Citation: Molecular Therapy — Methods & Clinical Development (2016) **3,** 16051; doi:10.1038/mtm.2016.51
Official journal of the American Society of Gene & Cell Therapy

www.nature.com/mtm

ARTICLE

PAX6 MiniPromoters drive restricted expression from rAAV in the adult mouse retina

Jack W Hickmott[1,2], Chih-yu Chen[1,3], David J Arenillas[1], Andrea J Korecki[1], Siu Ling Lam[1], Laurie L Molday[4], Russell J Bonaguro[1], Michelle Zhou[1], Alice Y Chou[1], Anthony Mathelier[1,2], Sanford L Boye[5], William W Hauswirth[5], Robert S Molday[4], Wyeth W Wasserman[1,2] and Elizabeth M Simpson[1,2,6]

Current gene therapies predominantly use small, strong, and readily available ubiquitous promoters. However, as the field matures, the availability of small, cell-specific promoters would be greatly beneficial. Here we design seven small promoters from the human paired box 6 (*PAX6*) gene and test them in the adult mouse retina using recombinant adeno-associated virus. We chose the retina due to previous successes in gene therapy for blindness, and the *PAX6* gene since it is: well studied; known to be driven by discrete regulatory regions; expressed in therapeutically interesting retinal cell types; and mutated in the vision-loss disorder aniridia, which is in need of improved therapy. At the *PAX6* locus, 31 regulatory regions were bioinformatically predicted, and nine regulatory regions were constructed into seven MiniPromoters. Driving Emerald GFP, these MiniPromoters were packaged into recombinant adeno-associated virus, and injected intravitreally into postnatal day 14 mice. Four MiniPromoters drove consistent retinal expression in the adult mouse, driving expression in combinations of cell-types that endogenously express *Pax6*: ganglion, amacrine, horizontal, and Müller glia. Two *PAX6*-MiniPromoters drive expression in three of the four cell types that express PAX6 in the adult mouse retina. Combined, they capture all four cell types, making them potential tools for research, and *PAX6*-gene therapy for aniridia.

Molecular Therapy — Methods & Clinical Development (2016) **3**, 16051; doi:10.1038/mtm.2016.51; published online 10 August 2016

INTRODUCTION

Gene therapy for ocular disorders is reaching the clinic, with long-term trials of gene therapies for Leber's congenital amaurosis reporting improvements in patients.[1,2] Promoters such as the chicken beta actin and cytomegalovirus are common in current gene therapies, particularly because they are strong, small, and well characterized.[3,4] However, in applications where restricted expression is desired, such as targeting transcription to a specific tissue,[5] or limiting it to particular cells,[6,7] a toolbox of specific promoters would be advantageous. Previously, we have developed Pleiades MiniPromoters (MiniPs; ~ 4 kb human regulatory sequences for tissue and cell-specific expression), using bioinformatics, and single copy knock-ins to the mouse genome.[8–10] Building on, and further refining these techniques, we are expanding the current toolbox by introducing new MiniPs from paired box 6 (*PAX6* (OMIM: 607108)).

Capturing the endogenous expression pattern of a human gene in a small promoter is challenging, considering that eukaryotic genes can be regulated by a large number of regulatory regions (RRs) megabases away from the transcriptional start site (TSSs).[11–14] However, it is possible to narrow the search field using high-throughput chromosome capture (Hi-C) data, which reflects physical interactions on a chromosome and can highlight a window in the genome within which regulation of a particular gene is likely to occur.[15] Working within such a window, resources such as FANTOM5 and ENCODE provide data predictive of TSSs and enhancers.[16–19] Segmentation tools such as Segway and ChromHMM are also helpful as they use features such as DNAase1 hypersensitivity and epigenetic markers to help predict enhancer and promoter regions.[17] Finally the JASPAR database further informs these predictions by supplying transcription factor binding sites (TFBSs), which are features of RRs.[20] In combination with a gene transfer platform such as recombinant adeno-associated virus (rAAV) to screen the designs *in vivo*, these bioinformatics tools may allow for the design, and rapid testing, of new custom MiniPs. Importantly, MiniPs developed using one research and therapeutic modality, such as rAAV, may be generally useful in other viruses, plasmids, and in a genomic context.[21–25]

PAX6 is a well-studied gene, potentially amenable to the development of MiniPs. Although PAX6 is expressed in a variety of tissues including the central nervous system (CNS), pancreas, and small intestine, it is best known as the essential transcription factor for

[1]Centre for Molecular Medicine and Therapeutics at the BC Children's Hospital, University of British Columbia, Vancouver, British Columbia, Canada; [2]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; [3]Graduate Program in Bioinformatics, University of British Columbia, Vancouver, British Columbia, Canada; [4]Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, Canada; [5]Department of Ophthalmology, College of Medicine, University of Florida, Gainesville, Florida, USA; [6]Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada. Correspondence: EM Simpson (simpson@cmmt.ubc.ca)
Received 1 April 2016; accepted 13 June 2016

panocular development in species as diverse as flies (*Drosophila melanogaster*), mice (*Mus musculus*), and humans.[26–30] In humans, loss-of-function mutations produce the ocular disorder aniridia (OMIM: 106210). Although named for the lack of iris, aniridia is panocular, with vision loss attributable to three main causes: (i) hypomorphic fovea, (ii) progressive corneal clouding, and (iii) progressive glaucoma.[31,32] Of these, glaucoma is the best managed, leaving hypomorphic fovea and corneal clouding in need of a vision-saving therapy, for which gene-based therapies are being developed. For instance, it has already been reported that postnatal normalization of PAX6 in a mouse model of aniridia can restore electrical activity in the retina and mouse visual behavior, even when administration begins at postnatal day 14 (P14).[33] However, this effect was achieved using a small molecule drug that allows read-through of specific nonsense mutations, which represent only ~12% of reported human *PAX6* mutations (http://lsdb.hgu.mrc.ac.uk/home.php?select_db=PAX6). Therefore, a large portion of the aniridia patient community stands to benefit from other approaches to gene augmentation, such as rAAV gene therapy.

One challenge for *PAX6* gene therapy is that expression of the endogenous protein is complex, and inappropriate PAX6 could be detrimental. Ectopic expression of *PAX6* orthologues in *D. melanogaster* and *Xenopus laevis* resulted in the formation of ectopic eyes.[26,34] Furthermore, transgenic mice carrying human *PAX6*, and transcribing in total 2.5x normal levels, were found to have abnormal ocular development resulting in microphthalmia.[35,36] Finally, expression is temporally regulated with, for example, broad and robust developmental expression being restricted to ganglion, amacrine, horizontal, and Müller glia cells in the adult retina.[37–44] At least 39 cis-regulatory elements have been verified *in vivo*.[11,12,45–58] Of these elements, those with known adult expression, and those amenable to "cut down" are of greatest interest for this work, as they would be suited for gene therapies administered after development, and compatible with the 4.9-kb packaging capacity of rAAV.[59]

In this work, we utilize bioinformatics to design MiniPs with specific expression suitable for use in rAAV-based studies. Building on the extensive knowledge of *PAX6* regulation, we identified 31 potential RRs and selected nine for testing in seven MiniPs. DNA synthesis allowed precise and prompt generation of MiniPs, and a "plug and play" rAAV-genome plasmid enabled rapid virus production and testing in mice. We expected to identify unique aspects of PAX6 expression, but were pleasantly surprised to find that between only two promoters, all of the adult retina cell types that endogenously express PAX6 were captured. Thus, we have developed *PAX6* MiniPs that target therapeutically interesting cell types, which may be of use for the gene therapy treatment of diseases afflicting the inner retina such as diabetic retinopathy,[60] glaucoma,[61] and recessive retinitis pigmentosa inner retinopathy,[62] as well as for ocular *PAX6*-gene therapy for aniridia.

RESULTS

Highly-interactive regulatory neighborhood revealed at the *PAX6* locus

Topologically associating domains (TADs), which are sub-regions of chromosomes defined by an elevated frequency of intraregional DNA–DNA interactions in Hi-C experiments, were examined from mouse J1 embryonic stem cells (mESCs), mouse cortex cells, human H1 embryonic stem cells (hESCs), and a human IMR90 fibroblast cell line.[18,19] All 39 published RRs of *PAX6* (listed in Supplementary Table S1) are situated within the *PAX6*-containing TAD in all cell types examined (Figure 1 and Supplementary Figure S1). We then developed a local clustering approach to search for highly interactive neighborhoods. This revealed that within the *PAX6*-containing TAD, there is a highly-interactive regulatory neighborhood containing all the *PAX6* TSSs. Although *Pax6* expression is not high in mouse cortex cells and is supressed in mESCs,[63] this highly-interactive regulatory neighborhood overlapped almost perfectly between the two cell types (Figure 1a; mm9 coordinates: chr2:105495781-105653515 for mouse cortex cells at 99.7 percentile and chr2:105501001-105652563 for the mESCs at 99.6 percentile). Lifting over the genomic coordinates of the regulatory neighborhood from mouse mm9 to the human hg19 genome assembly (Figure 1b), it was revealed that the mouse regulatory neighborhood overlapped with the highly-interactive regulatory neighborhood similarly identified in the human data (overlaps of 98.7 and 100 percent for hESCs and the IMR90 fibroblast cell line respectively). Spanning from the 5′ end of *Pax6os1* to the last four exons of *Elp4* on the 3′ end, the <160 kb *Pax6* highly-interactive regulatory neighborhood overlaps with 33 (85%) previously published RRs. The rest of published RRs (15%) were located within a weaker interacting region situated between *Pax6* and the *Rcn1* promoter (Figure 1).

PAX6 is transcribed from at least three promoters

The *PAX6* exon structure was defined using the 10 different *PAX6* transcripts reported in UCSC (hg19 assembly; https://genome.ucsc.edu/cgi-bin/hgGateway), which had also been presented by at least one of the following resources: Protein Data Bank (www.rcsb.org/pdb/home/home.do), RefSeq (www.ncbi.nlm.nih.gov/refseq), SwissProt (www.uniprot.org/), or CCDS (www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi) (Supplementary Figure S2a). This complexity is the product of alternative splicing and the use of multiple promoters.[56] CAGE data from the FANTOM5 consortium supports a three-promoter model for *PAX6*, with transcription being driven from P0, P1, and Pα (Supplementary Figure S2b).[16] Interestingly, this CAGE data does not indicated the existence of a promoter P4.[48] For this analysis, the CAGE data was curated into three groups, after cancer and induced pluripotent stem cell reads were removed, producing: "All human tissues", "CNS tissues", and "ocular tissues" groupings. The all human tissues group contains all of the data presented in the other two groups, while the CNS and ocular tissues groups are mutually exclusive. Interestingly, in the "all humans tissues" group,[16] *PAX6* expression was predominantly driven by promoter P1, with P0 initiating proportionally fewer transcripts, and Pα only producing a small minority of the transcripts, which failed to exceed baseline in the CNS tissues. Additionally, while the transcripts initiated by P0 and Pα start from focused TSSs, the transcripts initiated by P1 appear to originate from a range of TSSs spanning more than 300 bp. Combining the 10 mRNA transcripts with TSSs results in a complex model of *PAX6* where each promoter drives at least two different isoforms of the mRNA (Supplementary Figure S2c).

Thirty-one RRs predicted by bioinformatic analysis of *PAX6*

We conducted an initial literature and database assessment of *PAX6*, and created entries for the human and mouse genes in the Transcription Factor Encyclopedia (www.cisreg.ca/cgi-bin/tfe/home.pl), and deposited regulatory data in PAZAR (www.pazar.info). Later we developed a computational approach to predict RRs within the highly-interactive regulatory neighborhood at *PAX6*. In brief, regulatory potential was computed using three combined criteria: conservation, ChIP-seq supported TFBS, and predicted RRs from segmentation methods (Segway

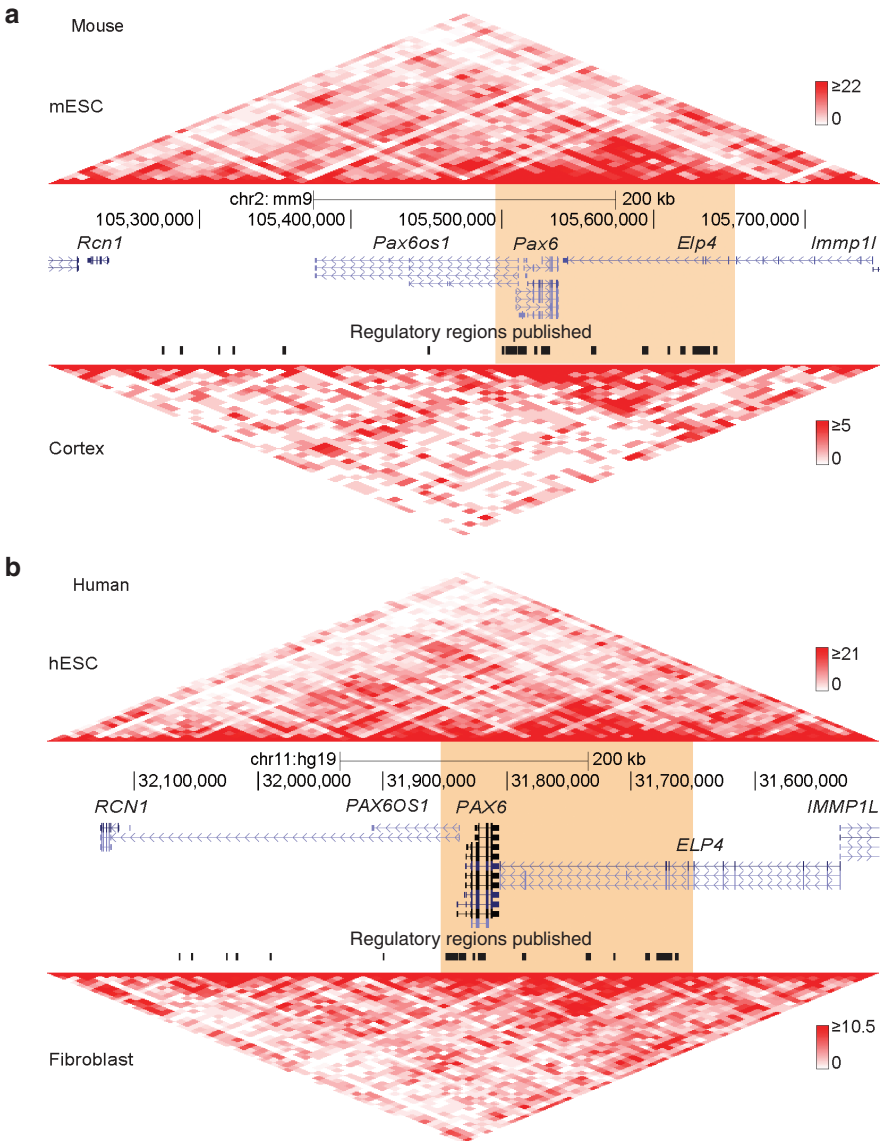

Figure 1 A paired box 6 (*PAX6*) containing highly-interactive regulatory neighborhood, encompassing the majority of previously published PAX6 regulatory regions (RRs), is revealed by mouse and human chromosome interaction data. Topologically associating domains (TADs) and high-throughput chromosome capture (Hi-C) data are indicative of interactions between distal RRs and promoters. Presented are two-dimensional heatmap visualizations of previously published Hi-C chromatin interaction datasets.[18,19] The interaction strength is indicated by color, ranging from red (set as ≥90th percentile of counts) to white (no observed interactions). These values correspond to the number of interacting segments observed in the DNA sequences for pairings of 10-kb bins. A *PAX6* containing highly-interactive regulatory neighborhood computed from mouse cortex cells is highlighted in orange. Gene transcripts are indicated in blue and the "Regulatory regions published" displays our curation of all previously published *PAX6* RRs as black rectangles. (**a**) Visualization of datasets from mouse cells: top, mouse embryonic stem cell (mESC) line J1; bottom, adult C57BL/6NCrl mouse cortex. The 90th percentile is 22 and five counts for mESC and cortex respectively. The displayed segment corresponds to 105,200,001–105,750,000 on Chromosome 2, mm9 assembly. (**b**) Visualization of datasets from human cells: top, human embryonic stem cell (hESC) line H1; bottom, human fibroblast cell line IMR90. The 90th percentile is 21 and 10.5 counts for hESC and IMR90 respectively. The position of the human *PAX6* containing highly-interactive regulatory neighborhood is matched by position using the lift-over function of the UCSC Genome Browser to the region defined in mice. The displayed segment corresponds to 32,170,000–31,500,001 on Chromosome 11, hg19 assembly.

and ChromHMM).[17] Regulatory prediction scores (RPS) for these three criteria were computed by applying a 200 bp sliding window to the *PAX6* highly interactive neighborhood. Each of the three criteria contributed a low, medium, or high RPS of 0, 0.5, or 1.0 respectively, for a maximum score of 3.0 (Figure 2, Supplementary Table S2). All overlapping and immediately adjacent (book-ended) windows with scores ≥ two were merged, which produced 31 RRs predicted to have high regulatory potential (Supplementary Figure S3). Of the 31 regions, 19 overlap with one or more previously published regulatory elements.

Seven *PAX6* MiniPromoters were constructed from nine bioinformatically predicted RRs

Of the 31 predicted RRs, nine were hand selected to be tested as components of MiniPs based on the literature and bioinformatics information available at the time (before 1 January 2012). For this work, RRs with known biological function as established in transgenic mice were considered, however as our system explores expression from a viral genome, regions with a breadth of supporting data were preferred. However, RRs overlapping with elements such as the pancreatic enhancer, repressor element, CNS

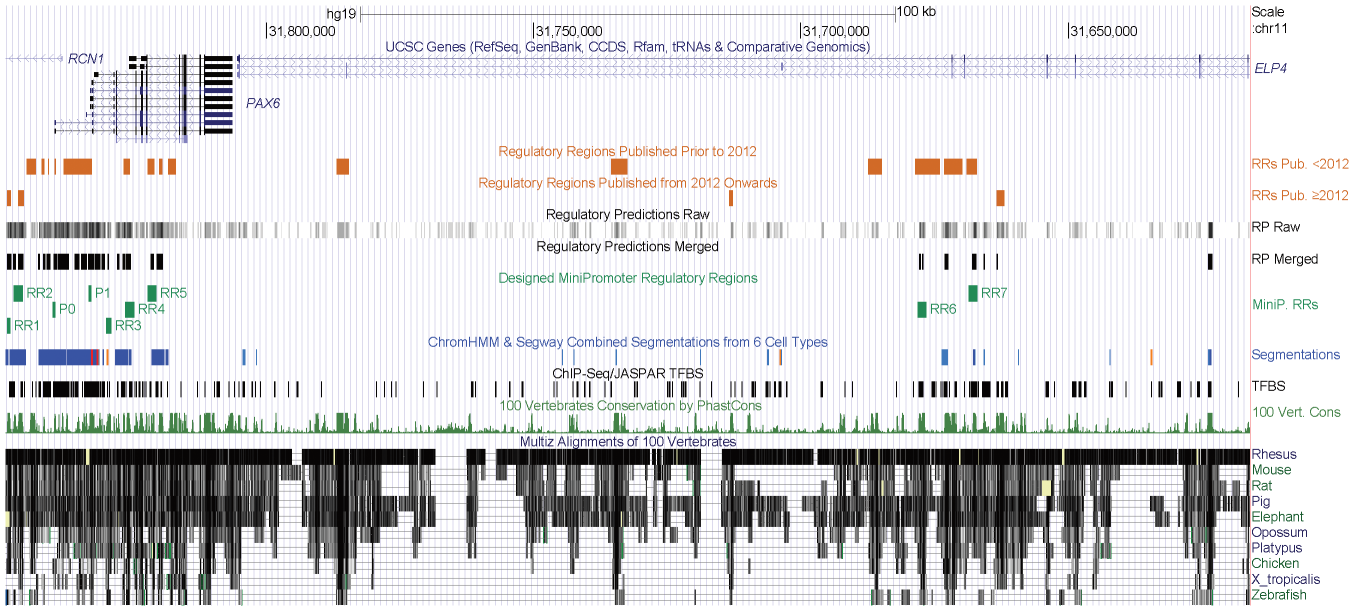

Figure 2 Bioinformatic analysis of the paired box 6 (*PAX6*) containing highly-interactive regulatory neighborhood revealed 31 putative regulatory regions (RRs). Visualization of the data used to predict RRs within the highly-interactive regulatory neighborhood (Chr11: 31,848, 751–31,616,062, hg19) specified in Figure 1. Tracks are described from top of figure to bottom. Transcripts are displayed, with black having a higher validation level than grey (details in Supplementary Figure S2). Orange rectangles in the RRs Pub tracks denote RRs manually curated from the scientific literature separated into those available prior to (pre-2012) and post experimental design (from 2012 to December 2015). White and black vertical lines in the regulatory prediction (RP) Raw track indicate low to high scoring regions respectively. Black rectangles in the RP Merged track represent high scoring regions (≥ 2.0) that were merged into 31 predicted RRs. Green rectangles in the MiniP RRs track represent the 12 RRs selected and manually refined to produce nine RRs for testing in MiniPromoters (RR1-RR7, P0, and P1). Data used for the RP included predicted classifications from the ChromHMM and Segway segmentation tools (red, promoter region including transcription start site(s); blue, enhancer; azure, weak enhancer or open chromatin cis regulatory element; orange, predicted promoter flanking region), ChIP-seq-supported transcription factor binding sites (TFBS, black rectangles), and phastCons conservation scores based on 100 vertebrate genomes (green histogram). Genome sequence similarity plots for a hand-selected set of 10 species are displayed.

element, CE2, and HS5^{+} were excluded from selection if they overlapped with previously published RRs that drive expression only during development, or exclusively outside of the retina.[47,50,53,57] Conversely, RRs overlapping with regions that have been previously shown to drive expression in the adult retina were favored producing: P0 and P1, which overlap with promoters 0 and 1 respectively;[54,56] RR4 with promoter α, the neural retina enhancer, a promoter α enhancer, and ele4H;[47,55,56] RR5 with CE1;[48] RR6 with HS234Z;[12] and RR7 with HS6 (ref. 50) (Supplementary Figure S3). Three RRs (RR2, RR5, and RR6) were formed by connecting two high scoring RRs with the small (≤500 bp) highly conserved sequence between them. The final sequence of P0 was determined by first aligning a previously described P0 sequence from mice to the human genome, and then trimming it down to 454 bp based on conservation.[64] A core promoter based on 454 bp of the sequence of P1 was designed by lengthening a smaller, previously tested, promoter sequence.[58] The 3′ end was extended to just before the 3′ end of exon one, and the 5′ end was extended to reach the final size of 454 bp.

Since the viral packaging size of rAAV is only ~4.9 kb,[59] the maximum size of a *PAX6* MiniPromoter was set at 2.15 kb for this study, leaving room in the rAAV genome for reporter constructs such as Emerald GFP (EmGFP, 720 bp) and other elements such woodchuck hepatitis virus posttranscriptional regulatory element (WPRE, 587 bp) a poly adenosine tail (222 bp), and inverted terminal repeats. Subtracting 454 bp for the core promoter RR included in each MiniPromoter, 1,696 bp was reserved for each of the seven remaining RRs. Taking conservation into consideration, the size of each RR was maximized to provide the best chance of capturing important regulatory sequences. The final sequence of each MiniPromoter contains at least one of RR1-RR7 and either P0 or P1 (Figure 3a).

Four *PAX6* MiniPromoters drive consistent EmGFP expression in *PAX6* expressing cells

All seven MiniPs were synthesized and cloned into a rAAV genome containing a chimeric intron, EmGFP reporter, WPRE mut6,[65] SV40 polyA sequence, and AAV2 inverted terminal repeats (Figure 3b). EmGFP was selected because it has been shown to fluoresce brighter than Enhanced GFP.[66] The viral genomes were packaged into rAAV2(Y272F, Y444F, Y500F, Y730F, and T491V) (hereafter referred to as rAAV2(QuadYF+TV) and administered by intravitreal injection into P14 mice. Using a GFP antibody to enhance the fluorescent signal and better visualize cellular processes, all MiniPs drove detectable EmGFP expression (Figure 4). Intravenous injections at P4 of rAAV9 were used to compare quantitatively Ple254 and Ple255 with the ubiquitous promoter smCBA (Supplementary Figure S4). This methodology does not label the outer retina well,[67] but does show the restricted expression of the two *PAX6* MiniPs, and that overall expression strength was similar to the ubiquitous promoter.

Four MiniPs (Ple254, Ple255, Ple259, and Ple260) drove consistent expression in the inner nuclear and ganglion cell layers of the adult mouse retina (Figures 4–8). Images of the same MiniP presented in Figure 4 and Figure 5 come from different mice, and are representative of consistent results for at least four of the five eyes injected per construct. For Ple254, an inconsistent observation of expression in the outer nuclear layer was also seen. Colocalization of GFP and PAX6 immunofluorescent staining suggested that, Ple254, Ple255,

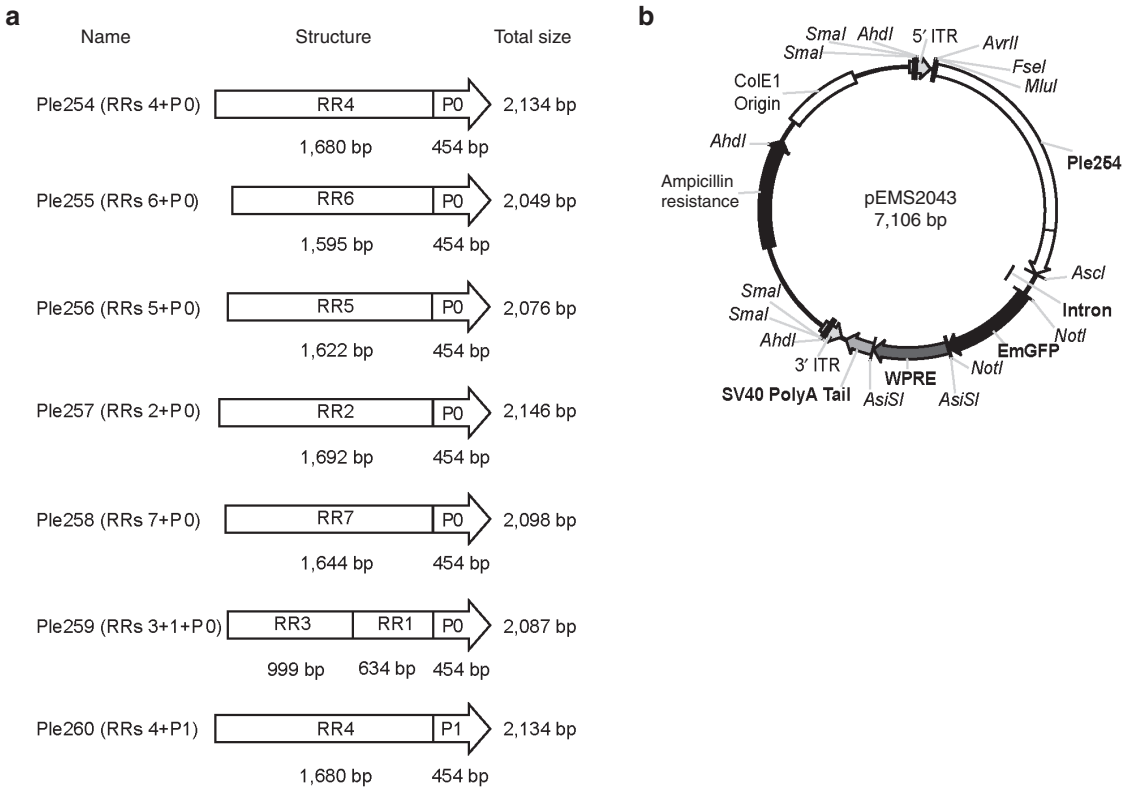

Figure 3 Paired box 6 (*PAX6*) MiniPromoters, concatenations of regulatory elements found within the *PAX6* containing highly-interactive regulatory neighborhood, were cloned into a custom recombinant adeno-associated virus (rAAV) genome. (**a**) Seven MiniPromoters, named Ple254-Ple260, were designed by concatenating seven hand selected regulatory regions (RRs) with either the P0 (Ple254-Ple259) or P1 (Ple260) core promoter sequence. Ple254 and Ple260 are related in that they both contain RR4, but differ in that they contain P0 and P1 respectively. (**b**) Custom rAAV plasmid backbone streamlined assembly of viral genomes. A plasmid backbone was assembled to facilitate easy cloning of promoters and reporters into a rAAV genome. A representative viral genome (pEMS2043) contains a 5′ inverted terminal repeat (light gray arrow pointing clockwise) restriction sites (*Avr*II, *Fse*I, *Mlu*I, and *Asc*I), a representative MiniPromoter (Ple254; white arrow with black outline), chimeric intron, Emerald GFP (EmGFP) reporter open reading frame (ORF) (black arrow) flanked by *NotI* sites (the 5′ *NotI* site forms a Kozak sequence with the 5′ end of the reporter construct), Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE; dark gray arrow) flanked by *Asi*SI restriction sites, and a 3′ inverted terminal repeat (light gray arrow pointing counter clockwise). The plasmid also carries an ampicillin resistance gene (black arrow) and a ColE1 origin of replication (gray rectangle).

Ple259, and Ple260 consistently drove EmGFP expression in patterns that overlap with the expression of PAX6 in the adult mouse retina (Figure 5). More specifically, colocalization of GFP immunofluorescence with antibody staining for Brn-3 (a ganglion cell marker) and Syntaxin (an amacrine cell marker) revealed that Ple254, Ple255, Ple259, and Ple260 drove expression in ganglion and amacrine cells, cell types that endogenously express PAX6 (Figures 5–8). Ple255 also drove expression in horizontal cells, as initially identified by GFP immunostaining of cells in in the inner nuclear layer with processes that extended through the outer plexiform layer (Figure 4), and subsequently by colocalization of GFP immunofluorescent staining with PAX6 (Figure 5) and the horizontal cell marker Calbindin-D-28K (Figure 7). Ple259 also drove expression in Müller glia, as initially identified by GFP immunostaining of cells residing in the inner nuclear layer and process that extended from the outer nuclear layer through the inner nuclear layer to the ganglion cell layer (Figures 4 and 5), and subsequently by colocalization of GFP immunofluorescent staining with the Müller glia marker SOX9 (Figure 8).[41]

DISCUSSION

A highly-interacting local neighborhood at the *PAX6* locus defined the search area for predicting RRs. Chromatin interaction data has been shown to reflect the degree of interaction between pairs of fragments in the genome, including that of promoter and RRs.[15] Using public Hi-C datasets for mouse and human cells, we found all previously published RRs that drive the endogenous expression of PAX6 to be within the *PAX6*-containing TADs in all four cell types examined. The identification of a highly-interactive regulatory neighborhood encompassing the *PAX6* locus narrowed our focus to a region that contains 85% of the previously described *PAX6* RRs. This indicated an association between 3D proximity and regulatory targets, which further supports the potential application of chromatin interaction data in guiding the identification of novel RRs of *PAX6*. Thus, we used the boundaries of the highly-interactive regulatory neighborhood to focus our bioinformatics work, within which we predicted RRs for use in *PAX6* MiniP development.

PAX6 promoter analysis supported selection of P0 as the core promoter for MiniP development. CAGE data from the FANTOM5 consortium supports a three-promoter model of *PAX6*. In choosing amongst the *PAX6* core promoters, the CAGE data revealed that P1 dominates transcript initiation in both the CNS and ocular tissues examined. However, it was also noted that P1 initiates transcription over a 300 bp region, raising concerns regarding the size of the promoter region that would be needed. To test this, we defined a small P1 promoter and tested it successfully in Ple260. At P0, the CAGE data revealed less, but clearly presents transcription initiation in the relevant tissues, which was focal. In addition, a core mouse promoter had previously been used from this site. In combination

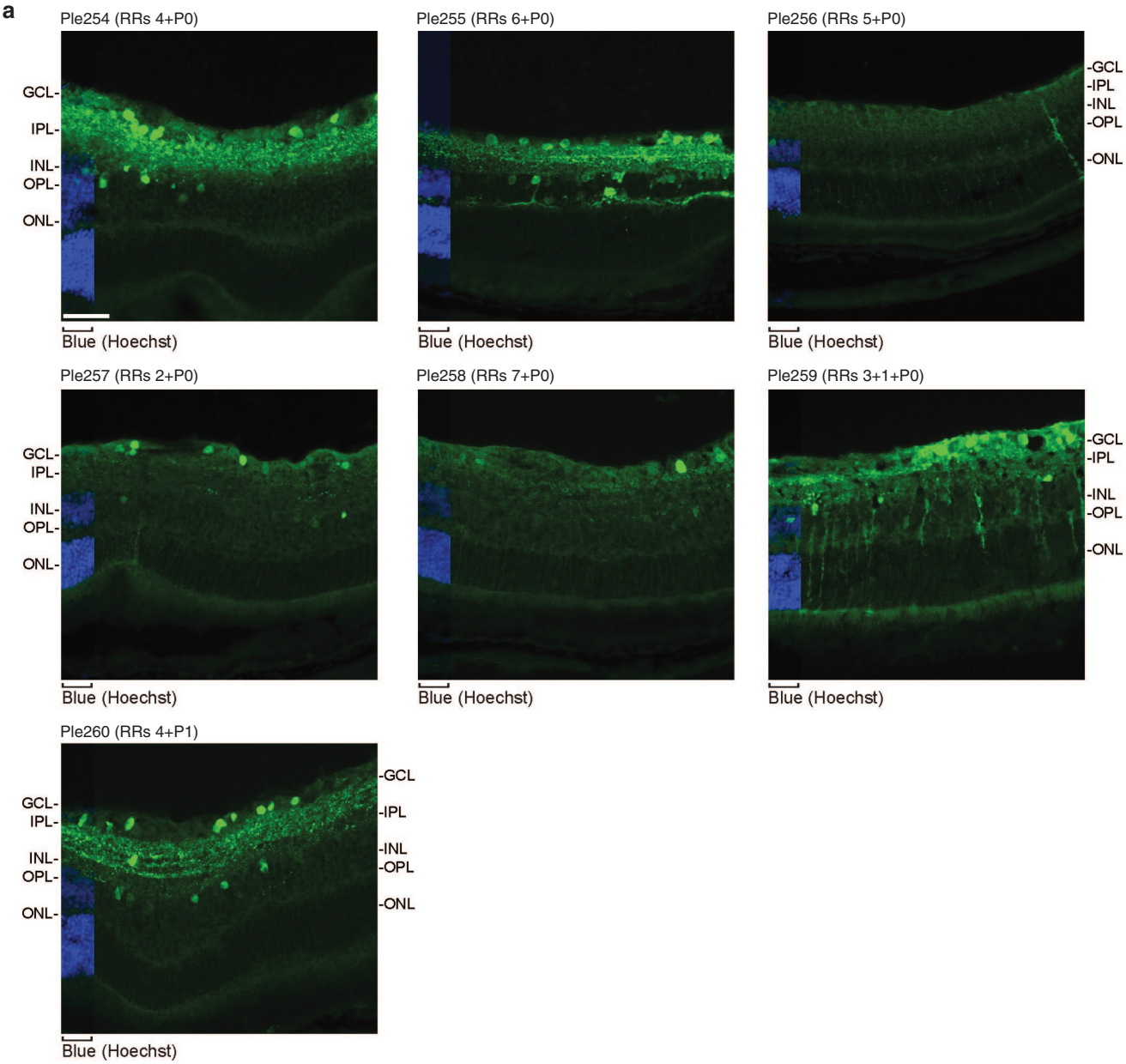

Figure 4 Four paired box 6 (*PAX6*) MiniPromoters drove consistent Emerald GFP (EmGFP) expression overlapping with PAX6 in the adult mouse retina. (**a**) Green (GFP antibody) labelling of mouse retinas transduced after intravitreal injection with rAAV2(QuadYF+TV) encoding *PAX6* MiniPromoters driving EmGFP expression revealed MiniPromoter expression patterns. Ple254, Ple255, Ple259, and Ple260 all drove consistent EmGFP expression in the ganglion cell layer (GCL) and inner nuclear layer (INL) whereas Ple256, Ple257, and Ple258 drove inconsistent rare EmGFP expression. IPL, inner plexiform layer; OPL, outer plexiform layer; scale bar, 50 µm (all images taken at same magnification).

with the α enhancer, this P0 mouse core promoter was found to drive expression in the retina of postnatal day 20 mice.[58,64] Thus, we conservatively chose and successfully defined a small human P0 promoter in Ple254, Ple 255, and Ple 259.

Bioinformatic analysis recapitulated and refined previously published *PAX6* RRs. During the Pleiades Promoter Project, MiniP design was largely based on RRs predicted using conservation, literature-based annotation, and early ChIP-chip data.[8,9] Expanding on these techniques we used recent computational tools such as Hi-C chromatin capture, ENCODE, and FANTOM5 data sets to explore and evaluate the *PAX6* locus.[8,17,68] Using the highly-interactive regulatory neighborhood as a guide, we predicted RRs for *PAX6* employing three criteria: conservation, TFBS, and predicted sequence classification (combined Segway and ChromHMM segmentation). This approach, while highlighting previously reported RRs such as CE1 and HS234Z, also revealed new potential RRs for future investigation.[12,48,54] Pre-empting some of this work, it has been reported recently that sequencing of the *PAX6* loci in elephant sharks revealed new RRs such as agCNE9 and agCNE11 which overlap with RR1 and RR2 respectively.[51] Thus, this combined bioinformatics approach has potential for predicting regulatory elements to gain a deeper insight into how other genes are regulated, or to guide the design of other tissue- and cell-type specific MiniPs.

Nine bioinformatically predicted RRs were used to make seven MiniPs for rapid screening in rAAV2(QuadYF+TV). A total of 31 RRs were predicted to have high regulatory potential, of which 19 overlapped one or more previously published RRs. While it may have seemed easier to use the previously published RRs, many of

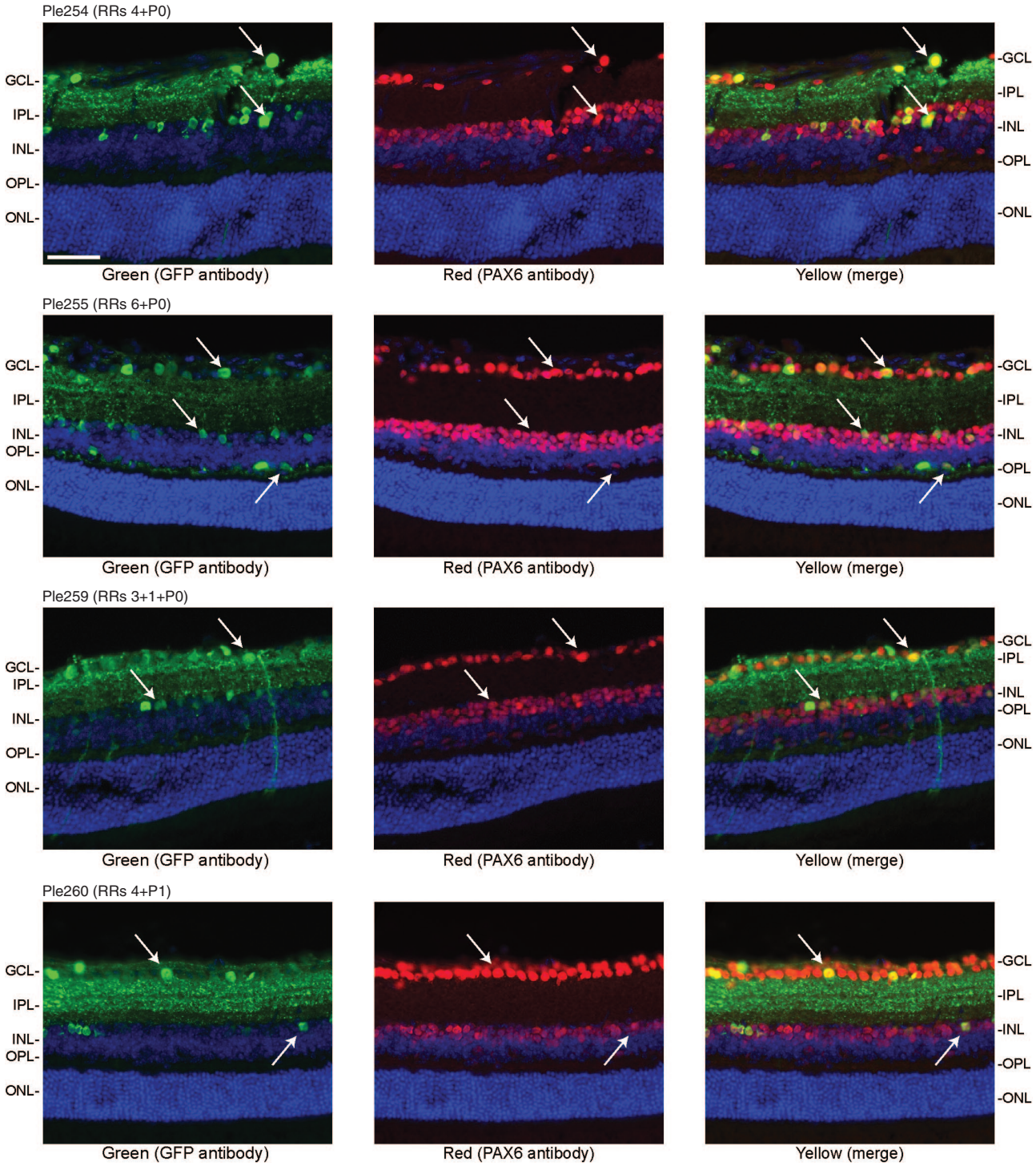

Figure 5 Ple254, Ple255, Ple259, and Ple260 drove expression overlapping with the expression pattern of paired box 6 (PAX6) in the mouse retina. Representative histological sections of mouse retinas transduced by intravitreally injected rAAV2(QuadYF+TV) encoding *PAX6* MiniPromoters driving EmGFP (Emerald GFP). Green (GFP antibody) and red (PAX6 antibody) colabelling revealed that all four MiniPromoters drive EmGFP expression overlapping with elements of the PAX6 expression profile in the ganglion cell layer (GCL) and inner nuclear layer (INL). IPL, inner plexiform layer; OPL, outer plexiform layer; ONL, outer nuclear layer; blue, Hoechst; arrows, examples of colabeled cells; scale bar, 50 µm (all images taken at same magnification).

these regions were unsuitable for this study. First, RRs that drove expression exclusively during development, or exclusively outside of the retina, were not suitable in the design of MiniPs for adult retinal expression. Second, MiniPs needed to be ~2 kb to be useful in the size-restricted rAAV genome. Consequently, new prediction of regulatory sequences was important to uncover elements of appropriate size for use in MiniPs, or emphasize the regions of previously described elements to allow for "cut down".

In our experiments, it was essential that the tropism of the virus serotype did not limit the expression profile of the MiniPs. The initial published characterization of rAAV2(QuadYF+TV) demonstrated that it can transduce cells across the entire retina, and in particular

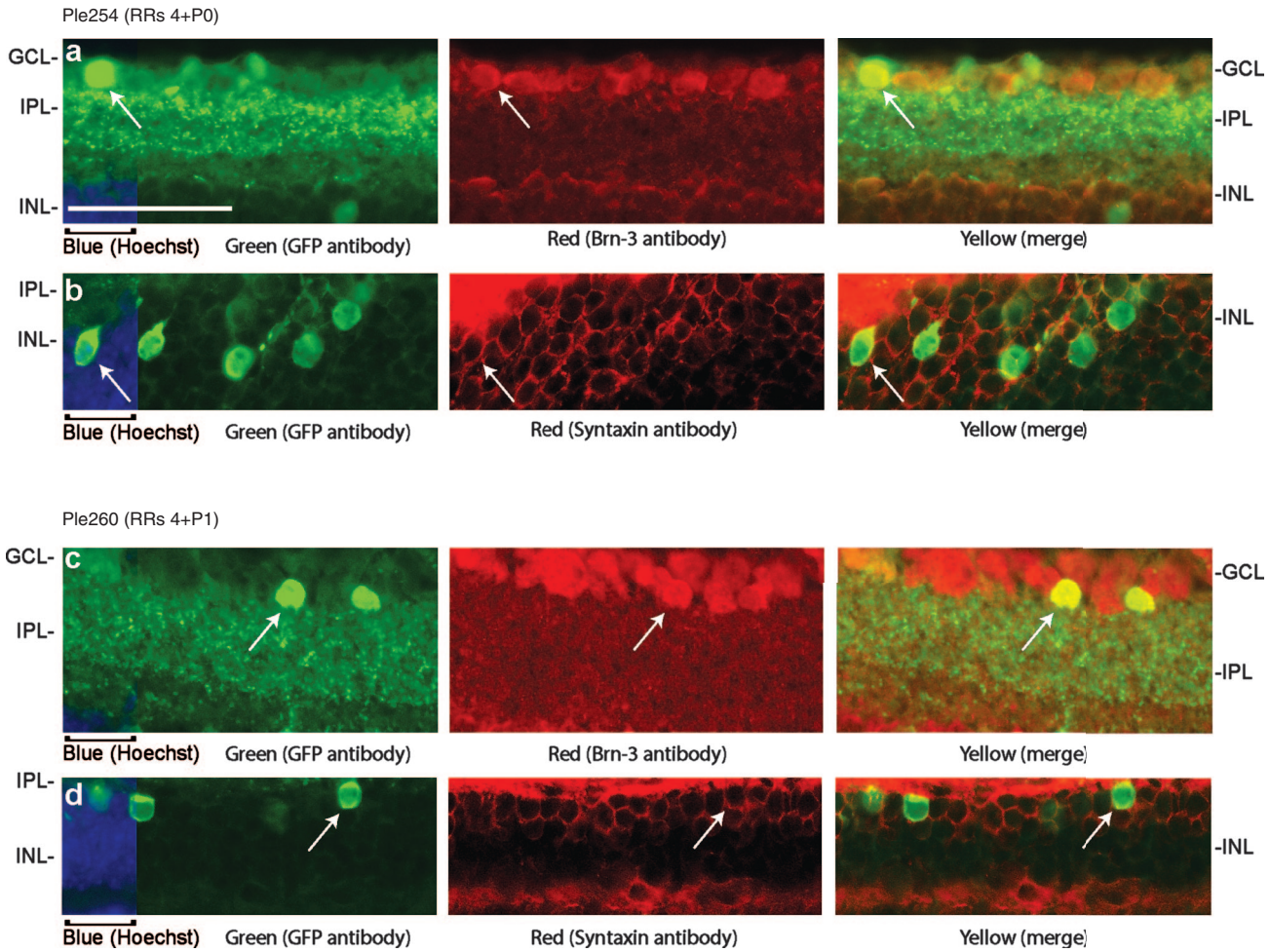

Figure 6 Ple254 and Ple260 drove EmGFP (emerald GFP) expression in mouse retinal ganglion cells and amacrine cells. (**a** and **c**) Green (GFP antibody) and red (Brn-3 antibody, retinal ganglion cell marker) colabelling revealed that both Ple254 and Ple260 drive expression in retinal ganglion cells. (**b** and **d**) Green (GFP antibody) and red (Syntaxin antibody, an amacrine cell marker) colabelling revealed that Ple254 and Ple260 drive expression in amacrine cells. GCL, ganglion cell layer; IPL, inner plexiform layer; INL, inner nuclear layer; OPL, outer plexiform layer; ONL, outer nuclear layer; blue, Hoechst; arrows, examples of colabeled cells; scale bar, 50 µm (all images taken at same magnification).

the photoreceptors.[24] Our recent work with MiniP Ple155 revealed that rAAV2(QuadYF+TV) can transduce bipolar cells.[7] Here we demonstrate transduction of ganglion, amacrine, Müller glia, and horizontal cells, which in combination with previous efforts, provides evidence that rAAV2(QuadYF+TV) is capable of transducing all major cell types of the retina. Thus, the MiniPs drive the restricted expression profile observed rather than the virus serotype. This is further supported by the change in expression profile observed between *PAX6* MiniPs, when RRs are shuffled while the injection technique, serotype, inverted terminal repeats, and all other viral components are held constant.

Four *PAX6* MiniPs drove consistent EmGFP expression from rAAV when delivered intravitreally, which overlaps with PAX6 expression in the adult mouse retina. Using rAAV2(QuadYF+TV) to test MiniPs directly in the mouse eye, a departure from the more time consuming method of testing MiniPs in genome-engineered mice, we found that four of the seven MiniPs drive expression that overlaps with that of PAX6. Rather than producing a promoter that drove expression in a single PAX6 expressing cell type, the objective of this study was to generate a promoter that best recapitulated the entirety of the expression profile of PAX6 in the adult retina. Toward this objective two *PAX6* MiniPs, Ple255 and Ple259, are particularly interesting in that each capture three of the four cell types that express PAX6, and together they capture the entirety of PAX6 expression in the adult mouse retina.[40,41] As both specificity and strength are important features in a MiniP, both promoters are appealing candidates for future optimization to capture the entire adult retinal expression pattern of PAX6, and for use in PAX6 research and gene therapy for aniridia and other ocular diseases.

MATERIALS AND METHODS

Chromatin interaction from Hi-C dataset

Publically available datasets for TADs and Hi-C experiments using the restriction enzyme *Hind*III in mESC, mouse cortex cells from eight-week old male C57BL/6NCrl mice, hESC, and human IMR90 fibroblast cells were accessed to explore chromatin interactions at the *PAX6* locus.[18,19] Summary files from Gene Expression Omnibus (GSE35156), which listed paired-end reads mapped to mouse mm9 and human hg18 genome assemblies were retrieved. We mapped the genomic coordinates of TADs and paired-end reads, originally defined on the hg18 human assembly, onto the hg19 build using the liftOver tool provided by the UCSC Genome Browser. Reads from duplicated Hi-C datasets of the same cell type were combined. Numbers of paired-end reads linking each possible pair of 10-kb bins were counted, and each 10-kb bin was set to overlap with 6 kb of the bin that came before it. The datasets were plotted as two-dimensional heat maps using the "HiTC" R-package (www.bioconductor.org/packages/release/bioc/html/HiTC.html, version 1.6.0, R version 3.0.2).[69]

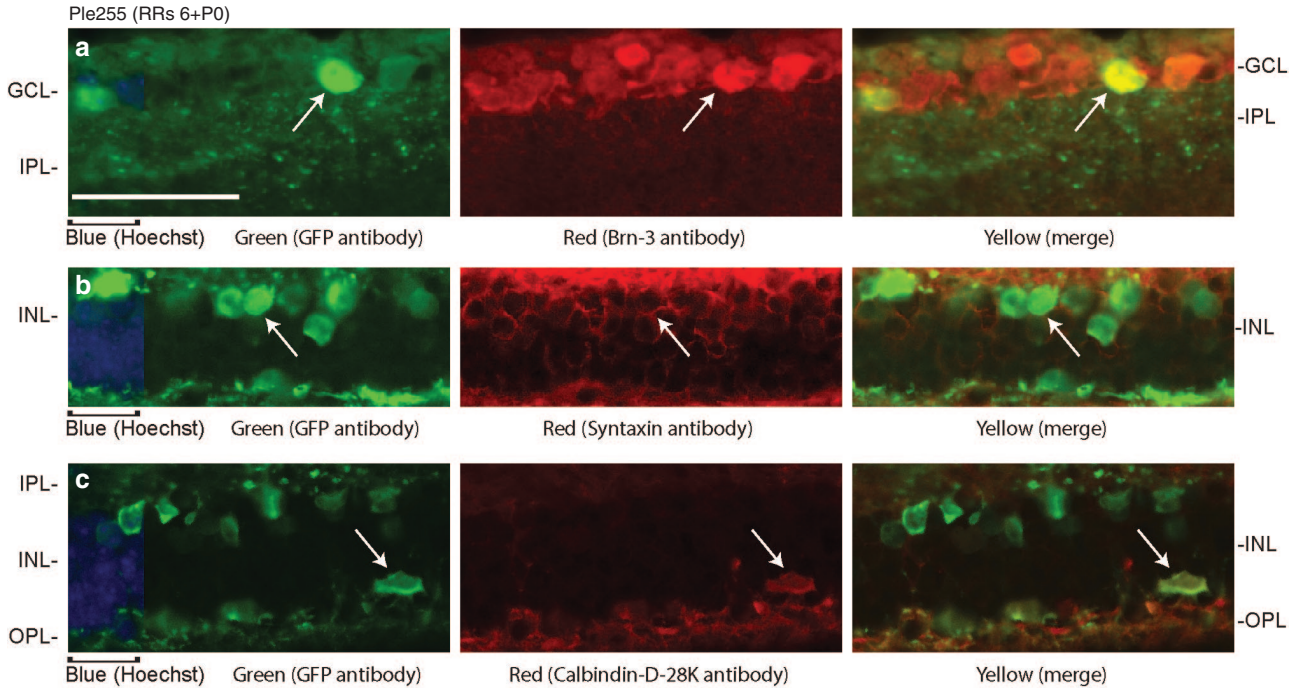

Figure 7 Ple255 drove EmGFP (emerald GFP) expression in mouse retinal ganglion, amacrine, and horizontal cells. (**a**) Green (GFP antibody) and red (Brn-3 antibody, a retinal ganglion cell marker) colabelling revealed that Ple255 drives expression in retinal ganglion cells. (**b**) Green (GFP antibody) and red (Syntaxin antibody, an amacrine cell marker) colabelling revealed that Ple255 drives expression in amacrine cells. (**c**) Green (GFP antibody) and red (Calbindin-D-28K antibody, a horizontal cell marker) colabelling revealed that Ple255 drives expression in horizontal cells. GCL, ganglion cell layer; IPL, inner plexiform layer; INL, inner nuclear layer; OPL, outer plexiform layer; ONL, outer nuclear layer; blue, Hoechst; arrows, examples of colabeled cells; scale bar, 50 µm (all images taken at same magnification).

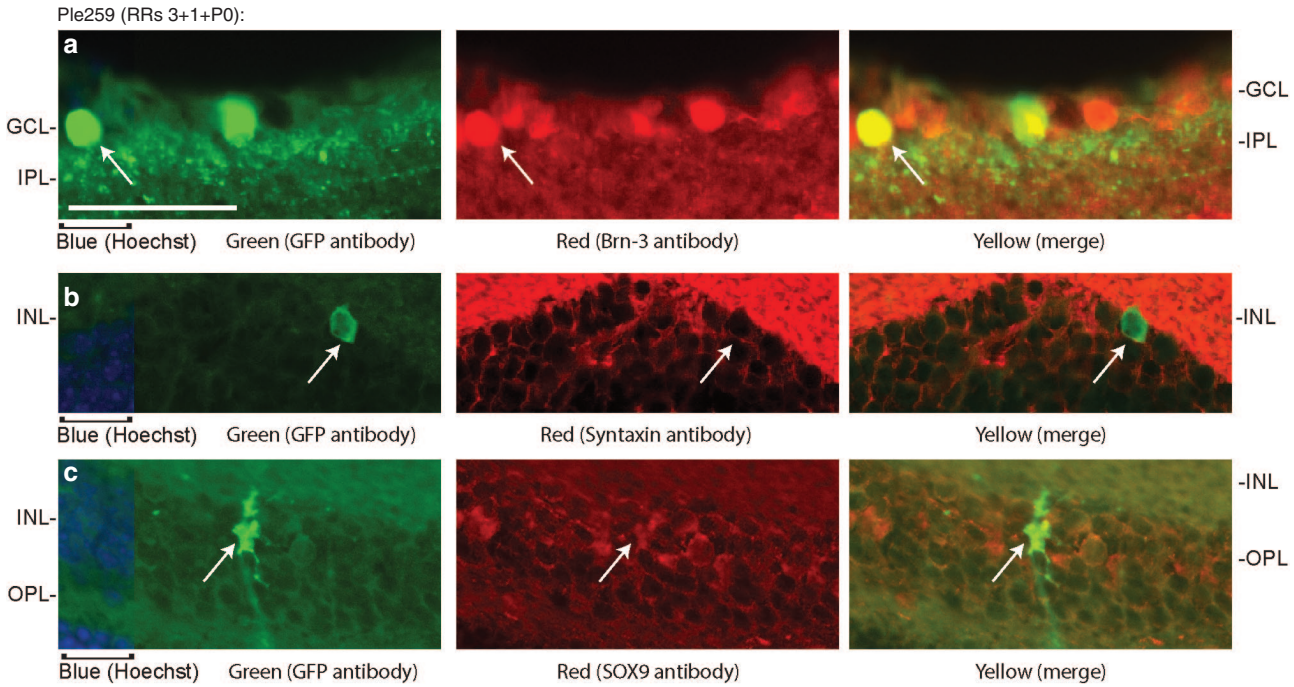

Figure 8 Ple259 drove EmGFP (emerald GFP) expression in mouse retinal ganglion, amacrine, and Müller glia cells. (**a**) Green (GFP antibody) and red (Brn-3 antibody, retinal ganglion cell marker) colabelling revealed that Ple259 drives expression in retinal ganglion cells. (**b**) Green (GFP antibody) and red (Syntaxin antibody, amacrine cell marker) colabelling revealed that Ple259 drives expression in amacrine cells. (**c**) Green (GFP antibody) and red (SOX9 antibody, Müller glia marker) colabelling revealed that Ple259 drives expression in Müller glia. GCL, ganglion cell layer; IPL, inner plexiform layer; INL, inner nuclear layer; OPL, outer plexiform layer; ONL, outer nuclear layer; blue, Hoechst; arrows, examples of colabeled cells; scale bar, 50 µm (all images taken at same magnification).

Local clustering approach to identify highly interactive neighborhoods

Highly interactive regions around *PAX6* in all cells were identified through a local neighborhood clustering approach. A search was initiated from the *PAX6* containing TAD of each corresponding cell type. In a sliding window analysis with each window containing 2*n*+1 of the 10-kb bins, where *n* is the number of bins extended from the center bin of each given window in both directions, we summed the total interactions for the 2*n*+1 consecutive bins. We determined the maximum interaction sum among all analyzed windows containing all *PAX6* TSSs (as specified in UCSC Genes annotation), and reported the percentile of this observed sum relative to the distribution of all sums observed for all windows of the same size within the TAD. We report this percentile (0–100) as the interactive score. The procedure was repeated for *n* from 5–45 for each cell type, and the highly interactive neighborhood was defined as the window of size 2*n*+1 with the highest interactive score.

PAX6 transcription start sites

The promoter structure of *PAX6* was delineated using capped 5′ mRNA end positions determined from FANTOM5 cap analysis gene expression (CAGE) data.[16] *PAX6* CAGE data was collected for all available human tissues using the ZENBU data explorer (http://fantom.gsc.riken.jp/zenbu/, accessed January 2014). The CAGE data was then manually curated to exclude reads from cancer cells and induced pluripotent stem cell experiments, generating the "All human tissues" group. From the "All human tissues" reads, the CNS tissues (excluding the neural retina) were selected and copied into the "CNS tissues group", similarly reads from ocular tissues (including the neural retina) were selected and copied into "Ocular tissues" group. The resulting "CNS tissues" and "Ocular tissues" groups contained mutually exclusive subsets of the "All human tissues" group.

Computational prediction of RRs

The regulatory potential of regions within the *PAX6* regulatory domain (chr11:31,616,062-31,848,751 on the hg19 assembly) was computed using three combined criteria: conservation, ChIP-seq supported TFBS, and predicted RRs from segmentation methods (Segway and ChromHMM).[17] RPS for these three criteria were computed by applying a 200 bp sliding window to the *PAX6* highly interactive neighborhood with a step size of 100 bp (n = 2,325). Each of the three criteria contributed a low, medium, or high RPS of 0, 0.5, or 1.0 respectively, for a maximum score of 3.0; details of the individual component scoring follow.

Conservation scores were computed using the 100 vertebrate phastCons data from the UCSC genome browser.[70] For each window the mean phastCons score was calculated. The low, medium, and high score thresholds were determined by taking the distribution of the mean phastCons scores for each of the 2,325 scoring windows and applying a Gaussian mixture model using the R-statistics package (www.r-project.org, version 3.1.2) with the mixtools library (http://cran.r-project.org/web/packages/mixtools/index.html, version 1.0.1). The data were consistent with two overlapping distributions. The right-hand distribution was consistent with selective pressure. The mean and standard deviation of this distribution was used to compute the score thresholds with a low, medium, and high RPS corresponding to mean phastCons scores ≤ 0.17, between 0.17 and 0.79, and > 0.79 respectively.

Predicted TFBSs within ChIP-seq peaks were retrieved from the MANTA database.[71] Precisely, TFBS were defined by scanning the peaks from a set of 477 TF ChIP-seq experiments from ENCODE (http://www.encodeproject.org, accessed March 2014)[72] and PAZAR (www.pazar.info, accessed March 2014)[73] with the corresponding transcription factor binding profiles were retrieved from the JASPAR database of transcription factor binding site profiles (http://jaspar.genereg.net/, version 5.0).[20,73–75] All positions within ChIP-seq peaks with a relative profile score ≥85% were recorded. The count of TFBS within each scoring window was used to assign the RPS. Windows with either no TFBS or >10 TFBS were assigned a low RPS. The rationale for the latter is based on the concept that too many binding sites are suggestive of nonspecific binding properties.[76] Windows with a count of 1–5 TFBS were assigned a medium RPS and windows with 6–10 TFBS were assigned a high RPS. Similar to the phastCons scoring method, the choice of threshold on the number of predicted TFBS used to assign low, medium, or high RPS was determined based on the distribution of the TFBS counts.

The combined Segway and ChromHMM segmentation data[17] was obtained from the ENCODE project at UCSC (http://genome.ucsc.edu/ENCODE/downloads.html, accessed March 2014). All segments within the *PAX6* regulatory domain predicted to be either weak enhancers, enhancers, promoter flanking regions, or TSS were used. Windows overlapping at least one enhancers or TSS element were assigned a high RPS. Windows overlapping at least one weak enhancers or promoter flanking regions element (but no enhancers or TSS elements) were assigned a medium RPS, and windows with no overlapping elements received a low RPS. For each of six profiled cell types (GM12878, H1-hESC, HeLa-S3, HepG2, HUVEC, and K562), the presence of enhancers, promoters, and TSSs were documented.

Regulatory region selection and MiniPromoter design

Predicted RRs were hand selected for testing in rAAV as part of a MiniP. From the pool of predicted RRs, regions were excluded from selection if they overlapped with previously characterized RRs that drive expression exclusively outside of the eye, or exclusively during development. Conversely, regions were favored if they overlapped with previously published elements known to drive expression in the adult retina or were previously untested. The boundaries of each regulatory region were determined by conservation, where the edges were extended in a direction consistent with conservation up to the maximum size, or the loss of conservation constituted the boundary of the RR.

Cloning of the rAAV backbone and viral genomes

The expression cassette (*Eco*RI site, multiple cloning site, *Mlu*I site, Ple251 MiniP, *Asc*I site, pCI (chimeric intron), *Not*I site, icre open reading frame, *Not*I site, *Asi*SI site, a mutant WPRE mut6 (ref. 65) sequence, *Asi*SI site, SV40 poly-adenosine tail, and *Sal*I *site*) was synthesized (DNA2.0, Menlo Park, CA) and cloned into the *Eco*RI/*Sal*I sites of P2393 (pENN.AAV.tMCK.Pl.ffluc.bgh; University of Pennsylvania, Philadelphia, PA). An EmGFP open reading frame, sequence from Vivid ColorspcDNA6.2/N-emGFP-DEST (Life Technologies, Carlsbad, CA), was synthesized (DNA2.0) with *Not*I sites on the 5′ and 3′ ends of the construct.[66] Icre was removed by *Not*I digest and EmGFP was ligated into the *Not*I sites to produce an rAAV backbone carrying the EmGFP reporter. Ple254, Ple255, Ple256, Ple257, Ple258, and Ple260 were synthesized (DNA2.0), and Ple259 was synthesized (Integrated DNA Technologies, Coralville, IA), all with *Mlu*I and *Asc*I restriction sites on the 5′ and 3′ ends respectively. The rAAV backbone and MiniPs were double digested with *Mlu*I and *Asc*I, and the MiniPs were subsequently ligated into the rAAV backbone.

Packaging of viral genomes into rAAV2(Y272F, Y444F, Y500F, Y730F, and T491V) and rAAV9

rAAV genome vector plasmids were packaged into a capsid variant of rAAV2 with four tyrosine to phenylalanine mutations (Y272F+Y444F+Y500F+Y730F) and a threonine to valine (T491V) mutation, referred as rAAV2(QuadYF+TV).[24] Packaging, purification, and the titre was measured at the University of Florida Retinal Gene Therapy Group vector lab as previously described.[77,78] The virus was suspended in Balanced Salt Solution (Alcon Canada, Mississauga, Canada) + 0.014% tween 40 producing a minimum titre of 10^{13} viral genomes/ml (vg/ml) as determined by quantitative polymerase chain reaction (qPCR). Virus was then shipped to the University of British Columbia on dry ice and stored at −80°C upon arrival. Additionally, a subset of the MiniPs (Ple254 and Ple255) and a ubiquitous control (smCBA) were packaged into rAAV9 at the Vector Core at the University of Pennsylvania (Philadelphia, PA).

Production of postnatal day 4 and 14 mice

Virus was injected into P14 B6129F1 hybrid mice generated by mating C57BL/6J (Jackson Laboratory, JAX Stock # 000664, Bar Harbor, ME) dams and 129S1/SvImJ (JAX Stock # 002448) sires. Mating cages were monitored daily for newborn pups, starting 18 days after the dam and sire were setup. The day the pups were found was recorded as postnatal day 0 (P0). Mouse pups were then left undisturbed in the mating cage with their parents until the day of injection, P4 or P14.

Intravitreal and intravenous injection of mouse pups

For intravitreal injections, virus was diluted to 8.22×10^{12} vg/ml with BSS (Alcon Canada, Mississauga, Canada) and further diluted to 5×10^{12} vg/ml with BSS + 0.05% Fast Green (Sigma Aldrich, Catalog# F7252, St. Louis, MO). For each injection, 5 µl of virus was loaded into a BD Ultra-Fine II insulin syringe (Becton Dickinson, Catalog# 328289, Franklin Lakes, NJ). P14 mice

were anesthetized with isoflurane and placed under a dissecting microscope. The right eye was covered in Refresh Lacri-lube (Allergen, Dublin, Ireland) and the left eye was washed with Eye Stream (Alcon Canada) and treated with one or two drops of Alcaine Eye Drops (Alcon Canada). Mice were then ear notched for identification. Next a BD 26G3/8 hypodermic needle (Becton Dickinson Cat# 305110) was used to make an aperture through the conjunctiva adjacent to the limbus on the nasal side of the eye. Finally, the insulin syringe was inserted through the sclera on the temporal side of the eye, into the intravitreal space, and 5 µl of rAAV solution (5×10^{12} vg/ml) was administered.

For intravenous injections, viruses were diluted to 1×10^{13} vg/ml in phosphate-buffered saline +0.05% Fast Green (Sigma Aldrich) and 50 µl were injected into the superficial temporal vein using a 30-gauge needle and 1 cc syringe. Mouse pups were then tattooed for identification and returned to their cage.

Tissue harvesting, sectioning, fluorescent antibody staining, and epifluorescence quantification

Eyes were collected 40 days (intravitreal injections) and 28 days (intravenous injections) post injection, fixed in 4% paraformaldehyde for 2 hours at 4˚C, rinsed with phosphate buffer (pH 7.4) and dehydrated in 25% sucrose overnight at 4 ˚C. Eyes were embedded in Tissue-Tek O.C.T. compound (Sakura Finetek, Catalog# 4583, Torrance, CA) and 16 µm (intravitreal injections) or 20 µm (intravenous injections) sections were cut with a Microm HM550 cryostat (Thermo Scientific, Waltham, MA).

For fluorescent antibody staining, sections were blocked for 30 minutes at room temperature in 10% bovine serum albumin (Sigma Aldrich, Catalog#A7906 St. Louis, MO) + 0.3% Triton X-100 (Sigma Aldrich, Catalog# T8787, St. Louis, MO). Once blocked, sections were incubated in primary antibody stain (GFP antibody (1:100; AVES, Catalog# GFP-1020, Tigard, OR)), PAX6 antibody (1:100; Covance Cat# PRB-278P, Princeton, NJ), Brn-3 antibody (1:100; Santa Cruz Biotechnology Catalog# sc-28595, Dallas, TX), Syntaxin antibody (1:100; Sigma-Aldrich Catalog# S0664, St. Louis, MO), Calbindin-D-28K antibody (1:100; Sigma-Aldrich Catalog# C9848, St. Louis, MO), or SOX9 (1:100; Millipore, Catalog# ABE571, Billerica, MA) in phosphate buffer containing 2.5% bovine serum albumin with 0.1% Triton X-100) at room temperature for 2 hours. Next, sections were rinsed 3 times for 5 minutes each in phosphate buffer and stained with a secondary antibody (either Alexa594 conjugated goat antirabbit immunoglobulin, Alexa594 conjugated goat antimouse immunoglobulin, or Alexa488 conjugated goat antichicken immunoglobulin (1:1,000; Molecular Probes Catalog numbers A-11012, A-11005, and A-11039 respectively, Eugene, OR)) and counter stained with Hoechst (1:1,000 from 2 µg/ml, Sigma-Aldrich Cat# 881405, St. Louis, MO) for 1 hour at RT. Sections were given 3 five-minute washes and were mounted with ProLong Gold Antifade Mountant (Life Technologies, Catalog# P36930, Carlsbad, CA). Sections were imaged on a Bx61 Microscope (Olympus America, Centre Valley, PA) using cellSens software (Olympus America, Centre Valley, PA) at either 10× magnification (Figures 4 and 5) or 20× magnification (Figures 6–8). Raw image files were converted to composite tagged image file format (TIFF) files using imageJ software (http://imagej.nih.gov/ij/, version 1.48)) with the Bio-Formats plugin (http://www.openmicroscopy.org/site/support/bio-formats5.1/users/imagej/). Images were imported and the multichannel images were exported as composite, or single color, TIFF images. To produce images with only partial blue and green overlay (Figures 4,6 and 7 and Supplementary Figure S4) both a blue green composite TIFF, and a single green channel image TIFF, from the same source image were produced. The images were then aligned with the composite image arranged above the single channel image. A clipping mask was then applied to crop the blue and green composite image so that it was only visible along the leftmost boarder of the green image.

For EmGFP cell epifluorescence quantification, sections were stained with either the Brn-3 or Syntaxin antibodies and prepared for imaging as described above. Quantification was performed on sections from three separate mice per MiniP using imageJ software as previously described.[79,80] Briefly, individual cells expressing EmGFP and colabeled with either Brn-3 or Syntaxin were traced by hand. Integrated density and mean value measurements were recorded for the green (EmGFP) channel for each cell. Four background measurements were also taken by tracing and measuring nonepifluorescent regions of similar size, adjacent to epifluorescent cells. Cell epifluorescent measurements were then calculated as the integrated density of the cell minus the product of the area of the cell and the mean epifluorescence of background readings. All images were taken at the same Fluorescein isothiocyanate exposure and gain settings.

Materials and data availability

All MiniPromoter constructs and rAAV genome plasmids have been made available to the research community through AddGene (www.addgene.org).

CONFLICT OF INTEREST

J.W.H., R.S.M., W.W.W., E.M.S., and the University of British Columbia are applying for patent protection for the *PAX6* MiniPs. W.W.H. and the University of Florida have a financial interest in the use of rAAV therapies, and own equity in a company (AGTC) that might, in the future, commercialize some aspects of this work.

ACKNOWLEDGMENTS

This work was supported by an Aniridia Multi Investigator Grant from the Sharon Stewart Aniridia Research Trust (20R64586 to E.M.S.), the Canadian Institutes of Health Research (MOP-119586 to E.M.S., MOP-82875 to W.W.W.), Brain Canada and the Quebec Consortium for Drug Discovery (20R23808 to E.M.S.), the National Science and Engineering Research Council (RGPIN355532-10 to W.W.W.), and the National Institutes of Health (1R01GM084875 to W.W.W., 1R01EY002422 to R.S.M., P30EY021721 to W.W.H.). Additionally, we acknowledge salary support from the University of British Columbia Four Year Doctoral Fellowship and Graduate Student Initiatives (to J.W.H.); the Canadian Institutes of Health Research Canadian Graduate Scholarships (to J.W.H.); the Child and Family Research Institute and BC Children's Hospital Foundation (to A.M.); and The Natural Sciences and Engineering Research Council of Canada (to C.Y.C.). The authors thank: Xin Cynthia Ye and Cindy Zhang for their work entering PAX6 into the PAZAR database (www.pazar.info); the FANTOM consortium for providing prepublication access to CAGE data; Yifeng Li for conducting preliminary analysis of TFBSs in RR4, RR6, RR1, and RR3; Jingsong Wang and the Child and Family Research Institute Imaging Core for microscopy support; Dora Pak for administrative support; Miroslav Hatas for information technology and systems support; and T.C. Lengyell for intravenous rAAV injections.

REFERENCES

1. Bainbridge, JW, Mehat, MS, Sundaram, V, Robbie, SJ, Barker, SE, Ripamonti, C *et al.* (2015). Long-term effect of gene therapy on Leber's congenital amaurosis. *N Engl J Med* **372**: 1887–1897.
2. Jacobson, SG, Cideciyan, AV, Roman, AJ, Sumaroka, A, Schwartz, SB, Heon, E *et al.* (2015). Improvement and decline in vision with gene therapy in childhood blindness. *N Engl J Med* **372**: 1920–1926.
3. Boshart, M, Weber, F, Jahn, G, Dorsch-Häsler, K, Fleckenstein, B and Schaffner, W (1985). A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* **41**: 521–530.
4. Miyazaki, J, Takaki, S, Araki, K, Tashiro, F, Tominaga, A, Takatsu, K *et al.* (1989). Expression vector system based on the chicken beta-actin promoter directs efficient production of interleukin-5. *Gene* **79**: 269–277.
5. Chuah, MK, Petrus, I, De Bleser, P, Le Guiner, C, Gernoux, G, Adjali, O *et al.* (2014). Liver-specific transcriptional modules identified by genome-wide in silico analysis enable efficient gene therapy in mice and non-human primates. *Mol Ther* **22**: 1605–1613.
6. Xiong, J, Sun, WJ, Wang, WF, Liao, ZK, Zhou, FX, Kong, HY *et al.* (2012). Novel, chimeric, cancer-specific, and radiation-inducible gene promoters for suicide gene therapy of cancer. *Cancer* **118**: 536–548.
7. Scalabrino, ML, Boye, SL, Fransen, KM, Noel, JM, Dyka, FM, Min, SH *et al.* (2015). Intravitreal delivery of a novel AAV vector targets ON bipolar cells and restores visual function in a mouse model of complete congenital stationary night blindness. *Hum Mol Genet* **24**: 6229–6239.
8. Portales-Casamar, E, Swanson, DJ, Liu, L, de Leeuw, CN, Banks, KG, Ho Sui, SJ *et al.* (2010). A regulatory toolbox of MiniPromoters to drive selective expression in the brain. *Proc Natl Acad Sci USA* **107**: 16589–16594.
9. de Leeuw, CN, Dyka, FM, Boye, SL, Laprise, S, Zhou, M, Chou, AY *et al.* (2014). Targeted CNS delivery using human minipromoters and demonstrated compatibility with adeno-associated viral vectors. *Mol Ther Methods Clin Dev* **1**: 5.
10. de Leeuw, CN, Korecki, AJ, Berry, GE, Hickmott, JW, Lam, SL, Lengyell, TC *et al.* (2016). rAAV-compatible MiniPromoters for restricted expression in the brain and eye. *Mol Brain* **9**: 52.
11. Kleinjan, DA, Seawright, A, Mella, S, Carr, CB, Tyas, DA, Simpson, TI *et al.* (2006). Long-range downstream enhancers are essential for Pax6 expression. *Dev Biol* **299**: 563–581.
12. Kleinjan, DA, Seawright, A, Schedl, A, Quinlan, RA, Danes, S and van Heyningen, V (2001). Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. *Hum Mol Genet* **10**: 2049–2059.

13. Bhatia, S, Bengani, H, Fish, M, Brown, A, Divizia, MT, de Marco, R *et al.* (2013). Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am J Hum Genet* **93**: 1126–1134.
14. Amano, T, Sagai, T, Tanabe, H, Mizushina, Y, Nakazawa, H and Shiroishi, T (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell* **16**: 47–57.
15. Sanyal, A, Lajoie, BR, Jain, G and Dekker, J (2012). The long-range interaction landscape of gene promoters. *Nature* **489**: 109–113.
16. Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Lassmann, T. *et al.* (2014) *A promoter-level mammalian expression atlas*. *Nature* **507**, 462–470.
17. Hoffman, MM, Ernst, J, Wilder, SP, Kundaje, A, Harris, RS, Libbrecht, M *et al.* (2013). Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res* **41**: 827–841.
18. Dixon, JR, Selvaraj, S, Yue, F, Kim, A, Li, Y, Shen, Y *et al.* (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376–380.
19. Shen, Y, Yue, F, McCleary, DF, Ye, Z, Edsall, L, Kuan, S *et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**: 116–120.
20. Mathelier, A, Zhao, X, Zhang, AW, Parcy, F, Worsley-Hunt, R, Arenillas, DJ *et al.* (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* **42** (Database issue): D142–D147.
21. Kügler, S, Kilic, E and Bähr, M (2003). Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area. *Gene Ther* **10**: 337–347.
22. McLean, JR, Smith, GA, Rocha, EM, Hayes, MA, Beagan, JA, Hallett, PJ *et al.* (2014). Widespread neuron-specific transgene expression in brain and spinal cord following synapsin promoter-driven AAV9 neonatal intracerebroventricular injection. *Neurosci Lett* **576**: 73–78.
23. Young, JE, Vogt, T, Gross, KW and Khani, SC (2003). A short, highly active photoreceptor-specific enhancer/promoter region upstream of the human rhodopsin kinase gene. *Invest Ophthalmol Vis Sci* **44**: 4076–4085.
24. Kay, CN, Ryals, RC, Aslanidi, GV, Min, SH, Ruan, Q, Sun, J *et al.* (2013). Targeting photoreceptors via intravitreal delivery using novel, capsid-mutated AAV vectors. *PLoS One* **8**: e62097.
25. Washbourne, P and McAllister, AK (2002). Techniques for gene transfer into neurons. *Curr Opin Neurobiol* **12**: 566–573.
26. Halder, G, Callaerts, P and Gehring, WJ (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science* **267**: 1788–1792.
27. Hill, RE, Favor, J, Hogan, BL, Ton, CC, Saunders, GF, Hanson, IM *et al.* (1991). Mouse small eye results from mutations in a paired-like homeobox-containing gene. *Nature* **354**: 522–525.
28. Hoge, MA. (1915). Another gene in the fourth chromosome of Drosophila. *The American Naturalist* **49**: 47–49.
29. Roberts, RC. (1967). Small-eyes, a new dominant mutant in the mouse. *Genet Res* 9: 121–122.
30. Glaser, T, Walton, DS and Maas, RL (1992). Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene. *Nat Genet* **2**: 232–239.
31. Hingorani, M, Hanson, I and van Heyningen, V (2012). Aniridia. *Eur J Hum Genet* **20**: 1011–1017.
32. Lee, H, Khan, R and O'Keefe, M (2008). Aniridia: current pathology and management. *Acta Ophthalmol* **86**: 708–715.
33. Gregory-Evans, CY, Wang, X, Wasan, KM, Zhao, J, Metcalfe, AL and Gregory-Evans, K (2014). Postnatal manipulation of Pax6 dosage reverses congenital tissue malformation defects. *J Clin Invest* **124**: 111–116.
34. Chow, RL, Altmann, CR, Lang, RA and Hemmati-Brivanlou, A (1999). Pax6 induces ectopic eyes in a vertebrate. *Development* **126**: 4213–4222.
35. Schedl, A, Ross, A, Lee, M, Engelkamp, D, Rashbass, P, van Heyningen, V *et al.* (1996). Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities. *Cell* **86**: 71–82.
36. Manuel, M, Pratt, T, Liu, M, Jeffery, G and Price, DJ (2008). Overexpression of Pax6 results in microphthalmia, retinal dysplasia and defective retinal ganglion cell axon guidance. *BMC Dev Biol* **8**: 59.
37. Grindley, JC, Davidson, DR and Hill, RE (1995). The role of Pax-6 in eye and nasal development. *Development* **121**: 1433–1442.
38. Hsieh, YW and Yang, XJ (2009). Dynamic Pax6 expression during the neurogenic cell cycle influences proliferation and cell fate choices of retinal progenitors. *Neural Dev* **4**: 32.
39. Oron-Karni, V, Farhy, C, Elgart, M, Marquardt, T, Remizova, L, Yaron, O *et al.* (2008). Dual requirement for Pax6 in retinal progenitor cells. *Development* **135**: 4037–4047.
40. de Melo, J, Qiu, X, Du, G, Cristante, L and Eisenstat, DD (2003). Dlx1, Dlx2, Pax6, Brn3b, and Chx10 homeobox gene expression defines the retinal ganglion and inner nuclear layers of the developing and adult mouse retina. *J Comp Neurol* **461**: 187–204.
41. Roesch, K, Jadhav, AP, Trimarchi, JM, Stadler, MB, Roska, B, Sun, BB *et al.* (2008). The transcriptome of retinal Müller glial cells. *J Comp Neurol* **509**: 225–238.
42. Davis, JA and Reed, RR (1996). Role of Olf-1 and Pax-6 transcription factors in neurodevelopment. *J Neurosci* **16**: 5082–5094.
43. Nishina, S, Kohsaka, S, Yamaguchi, Y, Handa, H, Kawakami, A, Fujisawa, H *et al.* (1999). PAX6 expression in the developing human eye. *Br J Ophthalmol* **83**: 723–727.
44. Aota, S, Nakajima, N, Sakamoto, R, Watanabe, S, Ibaraki, N and Okazaki, K (2003). Pax6 autoregulation mediated by direct interaction of Pax6 protein with the head surface ectoderm-specific enhancer of the mouse Pax6 gene. *Dev Biol* **257**: 1–13.
45. Bhatia, S, Monahan, J, Ravi, V, Gautier, P, Murdoch, E, Brenner, S *et al.* (2014). A survey of ancient conserved non-coding elements in the PAX6 locus reveals a landscape of interdigitated cis-regulatory archipelagos. *Dev Biol* **387**: 214–228.
46. Griffin, C, Kleinjan, DA, Doe, B and van Heyningen, V (2002). New 3′ elements control Pax6 expression in the developing pretectum, neural retina and olfactory region. *Mech Dev* **112**: 89–100.
47. Kammandel, B, Chowdhury, K, Stoykova, A, Aparicio, S, Brenner, S and Gruss, P (1999). Distinct cis-essential modules direct the time-space pattern of the Pax6 gene activity. *Dev Biol* **205**: 79–97.
48. Kleinjan, DA, Seawright, A, Childs, AJ and van Heyningen, V (2004). Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. *Dev Biol* **265**: 462–477.
49. Marsich, E, Vetere, A, Di Piazza, M, Tell, G and Paoletti, S (2003). The PAX6 gene is activated by the basic helix-loop-helix transcription factor NeuroD/BETA2. *Biochem J* **376**: 707–715.
50. McBride, DJ, Buckle, A, van Heyningen, V and Kleinjan, DA (2011). DNaseI hypersensitivity and ultraconservation reveal novel, interdependent long-range enhancers at the complex Pax6 cis-regulatory region. *PLoS One* **6**: e28616.
51. Ravi, V, Bhatia, S, Gautier, P, Loosli, F, Tay, BH, Tay, A *et al.* (2013). Sequencing of Pax6 loci from the elephant shark reveals a family of Pax6 genes in vertebrate genomes, forged by ancient duplications and divergences. *PLoS Genet* **9**: e1003177.
52. Williams, SC, Altmann, CR, Chow, RL, Hemmati-Brivanlou, A and Lang, RA (1998). A highly conserved lens transcriptional control element from the Pax-6 gene. *Mech Dev* **73**: 225–229.
53. Wu, D, Li, T, Lu, Z, Dai, W, Xu, M and Lu, L (2006). Effect of CTCF-binding motif on regulation of PAX6 transcription. *Invest Ophthalmol Vis Sci* **47**: 2422–2429.
54. Xu, ZP and Saunders, GF (1997). Transcriptional regulation of the human PAX6 gene promoter. *J Biol Chem* **272**: 3430–3436.
55. Xu, ZP and Saunders, GF (1998). PAX6 intronic sequence targets expression to the spinal cord. *Dev Genet* **23**: 259–263.
56. Xu, PX, Zhang, X, Heaney, S, Yoon, A, Michelson, AM and Maas, RL (1999). Regulation of Pax6 expression is conserved between mice and flies. *Development* **126**: 383–395.
57. Zhang, X, Rowan, S, Yue, Y, Heaney, S, Pan, Y, Brendolan, A *et al.* (2006). Pax6 is regulated by Meis and Pbx homeoproteins during pancreatic development. *Dev Biol* **300**: 748–757.
58. Zheng, JB, Zhou, YH, Maity, T, Liao, WS and Saunders, GF (2001). Activation of the human PAX6 gene through the exon 1 enhancer by transcription factors SEF and Sp1. *Nucleic Acids Res* **29**: 4070–4078.
59. Dong, JY, Fan, PD and Frizzell, RA (1996). Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. *Hum Gene Ther* **7**: 2101–2112.
60. Dominguez, JM II, Hu, P, Caballero, S, Moldovan, L, Verma, A, Oudit, GY *et al.* (2016). Adeno-associated virus overexpression of angiotensin-converting enzyme-2 reverses diabetic retinopathy in type 1 diabetes in mice. *Am J Pathol* **186**: 1688–1700.
61. Martin, KR, Quigley, HA, Zack, DJ, Levkovitch-Verbin, H, Kielczewski, J, Valenta, D *et al.* (2003). Gene therapy with brain-derived neurotrophic factor as a protection: retinal ganglion cells in a rat glaucoma model. *Invest Ophthalmol Vis Sci* **44**: 4357–4365.
62. Aleman, TS, Soumittra, N, Cideciyan, AV, Sumaroka, AM, Ramprasad, VL, Herrera, W *et al.* (2009). CERKL mutations cause an autosomal recessive cone-rod dystrophy with inner retinopathy. *Invest Ophthalmol Vis Sci* **50**: 5944–5954.
63. Kaspi, H, Chapnik, E, Levy, M, Beck, G, Hornstein, E and Soen, Y (2013). Brief report: miR-290-295 regulate embryonic stem cell differentiation propensities by repressing Pax6. *Stem Cells* **31**: 2266–2272.
64. Bäumer, N, Marquardt, T, Stoykova, A, Ashery-Padan, R, Chowdhury, K and Gruss, P (2002). Pax6 is required for establishing naso-temporal and dorsal characteristics of the optic vesicle. *Development* **129**: 4535–4545.
65. Zanta-Boussif, MA, Charrier, S, Brice-Ouzet, A, Martin, S, Opolon, P, Thrasher, AJ *et al.* (2009). Validation of a mutated PRE sequence allowing high and sustained transgene expression while abrogating WHV-X protein synthesis: application to the gene therapy of WAS. *Gene Ther* **16**: 605–619.
66. Teerawanichpan, P, Hoffman, T, Ashe, P, Datla, R and Selvaraj, G (2007). Investigations of combinations of mutations in the jellyfish green fluorescent protein (GFP) that afford brighter fluorescence, and use of a version (VisGreen) in plant, bacterial, and animal cells. *Biochim Biophys Acta* **1770**: 1360–1368.
67. Byrne, LC, Lin, YJ, Lee, T, Schaffer, DV and Flannery, JG (2015). The expression pattern of systemically injected AAV9 in the developing mouse retina is determined by age. *Mol Ther* **23**: 290–296.
68. Andersson, R, Gebhard, C, Miguel-Escalada, I, Hoof, I, Bornholdt, J, Boyd, M *et al.*; FANTOM Consortium. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* **507**: 455–461.

69. Servant, N, Lajoie, BR, Nora, EP, Giorgetti, L, Chen, CJ, Heard, E *et al.* (2012). HiTC: exploration of high-throughput 'C' experiments. *Bioinformatics* **28**: 2843–2844.
70. Pollard, KS, Hubisz, MJ, Rosenbloom, KR and Siepel, A (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**: 110–121.
71. Mathelier, A, Lefebvre, C, Zhang, AW, Arenillas, DJ, Ding, J, Wasserman, WW *et al.* (2015). Cis-regulatory somatic mutations and gene-expression alteration in B-cell lymphomas. *Genome Biol* **16**: 84.
72. ENCODE Project Consortium. (2004) The ENCODE (ENCyclopedia Of DNA Elements) project. *Science* **306**: 636–640.
73. Portales-Casamar, E, Kirov, S, Lim, J, Lithwick, S, Swanson, MI, Ticoll, A *et al.* (2007). PAZAR: a framework for collection and dissemination of cis-regulatory sequence annotation. *Genome Biol* **8**: R207.
74. Portales-Casamar, E, Thongjuea, S, Kwon, AT, Arenillas, D, Zhao, X, Valen, E *et al.* (2010). JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res* **38**(Database issue): D105–D110.
75. Portales-Casamar, E, Arenillas, D, Lim, J, Swanson, MI, Jiang, S, McCallum, A *et al.* (2009). The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. *Nucleic Acids Res* **37**(Database issue): D54–D60.
76. Worsley Hunt, R and Wasserman, WW (2014). Non-targeted transcription factors motifs are a systemic component of ChIP-seq datasets. *Genome Biol* **15**: 412.
77. Zolotukhin, S, Potter, M, Zolotukhin, I, Sakai, Y, Loiler, S, Fraites, TJ Jr *et al.* (2002). Production and purification of serotype 1, 2, and 5 recombinant adeno-associated viral vectors. *Methods* **28**: 158–167.
78. Jacobson, SG, Acland, GM, Aguirre, GD, Aleman, TS, Schwartz, SB, Cideciyan, AV *et al.* (2006). Safety of recombinant adeno-associated virus type 2-RPE65 vector delivered by ocular subretinal injection. *Mol Ther* **13**: 1074–1084.
79. Burgess, A, Vigneron, S, Brioudes, E, Labbé, JC, Lorca, T and Castro, A (2010). Loss of human Greatwall results in G2 arrest and multiple mitotic defects due to deregulation of the cyclin B-Cdc2/PP2A balance. *Proc Natl Acad Sci USA* **107**: 12564–12569.
80. McCloy, RA, Rogers, S, Caldon, CE, Lorca, T, Castro, A and Burgess, A (2014). Partial inhibition of Cdk1 in G 2 phase overrides the SAC and decouples mitotic events. *Cell Cycle* **13**: 1400–1412.

Supplementary Information accompanies this paper on the *Molecular Therapy—Methods & Clinical Development* website (http://www.nature.com/mtm)