Temporal regulation of green and red cone specification in human retinas and retinal organoids

Sarah E. Hadyniak¹, Kiara C. Eldred^{1,°}, Boris Brennerman¹, Katarzyna A. Hussey¹, Rajiv C. McCoy¹, Michael E. G. Sauria¹, James A. Kuchenbecker², Maureen Neitz², Jay Neitz², James Taylor^{1,†}, and Robert J. Johnston, Jr.^{1,*}

¹ Department of Biology, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, United States

² Department of Ophthalmology, University of Washington, Seattle, WA, 98105, United States

[°] Current address: Department of Biological Structure, University of Washington, Seattle, WA, 98105, United States

[†] Deceased

*Corresponding author. Email: robertjohnston@jhu.edu

One sentence summary: Temporal specification of human cones

Key words: human retina, retinal organoid, photoreceptor, red cones, green cones, temporal, stochastic, opsin, retinoic acid, NR2F2

Abstract

Trichromacy is unique to primates among mammals, enabled by specification of blue (short/S), green (medium/M), and red (long/L) cones. During retinal development in humans and Old World monkeys, cones make a poorly understood choice between M and L cone subtype fates. Two nonexclusive models have been proposed: a stochastic model, in which a regulatory DNA element randomly loops to either the *M-* or *L-opsin* promoter to drive expression, and a temporal model, in which cone subtypes are generated in a developmental progression. Here we present data that support a temporal mechanism for M and L cone specification. *M-opsin* is expressed prior to *L-opsin* during fetal retinal development. Natural variation in the ratios of M and L cone subtypes is associated with a polymorphism in the *NR2F2* gene, a mediator of retinoic acid (RA) signaling. RA-synthesizing enzymes are highly expressed in early development and then decrease, suggesting high RA signaling early. Human retinal organoids grown in standard culture conditions were enriched for L cones, whereas early addition of RA yielded M cone-rich organoids. Our data suggest that M cones are generated before L cones and that RA signaling induces M cones and suppresses L cones early during development. These studies advance human retinal organoids as a model to study developmental processes unique to humans and primates.

Introduction

Trichromacy in humans is enabled by three subtypes of cone photoreceptors that express opsin photopigments sensitive to short (S), medium (M), or long (L) wavelengths of light¹. Among mammals, only humans, Old World primates, and some New World monkeys have retinas with both M and L cones²⁻⁸. The only known difference between M and L cones is the expression of their respective opsin photopigment⁹⁻¹¹. The *OPN1LW* (*L-opsin*) and *OPN1MW* (*M-opsin*) genes lie in an array on the X chromosome under the control of a shared regulatory DNA element called a locus control region (LCR)¹²⁻¹⁴. Though the number and ordering of *OPN1LW* and *OPN1MW* genes can vary, the LCR regulatory element is most commonly found upstream of *OPN1LW* followed by two copies of *OPN1MW* (**Fig. 1A**)¹⁵. Expression analyses suggested that, no matter the arrangement, only the first two opsin genes in the array are expressed in the human eye^{9,12,16-19}. M and L cones have been technically challenging to distinguish due to the high sequence and structural similarity between the M- and L-opsins, and thus, the mechanisms underlying this fate decision have remained largely mysterious. Here, we examined the timing of *M*- and *L-opsin* mRNA expression during development, associated differences in M and L cone ratios with naturally occurring genomic polymorphisms in human populations, and tested mechanisms underlying specification of M and L cones in human retinal organoids to understand the developing human retina.

Two nonexclusive mechanisms have been proposed to control M and L cone generation: a stochastic model and a temporal model. In the stochastic model, the LCR randomly loops to either the *M-opsin* promoter or the *L-opsin* promoter, stably driving exclusive expression of that opsin^{1,13,20}. The stochastic model is based on transgene

reporter experiments in mice, which showed that the probability of expression from the *M*- or *L-opsin* promoters was dependent on the order and distance between the LCR and opsin promoters¹³. In the temporal model, M cones are generated before L cones. Functional microspectrophotometry, multifocal-ERG imaging, and expression studies suggested that, compared to the early-developing central retina, the late-developing peripheral retina has an increased proportion of L cones^{11,21-33}. In this study, we examined human retinas and retinal organoids during development to assess these mechanisms.

Organoids provide a powerful system to investigate the molecular mechanisms controlling fate specification during human development^{34,35}. We previously showed that human retinal organoids generate S cones before M and L cones with similar gene expression patterns and cell morphologies. We experimentally manipulated organoids to identify a critical role for thyroid hormone signaling in this step of cone subtype specification³⁶. We thus sought to probe the organoid system to understand the primate-specific fate decision between M and L cones.

We found that *M-opsin* is expressed earlier than *L-opsin* in human retinas. We studied natural variation in M and L cone ratios in the human population and identified an association with a variant in *NR2F2*, a mediator of retinoic acid (RA) signaling. We analyzed a fetal RNA-seq dataset to assess expression of RA regulatory genes and found high expression of RA-synthesizing enzymes early followed by a significant decrease, suggesting high RA signaling early and low RA signaling late during development. To distinguish between the highly similar *M*- and *L-opsin* mRNA transcripts, we developed an *in situ* hybridization approach and examined expression in human retinal organoids. Organoids grown in standard culture conditions were enriched for L cones, whereas

organoids supplemented with RA early in development displayed a high proportion of M cones. Our data suggest that M and L cones are specified in a temporal progression and that early RA signaling induces M cones and suppresses L cones in the developing human retina.

Results

***M*-opsin is expressed before *L*-opsin during human fetal retinal development**

To determine the developmental dynamics of cone subtype generation, we examined the timing of *M*- and *L*-opsin expression. In a stochastic model, we predicted that *M*- and *L*-opsin should be expressed simultaneously throughout development. In a temporal model, one opsin should be expressed before the other opsin.

Human *M*- and *L*-opsin mRNA transcripts exhibit ~98% sequence identity^{1,37,38}, which complicates expression analysis. We quantified *M*- and *L*-opsin expression in RNA-seq datasets by analyzing pileup counts, quantifying the total number of reads supporting *M*- versus *L*-opsin at 20 divergent nucleotides (**Fig. 1B**), and normalizing by the total number of reads aligned to these divergent sites.

We first tested this computational approach with the Weri-Rb-1 retinoblastoma cell line, which expresses *M*- and *L*-opsins at low levels³⁹. T3, the active form of thyroid hormone, induces *M*- and *L*-opsin expression in Weri-Rb-1 cells^{36,40}. We conducted RNA-seq and observed an increase in both *M*- and *L*-opsin in T3-treated Weri-Rb-1 cells compared to untreated cells (**Fig. S1A-B**), validating our approach to quantify *M*- and *L*-opsin expression.

We assessed the timing of *M*- and *L*-opsin expression during human retinal development by analyzing published RNA-seq data⁴¹. The recent availability of these data and the challenges in attaining fetal retinal samples likely prevented this type of analysis previously. We assessed expression in 12 samples from prenatal day 52/54 through day 136 postconception. *M*-opsin was strongly expressed from day 115 through day 136 (**Fig. 1C**). In contrast, minimal *L*-opsin expression was observed through day 136 (**Fig. 1C**). We analyzed three independent adult human retinal samples from a published dataset⁴² and observed expression of *M*- and *L*-opsin (**Fig. 1D**). These data suggest that *M*-opsin is expressed before *L*-opsin during human retinal development.

Natural variation in L:M cone ratios implicates retinoic acid signaling

We next sought to identify a regulatory mechanism controlling specification of M and L cones. L:M cone ratios vary widely among people with normal color vision, ranging from 6% to 94% L cones^{21,25,43-46}. We hypothesized that DNA sequence differences in the *L/M*-opsin gene locus or putative regulatory genes influence variation in L:M cone ratios^{47,48}. We assessed L:M cone ratios from 738 males with normal color vision using the flicker-photometric electroretinogram (FP-ERG). We focused our study on males because variation at the X-linked *L/M*-opsin gene locus is a primary site for mutations causing red-green color blindness, and an increase in variation could be revealed in hemizygous conditions⁴⁹⁻⁵². In this population, we observed ratios ranging from 15% to 110% L cones (**Fig. 2A**). The association between FP-ERG and % L is nonlinear and as 100% is approached, very small changes in measured spectral sensitivities are associated with relatively large changes in estimates of % L. Thus, near 100%, small

experimental errors led to some values that exceeded 100% upon normalization.

To identify the genetic basis of this variation, we performed an association study using the L:M cone ratio data and the targeted DNA sequencing of 21 gene regions, including the *L/M-opsin* gene locus (*OPN1LW/MW*), the *S-opsin* gene (*OPN1SW*), and regulatory genes with putative roles in cone subtype specification including *DIO1*, *PIAS3*, *RXRG*, *OTX1*, *NEUROD1*, *THRB*, *NR2F1/COUP-TFI*, *RORB*, *CYP26A1/C1*, *NRL*, *OTX2*, *DIO2*, *DIO3*, *ONECUT1*, *NR2E3*, *NR2F2/COUP-TFII*, *RARA*, *ONECUT2*, and *SALL3* (**Fig. S2**).

For each of the 21 gene regions, we isolated DNA for each gene of interest by targeting sequence intervals delineated by the end coordinate of the neighboring upstream gene to the start coordinate of the neighboring downstream gene using an oligo-based pull-down method (**Fig. S2**). If the neighboring upstream or downstream genes were less than 100 kb from the gene of interest, we expanded the pulled down region to include at least 100 kb of upstream and downstream sequence (**Fig. S2**). We generated custom-designed oligos and used a solution-based capture method for target enrichment based on the Roche Nimble Design software. We barcoded and pooled samples, which we sequenced with 125 bp paired-end reads. We mapped the reads, genotyped the samples, and evaluated associations with L:M ratios (**Fig. 2C**).

From this approach, we identified a significant association between SNP rs372754794 ($\beta = -18.54$, p-value = 1.67×10^{-5} , $R^2 = 0.0569$) and variation in L:M cone ratio (**Fig. 2B-C**). This SNP, which occurs within the targeted sequence spanning *NR2F2/COUP-TFII*, lies in a non-coding RNA immediately upstream of *NR2F2* (**Fig. 2D**). The minor (G) allele, which is associated with reduced L:M ratio, was nearly exclusive to

African American (and few African) subjects within our sample (MAF = 0.0359; **Fig. S3A**) and exhibited similar patterns of frequency differentiation in external data from the 1000 Genomes Project. We note that the association with L:M ratio was robust even when restricting analysis to African American subjects ($\hat{\beta} = -14.45$, p-value = 1.44×10^{-4} , $R^2 = 0.1052$), arguing against ancestry as a confounding factor (**Fig. S3B**). Within African American populations from the 1000 Genomes Project⁵³, rs372754794 exhibited only modest linkage disequilibrium (LD) with other nearby variants (**Fig. S3C**), such that the associated haplotype is confined to the *NR2F2-AS1* transcript.

NR2F2 is a nuclear receptor that mediates RA signaling⁵⁴. While studying bilateral symmetry in the vertebrate body plan, Vilhais-Neto and colleagues showed that Nr2f2 binding was enriched at a retinoic acid response element (RARE) and that Nr2f2 knockdown led to a reduction in RARE activity. Nr2f2 was also found at the *Rarb* promoter, a well-characterized RA target gene, and its recruitment increased in the presence of RA. Nr2f2 forms a complex with a retinoic acid receptor (Rar), Rere, and p300, and the recruitment of Rar, Rere, and p300 at the *Rarb* promoter is dependent on Nr2f2⁵⁵. Similarly, Litchfield and colleagues showed that RA increased NR2F2 binding to a RARE in *RARB2*, an RA target gene. RA-mediated activation of *Rarβ2* was dependent on NR2F2⁵⁶. NR2F2 mediates RA signaling in other contexts including lung morphogenesis⁵⁷, trans-differentiation of breast cancer cells⁵⁸, and metabolism in hepatocytes⁵⁹. NR2F2 also bound RAREs in the PEPCK⁶⁰ and Gck gene promoters⁵⁹. Additionally, RA increases NR2F2 expression^{56,61}. NR2F2 can be activated by high, likely nonphysiological, levels of RA⁶². These studies support a role for NR2F2 in regulating target genes in the RA signaling pathway.

The second strongest associated SNP (rs36102671), which did not achieve statistical significance upon multiple testing correction (Bonferroni correction p-value threshold = 1.73×10^{-5}), lies within the first intron of the *RARA* gene, which encodes a receptor for RA (**Fig. 2C**). The minor (A) allele, which segregates at low allele frequencies in globally diverse populations (**Fig. S3D**), is associated with decreased L:M ratio ($\hat{\beta} = -15.95$, p-value = 6.54×10^{-5}). Within European populations, LD at this locus extends over an ~200 Kb region of chromosome 17, including *RARA* itself, as well as *MSL1*, *CASC3*, *MIR6866*, *MIR6867*, *RAPGEFL1*, *WIPF2*, and *CDC6* ($R^2 > 0.7$). In external data from the Genotype Tissue Expression Project (GTEx), the SNP also exhibits an association with altered splicing of *RARA* based on RNA sequencing of whole blood (**Fig. S3E**)⁶³.

While validation and further dissection of these association results requires investigation in a larger cohort, the consistent appearance of RA-related candidates pointed to a potential role for RA signaling in the specification of M and L cone subtypes in humans.

RA-synthesis genes are highly expressed early and decrease during development

Considering the association between M/L cone ratios and the natural variant in *NR2F2*, we analyzed a published dataset⁴¹ to assess expression of RA pathway genes during early human retinal development. The nuclear receptor genes *NR2F2* and *RARA* are expressed at consistent levels (**Fig. 2E, S3F**). Aldehyde Dehydrogenases (ALDHs) convert retinaldehyde to synthesize RA. *ALDH1A1* and *ALDH1A3* are expressed early in development and decrease over time, whereas *ALDH1A2* is expressed at very low levels (**Fig. 2F**). CYP26 enzymes catalyze the degradation of RA. *CYP26A1* is steadily

expressed, whereas *CYP26B1* and *CYP26C1* are not expressed (**Fig. 2G**). The high expression of two RA-synthesis enzymes early and the consistent expression of receptors and RA-degradatory enzymes suggested that RA signaling is high early and decreases during retinal development. Coupled with the observation that *M-opsin* is expressed before *L-opsin*, we hypothesized that high RA signaling early promotes M cone fate and suppresses L cone fate.

Visualization of *M*- and *L-opsin* expression

To test the role of RA signaling, we first needed to develop a method to visualize *M*- and *L-opsin* mRNA expression and observe M and L cones in human retinal and organoid tissue. The opsin protein sequences are ~96% identical and the cDNA sequences are ~98% identical, preventing the generation of antibodies and large fragment *in situ* hybridization reagents that distinguish *M*- and *L-opsin* expression *in vivo*⁶⁴.

To directly visualize *M*- and *L-opsin* expression in fixed retinas and organoids, we designed an oligo-based *in situ* hybridization approach. We generated a 40-nucleotide duplex probe for *M-opsin* and a 42-nucleotide duplex probe for *L-opsin* with partial overlap that target a region with 3 differential nucleotides in the *M*- and *L-opsin* mRNAs (**Fig. 1B, 3A**). We utilized colorimetric *in situ* hybridization to visualize *M-opsin* in blue and *L-opsin* in red.

We next tested the specificity of this approach to distinguish *M*- and *L-opsin* mRNA. HEK293 cells transfected with no plasmid showed no expression (**Fig. 3B, S4**). Cells transfected with *M-opsin* cDNA showed expression of *M-opsin* but not *L-opsin* (**Fig.**

3C, S4). Cells transfected with *L-opsin* cDNA displayed expression of *L-opsin* but not *M-opsin* (**Fig. 3D, S4**). Cells transfected with either *M-opsin* cDNA or *L-opsin* cDNA and then mixed into a single sample showed individual expression of *M-opsin* only or *L-opsin* only (**Fig. 3E, S4**). Finally, cells co-transfected with both *M-opsin* and *L-opsin* cDNA displayed co-expression of both *M-* and *L-opsin* (**Fig. 3F, S4**). Therefore, this *in situ* hybridization method distinguishes between the highly similar *M-* and *L-opsin* mRNAs in cultured cells.

We tested this approach in sectioned, flash-frozen adult human retinas. We observed exclusive expression of *M-* and *L-opsins* in individual cones (**Fig. 3G**). Thus, this *in situ* hybridization method unambiguously distinguishes *M-* and *L-opsin* mRNA.

RA promotes M cones and suppresses L cones early in human retinal organoids

We next sought to test our hypothesis that high RA signaling early promotes M cone fate and suppresses L cone fate in human retinal organoids. Differentiation of human retinal organoids involves addition of all-trans RA (hereafter referred to as RA) on days 20-43 to promote early retinal patterning (**Fig. S5**). RA was not added from day 43 to day 200, a timeframe which includes the end of primitive retina differentiation and the complete duration of cell fate specification (**'No RA'**).

We first used our *in situ* hybridization approach to examine the timing of M and L cone generation during human retinal organoid development. At day 120, we observed very few M and L cones (**Fig. 4A, S6A**). We first observed significant numbers of M and L cones on day 140 (**Fig. 4A, S6A**). From day 140 through day 200, organoids were enriched for L cones (**Fig. 4A-B, S6A**). These observations suggested that human retinal

organoids differentiated with this protocol are missing developmental cues to generate large populations of M cones before L cones as in human fetal development.

To test whether addition of RA induced M cone generation in retinal organoids, we added 1.0 μ M RA over different timeframes and assessed M and L cones at day 200 (**Fig. 4C**). Organoids grown in supplemental RA throughout development failed to differentiate, yielding minimal M or L cones (N=6). Addition of RA early in development from days 43 to 130 yielded organoids with almost exclusively M cones at day 200 (98.35% M, 1.65% L, 0% co-expressing; **Fig. 4D-E; 'Early RA'**). To corroborate this observation, we conducted RNA-seq on ESC-derived organoids grown in 'Early RA' conditions and observed high *M-opsin* and minimal *L-opsin* expression (**Fig. S1C**). We also analyzed our previously published RNA-seq data on iPSC-derived organoids grown in 'Early RA' conditions and observed near exclusive expression of *M-opsin* (**Fig. S1D**). Addition of RA late in development from days 130 to 200 yielded L cone-enriched organoids at day 200 (6.35% M, 92.56% L, 1.10% co-expressing; **Fig. 4D-E; 'Late RA'**), similar to organoids grown with no additional RA. We observed no significant differences in overall M/L cone densities at day 200 across experimental conditions (**Fig. S6B**). Together, these data suggest that RA is sufficient to induce M cones and suppress L cones early in retinal organoid development.

Discussion

We investigated the regulation of M and L cone subtype specification in the human retina. Our gene expression analysis suggested that *M-opsin* is expressed before *L-opsin* in human retinas. Our association study implicated a role for RA signaling in M versus L

cone subtype specification. ALDH genes, which code for RA synthesis proteins, are highly expressed early in retinal development and then decrease over time, suggesting that RA signaling is high early and decreases during retinal development. Human retinal organoids are L cone-rich in the absence of supplemental RA. RA is sufficient to induce M cones and suppress L cones early in organoid development. Our data support the temporal specification of M cones followed by L cones and a role for RA signaling early to promote M cones and suppress L cones (**Fig. 4F-G**).

The main models for M and L cone subtype specification involved stochastic and temporal mechanisms^{1,13,20-23,25,27,33}. Previously, reporter constructs testing regions of the *L/M-opsin* gene locus in the mouse retina suggested that the distance and order of the *L/M-opsin* promoters to the LCR determined the frequency of expression¹³. These results supported a model that cone specification was controlled by a stochastic mechanism in which the LCR randomly loops to the *M-opsin* or *L-opsin* promoter to determine expression state and cone subtype¹³. The temporal model was based on geographic distributions of M and L cones in the human retina, with higher proportions of L cones in the late-developing peripheral retina compared to the early-developing central retina²¹⁻³³. Our studies provide direct experimental evidence for the temporality of M and L cone generation and show that RA signaling regulates this mechanism. We speculate that temporal regulation by RA signaling is a main developmental driver that acts in conjunction with stochasticity in gene regulation to specify cone subtypes.

Our observations suggest that RA signaling influences cone fates before the expression of *M-* and *L-opsin* mRNA. In human retinal organoids, *M-* and *L-opsin* mRNA expression is first observed by day 140 (**Fig. 4A, S1D**), whereas addition of RA before

day 130 induces M cones and suppresses L cones (**Fig. 4D**). In human fetal retinas, *M-opsin* mRNA expression is readily observed on day 115 (**Fig. 1C**), whereas expression of RA synthesis genes (i.e. ALDHs) is highest on days 53 and 57 and then decreases (**Fig. 2F**). RA signaling may act at the gene locus to regulate opsin expression directly or function upstream to influence M and L cone fates.

Temporal regulation of RA signaling is important for the differentiation of photoreceptors during retinal development in a variety of species⁶⁵⁻⁶⁸. The arrangement of the LCR regulatory element and *L-* and *M-opsin* genes in humans resembles an independently evolved opsin gene array in zebrafish⁶⁹. In zebrafish, a regulatory element called the locus activating region (LAR) lies upstream of the *opn1lw1* and *opn1lw2* opsin genes⁷⁰. Expression of these two long-wavelength-detecting opsins defines cone subtypes: LWS1 cones express *opn1lw1* and LWS2 cones express *opn1lw2*. Specification of these two cone subtypes is temporally distinct, with *opn1lw2* expressed first during embryogenesis and *opn1lw1* expressed later after hatching. This temporal switch is controlled by retinoic acid (RA) signaling, which promotes *opn1lw1* and inhibits *opn1lw2* late in development⁷¹. This late role for RA signaling in promoting expression of the proximal opsin in zebrafish contrasts with our findings that RA signaling functions early to promote expression of the distal opsin in humans. These observations suggest differences in evolution of both the regulation of RA signaling and the *cis*-regulatory logic controlling opsin expression in these species.

Studying primate development poses specific challenges including limited access to tissue, a lack of a controlled genetic background, and an inability to genetically or pharmacologically manipulate developing retinal tissue. Commonly used vertebrate

model organisms such as mice, fish, and chickens have provided numerous insights into cell fate specification during vertebrate retinal development. However, differences between these systems and the primate visual system prevent interrogation of questions that can only be addressed in developing human tissue. Our studies show that human retinal organoids are a powerful model system to investigate developmental processes that are unique to humans and Old World monkeys.

Funding

S.E.H. was an F31 Ruth L. Kirschstein NRSA Fellow and supported by the National Eye Institute under grant 5F31EY029157-02. K.C.E. was a Howard Hughes Medical Institute Gilliam Fellow and was supported by the National Science Foundation Graduate Research Fellowship Program under grant 1746891. K.A.H. was supported by the National Science Foundation Graduate Research Fellowship Program under grant 1746891. R.C.M. was supported by R35GM133747 from NIH/NIGMS. M.N. and J.N. were supported by NEI grants R01EY021242 and P30EY001730, and unrestricted funds from Research to Prevent Blindness to ophthalmology at UW. R.J.J. was supported by the National Eye Institute R01EY030872 and the BrightFocus foundation G2019300.

References

- 1 Nathans, J., Thomas, D. & Hogness, D. S. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* **232**, 193-202, doi:10.1126/science.2937147 (1986).
- 2 Yokoyama, S. & Radlwimmer, F. B. The "five-sites" rule and the evolution of red and green color vision in mammals. *Mol Biol Evol* **15**, 560-567, doi:10.1093/oxfordjournals.molbev.a025956 (1998).
- 3 Vollrath, D., Nathans, J. & Davis, R. W. Tandem array of human visual pigment genes at Xq28. *Science* **240**, 1669-1672, doi:10.1126/science.2837827 (1988).
- 4 Dulai, K. S., von Dornum, M., Mollon, J. D. & Hunt, D. M. The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates. *Genome Res* **9**, 629-638 (1999).
- 5 Jacobs, G. H. & Neitz, J. Color vision in squirrel monkeys: sex-related differences suggest the mode of inheritance. *Vision Res* **25**, 141-143, doi:10.1016/0042-6989(85)90088-4 (1985).
- 6 Mollon, J. D., Bowmaker, J. K. & Jacobs, G. H. Variations of colour vision in a New World primate can be explained by polymorphism of retinal photopigments. *Proc R Soc Lond B Biol Sci* **222**, 373-399, doi:10.1098/rspb.1984.0071 (1984).
- 7 Onishi, A. *et al.* Dichromatism in macaque monkeys. *Nature* **402**, 139-140, doi:10.1038/45966 (1999).
- 8 Onishi, A. *et al.* Variations in long- and middle-wavelength-sensitive opsin gene loci in crab-eating monkeys. *Vision Res* **42**, 281-292, doi:10.1016/s0042-6989(01)00293-0 (2002).

- 361 9 Yamaguchi, T., Motulsky, A. G. & Deeb, S. S. Visual pigment gene structure and
362 expression in human retinae. *Hum Mol Genet* **6**, 981-990,
363 doi:10.1093/hmg/6.7.981 (1997).
- 364 10 Peng, Y. R. *et al.* Molecular Classification and Comparative Taxonomics of Foveal
365 and Peripheral Cells in Primate Retina. *Cell* **176**, 1222-1237 e1222,
366 doi:10.1016/j.cell.2019.01.004 (2019).
- 367 11 Hagstrom, S. A., Neitz, M. & Neitz, J. Cone pigment gene expression in individual
368 photoreceptors and the chromatic topography of the retina. *J Opt Soc Am A Opt*
369 *Image Sci Vis* **17**, 527-537, doi:10.1364/josaa.17.000527 (2000).
- 370 12 Winderickx, J., Battisti, L., Motulsky, A. G. & Deeb, S. S. Selective expression of
371 human X chromosome-linked green opsin genes. *Proc Natl Acad Sci U S A* **89**,
372 9710-9714, doi:10.1073/pnas.89.20.9710 (1992).
- 373 13 Smallwood, P. M., Wang, Y. & Nathans, J. Role of a locus control region in the
374 mutually exclusive expression of human red and green cone pigment genes. *Proc*
375 *Natl Acad Sci U S A* **99**, 1008-1011, doi:10.1073/pnas.022629799 (2002).
- 376 14 Wang, Y. *et al.* A locus control region adjacent to the human red and green visual
377 pigment genes. *Neuron* **9**, 429-440, doi:10.1016/0896-6273(92)90181-c (1992).
- 378 15 Drummond-Borg, M., Deeb, S. S. & Motulsky, A. G. Molecular patterns of X
379 chromosome-linked color vision genes among 134 men of European ancestry.
380 *Proc Natl Acad Sci U S A* **86**, 983-987, doi:10.1073/pnas.86.3.983 (1989).
- 381 16 Neitz, M., Bollinger, K. & Neitz, J. in *Normal and defective colour vision* (eds J.
382 D. Mollon, Joel Pokorny, & K. Knoblauch) 318-327 (Oxford University Press,
383 2003).

- 17 Winderickx, J. *et al.* Defective colour vision associated with a missense mutation in the human green visual pigment gene. *Nat Genet* **1**, 251-256, doi:10.1038/ng0792-251 (1992).
- 18 Hayashi, T., Motulsky, A. G. & Deeb, S. S. Position of a 'green-red' hybrid gene in the visual pigment array determines colour-vision phenotype. *Nat Genet* **22**, 90-93, doi:10.1038/8798 (1999).
- 19 Bollinger, K., Sjöberg, S. A., Neitz, M. & Neitz, J. Topographical cone photopigment gene expression in deutan-type red-green color vision defects. *Vision Res* **44**, 135-145, doi:10.1016/j.visres.2003.09.006 (2004).
- 20 Wang, Y. *et al.* Mutually exclusive expression of human red and green visual pigment-reporter transgenes occurs at high frequency in murine cone photoreceptors. *Proc Natl Acad Sci U S A* **96**, 5251-5256, doi:10.1073/pnas.96.9.5251 (1999).
- 21 Hagstrom, S. A., Neitz, J. & Neitz, M. Variations in cone populations for red-green color vision examined by analysis of mRNA. *Neuroreport* **9**, 1963-1967, doi:10.1097/00001756-199806220-00009 (1998).
- 22 Neitz, M., Balding, S. D., McMahon, C., Sjöberg, S. A. & Neitz, J. Topography of long- and middle-wavelength sensitive cone opsin gene expression in human and Old World monkey retina. *Vis Neurosci* **23**, 379-385, doi:10.1017/S095252380623325X (2006).
- 23 Bowmaker, J. K., Parry, J. W. L., and Mollon, J.D. The arrangement of L and M cones in human and a primate retina. *Normal and defective colour vision*, 39-50 (2003).

- 24 Bumsted, K. & Hendrickson, A. Distribution and development of short-wavelength
cones differ between Macaca monkey and human fovea. *J Comp Neurol* **403**, 502-
516 (1999).
- 25 Carroll, J., Neitz, J. & Neitz, M. Estimates of L:M cone ratio from ERG flicker
photometry and genetics. *J Vis* **2**, 531-542, doi:10.1167/2.8.1 (2002).
- 26 Cornish, E. E., Xiao, M., Yang, Z., Provis, J. M. & Hendrickson, A. E. The role of
opsin expression and apoptosis in determination of cone types in human retina.
Exp Eye Res **78**, 1143-1154, doi:10.1016/j.exer.2004.01.004 (2004).
- 27 Kuchenbecker, J. A., Sahay, M., Tait, D. M., Neitz, M. & Neitz, J. Topography of
the long- to middle-wavelength sensitive cone ratio in the human retina assessed
with a wide-field color multifocal electroretinogram. *Vis Neurosci* **25**, 301-306,
doi:10.1017/S0952523808080474 (2008).
- 28 La Vail, M. M., Rapaport, D. H. & Rakic, P. Cytogenesis in the monkey retina. *J*
Comp Neurol **309**, 86-114, doi:10.1002/cne.903090107 (1991).
- 29 Martin, P. R., Grunert, U., Chan, T. L. & Bumsted, K. Spatial order in short-
wavelength-sensitive cone photoreceptors: a comparative study of the primate
retina. *J Opt Soc Am A Opt Image Sci Vis* **17**, 557-567,
doi:10.1364/josaa.17.000557 (2000).
- 30 Mollon, J. D. & Bowmaker, J. K. The spatial arrangement of cones in the primate
fovea. *Nature* **360**, 677-679, doi:10.1038/360677a0 (1992).
- 31 Otake, S., Gowdy, P. D. & Cicerone, C. M. The spatial arrangement of L and M
cones in the peripheral human retina. *Vision Res* **40**, 677-693, doi:10.1016/s0042-
6989(99)00202-3 (2000).

- 430 32 Roorda, A., Metha, A. B., Lennie, P. & Williams, D. R. Packing arrangement of the
431 three cone classes in primate retina. *Vision Res* **41**, 1291-1306,
432 doi:10.1016/s0042-6989(01)00043-8 (2001).
- 433 33 Xiao, M. & Hendrickson, A. Spatial and temporal expression of short, long/medium,
434 or both opsins in human fetal cones. *J Comp Neurol* **425**, 545-559 (2000).
- 435 34 Nakano, T. *et al.* Self-formation of optic cups and storable stratified neural retina
436 from human ESCs. *Cell Stem Cell* **10**, 771-785, doi:10.1016/j.stem.2012.05.009
437 (2012).
- 438 35 Wahlin, K. J. *et al.* Photoreceptor Outer Segment-like Structures in Long-Term 3D
439 Retinas from Human Pluripotent Stem Cells. *Sci Rep* **7**, 766, doi:10.1038/s41598-
440 017-00774-9 (2017).
- 441 36 Eldred, K. C. *et al.* Thyroid hormone signaling specifies cone subtypes in human
442 retinal organoids. *Science* **362**, doi:10.1126/science.aau6348 (2018).
- 443 37 Neitz, J. & Neitz, M. The genetics of normal and defective color vision. *Vision Res*
444 **51**, 633-651, doi:10.1016/j.visres.2010.12.002 (2011).
- 445 38 Verrelli, B. C. & Tishkoff, S. A. Signatures of selection and gene conversion
446 associated with human color vision variation. *Am J Hum Genet* **75**, 363-375,
447 doi:10.1086/423287 (2004).
- 448 39 Shaaban, S. A. & Deeb, S. S. Functional analysis of the promoters of the human
449 red and green visual pigment genes. *Invest Ophthalmol Vis Sci* **39**, 885-896
450 (1998).

- 40 Liu, Y., Fu, L., Chen, D. G. & Deeb, S. S. Identification of novel retinal target genes of thyroid hormone in the human WERI cells by expression microarray analysis. *Vision Res* **47**, 2314-2326, doi:10.1016/j.visres.2007.04.023 (2007).
- 41 Hoshino, A. *et al.* Molecular Anatomy of the Developing Human Retina. *Dev Cell* **43**, 763-779 e764, doi:10.1016/j.devcel.2017.10.029 (2017).
- 42 Pinelli, M. *et al.* An atlas of gene expression and gene co-regulation in the human retina. *Nucleic Acids Res* **44**, 5773-5784, doi:10.1093/nar/gkw486 (2016).
- 43 Dartnall, H. J., Bowmaker, J. K. & Mollon, J. D. Human visual pigments: microspectrophotometric results from the eyes of seven persons. *Proc R Soc Lond B Biol Sci* **220**, 115-130, doi:10.1098/rspb.1983.0091 (1983).
- 44 Hofer, H., Carroll, J., Neitz, J., Neitz, M. & Williams, D. R. Organization of the human trichromatic cone mosaic. *J Neurosci* **25**, 9669-9679, doi:10.1523/JNEUROSCI.2414-05.2005 (2005).
- 45 McMahon, C., Carroll, J., Awua, S., Neitz, J. & Neitz, M. The L:M cone ratio in males of African descent with normal color vision. *J Vis* **8**, 5 1-9, doi:10.1167/8.2.5 (2008).
- 46 Roorda, A. & Williams, D. R. The arrangement of the three cone classes in the living human eye. *Nature* **397**, 520-522, doi:10.1038/17383 (1999).
- 47 McMahon, C., Neitz, J. & Neitz, M. Evaluating the human X-chromosome pigment gene promoter sequences as predictors of L:M cone ratio variation. *J Vis* **4**, 203-208, doi:10.1167/4.3.7 (2004).

- 48 Gunther, K. L., Neitz, J. & Neitz, M. Nucleotide polymorphisms upstream of the X-chromosome opsin gene array tune L:M cone ratio. *Vis Neurosci* **25**, 265-271, doi:10.1017/S0952523808080280 (2008).
- 49 Neitz, J. & Jacobs, G. H. Polymorphism of the long-wavelength cone in normal human colour vision. *Nature* **323**, 623-625, doi:10.1038/323623a0 (1986).
- 50 Nathans, J. *et al.* Molecular genetics of human blue cone monochromacy. *Science* **245**, 831-838, doi:10.1126/science.2788922 (1989).
- 51 Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B. & Hogness, D. S. Molecular genetics of inherited variation in human color vision. *Science* **232**, 203-210, doi:10.1126/science.3485310 (1986).
- 52 Merbs, S. L. & Nathans, J. Absorption spectra of the hybrid pigments responsible for anomalous color vision. *Science* **258**, 464-466, doi:10.1126/science.1411542 (1992).
- 53 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
- 54 Zhang, R., Wang, Y., Li, R. & Chen, G. Transcriptional Factors Mediating Retinoic Acid Signals in the Control of Energy Metabolism. *Int J Mol Sci* **16**, 14210-14244, doi:10.3390/ijms160614210 (2015).
- 55 Vilhais-Neto, G. C. *et al.* Rere controls retinoic acid signalling and somite bilateral symmetry. *Nature* **463**, 953-957, doi:10.1038/nature08763 (2010).
- 56 Litchfield, L. M. *et al.* Identification and characterization of nucleolin as a COUP-TFII coactivator of retinoic acid receptor beta transcription in breast cancer cells. *PLoS One* **7**, e38278, doi:10.1371/journal.pone.0038278 (2012).

- 57 Malpel, S., Mendelsohn, C. & Cardoso, W. V. Regulation of retinoic acid signaling during lung morphogenesis. *Development* **127**, 3057-3067 (2000).
- 58 Prahalad, P., Dakshanamurthy, S., Ressom, H. & Byers, S. W. Retinoic acid mediates regulation of network formation by COUP-TFII and VE-cadherin expression by TGFbeta receptor kinase in breast cancer cells. *PLoS One* **5**, e10023, doi:10.1371/journal.pone.0010023 (2010).
- 59 Li, R. *et al.* A RARE of hepatic Gck promoter interacts with RARalpha, HNF4alpha and COUP-TFII that affect retinoic acid- and insulin-induced Gck expression. *J Nutr Biochem* **25**, 964-976, doi:10.1016/j.jnutbio.2014.04.009 (2014).
- 60 Scribner, K. B., Odom, D. P. & McGrane, M. M. Nuclear receptor binding to the retinoic acid response elements of the phosphoenolpyruvate carboxykinase gene in vivo: effects of vitamin A deficiency. *J Nutr Biochem* **18**, 206-214, doi:10.1016/j.jnutbio.2006.03.011 (2007).
- 61 Achatz, G. *et al.* Functional domains of the human orphan receptor ARP-1/COUP-TFII involved in active repression and transrepression. *Mol Cell Biol* **17**, 4914-4932, doi:10.1128/mcb.17.9.4914 (1997).
- 62 Kruse, S. W. *et al.* Identification of COUP-TFII orphan nuclear receptor as a retinoic acid-activated receptor. *PLoS Biol* **6**, e227, doi:10.1371/journal.pbio.0060227 (2008).
- 63 Consortium, G. T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330, doi:10.1126/science.aaz1776 (2020).
- 64 Pettan-Brewer, C., Fu, L. & Deeb, S. S. Distinguishing L from M photopigment coding sequences by hybridization to novel locked nucleic acid (LNA)

518 oligonucleotide probes. *Vis Neurosci* **25**, 283-287,
519 doi:10.1017/S0952523808080607 (2008).

520 65 Sen, J., Harpavat, S., Peters, M. A. & Cepko, C. L. Retinoic acid regulates the
521 expression of dorsoventral topographic guidance molecules in the chick retina.
522 *Development* **132**, 5147-5159, doi:10.1242/dev.02100 (2005).

523 66 Hyatt, G. A., Schmitt, E. A., Marsh-Armstrong, N. R. & Dowling, J. E. Retinoic acid-
524 induced duplication of the zebrafish retina. *Proc Natl Acad Sci U S A* **89**, 8293-
525 8297, doi:10.1073/pnas.89.17.8293 (1992).

526 67 Wilson, J. G., Roth, C. B. & Warkany, J. An analysis of the syndrome of
527 malformations induced by maternal vitamin A deficiency. Effects of restoration of
528 vitamin A at various times during gestation. *Am J Anat* **92**, 189-217,
529 doi:10.1002/aja.1000920202 (1953).

530 68 da Silva, S. & Cepko, C. L. Fgf8 Expression and Degradation of Retinoic Acid Are
531 Required for Patterning a High-Acuity Area in the Retina. *Dev Cell* **42**, 68-81 e66,
532 doi:10.1016/j.devcel.2017.05.024 (2017).

533 69 Chinen, A., Hamaoka, T., Yamada, Y. & Kawamura, S. Gene duplication and
534 spectral diversification of cone visual pigments of zebrafish. *Genetics* **163**, 663-
535 675 (2003).

536 70 Tsujimura, T., Hosoya, T. & Kawamura, S. A single enhancer regulating the
537 differential expression of duplicated red-sensitive opsin genes in zebrafish. *PLoS*
538 *Genet* **6**, e1001245, doi:10.1371/journal.pgen.1001245 (2010).

539 71 Mitchell, D. M. *et al.* Retinoic Acid Signaling Regulates Differential Expression of
540 the Tandemly-Duplicated Long Wavelength-Sensitive Cone Opsin Genes in
541 Zebrafish. *PLoS Genet* **11**, e1005483, doi:10.1371/journal.pgen.1005483 (2015).
542
543

Materials and Methods

Cell line maintenance

H7 ESC (WA07, WiCell) and episomal-derived EP1.1 iPSC were used for retinal organoid differentiation. Pluripotency of EP1.1 cells was evaluated previously with antibodies for NANOG, OCT4, SOX2, and SSEA2¹. Stem cells were maintained in mTeSR™ 1 (85857, Stem Cell Technologies) on 1% (v/v) Matrigel-GFR™ (354230, BD Biosciences) coated dishes and grown at 37°C in a HERAcell 150i or 160i 10% CO₂ and 5% O₂ incubator (Thermo Fisher Scientific). Cells were passaged every 4-5 days according to confluence as in Wahlin *et. al*¹. Cells were passaged with Accutase (SCR005, Sigma) for 7-12 minutes to be dissociated to single cells. Cells in Accutase were added 1:2 to mTeSR™ 1 plus 5 µM Blebbistatin (Bleb, B0560, Sigma), pelleted at 150g for 5 minutes, and suspended in mTeSR™ 1 plus Bleb and plated at 5,000-15,000 cells per well in a 6-well plate. Cells were fed with mTeSR™ 1 48 hours following passage and every 24 hours until passaged again. To minimize cell stress, no antibiotics were used.

Weri-Rb-1 retinoblastoma cells were obtained from ATCC and maintained in RPMI 1640 Medium (11875135, Gibco) + 10% Fetal Bovine Serum (16140071, Gibco) + 1X Penicillin-Streptomycin (30-002-CI, Corning) at 37°C in a HERAcell 150i or 160i 5% CO₂ incubator (Thermo Fisher Scientific). Cells were passaged every 4 days at ~1 x 10⁵ – 2 x 10⁶ cells/mL in uncoated flasks by pelleting at 150g for 5 minutes and resuspending in fresh media.

HEK293 cells were a gift from the Kuruvilla lab at Johns Hopkins University. Cells were maintained in DMEM 4.5 g/L D-glucose, L-glutamine (11965084, Gibco) + 10% Fetal Bovine Serum (16140071, Gibco) + 1X Penicillin-Streptomycin (30-002-CI, Corning) at 37°C in a HERAcult 150i or 160i 5% CO₂ incubator (Thermo Fisher Scientific). Cells were split 1:4 – 1:10 every 2-3 days when approximately 80% confluent and kept in 6-well dishes or uncoated flasks using Trypsin (0.25%)-EDTA (0.02%) in HBSS (118-093-721, Quality Biological).

Cell lines were tested monthly for mycoplasma using MycoAlert (LT07, Lonza).

Cell Culture Media

Stem Cell media: mTeSR1 (85857, StemCell Technologies)

E6 supplement: 970 µg/mL Insulin (11376497001, Roche), 535 µg/mL holo-transferrin (T0665, Sigma), 3.20 mg/mL L-ascorbic acid (A8960, Sigma), 0.7 µg/mL sodium selenite (S5261, Sigma).

BE6.2 media for early retinal differentiation: 2.5% E6 supplement (above), 2% B27 Supplement (50X) minus Vitamin A (12587010, Gibco), 1% Glutamax (35050061, Gibco), 1% NEAA (11140050, Gibco), 1mM Pyruvate (11360070, Gibco), and 0.87 mg/mL NaCl in DMEM (11885084, Gibco).

LTR (Long-Term Retina) media: 25% F12 (11765062, Gibco) with 2% B27 Supplement (50X) (17504044, Gibco), 10% heat inactivated FBS (16140071, Gibco), 1mM Sodium Pyruvate, 1% NEAA, 1% Glutamax and 1 mM taurine (T-8691, Sigma) in DMEM (11885084, Gibco).

RPMI + supplement media: 10% heat inactivated FBS (16140071, Gibco), 2.5% penicillin (30-002-CI, Corning) in RPMI Medium 1640 (11875135, Gibco).

HEK293 + supplement media: 10% heat inactivated FBS (16140071, Gibco), 2.5% penicillin (30-002-CI, Corning) in DMEM 4.5g/L D-glucose, L-glutamine (11965084, Gibco).

Retinoic acid treatment: For organoids, 1.04 μ M all-trans retinoic acid (ATRA; R2625; Sigma) in LTR.

Thyroid hormone treatment: For Weri-Rb1 cells, 100 nM T3 (T6397, Sigma) in RPMI + supplement media.

Organoid differentiation

Organoids were differentiated from H7 WA07, H7iCas9 ESCs, or EP1.1 iPSCs as described in Eldred *et al.* 2018 with minor variations (**Fig. S5**)².

Pluripotent stem cells were well-maintained. Cultures with minimal to no spontaneous differentiation were used for aggregation. To aggregate, cells were passaged in Accutase (SCR005, Sigma) at 37°C for 12 min to ensure complete dissociation. Cells were seeded in 50 μ Ls of mTeSR1 at 3,000 cells/well into 96-well ultra-low adhesion round bottom Lipidure coated plates (51011610, NOF) or ultra-low attachment microplate (7007, Corning). Cells were placed in hypoxic conditions (10% CO₂ and 5% O₂) for 24 hours to enhance survival. Cells naturally aggregated by gravity over 24 hours.

On day 1, cells were moved to normoxic conditions (5% CO₂). On days 1-3, 50 μ Ls of BE6.2 media containing 3 μ M Wnt inhibitor (IWR1e: 681669, EMD Millipore) and

1% (v/v) Matrigel were added to each well. On days 4-9, 100 μ Ls of media were removed from each well, and 100 μ Ls of media were added. On days 4-5, BE6.2 media containing 3 μ M Wnt inhibitor and 1% Matrigel was added. On days 6-7, BE6.2 media containing 1% Matrigel (354230, BD Biosciences) was added. On days 8-9, BE6.2 media containing 1% Matrigel and 100 nM Smoothed agonist (SAG: 566660, EMD Millipore) was added.

On day 10, aggregates were transferred to 15 mL tubes, rinsed 3X in 5mL DMEM (11885084, Gibco), and resuspended in BE6.2 with 100 nM SAG in untreated 10 cm polystyrene petri dishes. From this point on, media was changed every other day. Aggregates were monitored and manually separated if stuck together or to the bottom of the plate.

On days 13-16, LTR media with 100 nM SAG was added. On day 16, retinal vesicles were manually dissected using sharpened tungsten needles. After dissection, cells were transferred into 15 mL tubes and washed 2X with 5 mLs of DMEM. On days 16-20, cells were maintained in LTR and washed 2X with 5 mLs of DMEM, before being transferred to new plates to wash off dead cells. To increase survival and differentiation, 1.04 μ M all-trans retinoic acid (ATRA; R2625; Sigma) was added to LTR medium from days 20-43. Additional time windows of 1.04 μ M were added depending on experimental conditions. 10 μ M gamma-secretase inhibitor (DAPT, 565770, EMD Millipore) was added to LTR from days 28-42. Organoids were grown at low density (10-20 per 10 cm dish) to reduce aggregation. Periodically, organoids were culled from the plate based on absence of clear laminal structure indicating proper retinal organoid growth.

RNA-Seq experiments

EP1 iPSC-derived organoids were previously grown and analyzed for Eldred *et al.* 2018 at time points ranging from day 10 to day 250 of differentiation². Samples were taken at day 10 (N=3), day 20 (N=2), day 35 (N=3), day 69 (N=3), day 111 (N=3), day 128 (N=3), day 158 (N=2), day 173 (N=3), day 181 (N=3), day 200 (N=3), and day 250 (N=3).

H7 ESC-derived organoids were taken at day 329 (N=3).

Weri-Rb-1 samples were grown in control RPMI + supplement media or T3-treated RPMI + supplement media for four days (N=1).

RNA from individual samples was extracted using the Zymo Direct-zol RNA Microprep Kit (R2062, Zymo Research) according to manufacturer's instructions. Libraries were prepared using the Illumina TruSeq stranded mRNA kit and sequenced on an Illumina NextSeq 500 with single 75 bp reads.

RNA sequencing analysis

Quantification of Opsin Expression in Human Fetal Retina Samples from^{3,4}:

Expression of opsin genes was directly quantified from FASTQ files using Kallisto 0.44.0. The kallisto index was generated from the Gencode V27 human transcriptome, with the default kmer setting of 31. Reads were quantified in single-ended mode with a specified length of 75 and standard deviation of 10.

Analysis of Opsin Pileups: To examine the reads mapped to each opsin directly, reads were aligned to the opsin transcripts using Bowtie 2.3.4.3, and aligned

files were processed with samtools 1.7 to produce pileups around selected locations in the transcripts. For the quantification of human samples, a Bowtie2 index was generated from sequences ENST00000595290.5 and ENST00000369951.8 using the default settings. Pileups were generated for transcript positions given in Table 1:

Exon	Base		Transcript position		Exon Position	Absolute Position	
	LW	MW	LW	MW		LW	MW
2	C	T	254	276	81	154150737	154187851
	A	G	360	382	187	154150843	154187957
	A	G	391	413	218	154150874	154187988
	C	A	407	429	234	154150890	154188004
4	T	C	749	771	110	154154684	154191798
	G	A	757	779	118	154154692	154191806
	C	G	758	780	119	154154693	154191807
	T	C	759	781	120	154154694	154191808
	A	G	766	788	127	154154701	154191815
5	A	G	880	902	75	154156369	154193483
	T	C	883	905	78	154156372	154193486
	T	G	885	907	80	154156374	154193488
	G	A	888	910	83	154156377	154193491
	A	T	890	912	85	154156379	154193493
	G	T	895	917	90	154156384	154193498
	C	A	901	931	104	154156398	154193512
	A	G	905	935	108	154156402	154193516
	T	C	940	970	143	154156437	154193551
	G	C	944	974	147	154156441	154193555
	A	T	978	1008	181	154156475	154193589

Association Study

Subjects: All experiments involving human subjects were conformed to the principles expressed in the Declaration of Helsinki and were approved by the University of Washington Human Subjects Review Committee. Subjects were 738 males who had previously passed several standard tests for color vision deficiency, including Ishihara's 24-plate test, Richmond HRR 2002 edition, and the Neitz Test of Color Vision. Subjects identified themselves as being either of African or African American ancestry (N=147), Asian ancestry (N=108), Caucasian ancestry (N=362), Latino ancestry (N=23), Mixed (N=62), Native American ancestry (N=13), South Asian ancestry (N=9), or Other/Not Specified (N=14). The proportion of L cones, expressed as the percentage of L plus M cones that were L, were determined for each subject from a combination of genetic analysis and the flicker photometric electroretinogram (FP-ERG) as described in^{5,6}. Briefly, the spectral sensitivity of the individual was measured using control and test lights of specific wavelengths to illuminate a small portion of the retina. These conditions were designed to eliminate rod and S-cone contributions. While the test lights were flickered on and off, electrodes recorded the ERG neuronal responses to the stimulus. The spectral sensitivity was determined by adjusting the intensity of the test light until the ERG signal produced exactly matched that produced by the fixed intensity reference light. A correction factor was applied to the estimated proportion of L cones to account for the 1.5 times greater contribution to the FP-ERG signal for each M cone compared to each L cone⁷. These data combined with genetic analysis of the L and M cone genes to determine the spectral sensitivity of each opsin allowed for the determination of the L:M cone ratio.

DNA preparation, pull-down, and sequencing:

DNA was isolated from buccal swabs or whole blood using PureGene DNA extraction kits as described in⁸. DNA was then diluted to 10 ng/μL in low ETDA (USB corporation, #75793). DNA was sheared to an average length of 250 bps using the Bioruptor® Pico (Diagenode). Samples were stored on ice for 10-15 min prior to sonication. Sonication was performed for 7 cycles of on for 15 seconds, off for 90 seconds. After 3 cycles, the samples were briefly centrifuged before continuing with sonication for the remaining 4 cycles. Library preparation and hybridization pull-downs were performed according to the SeqCap EZ HyperCap Workflow User's Guide, version 2.0 following all recommended instructions. Libraries were created using the KAPA Hyper Prep Kit (KAPA Biosystems, #KK8504), using the KAPA Dual-Indexed Adapter Set (15 mM, KAPA Biosystems/Roche, #08 278 555 702). Probes for the pull-down were designed according to the Roche Nimble Design software, targeting the gene regions listed in (**Fig. S2**). Samples were sequenced on a Nova-seq, high-output, with 125 bp paired-end reads.

Analysis:

Paired-end reads for 738 samples were aligned to the human genome (build hg38) using BWA mem (v0.7.15) using default parameters. Duplicate reads were marked using Picard (v2.9.2). Each region was subdivided into nonoverlapping 1 kb windows starting at the upstream boundary using bedtools (v2.29.2). Mapped reads were genotyped on a region-by-region basis using Freebayes-parallel (v1.2.0), the argument '--use-best-n-alleles 4', and, for the region on chromosome X, '-p 1'. Variant calls were filtered for quality greater than 10 times the number of additional observations, at least one alternative allele read on each strand,

and at least two reads supporting the alternate allele balanced up- and downstream using vcflib (v1.0.0). Duplicate variants were then removed. Variants were filtered with PLINK (v1.90b6.12) for a minor allele frequency of at least 3.5% and a per-variant genotyping rate of at least 90%. Variants were pruned using LDAK (v5.0) with a linkage disequilibrium cutoff of $R^2 < 0.75$ within 100 kb windows. A kinship matrix was calculated from the pruned variant set using LDAK, ignoring weights and a power of -1. P-values were calculated in LDAK from the pruned variant set, using the first five genotype eigenvectors as covariates to control for potential population stratification. Significant variants were defined using an alpha of 0.05 after Bonferroni correction.

Transfection of opsin cDNA into HEK293 cells

HEK293 cells were passaged 24 hours prior to be <60% confluent on the day of transfection. Transfection was done with the Lipofectamine® 3000 transfection kit (L3000001, Invitrogen). For a 10 cm dish, 7.5 µL lipofectamine 3000 was mixed with 125 µL Opti-MEM (31985062, Thermo Fisher Scientific) at room temperature for 5 minutes. 5 µL of P3000, 0.5-1 µg DNA, and 125 µL Opti-MEM was mixed at room temperature for 5 minutes. The two solutions were mixed and incubated at room temperature for 10 minutes. Media was aspirated from HEK293 cells and replaced with 8 mL of prewarmed Opti-MEM. Lipofectamine mixture was mixed onto cells and incubated at 37°C in a HERAcCell 150i or 160i 5% CO₂ incubator for 4-6 hours. Transfection media was aspirated and replaced with HEK293 maintenance media (DMEM/10% FBS/1X Penn/Strep) and allowed to recover overnight. Cells were allowed to adhere to slides coated with 10% Poly-L-lysine solution (0.1% w/v in H₂O) (P8920,

Sigma) for 1-3h at 37°C in a HERAcCell 150i or 160i 5% CO₂ incubator (Thermo Fisher Scientific). Cells were washed in 1x PBS and fixed in 10% neutral buffered formalin (HT501128, Sigma) for 30 minutes. Slides were either used immediately for RNA *in situ* or dried and stored at -80°C.

Plasmids containing cDNA for *OPN1MW* and *OPN1LW* were purchased from GenScript: NM_000513.2 (*OPN1MW*) and NM_020061.5 (*OPN1LW*).

Human retina and organoid preparation and cryosectioning

Human retina: Donor sample was flash frozen on dry ice 10.9 hours postmortem and stored at -80°C. The human eye was allowed to come to room temperature in 1X PBS and the retina was dissected out of the eye. The retina was fixed for 45 minutes in 10% neutral buffered formalin (HT501128, Sigma) and washed in 1X PBS. Subsections of retina were mounted in Tissue-Tek O.C.T. compound (4583, Sakura), placed on dry ice to freeze, and stored at -80°C. The retina was sectioned in 10 µm sections. Slides were air dried for 6 hours to overnight with a post-fixation step of 15 minutes in 10% neutral buffered formalin (HT501128, Sigma) and washed in 1X PBS. Slides were dried and stored at -80°C for less than 3 months before use.

Organoid: Organoids were fixed for 45 minutes in 10% neutral buffered formalin (HT501128, Sigma) and washed in 1X PBS. Organoids were placed in a 25% sucrose in 0.1 M phosphate buffer solution overnight, and then mounted in Tissue-Tek O.C.T. compound (4583, Sakura), placed on dry ice to freeze, and stored at -80°C. Organoids were sectioned in 10 µm sections. Slides were air dried for 6 hours to overnight with a post-fixation step of 15 minutes in 10% neutral buffered formalin (HT501128, Sigma)

and washed in 1X PBS. Slides were dried and stored at -80C for less than 3 months before use.

Organoid cell lines

No RA: H7 ESC (N=3)

Early RA: H7 ESC (N=1), H7 iCas9 ESC (N=2)

Late RA: H7iCas9 ESC (N=5)

RNA *in situ* hybridization

BaseScope RNA *in situ* hybridization was performed according to manufacturer's instructions and modified with several changes. Probe sequences were designed by ACD Biotechne based on *OPN1MW* and *OPN1LW* mRNA sequences *NM_000513.2* *NM_020061.5* from the human genome hg38.

Sections were allowed to come to room temperature from storage at -80C and rehydrated in 1X PBS. Samples were pretreated according to manufacturer's instructions: 10 minutes of RNAscope Hydrogen Peroxide followed by a wash in dH₂O and then 2 washes in 1X PBS.

RNAscope Protease III was applied at a 1:15 dilution in 1X PBS for 15 minutes in a humid chamber for HEK293 cells. RNAscope Protease IV was applied for 20 minutes in a humid chamber for organoids and human eye samples. Samples were washed in 1X PBS twice.

Probes were added at the manufacturer suggested concentration to samples in

the HybEZ Humidity Control rack with lid and insert into the HybEZ oven for 2 hours at 40C. Samples were washed twice in 1x RNAscope wash buffer for 2 min.

Amplification and development washes were done with the manufacturer reagents without change to the recommended concentrations. All washes were done in the HybEZ Humidity Control rack with lid and insert, either at room temperature (RT) or at 40C in the HybEZ oven. Each wash was done at the following temperatures and time lengths:

Reagent	Temperature (°C)	Time (min)
AMP 1	40	30
AMP 2	40	30
AMP 3	40	15
AMP 4	40	30
AMP 5	40	30
AMP 6	40	15
AMP 7	RT	30
AMP 8	RT	15
RED-A/B	RT	10
AMP 9	40	15
AMP 10	40	15
AMP 11	RT	30
AMP 12	RT	15
GREEN-A/B	RT	10

2X 2 min washes in 1X RNAscope Buffer were conducted between each reagent wash. The final wash was in tap water. Slides were baked for a minimum of 30 minutes at 65C in the HybEZ oven. Samples were preserved with VectaMount (Vector Laboratories, H-5000) mounting media and a sealed coverslip.

Imaging

HEK293: Cells were imaged using an EVOS XL Core Cell imaging system (ThermoFisher).

Organoids and human retina: Samples were imaged using a Zeiss LSM 980 color camera system.

Quantification

HEK293: All individual cells were counted manually.

Organoids: All serially sectioned organoids were imaged and counted manually. Organoids that had fewer than 150 cones ($n \leq 150$) were removed from our analysis.

Statistical tests are listed in figure legends. All error bars represent SEM.

Methods References

- 1 Wahlin, K. J. *et al.* Photoreceptor Outer Segment-like Structures in Long-Term 3D Retinas from Human Pluripotent Stem Cells. *Sci Rep* **7**, 766, doi:10.1038/s41598-017-00774-9 (2017).
- 2 Eldred, K. C. *et al.* Thyroid hormone signaling specifies cone subtypes in human retinal organoids. *Science* **362**, doi:10.1126/science.aau6348 (2018).
- 3 Pinelli, M. *et al.* An atlas of gene expression and gene co-regulation in the human retina. *Nucleic Acids Res* **44**, 5773-5784, doi:10.1093/nar/gkw486 (2016).
- 4 Hoshino, A. *et al.* Molecular Anatomy of the Developing Human Retina. *Dev Cell* **43**, 763-779 e764, doi:10.1016/j.devcel.2017.10.029 (2017).
- 5 Carroll, J., Neitz, J. & Neitz, M. Estimates of L:M cone ratio from ERG flicker photometry and genetics. *J Vis* **2**, 531-542, doi:10.1167/2.8.1 (2002).
- 6 McMahon, C., Carroll, J., Awua, S., Neitz, J. & Neitz, M. The L:M cone ratio in males of African descent with normal color vision. *J Vis* **8**, 5 1-9, doi:10.1167/8.2.5 (2008).
- 7 Hofer, H., Carroll, J., Neitz, J., Neitz, M. & Williams, D. R. Organization of the human trichromatic cone mosaic. *J Neurosci* **25**, 9669-9679, doi:10.1523/JNEUROSCI.2414-05.2005 (2005).
- 8 Gunther, K. L., Neitz, J. & Neitz, M. Nucleotide polymorphisms upstream of the X-chromosome opsin gene array tune L:M cone ratio. *Vis Neurosci* **25**, 265-271, doi:10.1017/S0952523808080280 (2008).

Figure 1

829 **Fig. 1. *M-opsin* is expressed before *L-opsin* in fetal human retinas.**

830 **(A)** Human *L/M-opsin* locus. LCR = locus control region; *OPN1LW* (red) = *L-opsin*,

831 *OPN1MW* (green) = *M-opsin*. * indicates that the copy number of *OPN1MW* genes is

832 variable in the human population.

833 **(B)** Alignment of human *M-opsin* and *L-opsin* mRNA. Red bar = nucleotide difference.

834 Horizontal pink and blue lines = location of the *in situ* probes (as in **Fig. 3A**).

835 **(C-D)** *M-* and *L-opsin* expression. Heatmap indicates total pileup count normalized to

836 total read count. Each square indicates one nucleotide difference in *M-* or *L-opsin*.

837 **(C)** Human embryonic and fetal retinas analyzed from ⁴¹ (n = 1 for each timepoint;

838 except 94-2 where n = 2). 52/54 = exact date is unclear.

839 **(D)** Human adult retinas (n = 1 for each timepoint) analyzed from ⁴².

Figure 2

840 **Fig. 2. SNPs in RA signaling regulators are associated with changes in L:M cone**
841 **ratios.**

842 **(A)** Histogram of the ratios of % L from 738 human males with normal color vision.

843 **(B)** Violin plots of % L cone ratios for the genotypes of the significant polymorphism in
844 *NR2F2*.

845 **(C)** Manhattan plot of the genetic variant p-values. Genetic variants above the gray line
846 (Bonferroni corrected threshold) are considered significant ($p < 0.05$).

847 **(D)** *NR2F2* gene locus. *NR2F2* gene (pink); *NR2F2* antisense RNA (dark pink). Yellow
848 arrow denotes the location of the SNP.

849 **(E-G)** Human fetal retina expression for **(E)** *NR2F2*, **(F)** *ALDH1A1*, *A2*, and *A3*, and **(G)**
850 *CYP26A1*, *B1*, and *C1* from ⁴¹. CPM = log counts per million.

Figure 3

Fig 3. Visualization of *M*- and *L*-opsin mRNA.

(A) Alignment of portions of exon 5 from *M*- and *L*-opsins. *in situ* hybridization probe

binding sites are indicated by blue (*M*-opsin) and pink (*L*-opsin) boxes. Green

arrowheads indicate 3 nucleotide differences.

(B-F) HEK293 cells probed for both *M*-opsin and *L*-opsin. Insets indicate a schematic of

the transfected plasmid; blue = *M*-opsin cDNA; pink = *L*-opsin cDNA.

(B) No cDNA was transfected.

(C) *M*-opsin cDNA was transfected.

(D) *L*-opsin cDNA was transfected.

(E) Either *M*- or *L*-opsin cDNA was transfected independently and then the cells were

mixed.

(F) *M*- and *L*-opsin cDNAs were co-transfected.

(G) 10 μ m section of adult human retina probed for both *M*-opsin and *L*-opsin; blue = *M*-

opsin; pink = *L*-opsin.

Figure 4

Fig. 4. RA signaling induces *M-opsin* and inhibits *L-opsin* early in human retinal organoids.

(A) *M-opsin* (blue) and *L-opsin* (pink) expression in ‘No RA’ human retinal organoids.

White outlines indicate *M*- or *L-opsin*-expressing cells.

(B) Quantification of A. For day 120, organoids had fewer than 150 M or L cones (N=3)

(Fig. S6A). For day 140, 160, 180, and 200, organoids had greater than 150 M or L

cones **(Fig. S6A)**. For day 140, N=4; for day 160, N=3; for day 180, N=3; and for day

200, N=3. Tukey’s multiple comparisons test: no significance. Error bars indicate SEM.

Data for ‘No RA Day 200’ are from the same experiment as in **Fig. 4E**.

(C) Yellow bars indicate temporal windows of RA addition. ‘No RA’ = no RA addition

after day 43. ‘Early RA’ = RA addition on days 43-130. ‘Late RA’ = RA addition on days

130-200.

(D) *M-opsin* (blue) and *L-opsin* (pink) expression in organoids grown in ‘No RA’, ‘Early

RA’, and ‘Late RA’ conditions. White outlines indicate *M*- or *L-opsin*-expressing cells.

(E) Quantification of D. For ‘No RA’, N = 3; for ‘Early RA’, N=3; and for ‘Late RA’, N=5.

Tukey’s multiple comparisons test: ‘No RA’ *L-opsin* versus ‘Early RA’ *L-opsin* $P <$

0.0001; ‘Early RA’ *L-opsin* versus ‘Late RA’ *L-opsin* $P < 0.0001$. Error bars indicate

SEM. Data for ‘No RA Day 200’ are from the same experiment as in **Fig. 4B**.

(F) Summary of M:L cone ratios in human retinas and retinal organoids.

(G) Model for temporal specification of human cone cell subtypes.

Supplemental Fig. 1. *M-opsin* and *L-opsin* mRNA expression in retinoblastoma cells and human retinal organoids.

(A) Alignment of human *M-opsin* and *L-opsin* mRNA. Red bar = nucleotide difference.

(B to D) *M-* and *L-opsin* expression. Heatmap indicates total pileup count normalized to total read count. Each square indicates one nucleotide difference in *M-* or *L-opsin*.

(B) Control- and T3-treated Weri-Rb-1 retinoblastoma cells (N=1 for each timepoint).

(C) Human ESC-derived retinal organoids. RA treatment = 'Early' days 43-130. (N=3).

(D) Human iPSC-derived retinal organoids. RA treatment = 'Early' days 43-130. (N=3 for each timepoint; except day 20 and 158 where N=2) analyzed from ³⁶.

Supplemental 2

List of genes in association study and gene coordinates

Gene of Interest	chr#	Start	End
DIO1	chr1	53879779	53926574
PIAS3	chr1	145832505	145876560
RXRG	chr1	165346715	165499575
OTX1	chr2	63035597	63131383
NEUROD1	chr2	181647107	181693113
THRB	chr3	23970618	24697919
NR2F1 (COUP-TFI)	chr5	93319359	93628069
OPN1SW	chr7	128759936	128801444
RORB	chr9	74405551	74732495
CYP26A1/C1	chr10	93049490	93167890
NRL	chr14	24068098	24141275
OTX2	chr14	56755757	56970833
DIO2	chr14	80137530	80506478
DIO3	chr14	101447117	101623452
ONECUT1	chr15	52686989	52880012
NR2E3	chr15	71773279	71832289
NR2F2 (COUP-TFII)	chr15	96074584	96352953
RARA	chr17	40294657	40370655
ONECUT 2	chr18	57358933	57558283
SALL3	chr18	78908993	79079394
LW/MW gene region	chrX	154087755	154305671

894 **Supplemental Fig. 2. Genes tested in association study.**

895 Gene names and corresponding chromosome coordinates for assembly hg38.

Supplemental 3

A

C

B

D

E

F

Supplemental Fig. 3. Association study

(A) Minor (G) allele ratio by ancestry.

(B) Association of L:M ratio and rs372754794 restricted to ancestry.

(C) Local LD plot for rs37275494 restricted to African American (ASW) populations.

(D) Association of L:M ratio and rs36102671 restricted to ancestry.

(E) rs36102671 association with altered splicing of *RARA* in whole blood from the Genotype Tissue Expression Project (GTEx).

(F) Human fetal retina expression for *RARA* from ⁴¹. CPM = log counts per million.

Supplemental 4

M- and L-opsin staining

Transfection Condition		M +	L +	Both +
	Neg. Control	0	2	0
	M-opsin	16,477	3	0
	L-opsin	3	7,047	0
	M- or L-opsin	230	463	0
	M- and L-opsin	1,159	608	2,007

904 **Supplemental Fig. 4. Quantification of HEK293 cells expressing *M-opsin* and/or *L-***
905 ***opsin* expression. Quantification of Fig. 3B-F (N=3 for each condition).**

Supplemental 5

906 **Supplemental Fig. 5. Human retinal organoid protocol.** Standard conditions for human
907 retinal organoid differentiation, adapted from ³⁶.

Supplemental 6

Supplemental Fig. 6. Total M/L cone densities per organoid.

(A-B) Quantification of total M/L cones per organoid. RA treatment conditions and organoid age (as in **Fig. 4**) are listed on the X axis. Error bars indicate SEM. Individual circles represent individual trials. 'No RA Day 200' repeated on both graphs. No significant difference using Tukey's multiple comparisons test.