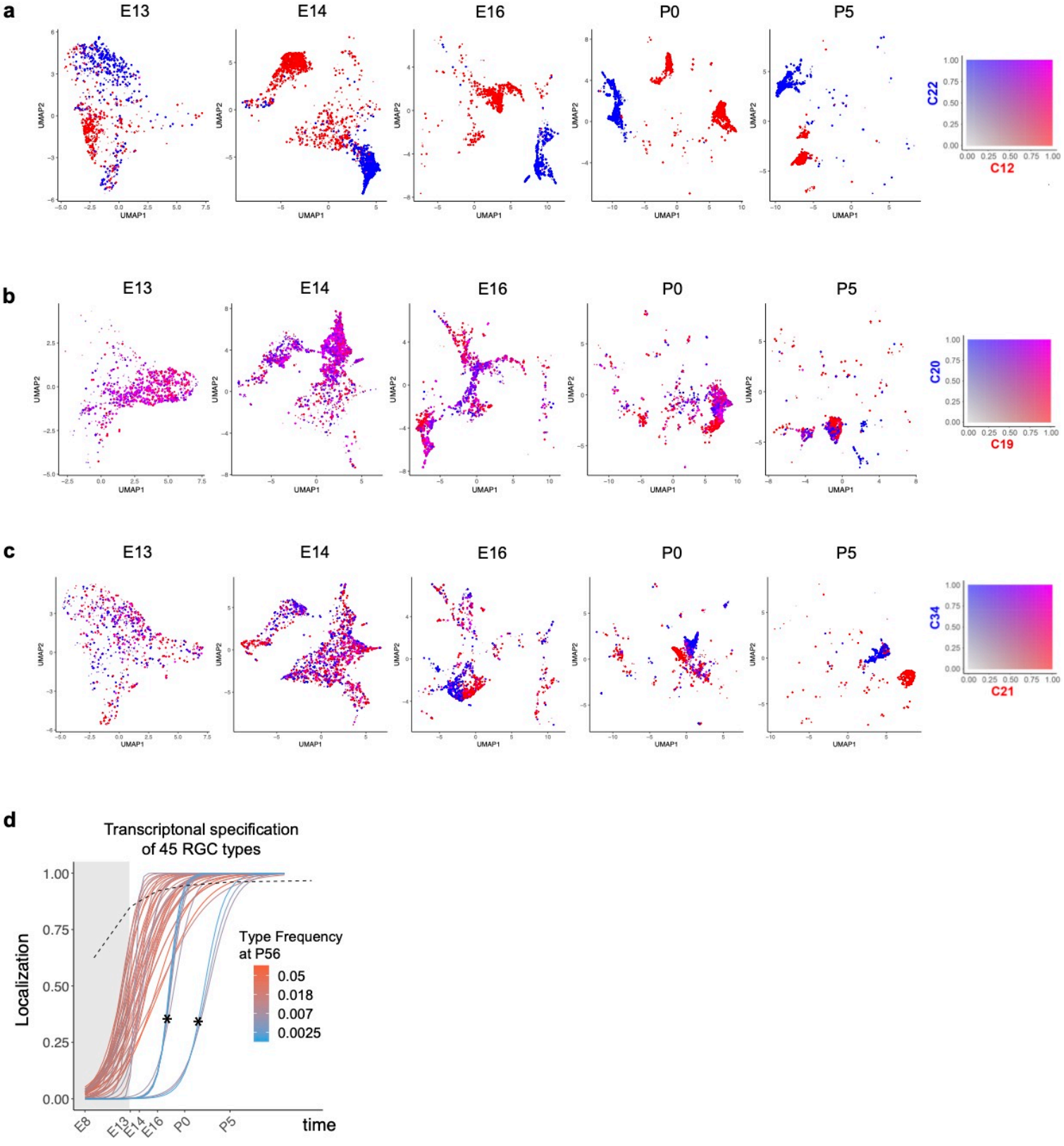
Supplementary Fig. 5



Supplementary Figure 5. Temporal correspondences between transcriptomic clusters evaluated using Waddington Optimal Transport, related to Figure 4

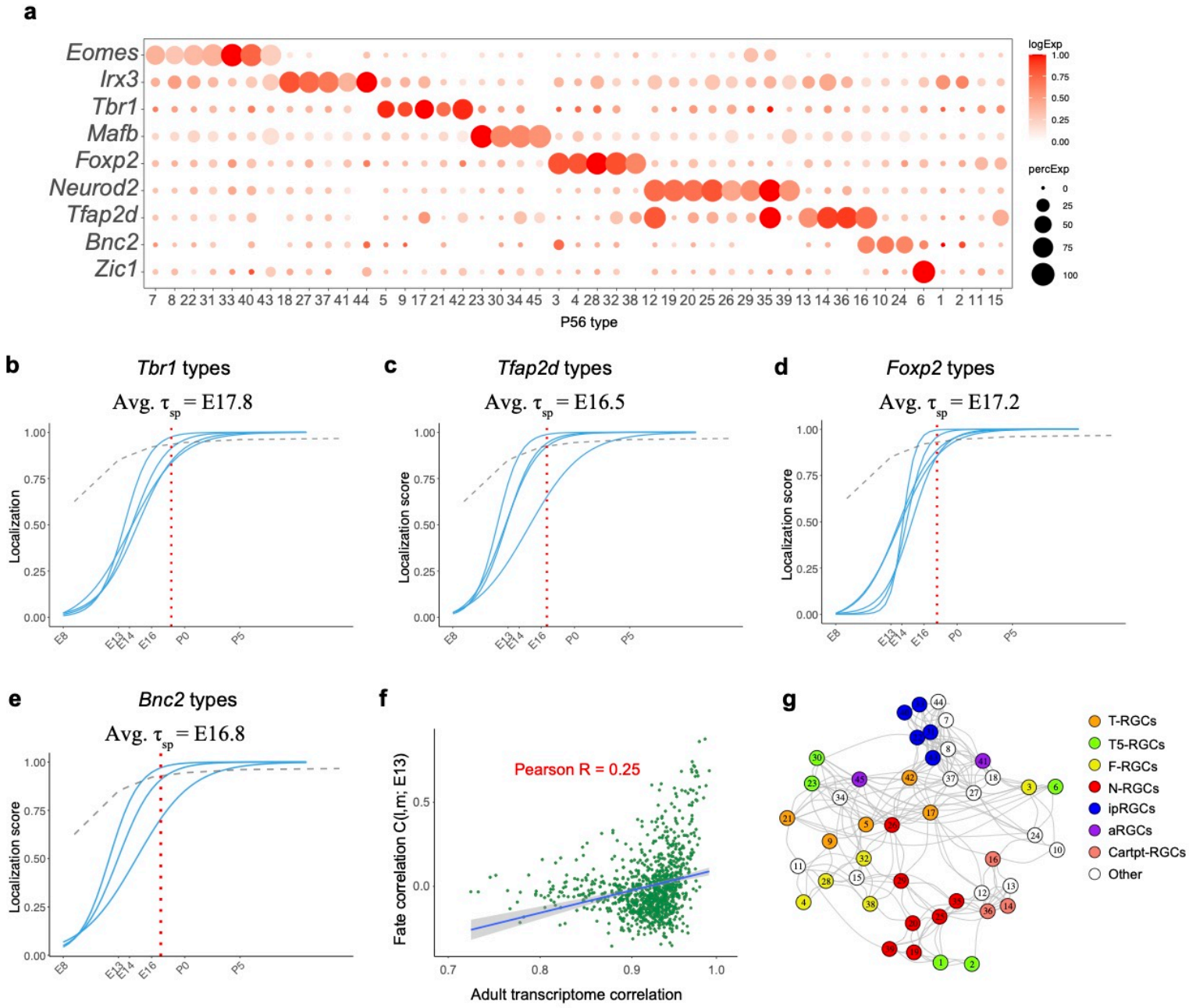
a-e. Average temporal couplings at the level of clusters. Panels correspond to the pairs E14-E13 (a), E16-E14 (b), P0-E16 (c), P5-P0 (d), P56-P5 (e), respectively. In each case, the WOT inferred transport map was grouped along rows and columns based on transcriptomic cluster, and the elements were summed within each group. The resulting matrix was normalized such that each row sums to 100%. These matrices strongly resemble those in **Figs. 3d-h**, as confirmed by the high values of the Pearson correlation coefficient (top, all $p \geq 0.92$).

Supplementary Figure 6



Supplementary Figure 6. Fate decoupling and temporal specification of RGC types.
a-c. Examples of fate decoupling. Panels from left to right correspond to ages E13-P5 with precursor RGCs shown on a reduced dimensional UMAP representation as in **Figs. 2c-g**. Each RGC is colored using a biaxial color scale (legend) based on its predicted fate values. C12 and C22 are transcriptomically distinct as early as E13 (a). C19 and C20 exhibit high fate coupling at all embryonic ages and are only decoupled at P5 (b). C21 and C34 decouple around E16 (c).
d. Same as **Fig. 6h** showing specification curves for RGC types, but in this case each curve is colored based on adult frequency (colorbar, right). The 7 curves marked by asterisks correspond to late-specified types. As can be seen from their colors, they are also among the types with the lowest frequency (<0.3%), which may result in the dropout of the corresponding precursors because of sampling fluctuations. Such dropouts at earlier time points give the appearance of late specification. Because of this issue, we exclude them from our analysis.

Supplementary Figure 7



Supplementary Figure 7. Transcription factor (TF) based subgroups.

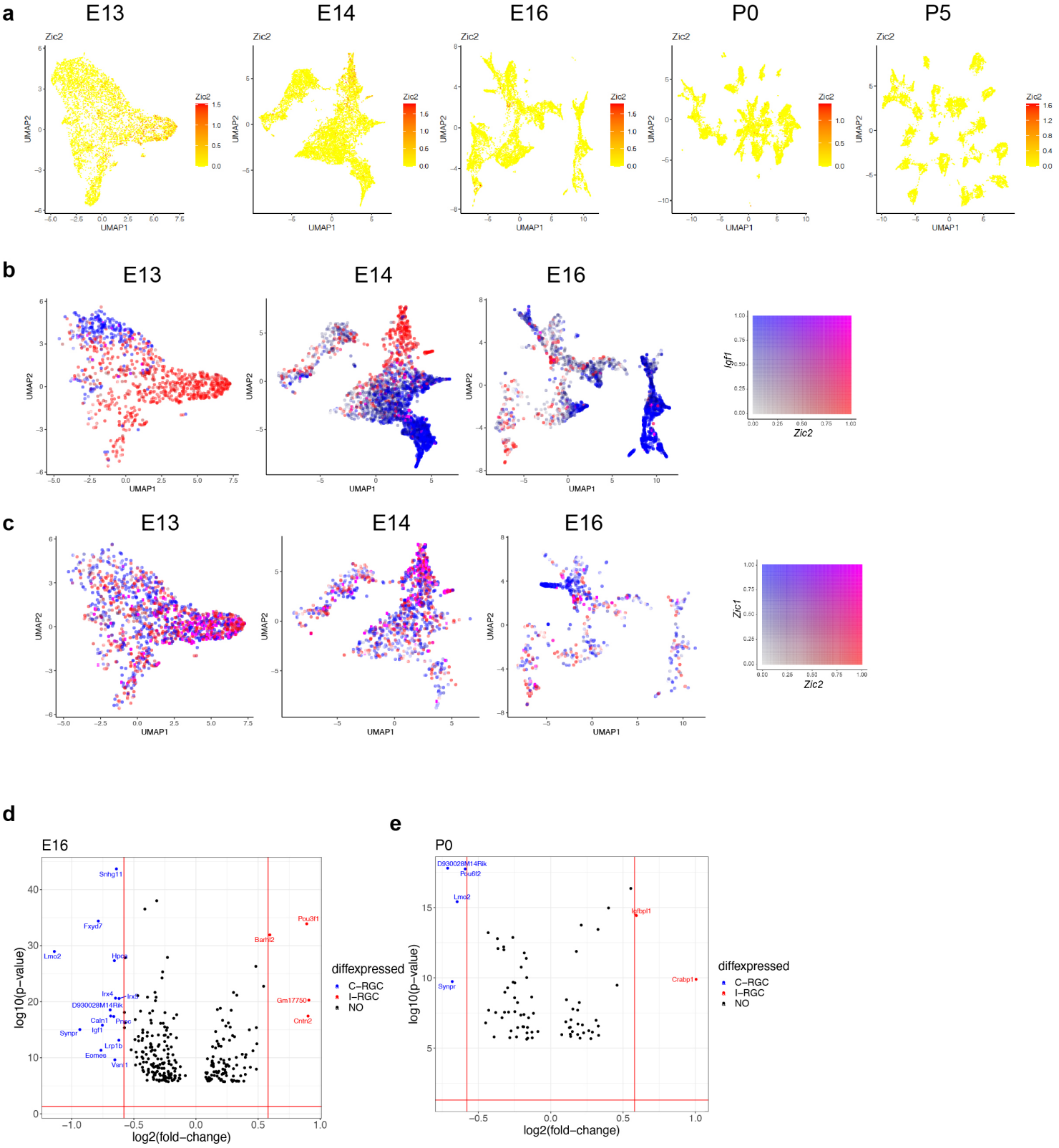
a. Dotplot showing mutually exclusive patterns of expression of TFs that mark groups of RGC types. In addition to the 8 TFs shown in **Fig. 6a**, we highlight *Zic1*, which selectively labels C6. Selectively expressed TFs could not be identified for 4 types (C1, C2, C11, and C15).

b-e. Localization curves (as in **Fig. 5g**) for *Tbr1*+ types (b), *Tfap2d*+ types (c), *Foxp2*+ types (d) and *Bnc2*+ types (e), Note that the "low frequency" types labeled in **Supplementary Fig. 6d** are not shown. The mean specification time τ_{sp} for each group is shown above the graphs.

f. Correlation of fate coupling at E13 with transcriptomic correlation at P56.

g. Same as **Fig. 6a**, with nodes corresponding to subclasses defined in Tran et al., 2019, which includes ipRGCs, alpha-RGCs and T5-RGCs.

Supplementary Figure 8



Supplementary Figure 8. Transcriptomic analysis of RGC laterality.

a. RGCs at E13, E14, E16, P0, and P5 colored by their expression of *Zic2*, an I-RGC marker. *Zic2* is expressed in a localized fashion until E16 and becomes undetectable beyond this age.

b. *Zic2*, an I-RGC marker and *Igf1*, a C-RGC marker, are expressed in mutually exclusive patterns at E13-E16.

c. *Zic2* and *Zic1,* two I-RGC markers, are co-expressed in subsets of RGCs at E13-E16.

d. Volcano plot showing DE genes (MAST test, $p < 10^{-6}$) between predicted I-RGCs and C-RGCs at E16. The x- and the y-axes show the fold-change and the p-value in log2- and log10- units, respectively. Dots represent genes, with red and dots highlighting I- and C-RGC enriched genes respectively at fold-change > 1.5 and Bonferroni corrected p-value < 5e-5. The two vertical bars correspond to a fold-change of 1.5 in either direction.

e. Same as d, for P0