The genetics of normal and defective color vision

Jay Neitz*, Maureen Neitz

University of Washington, Dept. of Ophthalmology, Seattle, WA 98195, United States

ARTICLE INFO

Article history:
Received 3 July 2010
Received in revised form 25 November 2010
Available online xxxx

Keywords:
Color vision
Cone photoreceptor
Colorblindness
Cone mosaic
Opsin genes
Evolution
Comparative color vision
Cone photopigments
Circuitry

ABSTRACT

The contributions of genetics research to the science of normal and defective color vision over the previous few decades are reviewed emphasizing the developments in the 25 years since the last anniversary issue of Vision Research. Understanding of the biology underlying color vision has been vaulted forward through the application of the tools of molecular genetics. For all their complexity, the biological processes responsible for color vision are more accessible than for many other neural systems. This is partly because of the wealth of genetic variations that affect color perception, both within and across species, and because components of the color vision system lend themselves to genetic manipulation. Mutations and rearrangements in the genes encoding the long, middle, and short wavelength sensitive cone pigments are responsible for color vision deficiencies and mutations have been identified that affect the number of cone types, the absorption spectra of the pigments, the functionality and viability of the cones, and the topography of the cone mosaic. The addition of an opsin gene, as occurred in the evolution of primate color vision, and has been done in experimental animals can produce expanded color vision capacities and this has provided insight into the underlying neural circuitry.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Elucidating the neural machinery underlying normal and defective color perception is a difficult problem. It requires an understanding of mechanisms and events at the level of genes, proteins, neurons and circuits; all of which are relatively inaccessible. Most of the available information 25 years ago was measurements of the perceptual responses to visual stimuli and electrical responses of neurons in the visual pathway. The questions in color vision concern the mechanisms responsible for the appearance, detection and discriminability of stimuli of varying wavelength composition. The amount of information about the biology underlying color vision has been greatly increased through the application of the tools of molecular genetics; however, our understanding is still incomplete. In this review, we have made an effort to highlight new results that have come to light. Where possible, we have included new hypotheses that seek to incorporate new discoveries of the last few decades and we have tried to indicate future research directions.

This review focuses on the impact of genetic research on understanding mechanisms of normal and defective color vision. For other aspects of color vision, please refer to the other review articles in this issue. In addition, for a recent review emphasizing human psychophysics, we refer the reader to Stockman and Brainard (2010), for one highlighting primate physiology, Solomon and Lennie (2007), and for comparative primate color vision, Jacobs (2008).

2. Genes and photopigments

In 1986, Nathans and colleagues isolated and sequenced the genes encoding the human long wavelength (L), middle wavelength (M) and short wavelength (S) sensitive cone opsins and took the first steps toward testing the long-held, two-part hypothesis that (1) variation in the amino acid sequences of the cone opsins are responsible for the spectral differences among the photopigments that all share the same 11-cis retinal chromophore, and (2) alterations in the cone opsin genes underlie inherited color vision deficiencies (Nathans, Piantanida, Eddy, Shows, & Hgness, 1986; Nathans, Thomas, & Hgness, 1986). Findings from these studies both confirmed what previous genetic studies had suggested, and they produced several surprises.

As predicted by inheritance patterns of red–green (Waaler, 1968) and blue–yellow (Kalmus, 1955) color vision deficiencies, the genes for human long-wavelength (L) and middle-wavelength (M) cone opsins were localized to the X-chromosome at Xq28, and the gene for the short-wavelength (S) cone opsin to an autosome, chromosome 7 at 7q32 (Nathans, Piantanida et al., 1986). The official genetic designations for the L, M and S opsin genes are OPN1LW, OPN1MW, and OPN1SW, respectively. OPN1LW and OPN1MW are arranged in a tandem array (Nathans, Thomas et al., 1986; Vollrath, Nathans, & Davis, 1988). Among individuals with normal color vision there is variability in the number of
OPN1LW and OPN1MW genes per X-chromosome array, with more variability in the number of OPN1MW than in OPN1LW genes; thus, contrary to expectation, most people with normal color vision do not have just one L and one M gene (Drummond-Borg, Deeb, & Motulsky, 1989; Macke & Nathans, 1997; Nathans, Plantanida et al., 1986; Nathans, Thomas et al., 1986; Neitz & Neitz, 1995; Neitz, Neitz, & Grishok, 1995; Ueyama et al., 2001). OPN1LW and OPN1MW are nearly identical to one another, sharing more than 98% nucleotide sequence identity, whereas they share only about 40% nucleotide sequence identity with OPN1SW, indicating that OPN1LW and OPN1MW arose via a relatively recent gene duplication (Nathans, Thomas et al., 1986). Because of their similarity, the L and M opsin genes are prone to unequal homologous recombination, which has profound implications for visual function.

When Nathans and colleagues began their pioneering work, it was expected that all people with normal color vision would share the same L and the same M pigment. However, as a consequence of recombination that has intermixed the L and M opsin genes, there is variation in the amino acid sequences of both L and M opsins among individuals with normal color vision (Neitz, Neitz, and Grishok, 1995, Winderickx, Battisti, Hibibya, Motulsky, & Deeb, 1993). Several studies have identified amino acid differences that shift the spectral peaks of the L and M cone photopigments and have correlated color vision behavior to variation in the L and M opsin genes (Neitz, Neitz, & Jacobs, 1995; Neitz, Neitz, & Kainz, 1996; Neitz et al., 2004; Sanocki, Shevell, & Winderickx, 1993; Sharpe et al., 1998; Shevell, He, Kainz, Neitz, & Neitz, 1998; Winderickx, Lindsey et al., 1992).

All eutherian mammalian pigments share the same 11-cis retinal chromophore (Wald, 1967, 1968). It had been pretty well agreed that binding of the chromophore to opsin red-shifted the chromophore’s absorption spectrum, and that amino acid sequence differences among the opsins were responsible for the spectral characteristics of each of the cone pigments (Chen et al., 1989; Kosower, 1988; Wald, 1967). More recent technical innovations made it possible to measure spectral sensitivities of individual cone classes (Baylor, Nunn, & Schnapf, 1987; Dartnall, Bowmaker, & Mollon, 1983; Kraft, Neitz, & Neitz, 1998; Schnapf, Kraft, & Baylor, 1987) and to evaluate the effects of amino acid sequence differences on spectral sensitivity (Asenjo, Rim, & Oprian, 1994; Carroll, Neitz, & Neitz, 2002; Merbs, 1992; Merbs & Nathans, 1992, 1993; Neitz, Neitz, & Jacobs, 1995b; Sharpe et al., 1998; Stockman, Sharpe, Merbs, & Nathans, 2000).

Fig. 1 summarizes what is known about spectral tuning of the human L and M cone pigments. The L and M opsin genes each have six exons. The first and sixth exons do not typically vary among or between L and M opsin genes. Exon 5 encodes amino acid dimorphisms at positions 277 and 285 that together are responsible for the majority of the spectral difference between human L and M pigments. Exons 2, 3 and 4 also encode amino acid dimorphisms that produce additional smaller spectral shifts. There is considerable normal variability in the amino acid sequences of the L and M pigments, which in turn produces variability in the absorption spectra (Fig. 1A and B). The shortest normal M pigment variant is the most common. There is more normal variability in the L pigment and versions with one of two different spectral sensitivities occur with high frequency (Neitz & Jacobs, 1986). The longest L pigment, with a peak of approximately 559 nm, is somewhat more common than the version with the slightly shorter spectral peak (555.5 nm). The identities of the seven spectral tuning codons for the gene encoding the longest normal L and shortest normal M pigment are shown in Fig. 1C along with the consequences for spectral sensitivity of substituting each exon from the L pigment into the M pigment and vice versa. Substituting L exons individually into the M pigment tends to produce smaller spectral shifts compared to substituting M exons into L pigments. Exon 2 encoded differences

![Fig. 1. Spectral tuning of human L and M cone photopigments. (A) L-class pigments absorb maximally near 560 nm. (B) M-class pigments absorb maximally near 530 nm. (C) Genes encoding the L and M opsin genes each have six exons represented by narrow white bars and numbered 1–6. The colored regions indicate the introns in the L and M opsin genes. The genes are drawn to scale. Codons that specify amino acids involved in spectral tuning are indicated using the single letter amino acid code as follows: Y = tyrosine, T = threonine, A = alanine, I = isoleucine, S = serine, F = phenylalanine. The magnitude of the spectral shift in nanometers (nm) produced by the indicated amino acid differences specified by each exon are indicated on the far right and far left.](image-url)
shift the spectra of human L pigments, but do not measurably shift the peak of M pigments. Exons 3 and 4 encode differences that shift the spectra of both human L and M pigments, with relatively slightly larger shifts when substitutions are made in L compared to M pigments (Asenjo et al., 1994; Merbs & Nathans, 1992).

Seven amino acids, all encoded by exon 5, are typically different between human L vs. M opsins, of which 3, indicated in Fig. 1C, are involved in spectral tuning (Asenjo et al., 1994; Merbs & Nathans, 1992; Neitz, Neitz, & Jacobs, 1991). L pigments have the amino acid tyrosine (Y) at positions 277 and 309, and threonine (T) at position 285; M pigments have phenylalanine (F) at positions 277 and 309, and alanine (A) at 285. Thus, the identity of exon 5 separates the pigments into two clusters, the L class and the M class (Fig. 1A and B). Depending on the amino acids at other tuning sites, pigments with Y277 and T285 have peak sensitivities as long as 559 nm and pigments with F277 and A285 have peak sensitivities as short as 530 nm. Together, the substitutions at positions 277 and 285 account for about 20 nm of the difference in peak sensitivity of human L vs. M pigments. F vs. Y at position 309 shifts the peak sensitivity by only about 1 nm (Asenjo et al., 1994).

Exons 2, 3 and 4 all encode amino acid differences responsible for variability within L-class and M-class pigments. Exon 2 encodes 3 amino acid polymorphisms at positions 65, 111, and 116. The dimorphism at position 116 is the spectrally active one of the three (Asenjo et al., 1994). L pigments with serine at position 116 (S116) are red-shifted by 2–3 nm compared to those with tyrosine. As introduced above, there are context effects such that size of the spectral shift produced by a particular substitution depends on the identity of the other amino acids in the pigment. Thus, even though a T116S substitution in an L-class pigment produces a shift, an S116T substitution in the M pigment does not produce a measureable shift.

Of the 5 dimorphic amino acid positions specified by exon 3, the most variable exons, the only serine (S) vs. alanine (A) at position 180 produces a measureable shift in the spectrum. Pigments with S180 have peak sensitivities at longer wavelengths relative to pigments with A180. As with all the amino acids that produce variability within L and M pigment classes, the absolute magnitude of the spectral shift is a little larger when an S180A substitution occurs in an L pigment vs. when an A180S substitution occurs in an M pigment (Asenjo et al., 1994; Merbs, 1992; Neitz et al., 1991). The variability in normal L pigments, introduced above, is evident in Rayleigh color matches (Neitz & Jacobs, 1986) and it is principally the result of the S/A difference at position 180 of the L pigment (Neitz, Neitz, & Jacobs, 1993; Piantanida & Gille, 1992; Wenderickx, Lindsey et al., 1992).

Exon 4 specifies dimorphisms at three amino acid positions, 230, 233 and 236, two of which have an identified role in spectral tuning. Pigments with isoleucine at position 230 and alanine at position 233 are red shifted compared to pigments with threonine at 230 and isoleucine at 233, and the magnitude of the spectral shifts are larger in L than in M pigments (Asenjo et al., 1994; Carroll et al., 2002; Merbs, 1992).

Because OPN1LW and OPN1MW genes are on the X-chromosome, females who are heterozygous, for example, having OPN1LW genes that encode two spectrally distinct L pigments, would have four different cone types – two different L, plus M and S – and thus the potential for having tetrachromatic color vision (Bosten, Robinson, Jordan, & Mollon, 2005; Jordan & Mollon, 1993, 1997). Prior to the discovery of variation in peak sensitivity of pigments underlying normal color vision, it had been appreciated that female carriers of a red–green color vision deficiency have the potential for tetrachromatic color vision (Nagy, MacLeod, Heyman, & Eisner, 1981). However, because of the normal variability in the L and M pigments more than half of all women express more than three spectrally different photopigments (Neitz, Kraft, & Neitz, 1998). There has been a persistent fascination with the possible existence of tetrachromatic females over the last quarter of a century, but experimental evidence has been mostly negative. Jordan and colleagues (Jordan, Deeb, Bosten, & Mollon, 2010) have conducted a study in which they were able to analyze DNA from nine subjects to confirm the genetic potential for tetrachromacy. Intriguingly, 1 of 24 obligate carriers of deuteranomaly exhibited tetrachromatic behavior on all their tests. However, they conclude that the overwhelming majority of carriers of color anomaly who have the potential to have four cone subtypes, do not exhibit four-dimensional color vision and that it is unlikely that anomalous trichromacy is maintained by an advantage to the carriers in discriminating colors. Human females who have an extra pigment most commonly have two L pigments, one with S180 and another with A180, in addition to S and M pigments. In female platyrhine monkeys, having two L pigments with this difference compared to having only a single X-encoded pigment is associated with very clearly demonstrable trichromacy vs. dichromacy (Rowe & Jacobs, 2007). It appears that the addition of an extra subpopulation of cones to a dichromat has a more dramatic effect on color vision than making the same addition to an already trichromatic retina. As discussed in later sections, the evolution of trichromacy from dichromacy may have occurred through opportunities afforded by specific features of the pre-existing circuitry in dichromats. This same circuitry in trichromats may be nearer its limits for supporting additional color vision capacity, and thus, in turn, may not lend itself so readily to tetrachromacy.

2.1. The genetics of color vision defects

Protan, deutan and tritan defects are characterized by the absence of a contribution to vision from L, M and S cones, respectively. Twenty-five years ago, the accepted model for the genetics of red–green color vision deficiencies (Piantanida, 1974; Pokorny & Smith, 1977) held that the dichromatic forms, protanopia and deuteranopia, were caused by the replacement of the normal M gene with one that encoded an L pigment for deuteranopia and the replacement of the normal L gene with one that encoded an M pigment for protanopia. These gene replacements were thought to cause the inappropriate expression of M opsin in L cones for protanopia, and of L opsin in M cones for deuteranopia. Anomalous trichromatic forms were thought to be caused by genes for anomalous photopigments replacing either the normal L or M pigments. Replacing one normal pigment with an anomalous one resulted in having peak sensitivities that were closer together than the peak sensitivities of the normal L and M cones. For each of the anomalous trichromacies, protanomaly and deuteranomaly, two forms, simple and extreme, were proposed to differ in the magnitude of the spectral difference between the anomalous pigment and the normal one with anomalous pigments being more shifted from normal in extreme forms.

Prior to the molecular genetics work, the genes for the L and M cone pigments were thought to exist at independent loci with another gene, glucose-6-phosphate dehydrogenase, located between them. It was believed that allelic diversity for the two independent genes was the cause for unrelated protan and deutan color vision deficiencies. To everyone’s great surprise, results from molecular genetics revealed that the L and M opsin genes were adjacent to one another with no intervening genes and the high frequency of recombination between these two genes is responsible for many peculiarities of red–green color vision including the extraordinarily high frequency of color vision defects.

Inherited color vision deficiencies can be explained by gene rearrangements that arose through unequal homologous recombination in females during meiosis (Nathans, Piantanida et al., 1986), as illustrated in Fig. 2. The DNA between the L and M genes is
opsins that will form L-class pigments. Which is determined by the spectral differences specified by the genes encoding the opsins. The second array produced has two L genes as its first gene in the array. The products are an X-chromosome with one opsins gene and a second with a tandem array of three opsins genes. A majority of human deuteranopes are “single gene dichromats” having been reduced to an L gene as the only opsins gene on the X-chromosome. The most frequent arrangement of opsins in humans with normal color vision is to have one L and two M genes arranged as shown in Fig. 2A. There are about 50 times more people with 2 M and one L genes than there are deuteranopes with one L gene, even though they would be produced in equal numbers from ancestors with one L and one M gene by the crossover mechanism. Selective pressure against color blindness, particularly in ancestors to modern humans, could explain the much lower frequency of deuteranopes.

Because the L and M genes are adjacent to each other on the X-chromosome and they are nearly identical, an L gene from the paternal chromosome can align with the M gene from the maternal chromosome, as shown in Fig. 2B. When X-chromosomes misalign, a crossover within the L gene on one X-chromosome and an M gene on the other X-chromosome produces two new arrays, each of which will cause a color vision defect when inherited by a male. One will cause a protan defect, the other will cause a deutan defect. In the protan-causing array the one remaining opsins gene is a hybrid between the parental L and M genes. As long as the hybrid has exons 5 from a parental M gene, the encoded photopigment will fall into the M-class (Fig. 1B). A male with a normal S-pigment and one X-chromosome pigment gene encoding an M pigment is an obligate protanope. The array associated with deutan color vision deficiency has a parental L gene as the first gene in the array. The second gene is a hybrid while the third gene is a parental M opsins gene. This array structure is the one most commonly found in deuteranomalous males, and it represented one of the most unexpected findings that has come from examining the molecular genetics associated with color vision deficiencies (Drummond-Borg et al., 1989; Nathans, Piantanida et al., 1986; Neitz, Neitz, & Kainz, 1996; Shevell et al., 1998). An initial explanation offered for the presence of apparently normal, intact M opsins genes in a color vision defect characterized by the absence of functional M cone contribution to vision was that the deutan hybrid gene and the normal M gene were co-expressed in the same cone, thereby shifting the spectral sensitivity of the “M-cone” toward that of the L cone. It was further hypothesized that the specific nature of the hybrid gene was important in determining the severity of the color vision defect (Nathans, Piantanida et al., 1986). Understanding why a gene array like that illustrated in Fig. 2B confers a color vision defect when both normal L and M genes required for normal color vision are present requires information about the mechanism responsible for expression of the X-chromosome opsins genes. The key to understanding the color vision genotype in this case has come from studies demonstrating that, except in rare cases (Sjoberg, Neitz, Balding, & Neitz, 1998), only the first two genes in the array are expressed (Bollinger, Sjoberg, Neitz, & Neitz, 2004; Hayashi, Motulsky, & Deeb, 1999; Neitz, Bollinger, & Neitz, 2003; Windericks, Battisti, Motulsky, & Deeb, 1992; Yamaguchi, Motulsky, & Deeb, 1997).

Research relating genotypes to phenotypes uncovered another unexpected finding – in humans there has been a tremendous amount of intermixing of L and M opsins genes. In the classical view of color blindness genetics, a deuteranomalous allele of the M pigment gene would have arisen by a single event in which a normal M gene was mutated. Thus, 94% of the males were envisioned to have a normal M gene and 6% had mutants. Similarly, 98% of males...
would have the normal L gene and 2% had mutants. However, the homology of photopigment genes and their arrangement in a tandem array allows for a high mutation rate. Thus, instead of being produced by a single mutation, many of the genotypes associated with color vision deficiencies are the result of a series of mutational steps away from ancestral gene arrays, with one normal L and one normal M gene. The generation of a gene array associated with protanomaly, as illustrated in Fig. 2C, is an example. In order to have the basis for protanomalous trichromacy as opposed to protanopia, protanomalous individuals must have at least two M genes encoding opsins that form two different pigments of the M-class (Fig. 1), but lack an L gene (as in Fig. 2C). There is no way to produce an array with multiple M and no L genes from parental arrays with one L and one M. One parent must have at least two M genes, as produced by the rearrangement in Fig. 2A. A subsequent crossover (Fig. 2C) can produce a gene array with multiple genes with different spectral sensitivities falling in the M class.

We assume that the present variety of opsins genes arose from an ancestral arrangement with one normal L and one normal M gene. Presumably, the misalignment required for a gene rearrangement involving an unequal crossover is a very rare event and, initially, the rate of mutations producing genetic variety would have been relatively low. However, once variety in gene copy number was introduced by early rearrangements, the generation of females heterozygous for copy number would have greatly accelerated the rate of rearrangement. As shown in Fig. 2C, for a female with one array with two genes and a second array with three genes, there is no perfect alignment. One of the two possible alignments pair an L with an M gene and if a crossover occurs when the genes are aligned in this way a gene rearrangement will result. Thus, in females with X-chromosomes with unequal opsin gene numbers, there would be misalignments in 50% of the meioses that would produce a gene rearrangement if a crossover occurs within the gene array.

There is evidence that a remarkable amount of rearrangement has occurred in the L and M opsin genes over the course of human history. The two most obvious indicators are the variability in copy number among the L and M genes and the polymorphism at position 180 in which nearly half of normal males in some samples have A180, the amino acid presumed to be characteristic of the primordial M gene. In addition to position 180, variants at other spectral tuning sites in exons 2 and 4 have also been identified among males with normal color vision. It appears that all possible spectral types within the L-class and within the M-class shown in Fig. 1 occur as normal variants. This is in striking contrast to the classic view, in which people with protan defects were viewed to have “the” normal M and a spectrally shifted “anomalous” L pigment and those with deutan color vision defects were believed to have “the” normal L and a spectrally shifted “anomalous” M pigment and the severity of the color vision deficiency was believed to be determined solely by the nature of the anomalous pigment.

Assuming that, at one time, the majority of human ancestors had the typical Old World primate arrangement of one L and one M opsin gene on the X-chromosome, a cladistic method of classifying species into groups (clades) based on gene sequences can be used to determine the most likely amino acid sequences encoded by a single ancestral L and a single ancestral M opsin gene. Comparison of the deduced amino acid sequences of L and M opsin genes, for example among Caucasian males with normal color vision, reveals that relatively few individuals have genes encoding the ancestral opsins. The majority of males with normal color vision have L genes that encode some amino acids of the ancestral M opsin and M genes that encode some of the amino acids of the ancestral L opsins (Verrelli & Tishkoff, 2004). Furthermore, there is complete overlap between hybrid pigments mediating color vision in people with normal color vision and in people with red-green color vision deficiencies, although the frequencies of some hybrids are higher in color deficiency compared to normal trichromacy (Carroll et al., 2002; Crognaele, Teller, Motulsky, & Deeb, 1998; Neitz et al., 1996; Sharpe et al., 1998; Winderickx et al., 1993).

Prior to knowledge of the cone opsin gene sequences and the molecular mechanism responsible for spectral tuning, a key to understanding color anomaly appeared to be the characterization of the anomalous pigments (DeMarco, Pokorny, & Smith, 1992), which, together with the putative normal L and M cone sensitivity functions, would comprise the characterization of the photopigment basis of anomalous trichromacy. The variability in normal pigments makes understanding the photopigment complement of color anomalous individuals more complicated; however, the discovery that the absorption spectra of primate photoreceptors assume a common shape when plotted on a normalized wavenumber axis (wavenumber divided by wavenumber of maximum sensitivity) (Baylor et al., 1987; Lamb, 1995; Mansfield, 1985), has greatly simplified the characterization of variant cone pigments. The implication is that all primate photopigments have a common shape that can be used to completely characterize the spectral properties of any human photopigment variant just from knowing its wavelength maximum. A comparison of the single pigment template curve (from Carroll, Bialozyński, Summerfelt, Neitz, & Neitz, 2000a, 2000b) to human cone fundamentals derived from color matching functions (Stockman, MacLeod, & Johnson, 1993) is shown in Fig. 3, and there is generally good agreement over the entire spectrum when values of 420 nm, 530 nm and 557.5 nm are used for the spectral peaks of the S, M and L pigments, respectively.

Fig. 3. Spectral sensitivities of human L, M and S cones plotted on a scale that is uniform in units of log of wavenumber. On this scale, all photopigments assume a common shape, described by a template curve (solid lines). The curve together with information about the spectral positions of the cone photopigments can be used to completely describe the photopigment basis for color vision in any individual. The template was derived by fitting an equation given at www.neitzvision.com to an amalgam of photopigment spectral sensitivity curves (Carroll, McMahon, Neitz, & Neitz, 2000). (A) The spectral peak of the template has been adjusted to fit cone fundamentals derived from color matching (Stockman & Brainard, 2010). (B) All the curves from (A) have been shifted to a best fit, illustrating the close similarity between the shapes of the L, M and S spectral sensitivities and the template. The slight differences in psychophysically derived fundamentals may derive, in part, from variation in the normal cone pigments.
Over recent years, there has been the most uncertainty about the spectral peak of the S cone. However, the results in Fig. 3 show that the spectral peak obtained from human psychophysics agrees pretty well with those from MSP (419 nm) (Dartnall, Bowmaker, & Mollon, 1983) and from reconstituting the human S-pigment in vitro (Fasick, Lee, & Oprian, 1999). We note that, theoretically, the human cone fundamentals derived from color matching functions should all have the identical shape when corrected for preretinal absorption and plotted on a log of wavenumber axis, and that shape should agree perfectly with the universal pigment template after appropriate corrections for optical density are made. The agreement is good, but not perfect. The discrepancies are likely due, in part, to the fact that the human cone fundamentals reflect average data from a number of humans who have different normal variants of M and L pigments. This might be expected to tend to broaden the cone fundamental curves. The fact that the human L fundamental is broader than either the human L or M curve is consistent with that idea, given that the normal human L pigments are the most variable.

Taken together, the spectral tuning data from Fig. 1 and the photopigment template curve of Fig. 3, it should be possible to completely characterize the photopigments of any person just from examination of the sequences of the photopigment genes they express. There is very good agreement between the deduced photopigment complement and color vision capabilities of people with color vision defects (Barbur, Rodriguez-Carmona, Harlow, Mancuso, Neitz & Neitz, 2008; Neitz et al., 1996, 2004; Shevell et al., 1998). An implication is that the technology is available to allow conventional color vision testing to be replaced by a genetic test.

The rest of this section summarizes what has been learned about genotype/phenotype relationships in red–green color vision deficiencies. Proton defects are characterized by the absence of an L cone contribution to vision. X-chromosome opsin gene arrays underlying the dichromatic form, protanopia, either have a single opsin gene that encodes an M pigment, or have multiple genes in which the first two encode opsins that produce M pigments that are identical in spectra (Jagla, Jägle, Hayashi, Sharpe, & Deeb, 2002; Nathans, Piantanida et al., 1986; Neitz, Neitz, He, & Shevell, 1999; Neitz et al., 2004). The anomalous trichromatic form, protanomaly, is associated with opsin gene arrays in which the first two genes encode opsins that produce M pigments that differ in spectra. Most commonly, the spectral separation in the M pigments is the result of amino acid differences at spectral tuning sites that shift the peak sensitivity of one pigment relative to the other. Ami-no acid substitutions have a generally smaller effect on the spectral sensitivity of M pigments relative to L pigments (Fig. 1), thus, individual protans are likely to have a smaller spectral separation between their two pigments drawn from the M-class compared to deutsans (described below), who draw two pigment from the L-class. This can explain why, as a group, protanomalous observers have somewhat poorer color discrimination than deuteranomalous observers. In addition to the amino acids that shift the spectrum, protanomalous color vision may also arise from amino acid differences that do not shift the relative peak sensitivity of the underlying M pigments, but instead increase the optical density of one pigment relative to the other (Neitz et al., 1999). In the latter case, color vision becomes dichromatic under conditions that bleach the pigments and equalize their optical densities.

Deutan defects are characterized by the absence of an L cone contribution to vision. X-chromosome opsin gene arrays underlying the dichromatic form, deuteranopia, either have a single opsin gene that encodes an L pigment, or have multiple genes in which the first two encode opsins that produce L pigments that are identical in spectra (Neitz et al., 1996). The anomalous trichromatic form, deuteranomaly, is associated with arrays in which the first two genes encode opsins that produce L pigments that differ in spectral sensitivity. Among individuals with deuteranomalous color vision, there is variation in phenotypes. For example, on the Hardy Rand and Rittler pseudoisochromatic plate test, in which the stimuli have a range of difficulty giving a measure of severity, deuteranomalous performance ranges from nearly normal to nearly as poor as a dichromat. There is a very good correlation between deuteranomalous behavior in color vision tests and the magnitude of the spectral separation between the underlying L pigments; however, in some cases when people have very small spectral differences between their pigments, discrimination performance as measured by the range of anomaloscope settings accepted as a match is better than might be predicted from the spectral separation of the underlying pigments (Barbur et al., 2008; Neitz et al., 1996, 2004; Shevell et al., 1998).

In recent years, it has been discovered that deleterious combinations of amino acids at the dimorphic positions can be produced as a consequence of inermixing of L and M opsins genes. One such combination at the exon 3 encoded amino acid positions is Leucine 153, Isoleucine 171, Alanine 174, Valine 178 and Alanine 180, abbreviated LIAVA (Barroll, Neitz, Hofer, Neitz, & Williams, 2004; Crognael et al., 2004; Mizrahi-Meissonnier, Merin, Banin, & Sharon, 2010; Neitz et al., 2004). The LIAVA combination is associated with the absence of cone function when specified either by an L or an M opsin gene. Males whose L opsin contains this deleterious combination are protanopes, males whose M opsin contains this combination are deuteranopes, males who have a single opsin gene on the X-chromosome that specifies this combination or for whom the first two opsin genes in the array specify this combination are blue cone monochromats. There is a perfect correlation between the opsin gene specifying this combination and color vision phenotype, despite the fact that the arrays differ dramatically in the number of opsins genes, and in the identity of the gene that specifies the LIAVA combination. This, together with the observations that the LIAVA combination has never been found in a person without vision problems and it has been found on a variety of genetic backgrounds, indicates that it is the causative mutation and rules out the possibility that the phenotype is caused by an unidentified mutation in linkage disequilibrium. Whether color vision defects caused by this deleterious combination are congenital or whether they develop over time as the photoreceptor becomes progressively non-functional is an open question.

A less frequent cause of inherited red–green color vision deficiencies has arisen through conventional mutational mechanisms that produce relatively rare, random mutations that cause the corresponding gene to encode an opsin that fails to form a functional photopigment or that prevents the gene itself from being transcribed (Bollinger, Bialozymski, Neitz, & Neitz, 1999; Carroll et al., 2010; Jagla et al., 2002; Neitz et al., 2004; Winderickx, Sanocki et al., 1992). The most common such mutation is the replacement of the cysteine residue at position 203 with the amino acid arginine (C203R).

Inherited tritan color vision deficiencies are rare in comparison to the inherited red–green color vision defects and they have an autosomal dominant inheritance pattern (Kalmus, 1955). In addition, the inheritance pattern shows incomplete penetrance meaning that some people who have the causative gene do not have the phenotype. Weitz and coworkers (Weitz, Went, & Nathans, 1992; Weitz, Miyake et al., 1992) showed that inherited tritan defects were caused by mutations in the 5 opsin gene that produce amino acid substitutions. They identified three amino acid substitutions in the 5 opsin gene as causing tritan defects: arginine substituted for glycine at position 79, proline substituted for serine at position 209, and serine substituted for proline at position 264. More recently, two new mutations – replacement of arginine at position 283 with glutamine, and replacement of leucine at position

56 with proline – have been found as causes of tritan defects (Baraas et al., 2007; Gunther, Neitz, & Neitz, 2003, 2006).

The genetic underpinnings of two more rare forms of inherited color vision loss have been elucidated. One, termed blue cone monochromacy or incomplete achromatopsia, results from mutations that prevent proper transcription of the X-chromosome cone opsin genes, or renders the encoded opsins non-functional, as will be discussed in detail below. The second is complete achromatopsia and is caused by mutations in genes that are normally expressed in all three cone types (Kohl, Jägle, Sharpe, & Wissinger, 2009; Thiadens et al., 2010). The most common mutations interfere with normal function of the cone photoreceptor cyclic GMP gated ion channel, which has two subunits, encoded by separate genes (CNGB3 and CNGB2). Mutations in the genes encoding the phototransduction proteins cone transducin (GNAT2 gene) and phosphodiesterase 6C (PDE6C) have also been found in association with achromatopsia. The most common cause of achromatopsia in humans is mutations in the CNGB3 gene, and a recent study using gene therapy in a canine model of a CNGB3 defect showed rescue of cone photoreceptor function (Komáromy et al., 2010).

3. Genes and the cone photoreceptor mosaic

The question of what governs the organization of the three cone types in the retinal cone mosaic is central to understanding the circuitry for color vision and evolutionary strategies for optimizing the mosaic organization for extracting visual information. Over the last 25 years, several developmental studies in humans and other Old World primates have examined this question (Bumsted & Hendrickson, 1999; Bumsted, Jasoni, Szél, & Hendrickson, 1997; Curcio, Sloan, Kalina, & Hendrickson, 1990; Curcio, Sloan, Packer, Hendrickson, & Kalina, 1987; Wikler & Rakic, 1991; Xiao & Hendrickson, 2000). In humans, immunohistochemistry has shown that S opsin is first detected in the fovea at about fetal week 10.9 and S cones cover the entire retina by about fetal week 19 (Bumsted & Hendrickson, 1999; Xiao & Hendrickson, 2000). The ability to detect L and M opsin immunologically occurs quite a bit later, with L/M cones first detected in the central retina at about fetal week 21.5, and extending over the entire retina by fetal weeks 34–37 (Xiao & Hendrickson, 2000). Because the L and M opsins are more than 96% identical in amino acid sequence, antibodies that recognize one, recognize both. In situ hybridization experiments in fetal human retinas using nucleic acid probes to label S and LM opsin mRNA give similar results with mRNA being detectable just shortly before opsin protein is detectable. As for antibody labeling, in situ hybridization methods are incapable of distinguishing between L and M cones. The significant lag in the appearance of LM vs. S opsin protein and mRNA during development suggests that differentiation of S cones is independently controlled from LM cones; however, because these methods do not distinguish between L and M cones, they shed no light on whether L and M cones are independently controlled.

Insight into the mechanism of differentiation of L vs. M cones has come from studies of New World primates in which trichromatic color vision evolved through allelic diversity of a single X-chromosome cone opsin gene locus rather than a gene duplication that placed two cone opsin genes on the same X-chromosome (Jacobs, Neitz, & Neitz, 1993). For example, in squirrel monkeys, there are three alleles of the X-chromosome cone opsin gene. One encodes an opsin that forms a pigment that is similar in spectral peak to the human L pigment, another is similar to the human M pigment, and a third has a spectral peak that is intermediate between human L and M. All males of the species are dichromatic, having only one X-chromosome, and thus having S cones and a single cone type that absorbs in the middle-to-long wavelengths.

Females have two X-chromosomes, and can either be homozygous or heterozygous for the X-chromosome opsin allele. If homozygous, they are dichromatic; however, females who carry both an allele for an L opsin and one for an M opsin have the equivalent of normal human color vision, having both L and M cones because the process of X-inactivation segregates the expression of the L and M opsin genes to separate populations of cones. The significance of this is that in trichromatic female squirrel monkeys, the difference between L and M cones is solely determined by the stochastic choice of which X-chromosome is retained as the active one. Variation in L:M cone ratio in female squirrel monkeys is similar to what is seen in human males with normal color vision. In squirrel monkeys the variation has been attributed to the stochastic process of X-inactivation, influenced by the number of cells present at the time of activation and other random factors in the inactivation process (Jacobs & Williams, 2006).

In humans, even though both L and M opsin genes reside on the X-chromosome, there is evidence that a stochastic mechanism also determines whether each individual cell expresses L vs. M opsin. In their early work to investigate the genetic mechanisms of blue cone monochromacy, Nathans and colleagues discovered a DNA element upstream of the L-opsin gene that is essential for transcription of the X-chromosome opsin genes (Nathans et al., 1989). The DNA element was given the name Locus Control Region, abbreviated LCR, and it is an enhancer that mediates cell-type specific expression of the X-chromosome opsin genes (Li, Timmers, Guy, Pang, & Hauswirth, 2007; Wang et al., 1992). The LCR is highly conserved and is present in all other mammalian species examined, the vast majority of which have a single X-chromosome opsin gene. Interactions between the opsin gene promoter and the LCR are thought to be required for opsin gene expression (Smallwood, Wang, & Nathans, 2002).

The gene duplication event that ultimately led to there being both L and M opsin genes on the same X-chromosome duplicated an ~40 kilobase pair segment that extends ~450 basepairs upstream of the opsin gene and includes the promoter, and it extends about 18 kilobase pairs downstream of the coding sequence. The LCR was not included in the duplication, and so in humans and other Old World primates, the L and M opsin genes must share the same enhancer, meaning only one of the X-chromosome cone opsin genes can be expressed at a time. Epigenetic modification of the opsin gene locus may play an important role in opsin gene expression and in determination of the fate of a nascent cone photoreceptor as an L vs. an M cone such that, during development, a competition between interactions that promote opsin gene expression and mechanisms that silence all genes that are not expressed as part of the cell’s final differentiated phenotype ultimately leaves each L/M cone with one of the X-chromosome opsin genes actively transcribed, while all others are silenced (Johnston & Desplan, 2008).

Although there is no experimental data demonstrating a role for epigenetic silencing, the observation that the ratio of L:M pigment messenger RNA changes over the course of development (Knoblauch, Neitz, & Neitz, 2006) has led us to suggest a model that can explain features of the topography of L and M cones in the retinal mosaic. We propose that during gestation, each cone cycles randomly between transcribing either the L or M opsin gene. According to the model, during development of the retina, at any given time a nascent cone is transcribing either an L or M opsin gene. At the time a gene is being expressed it is protected from silencing. This increases the probability that it will transcribe the same opsin in future cycles, while at the same time the non-transcribed genes are subject to epigenetic modifications that decrease their probabilities of being expressed in future cycles (Knoblauch et al., 2006). Probabilistic events result in one gene being expressed in each mature cone while all the others are silenced. Epigenetic
changes are passed onto daughter cells when a cell divides. The gene silencing mechanism can explain why the L/M cone ratio increases from the center of the retina out to the periphery (Hagstrom, Neitz, & Neitz, 1997; Neitz, Balding, McMahon, Sjoberg, & Neitz, 2006). The cells in the peripheral retina are born later than those in the center (Bumsted et al., 1997; Xiao & Hendrickson, 2000); if the M pigment genes are more prone to silencing and the epigenetic modifications are ongoing as the retina develops, the probability of expressing an M gene would decrease in the periphery. The heritability of the epigenetic changes introduces a slight nonrandomness compared to the idea that each cone independently makes a completely random choice to express M or L opsins. Cells near one another are the product of related cell divisions and share some epigenetic memory, making them more likely than expected by chance to express the same opsins. This explains the slight clumpiness that has been observed for the human L/M cone mosaic (Hofer, Carroll, Neitz, & Williams, 2005). Genetic changes in non-coding DNA in the region of the opsins genes are expected to influence how prone the DNA is to silencing and polymorphisms upstream of the opsins gene array have been found to be associated with individual differences in L/M cone ratio (Gunther, Neitz & Neitz, 2008; Stamatoyanopoulos et al., 2005). DNA is packaged as loops of chromatin in cells, and the most likely explanation for only two of the X-chromosome opsins genes being expressed is that the loop structures inhibit access of the LCR to opsins genes that are downstream of the first two genes in the array. The rare cases in which more than two genes from the array have been demonstrated to be expressed may be the result of mutations that affect the looping structure. Proximity of the LCR to the first gene in the array has been proposed to explain the greater number of L than M cones in the retinas of some people. However, there is a huge range of L/M cone ratio and in some groups, having more M than L cones is common. Thus, it appears that the details of how the DNA is looped in the nucleus is a more important determinant of the relative access of L vs. M genes to the LCR, and hence the cone ratio, than the linear distances along the DNA (McMahon, Carroll, Awua, Neitz & Neitz, 2008).

A variety of methods ranging from immunohistochemistry in post-mortem eyes to adaptive optics and retina densitometry in living eyes have been used to evaluate the relative number and distribution of S cones vs. L and M cones (Bumsted & Hendrickson, 1999; Hofer et al., 2005; Roorda, Metha, Lennie, & Williams, 2001). At 1 degree eccentric from the fovea, the average percentage of cones that were identified as S cones by adaptive optics was 5.72%, and the S cones are distributed in a relatively regular hexagonal array (Hofer et al., 2005). The relative ratio of L to M cones among individuals with normal color vision is highly variable. The first direct evidence for the relative distribution of L vs. M cones came from adaptive optics and retinal densitometry (Hofer et al., 2005; Roorda & Williams, 1999), and the results confirm reports of variation in the L/M ratio among humans with normal color vision from studies that used indirect methods (Carroll et al., 2002; Rushton & Baker, 1964, Kuchenbecker, Sahay, Tait, Neitz & Neitz, 2008; Mollon & Bowmaker, 1992; Neitz et al., 2006). In summary, the development of S cones and their arrangement in the adult retina compared to L and M cones all point to the identity of S cones as being distinctly different from L and M cones, with S cones being non-randomly distributed. In contrast, the arrangement and ratio of L and M cones are consistent with their arrangements being determined by a stochastic process such that the L and M cones represent a single receptor population differing only by which X-chromosome opsin gene is expressed.

Typically, mammals, including most New World primates, are dichromatic with a single opsin gene on the X-chromosome, and all cones that are not S cones express the available X-chromosome opsin gene. Thus, human single gene dichromats are expected to express, by default, the one remaining X-chromosome opsin gene in all the cones that would be M or L in a normal trichromat. However, in the case of protan or deutran color vision defects caused by individual missense amino acid substitutions or by the LIAVA deletional combination, it is expected that developing photoreceptors will transcribe and translate the mutant opsin gene. What affect this has on the mature adult cone mosaic has been investigated using adaptive optics and optical coherence tomography (OCT) for both the LIAVA mutation (Carroll et al., 2004) and the C203R mutation (Carroll et al., 2009). The adaptive optics images of the retina from a male with deuteranopia whose M opsin gene encoded the LIAVA combination show a mottled appearance of the cone mosaic. There are large regions lacking visible photoreceptors interspersed with the visible photoreceptors. The outer nuclear layer thickness of the LIAVA retina by OCT was within the normal range suggesting the areas where no cones were visible by adaptive optics nonetheless contained cones, albeit nonfunctional ones that cannot waveguide light back into the camera and appear dark. In contrast, adaptive optics images of the retina from a color deficient male with the C203R mutation in his M opsin gene showed a cone mosaic that was relatively undisturbed in appearance compared to the LIAVA retina, although cone density was reduced compared to control retinas. These results suggest that, as for mutations at the corresponding cysteine residue in rhodopsin, the cones that express the C203R mutant opsin degenerate. The appearance of the cone mosaic suggests that the degeneration must occur prior to maturation and packing of the cones into the fovea, which occurs postnatally. This would account for the reduced cone density but relatively regular packing arrangement.

Results from psychophysics and from electrotetrogntigraphy suggested there is residual S cone function in some tritanopes. It also seemed possible that complete vs. incomplete penetrance of inherited tritanopia was a function of whether the affected individual had mutations in both copies of the S opsin gene, or in just one copy (Weitz, Miyake et al., 1992; Weitz, Went et al., 1992). What has become clear more recently is that tritan color vision defects are analogous to retinitis pigmentosa caused by mutations in the gene encoding the rod pigment rhodopsin. Indeed, some of the mutations in the S opsin gene in tritan color vision deficiency occur at positions corresponding to positions in rhodopsin at which amino acid substitutions cause the rod photoreceptors to degenerate. In a recent study a father and his daughter, both of whom made tritan errors, were both shown to be heterozygous for a substitution of glutamine instead of the normally occurring arginine at position 283 of the S opsin. The father’s tritan phenotype was severe, whereas the daughter’s was relatively mild. Adaptive optics imaging revealed an absence of S cones in the father, but S cones were clearly present in the daughter. Taken together, the findings indicate that inherited tritan defects are associated with a progressive degeneration of the S cones, analogous to the degeneration of rods in retinitis pigmentosa (Baraas et al., 2007). This accounts for the apparent incomplete penetrance of the phenotype in that in early stages the S cones function and the tritan color vision defect does not manifest until S cones have reached a sufficient stage of degeneration.

It seemed equally possible that the rare disorder, blue cone monochromacy (BCM), which is associated with an absence of both L and M cone contribution to vision, could be due to mutations at the photoreceptor level or at a higher neural processing level until linkage mapping showed it to be linked to Xq28, the same chromosomal location as the L and M opsin genes (Lewis et al., 1987). Subsequent genetic analysis of affected individuals revealed a variety of different mutational mechanisms that give rise to BCM. One mechanism is the deletion of all but one opsin gene on the X-chromosome, with an inactivating mutation in the one gene remaining. Inactivating mutations similar to those found as rare
4. Genes and the circuitry for color vision

4.1. Background

Genetics play the central role in all of life’s processes, including the circuitry for color vision. In the last decade, research in a number of areas has shed light on the process of how genes operate to give rise to circuitry for color vision. As discussed above, mutations and deletions of the photopigment genes give rise to color vision defects with reduced color vision capacities. More recent work, which is a topic in the following sections, has focused on understanding the converse – how the addition of an extra opsin gene in a dichromat can give rise to an expansion in color vision capacities.

At the level of the ganglion cells and the lateral geniculate nucleus (LGN), the circuits that carry color information are spectrally opponent, firing action potentials to some wavelengths and being inhibited by others. In early diagrams illustrating the putative first stages of color processing, e.g. Wiesel and Hubel (1966), post synaptic elements were drawn as having selective contacts to L, M and S cones with excitatory elements connecting exclusively to one cone class and inhibitory elements to a second class (Fig. 4). The ideas that such selectivity in the wiring was necessary for color vision and was genetically specified were highlighted by proposals to explain inherited color vision deficiencies (Hurvich & Jameson, 1962).

In contrast to the selective contacts hypothesis, just over a quarter of a century ago, the opposite idea was forwarded that spectrally opponent cone cells arise by completely random connections to cone photoreceptors (Paulus & Kroger-Paulus, 1983). The “random wiring hypotheses” held that, rather than requiring genetic instructions, the opponent properties of the wiring could, in part, be a consequence of the centers of midget ganglion cells receiving input from a single cone and from features of the topography of the cone mosaic (Jusuf, Martin, & Grunert, 2006; Lennie, Haake, & Williams, 1991).

Twenty-five years ago, there was very little evidence to distinguish between the genetically-specified cone-selective vs. random wiring dichotomy. This is an example of a genetics and circuitry issue that has largely been resolved over the last decade. The main conclusion is that the circuitry for color vision arose from the interplay between genetically specified cone-type-specific connectivity for S vs. L/M cones and random, or mostly random, e.g., (Field et al., 2010) wiring for L vs. M connectivity. This difference in genetic specification of connections is the result of different evolutionary origins for the two types of connectivity.

4.2. Evolution of red–green color vision

All lower mammals are either monochromats or dichromats and Old World (OW) monkey and ape species are all trichromatic (Jacobs, 1981, 2004). Most New World (NW) monkeys have color vision that is an evolutionary intermediate between the dichromats and OW trichromats and their genetics and color vision has been a key discovery for answering questions about how new color vision capacities might arise from the addition of a photopigment gene.

Nothing evolves de novo. Everything new in biology arose by modification of some pre-existing system. The major questions have been, from what pre-existing circuitry did red–green color vision evolve and, what were the associated costs? The transition from dichromacy to trichromacy required the addition of a third cone class but what, if any, additional changes in the circuitry were required? Aspects of color vision in NW primates, as representatives of a transitional form of color vision between non-primates and the OW primates, have shed light on this question. NW monkey species that have what has been termed, “allelic trichromacy” include both individuals with trichromacy and with dichromacy as normal phenotypes (Boissinot et al., 1998; Jacobs, 1983; Mollon, Bowmaker, & Jacobs, 1984). In these animals, the circuitry for their two normal forms of color vision, di- and trichromatic, must share the same genetic instructions. Consider the squirrel monkey (Saimiri sciureus); two thirds of females are trichromats; however, all males and one third of the females are dichromats (Jacobs & Neitz, 1987). Thus, a large majority, a full two-thirds of the individuals of this species, are dichromats.

The simplest idea is that, together with X-inactivation, a single mutation like the one at position 277 seen in prosimians (Jacobs, Deegan, Tan, & Li, 2002) that blue-shifts the long wavelength pigment by about 12 nm, produced a mosaic of two subtypes of middle-to-long wavelength cone in the retinas of heterozygous female squirrel monkeys that, in turn, exploited pre-existing circuitry, to provide a new dimension of color vision. A few subsequent amino acid changes in the opsin to increase the spectral separation...
produced a subset of females with cone complements like modern human trichromats. The idea that the presence of three cone types alone provided rewards of trichromacy without additional subsequent genetic changes seems likely based on the principle that each individual genetic change must, in the words of Darwin (1859) be “useful to its possessor.” Having each individual mutation provide an advantage is the only way to overcome the compounded improbabilities of multiple highly unlikely mutations occurring in one animal (or one genetic line of animals). Moreover, a nascent mutation that subsequently benefited trichromats by refining the circuitry for red–green color vision would not have increased in frequency in the species if it proved to be a disadvantage to the larger dichromatic subpopulation. Thus, the most straightforward hypothesis is that full red–green color vision, similar to that experienced by human trichromats, emerged as the result of the evolution of three cone types in the retina without any further genetically coded modifications to the circuitry.

A related issue is whether there could be a balance between the relative selective advantages of dichromacy vs. trichromacy in the natural world. For example, such a balance has been offered as an explanation for the high frequency of color defective humans (Regan et al., 2001). However, in earlier sections we presented evidence that the high frequency of color defective humans is the result of an extraordinarily high mutation rate of those genes. NW primates are a different case; it appears that dichromacy is maintained in the majority of squirrel monkeys and related species simply because there are no males with the genes required for trichromacy (Jacobs & Neitz, 1985). In the absence of trichromatic genetic variants among the males, selection cannot act to increase the frequency of trichromacy in that gender. In contrast, natural selection can act on the allele frequencies in females to affect the relative number of dichromats vs. trichromats. Cebid monkeys in which there are 3 equally spaced spectral types of photopigments, for females, a balancing selection produced by the heterozygous advantage of trichromacy has produced three alleles at near equal frequency, maximizing the number of trichromats (Jacobs & Neitz, 1985; Mollon et al., 1984). The fact that selective pressure has produced the maximum number of trichromatic squirrel monkeys allowed by their genetic repertoire indicates that trichromacy in these monkeys has a strong selective advantage over dichromacy. One counterexample to the advantages of trichromacy over dichromacy is that human dichromats can perform better than trichromats in detecting luminance edges in the presence of strong masking chromatic contours (Morgan, Adam, & Mollon, 1992). The hidden digit designs of the Ishihara plates are a demonstration of this. This small subset of plates is intended to have a visible sign for colorblind people, but not be seen by normal trichromats. Dichromats seeing things that are invisible to trichromats is an attractive idea, especially to people with color vision defects; however, the masking effect is weak, and these particular plates are relatively ineffective in distinguishing people with color vision defects from normal. The lesson learned is that trichromacy provides powerful advantages with virtually no offsetting disadvantages.

Except for the extra opsin gene, there is no evidence for any differences between the dichromat and the trichromat. It is hard to reconcile this with the idea that there could be separate circuitry dedicated entirely to red–green color vision. Mollon (1989) popularized the now favored view that phylogenetically ancient neural machinery serves blue–yellow color vision, whereas a newly evolved subsystem serves red–green color vision. Certainly, red–green color vision depends on recently evolved subtypes of X-encoded cone opsins, and blue–yellow color vision is evolutionarily ancient based on its ubiquity across mammals; however, this does not necessarily imply the evolution of new circuitry dedicated to red–green color vision. Red–green color vision might represent an improvement in the function of circuitry that was already present in the dichromat. This might be analogous to the evolution of flight. A recent study suggests that the wings of the ancestor to modern birds, the archaeopteryx, could have been used for gliding but not flapping wing flight (Su et al., 2006; Xu et al., 2003). Flapping wings evolved as an improvement of wings used for gliding, not by the creation of a separate system. Similarly, trichromacy may represent an enhancement in the functions of pre-existing circuits. The emergent trichromats may have been able to make use of red–green opponent signals that were added to pre-existing circuitry with the addition of a third cone population rather than creating a separate system for red–green color vision.

Besides the opsin genes, there are no known differences in the genetic code between dichromats and trichromats. However, for other systems, during development, large differences in wiring of the visual system can result from differences in visual experience. Even though the circuitry for di- and trichromatic color vision share the same genetic instructions, this does not rule out the possibility that the differences in neural activity during development imposed by having different cone complements could be responsible for very different neural wiring in dichromats and trichromats including specialized circuitry for red–green color vision.

In the last decade, new genetic technology has been applied to questions about the neural plasticity of color vision. For the past 50 years, deprivation and ablation studies were the main tools for studying neural plasticity. However, recently, molecular genetics methods have made it possible to add inputs. In one series of experiments, genetic manipulations were used to generate trichromatic mice (Smallwood et al., 2003). “Knock-in” mice had the endogenous M-opsin (spectral peak = 511 nm) opsin gene replaced (Jacobs, Neitz, & Deegan, 1991) with one encoding a human L-opsin (spectral peak = 555.5 nm). The knock-in mice were mated to wild-type mice to produce heterozygous females in which X-inactivation produced two middle-to-long wavelength cone submosaics, and some of the heterozygous mice gained red–green color vision capacities (Jacobs, Williams, Cahill, & Nathans, 2007). Any color opponent signals in the genetically altered mice would have to arise by random connections to center-surround receptive fields in which, unlike primaries, both centers and surrounds contain several cones. Some ganglion cells in the transgenic mice would carry red–green, spectrally opponent signals because of random differences in the relative numbers of L and M cones in the center vs. the surround; however, the number of ganglion cells with strong opponency would be small, and as expected from this, red–green color vision in the mice was relatively weak. Nonetheless, this demonstrates that in a species lacking red–green color vision it is possible to get trichromatic behavior just by adding a new cone subpopulation. These experiments do not, however, address the question of how much the ability to extract red–green color vision depends on visual experience during development. That question was addressed by experiments in which gene therapy was used to add a third cone population in adult animals.

Gene therapy was performed on adult squirrel monkeys that were missing the L-opsin gene and were colorblind since birth (Mancuso et al., 2009). L-opsin was added to a subset of M cones, providing the receptive basis for trichromatic color vision. The addition of a third opsin in adult red–green color deficient primates was sufficient to produce trichromatic color vision behavior, demonstrating that trichromacy can arise from a single addition of a third cone class and it does not require a developmental process. The treated animals discriminated colors in the red and green range from each other and as different from gray. They could also discriminate among other color combinations that dichromats find impossible to tell apart. The ability to make new color discriminations was closely timed with the appearance of robust expression of the introduced opsin indicating that no rewiring or new circuitry was associated with the acquisition of red–green color vision.
Most midget ganglion cells in macaques are red–green spectrally opponent and a popular idea has been that the circuitry for red–green color vision evolved in an ancestral primate from the precursor to the midget ganglion cells that had been previously responsible for luminance contrast-based achromatic spatial vision. However, a new and different explanation has been offered for the newly acquired color vision in the treated squirrel monkeys—that they might be taking advantage of pre-existing blue–yellow color vision circuits (Mancuso et al., 2009; Mancuso, Neitz et al., 2010; Shapley, 2009). The treated monkeys were protanopes and after treatment, a subset of the cones co-expressed L-pigment with the native M pigment. Shapley (2009) explains that if some of the receptive fields remained wild type S vs M like the untreated protanope and other receptive fields opposed S cones to the newly introduced L-like cones making them essentially S vs L, the animal would have two blue–yellow systems with different spectral response properties that could be responsible for the observed trichromatic behavior. We agree with Shapley’s idea and have taken it one step farther suggesting that, not only does red–green color perception in the treated monkeys, in part, take advantage of the pre-existing blue–yellow system, the hue components of red–green color vision in all primates might represent an improvement in the function of the pre-existing blue–yellow system that was split in two by the addition of the third cone population (Mancuso et al., 2009). Understanding of the genetics and evolution of blue–yellow spectral opponency has been transformed over the last 25 years. In light of the recent suggestions of a relationship between blue–yellow and red–green hue systems in gene therapy treated monkeys, understanding the blue–yellow system may be critical to understanding the circuitry of hue mechanisms in general.

4.3. Evolution of blue–yellow opponency and S- and L/M cone specific connections

In terms of cone selectivity, primates can be considered to have two major cone classes: S and L/M. It is now abundantly clear that S and L/M cones in primates have separate circuitry, each with highly cone-type selective connections (Crook et al., 2009; Dacey & Lee, 1994; Dacey, Lee, Stafford, Pokorny, & Smith, 1996; Mariani, 1984; Martin & Grunert, 1999; Packer, Verweij, Li, Schnapf, & Dacey, 2010). The major types of bipolar, horizontal and ganglion cells likely to be relevant to color vision in primates are illustrated in Fig. 5. Evolution is conservative, with all vertebrates sharing many common features. Receptors homologous to human S cones, L/M cones and rods are nearly universal features of vertebrate visual systems and the evolutionary roots of the photoreceptor cell-type specific circuitry seen in primates likely predate vertebrates. The ancestors to human S vs. L/M opsins predated the appearance of eyes (Neitz, Carroll, & Neitz, 2001). A billion years before photopigments and photoreceptors served the function of vision, a primitive form of blue–yellow color vision was already in place driving circadian rhythms and vertical migration in one-celled organisms that predated bacteria (Lamb, Collin, & Pugh, 2007; Spudich & Spudich, 2008). UV light triggered archaebacteria to descend away from the damaging UV rays of midday, and the gentle orange light of dusk resulted in upward migration to collect long wavelengths for a primitive form of photosynthesis. Emerging millions of years later in evolution, the hagfish ‘eye’ continued to function as a circadian organ. These primitive, jawless, eel-shaped archaebacteria just as their likely mammalian homologues, the melanopsin-containing retinal ganglion cells (Berson, 2003; Koyanagi, Kubokawa, Tsukamoto, Shichida, & Terakita, 2005; Provencio, Jiang, WDeGrip, Hayes, & Rollag, 1998; Provencio et al., 2000; Rollag, Berson, & Provencio, 2003). Thus, a form of “blue–yellow” chromatic opponency may be one of the oldest sensory capacities, having originally evolved to signal the large spectral changes in the sky at dawn and dusk. These chromatic signals precisely mark the phase of the day-night cycle and provide a powerful cue for circadian entrainment. Since blue–yellow opponency is a characteristic of the primitive receptor systems responsible for circadian entrainment extending from archaebacteria to the parietal eye of reptiles, (Solessio & Enghbreton, 1993; Su et al., 2006) it is not surprising then that melanopsin ganglion cells in modern primates are blue–yellow spectrally opponent (Dacey et al., 2005). We assume that the circuitry for blue–yellow color vision evolved by adapting pre-existing components of the spectrally opponent functions of the circadian organs to new functions. The fact that biological mechanisms in which short- and long-wavelength lights have had opposing actions have been around since the emergence of animal life on earth makes it understandable why highly cone-type specific circuitry involving S vs. L/M cones would be inherent to the retina.

The discovery of the melanopsin ganglion cells is directly part of the legacy of Nathans’ original cloning of the opsin genes; it has been celebrated for its impact on circadian biology; however, it has been equally revolutionary for the field of color vision. Dermal melanophores migrate to the periphery of the pigment cells of frog skin (Xenopus laevis) in response to illumination and in an effort to identify the opsin responsible, Provencio et al. (1998) screened a melanophore cDNA library for opsin-like nucleotide sequences. The opsin identified was termed melanopsin. Subsequently, Provencio et al. identified the human homolog and showed that, in mammals, melanopsin expression is restricted to the retina (Provencio et al., 2000). Melanopsin-positive ganglion cells matched the anatomical characteristics of cells known to project to the primary circadian pacemaker of the hypothalamus (Gooley, Lu, Chou, Scammell, & Saper, 2001; Hattar, Liao, Takao, Berson, & Yau, 2002). The surprise for those of us studying color vision was that Dacey and colleagues (Dacey et al., 2005) determined that the melanopsin containing ganglion cells are the S-OFF spectrally opponent cells of the primate retina. Over many years of recording, no anatomically identified (M + L) – S cells have been identified in the macaque retina other than the melanopsin ganglion cells. Remarkably, application of the ON-pathway agonist, 2-amino-4-phosphonobutyric acid (AP-4), completely blocks the S-OFF light response in the primate melanopsin ganglion cells (Dacey, Peterson, & Robinson, 2002). This suggests the S-OFF signals are transmitted to the inner retina via the ON S-cone bipolar cell, and that a sign inversion of the ON signal in the inner retinal circuitry is critical for generation of the S-OFF response (Dacey et al., 2002). Flat contacts onto cones are associated with OFF-bipolar cells, and in electron micrographs, occasional flat contacts are observed on S-cone terminals (Calkins, 2001; Klug, Herr, Ngo, Sterling, & Schein, 2003). However, at the ganglion cell level the only physiologically identified (M + L) – S cells, the melanopsin ganglion cells, get their S-OFF response by sign inverting an input from S-cone ON-bipolar cells. The flat contacts onto S-cones observed anatomically appear to be capricious since no major physiological function resulting from them has been identified.

The discovery of the S-OFF melanopsin ganglion cell reveals a major role for S-cone ON bipolar cells to serve an ancient spectrally opponent system that is not related to hue perception. The melanopsin-expressing ganglion cells have the largest dendritic tree diameters of any primate retinal ganglion cell (Dacey et al., 2005). The only way cells with such large dendritic trees could have cone-opponent circuitry is through genetically-specified cone-specific connections fed through two sets of circuitry with opposing polarities, L/M-ON from L/M cone specific bipolar, and S-OFF through a sign reversing synapse from S cone specific bipolar (see Fig. 5). The S-OFF melanopsin ganglion cells have a S-ON.
counterpart, the small bistratified ganglion cells, with which they share input from the S-ON bipolar cell (Dacey, 1993; Dacey & Lee, 1994). As shown in Fig. 5, the S-ON small bistratified and S-OFF melanopsin ganglion cells appear to be partners in an ancient circuit sharing the same S-cone bipolar. The fact that the S-OFF component of this pair, the melanopsin ganglion cells, are not likely candidates for serving the yellow component of the blue-yellow hue system might raise doubt that S-ON small bistratified ganglion cells are the necessary input to a color system.

Fig. 5. Primate retinal ganglion cells that receive cone selective connections and have the potential for playing a role in color vision. Upper and lower panels show how the addition of a third cone type changes the chromatic inputs to different ganglion cells in the primate. The cell bodies of ganglion cells are drawn as diamonds. Bipolar cells have circular cell bodies. Horizontal cells bodies are hexagonal. For the trichromat (lower panel), four ON/OFF pairs of midget ganglion cells are drawn. Different combinations of cone connectivity distinguish the four ganglion cell pairs. Two ON/OFF pairs of ganglion cells, one pair with an L cone center and one with an M cone center, receive input from cones that make contacts with H2 horizontal cells, which contact nearby S cones. Two other ganglion cell pairs (an L center and an M center) receive input from cones that do not have the potential for significant S-cone input from the surround. Assuming that a small subset of ganglion cells receive S-cone input from the surround, the M cones with S in the surround give rise to an OFF-center ganglion cell with $(S+L) - M$ opponency and an ON-center ganglion cell with $M - (S+L)$ providing the potential retinal basis for a red and green, respectively, hue pathway. L cones with S in the surround give rise to an OFF-center ganglion cell with $(S+M) - L$ opponency and an ON-center ganglion cell with $L - (S+M)$ providing the potential retinal basis for a blue and yellow, respectively, hue pathway. ON midget ganglion cells with no S-cone input to the surround have $L-M$ opponency when the center cone is L and $M-L$ opponency when the center cone is M. M cones provide the center of one spectrally opponent ganglion cell but the surround of neighboring ganglion cells. If neighboring L-M and M-L are indiscriminately combined in the cortex, chromatic opponency for diffuse spots of light will cancel. This would make L-M and M-L ganglion cells the substrate for edge detectors that would also signal chromatic borders. Thus, the four pairs of midget ganglion cells could provide the retinal basis to serve red, green, blue and yellow hue perception and luminance/chromatic edge detection. One S-cone bipolar cell is illustrated. It connects specifically to an S cone. It provides the S-ON input to the small bistratified ganglion cell, which is drawn showing dendritic arbors in both the ON and OFF sublamina (labeled OFF-CENTER and ON-CENTER). The single S-cone bipolar cell also provides an S-OFF input via an inhibitory interneuron to the melanopsin ganglion cell (drawn in yellow with an “X” shaped dendritic arbor). Both the melanopsin ganglion cell and the small bistratified cells have large receptive fields so the ON component of the melanopsin and the OFF component of the small bistratified cell have $(S+L) - M$ cone inputs, giving them $L - M$ and $S - (L + M)$ spectral opponency, respectively. A comparison of the “DICHROMAT” top panel with the “TRICHROMAT” bottom panel shows how the spectral opponent properties of each of the 10 ganglion cells illustrated change when the retina is transformed from having two cone types to having three. These midget cells that are capable of only transmitting luminance information in the dichromat become L vs. M opponent in the trichromat. Putative midget ganglion cells with S vs. L inputs that could serve blue color vision are transformed into two pairs to serve blue-yellow and red-green color vision in the trichromat. An attempt was made to preserve some of the anatomical details of the retina in the cartoon. Cones and bipolar cells are shown with ribbon synapses. ON-bipolars make connections to the ribbon and terminate in the ON sublamina. OFF-bipolars are shown making more lateral connections representing flat contacts and they terminate in the OFF sublamina. The inset illustrates that horizontal cells make reciprocal synapses. Please cite this article in press as: Neitz, J., & Neitz, M. The genetics of normal and defective color vision. Vision Research (2011), doi:10.1016/j.visres.2010.12.002
cells could be directly responsible for the blue hue component. There are several other reasons that might eliminate S-ON ganglion cells as candidates for blue–yellow hue perception. If the firing of action potentials by the small bistratified cells were the neural correlate for the percept of blue, then we would expect to perceive blue whenever the small bistratified cell fires. This is not the case; unlike the midget ganglion cells that give relatively sustained ON or OFF responses, depending on the type, the small bistratified cells produce equally robust responses to the onset of short wavelength light and to the offset of long-wavelength lights presented in the dark (Crook et al., 2009). In contrast, short wavelength lights in the dark elicit blue percepts, but the offset of long-wavelength lights do not. A yellow light can produce a blue afterimage, however, one must look at a white stimulus after the offset of yellow; there is no blue sensation with the offset of a yellow light in the dark. A common percept associated with the offset of a bright yellow light in the dark is a yellow afterimage. Another major disconnect between the physiological properties of small bistratified cells and blue–yellow perception is that the small bistratified cells have S – (M + L) cone inputs. They fire action potentials to S-cone stimuli and they are inhibited by stimuli that selectively activate M or L cones. This does not match human blue–yellow perceptions for lights presented to the central retina. Percepts of blueness are elicited rather than inhibited by activation of M cones (De Valois & De Valois, 1993; Drum, 1989; Neitz & Neitz, 2008; Stockman & Brainard, 2010); so with respect to M-cone stimuli, the small bistratified cells behave opposite of what would be predicted if they served blue–yellow hue perception.

Genetics provide the final and most definitive test of the hypothesis that the S-ON small bistratified ganglion cell and the S-OFF melanopsin ganglion cell system are the substrate for blue–yellow color vision. Both of these ganglion cell types receive S-cone input from S-cone ON-bipolar cells via the metabotropic glutamate receptor, mGluR6. The hypothesis that these ganglion cells are the physiological substrates for blue–yellow perception predicts that people who have genetic mutations that render mGluR6 receptors non-functional will be unable to perform color discrimination and detection tasks based on S cones, resulting in deficits in blue–yellow color perception; however, people with genetic defects that eliminate function of mGluR6 receptors have none of these problems (Dryja et al., 2005).

Probably the major reason that the anatomically distinct S – (L + M) and (L + M) – S ganglion cell pathways have been thought to be substrates for blue–yellow color vision in the face of contradicting evidence is a perceived lack of other candidates. However, there have been reports (deMonasterio, 1979; Derrington, Krauskopf, & Lennie, 1984; Tailby, Solomon, & Lennie, 2008) of a set of underappreciated cells early in the visual pathway that are perfect candidates for providing hue perception. Our blue–yellow hue perception differs from that expected from S – (L + M) and (L + M) – S ganglion cells in that the direction of maximum activation of the perceptual hue mechanisms are “rotated in color space” (Webster, Miyahara, Malkoc, & Raker, 2000) relying on (S + M) – L and L – (S + M) circuitry in which M cone responses are added to S rather than being different from them. Similarly, our red–green hue perception is based on (S + L) – M and M – (S + L) mechanisms which are also “rotated” in color space compared to that expected from ganglion cells in which L and M cones are opposed. deMonasterio and colleagues (deMonasterio, Gouras, & Tolhurst, 1975) were the first to document the physiology of ganglion cells in the retina in which the L cone mechanism is opposed to the S and M cones, as expected for cells serving perceptual blue–yellow color vision. He also identified ganglion cells in which S and L cones are opposed to M, as required to explain red–green hue perception. A characteristic of our hue perception is that the circuitry for all four unique hues, red, green, blue and yellow relies on input from all three cones while a preponderance of ganglion and LGN cells receive input from only two cone types—L and M. deMonasterio commented that the finding of a “cortical trichromatic organization and of a subcortical dichromatic organization could indicate a central reorganization of the processing of color information”. However, his ganglion cell recordings showed “that many color-opponent ganglion cells in the rhesus monkey retina receive input from all three cone mechanisms, indicating that trichromatic interactions begin in the retina.” Modern multistage color models have continued to propose central reorganization of the processing of color information to explain the (S + M) vs. L and (S + L) vs. M inputs to perceptual hue channels (DeValois & DeValois, 1993; Stockman & Brainard, 2010). However, deMonasterio’s conclusions from a third of a century ago, that (S + M) vs. L and (S + L) vs. M pathways originate in the retina (as shown in Fig. 5) seem worth revisiting in the light of genetic results and the constraints imposed by evolution, as reviewed here.

Molecular genetics, physiology and anatomy have converged to indicate that in primates, there is a single bipolar cell type that is highly S cone specific and there are others that are highly L/M cone specific (Fig. 5) (Dacey, 1996; Li & DeVries, 2006; Martin & Gruner, 1999). The S vs. L/M cone specific connectivity severely restricts the possible ways in which S-cone input can be combined with L and M signals at early stages of color processing. Thus, cells like the small bistratified cells and the melanopsin ganglion cells that receive input from large numbers of L and M cones and combine

![Fig. 6. Cone weights of LGN cells recorded by (Tailby et al., 2008). Each dot represents the response properties of one LGN neuron. Normalized weights assigned to each cone type by the cells are plotted. The weight attached to M-cone input is plotted against that for the L-cone. The distance from the diagonals reflects the magnitude of S-cone input. Thus, a cell plotted at the origin would have only S-cone input and one on a diagonal line would have no S-cone input. Only those neurons that were determined to have significant S-cone input are represented. LGN neurons were found that represent all the spectrally opponent ganglion cells shown in Fig. 5. The cells have been color coded to represent their putative role in hue perception. The cells that plot in the upper left and lower right triangles have L vs. M opponency and either excitatory or inhibitory S-cone inputs as required to match the cone input to human hue perception. LGN cells with all the correct cone inputs to account for human hue perception were identified: M – (S + L) for green, (S + M) – L for blue, L – (S + M) for yellow and (S + L) – M for red; however, only one cell with the correct cone inputs to account for red percepts, (S + L) – M, was recorded. A threshold was used to decide which LGN cells to include as receiving S-cone input, it may be that other (S + L) – M cells were present but fell below the threshold. This seems reasonable because only 5% S cone contribution is required to account for normal hue perception. Redrawn from (Tailby et al. (2008)).](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002886)
S and L/M inputs via circuitry in the inner retina are constrained to have only S vs. (L + M) connectivity. From the known anatomical cone-type selective connections shown in Fig. 5, the only way to get retinal ganglion cells that combine inputs from all three cone types in the (S + L) vs. M and (S + M) vs. L combinations relevant to hue perception is via S-cone input to the surrounds of midget ganglion cells via H2 horizontal cells (Fig. 5). H2 horizontal cells (Dacey et al., 1996) receive a strong input from S cones and a weaker input from L and M cones and they provide the only known anatomical pathway for building ganglion cells that combine S with L or S with M cones, i.e., by providing S-cone input to the surrounds of midget bipolar cells so that the cone type in the receptive field center will be opposed to S plus the other cone type in the midget bipolar cell’s receptive field surround.

In summary, (1) S cone based hue perception in people whose S-cone bipolar cells are non-functional is indistinguishable from normal in color vision tests. (2) S vs. L/M cone specificity ensures that ganglion cell types receiving direct S signals via S-cone bipolar cells will have S vs. (L + M) circuitry and (3) (S + L) vs. M, (S + M) vs. L and S vs. (L + M) cell types have all been reported early in the retinostriate visual pathway. Examples of the different chromatic cell types are evident in the results of Tailby et al. (2008) which have been reproduced in Fig. 6. In primates, the midget ganglion cells are the major mediators of conscious perception making up more than 95% of the ganglion cells in the fovea (Dacey, 1994). The midget bipolar cells do not contact S cones (with the exception of the far periphery) (Field et al., 2010); thus, the great majority of midget bipolar cells do not carry S cone signals (Sun, Smithson, Zaidi, & Lee, 2006). H2 horizontal cells contact both S cones, which represent 5% of the total cone population, and L/M cones, which represent 95% of the cone population. Only the small subset of L and M cones that have strong H2 input and are in close proximity to S cones could have access to S cone signals that would be significant in the surrounds of midget ganglion cells. Thus, for midget ganglion cells in the central retina, if a small number have S-cone input, it would always be mixed with L or with M cone signals in the surround, opposed to the single L or M cone in the receptive field center (Fig. 7). This would provide the (S + L) vs. M and (S + M) vs. L combinations necessary for hue perception.

For the midget system (or its predecessor), occasional S cone signals could come via horizontal cells to an M/L cone pedicle that, in turn, is served by an ON- and an OFF-bipolar cell (Fig. 7). As has been described by Mancuso, Neitz et al. (2010) this could provide an S–ON signal to mediate the percept of blueness (via the OFF-bipolar) and an S–OFF signal to mediate the precept of yellowness (via the ON-bipolar), thus, providing parallel and complementary blue–yellow circuitry in which there is one S–ON for every S–OFF, and the pathways would have equal and opposite opponency. For the midget ganglion cells, since S-cone input would come from the surround, the OFF-bipolar cells could carry the S–ON signals. These would be unaffected by mutations in the mGluR6 gene that prevent ON-bipolar signaling from cones, explaining why S cone based vision, as assayed by standard color vision tests, is unaffected in patients who lack mGluR6-mediated signaling. S–OFF-bipolar cells would be functional in patients lacking functional mGluR6 receptors but S–ON midget cells would not. However, ON and OFF pathways have push–pull interactions mediated by crossover inhibition (Molnar & Werblin, 2007) at higher levels of the pathway; thus, as long as the S–ON side is intact, ON–OFF opponent interactions at higher levels would provide input to the S–OFF pathway.

A rationale in favor the H2 horizontal cell/midget ganglion cell based blue–yellow system hypothesis is that it provides the explanation for how red–green color vision could have arisen from dichromacy in evolution solely by the addition of a third cone population as we have proposed (Mancuso, Neitz et al., 2010) and is as illustrated in Fig. 7. It also explains how gene therapy can produce a full, extra dimension of color vision in adults without a developmental process. As discussed above, Shapley has proposed the logical solution that the addition of a third cone type in gene therapy treated monkeys may have split the pre-existing blue–yellow hue system in two. The small bistratified cells receive input from several dozen L/M cones. Thus, the addition of a randomly placed third cone type would produce a relatively homogenous population of
small bistratified cells with respect to mixed L/M cone contribution and there wouldn't be two populations. However, if blue–yellow color vision is based on a subset of midgets with S-cone input in the surrounds as proposed by Mancuso, Neitz et al. (2010), the addition of the third cone type would break the blue–yellow system cleanly into two subpopulations depending on whether the cone center is M or L, explaining both the trichromacy of the monkeys treated with gene therapy and the evolution of trichromacy in primates (Fig. 7). According to this idea, red–green hue perception represents an improvement in the pre-existing blue–yellow system in which the hue dimensions are enhanced. Nothing is lost or compromised in this split because visual detection and discrimination based on S cones becomes distributed across the two newly formed hue systems. Since both are hybrids of the dichromatic blue–yellow circuit, it would explain why S cones play an important role in both red–green and blue–yellow hue systems.

Trichromacy resulting from splitting the blue–yellow pathway is a new idea stimulated by the gene therapy results. Similarly, the report that mice gained trichromatic vision after having a new cone introduced stimulated ideas about the circuitry responsible. Makous (2007) proposed an idea much more in keeping with conventional thinking that the observed red–green color vision behavior in the mice could be the result of taking advantage of existing circuits responsible for extracting luminance contrast for spatial vision. As illustrated in Fig. 5, in the case of primates, there can be no doubt that pre-existing circuits responsible for spatial vision in dichromats carry L/M opponent signals after the addition of a third cone type. In dichromatic primate ancestors, if midgets comprised greater than 90% of the ganglion cells in the central retina, as they do in modern monkeys, they must have provided the major source for luminance contrast form vision. The addition of a third cone type would have made the majority of formerly achromatic midget cells in the central retina L vs M spectrally opponent in addition to being responsive to achromatic luminance contrast edges. If these L–M opponent ganglion cells are the basis for detecting red–green chromatic boundaries in the absence of luminance contrast at edges, it could represent an improvement in the pre-existing circuits for spatial vision enhancing their edge detecting capabilities to include pure red–green chromatic borders. For red–green opponent signals added in trichromats, it seems likely that both schools of thought are correct. L vs M signals could be used in both the pre-existing blue–yellow color vision system (e.g., as suggested by Shapely) and the achromatic luminance contrast spatial vision system (e.g., as suggested by Makous) for complementary functions that represent enhancements to each.

Makous (2007) suggested that the treated mice might “see” their own L/M cone mosaic. However, visualizing the stabilized L/M cone mosaic seems unlikely to be a problem for systems with emergent red–green color vision. A major transform for spatial vision in primary visual cortex is the properties of receptive fields. While most cells in the LGN respond well to small spots of light, cortical cells respond to moving oriented edges and bars. Responses to a stabilized L/M cone mosaic would be filtered out at this stage. The details of how the LGN receptive fields are transformed in the cortex is not completely clear; however, it is apparent that it involves combining the outputs from neighboring cells in the retina. Fig. 8 illustrates how this would happen in the classic feedforward model of Hubel and Wiesel (1962). Combining adjacent ON receptive fields indiscriminately adds together L–M and M–L receptive fields, canceling opponent responses to diffuse colored lights while producing cortical cells that respond well to luminance contrast edges. However, receptive fields constructed of red–green opponent ganglion cells would be expected to maintain responsiveness to moving red–green equiluminant edges. If so, such cortical cells could respond both to achromatic luminance and to red–green chromatic edges. This is the character of many cells observed in area VI. They respond to dark-light edges but not diffuse colored lights; however, they will respond to edges of one color opposed to another at relative intensities (Conway, Hubel, & Livingston, 2002; Gouras & Kruger, 1979; Hubel & Livingston, 1990; Johnson, Hawken, & Shapley, 2001, 2008). Such cells respond to borders defined by luminance contrast, but they also allow the visual system to extract form information when objects are differentiated from their backgrounds only by chromatic borders. Thus, when trichromacy evolved, the addition of the red–green opponent signals to the pre-existing edge detector system responsible for form vision would represent an enhancement in function. The only potential cost is that red–green chromatic edges could mask luminance edges. As mentioned above, this is the only consistently documented advantage of dichromacy over trichromacy.

The role in color vision for cells that respond to red–green equiluminant edges and luminance contrast has been uncertain. It may be unlikely that they contribute to hue perception because they appear to signal the presence of a chromatic edge, but not the difference between green–red as opposed to red–green boundaries (Conway et al., 2002; Hubel & Wiesel, 2005). If red–green isoluminant edge detection is handled by a high-resolution spatial contour detection system that was enhanced by the addition of L vs. M opponency to midget ganglion cells, the problem that the system does not provide hue information might be solved if hue perception is mediated by separate circuitry operating at lower spatial resolution that represents a subdivision of the pre-existing blue–yellow system. L vs. M opponency of the contour detectors does correspond to one of the cardinal directions that characterize threshold data obtained in color discrimination experiments (Krauskopf, Williams, & Heeley, 1982). In the past, the different spectral signatures obtained for psychophysical hue vs. detection tasks has been thought to represent two different stages of one color vision system (Stockman & Brainard, 2010). However, red–green.

Fig. 8. A possible scheme for explaining the chromatic properties of cortical cell receptive fields adapted from Hubel and Wiesel (Hubel & Wiesel, 1962). A number of LGN cells, of which four are illustrated, project upon a single cortical cell. The synapses are presumed to be excitatory in this “feed forward” model. In this model, a number of the inputs must be active at the same time in order to exceed the threshold of the cortical cell. Indiscriminate connectivity to L vs. M opponent cells in which L–M cells are always nearby M–L cells cancels opponent responses to diffuse colored lights, but responses to luminance edges are enhanced.
opponent information may be carried in two parallel pathways and may represent improvements to two pre-existing systems, one for spatial vision and the other for blue–yellow color vision rather than the creation of single hierarchical color vision system. There is agreement that explaining all aspects of human color perception requires the activities of the cones to be processed through multiple stages, including transformations at the level of the ganglion cells and additional cortical stages. Multistage, color models, include those of Judd (1949), DeVilos and DeVilos (1993), and Stockman and Brainard (2010). From the DeVilos and De Valois model as a starting point we have suggested that instead of being carried by a single pathway solely concerned with red–green color vision, red–green opponent signals in trichromats may be carried by two parallel pathways that preexisted in ancestral dichromatic primates, one for extracting contours relevant to spatial form and one that pre-existed for blue–yellow hue perception (Mancuso, Mauck, Kuchenbecker, Neitz, & Neitz, 2010). This can reconcile the different ideas recently proposed for how red–green color vision might arise from pre-existing circuits in genetically modified animals (Makous, 2007; Shapley, 2009). Evolution is purely opportunistic; by making use of different aspects of existing circuitry in which L vs. M opponent signals were added, evolution could have taken advantage of an opportunity to expand sensory capacities by a single genetic change at the level of the photopigments.

Finally, older multistage color models are primarily concerned with explaining our perceptual responses to isolated colored lights (S + M) vs. L cells, described above, that we propose could account for explaining our perceptual responses to isolated colored lights (S + M) vs. L cells, described above, that we propose could account for their center and to a different part of the spectrum in their surround (Hubel and Wiesel, 2005) and J. Kuchenbecker for discussions and assistance with the figures. A study of unusual Rayleigh matches in deutan deficiency. Vision Research, 39, 145–160.

References