



This is a repository copy of *Design of synthetic promoters for controlled expression of therapeutic genes in retinal pigment epithelial cells*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/171769/>

Version: Published Version

Article:

Johari, Y.B., Mercer, A.C., Liu, Y. et al. (2 more authors) (2021) Design of synthetic promoters for controlled expression of therapeutic genes in retinal pigment epithelial cells. *Biotechnology and Bioengineering*, 118 (5). pp. 2001-2015. ISSN 0006-3592

<https://doi.org/10.1002/bit.27713>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

ARTICLE

Design of synthetic promoters for controlled expression of therapeutic genes in retinal pigment epithelial cells

Yusuf B. Johari¹  | Andrew C. Mercer² | Ye Liu² | Adam J. Brown¹  | David C. James¹ 

¹Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, UK

²Research and Early Development, REGENXBIO Inc., Rockville, Maryland, USA

Correspondence

David C. James, Department of Chemical and Biological Engineering, University of Sheffield, Mappin St., Sheffield S1 3JD, UK.
 Email: d.c.james@sheffield.ac.uk

Funding information

REGENXBIO

Abstract

Age-related macular degeneration (AMD) associated with dysfunction of retinal pigment epithelial (RPE) cells is the most common cause of untreatable blindness. To advance gene therapy as a viable treatment for AMD there is a need for technologies that enable controlled, RPE-specific expression of therapeutic genes. Here we describe design, construction and testing of compact synthetic promoters with a pre-defined transcriptional activity and RPE cell specificity. Initial comparative informatic analyses of RPE and photoreceptor (PR) cell transcriptomic data identified conserved and overrepresented transcription factor regulatory elements (TFREs, 8–19 bp) specifically associated with transcriptionally active RPE genes. Both RPE-specific TFREs and those derived from the generically active cytomegalovirus-immediate early (CMV-IE) promoter were then screened in vitro to identify sequence elements able to control recombinant gene transcription in model induced pluripotent stem (iPS)-derived and primary human RPE cells. Two libraries of heterotypic synthetic promoters varying in predicted RPE specificity and transcriptional activity were designed de novo using combinations of up to 20 discrete TFREs in series (323–602 bp) and their transcriptional activity in model RPE cells was compared to that of the endogenous BEST1 promoter (661 bp, plus an engineered derivative) and the highly active generic CMV-IE promoter (650 bp). Synthetic promoters with a highpredicted specificity, comprised predominantly of endogenous TFREs exhibited a range of activities up to 8-fold that of the RPE-specific BEST1 gene promoter. Moreover, albeit at a lower predicted specificity, synthetic promoter transcriptional activity in model RPE cells was enhanced beyond that of the CMV-IE promoter when viral elements were utilized in combination with endogenous RPE-specific TFREs, with a reduction in promoter size of 15%. Taken together, while our data reveal an inverse relationship between synthetic promoter activity and cell-type specificity, cell context-specific control of recombinant gene transcriptional activity may be achievable.

KEYWORDS

cell-specific expression, gene therapy, RPE cells, synthetic biology, synthetic promoter, transcriptomics

Abbreviations: AAV, adeno-associated virus; AMD, age-related macular degeneration; BEST1, bestrophin-1; CMV, cytomegalovirus; GFP, green fluorescent protein; iPS, induced pluripotent stem; ORF, open reading frame; PR, photoreceptor; RPE, retinal pigment epithelium; TF, transcription factor; TFRE, transcription factor regulatory element; TSS, transcriptional start site.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Biotechnology and Bioengineering* Published by Wiley Periodicals LLC

1 | INTRODUCTION

The retinal pigment epithelium (RPE) is a multifunctional monolayer of neuroepithelium-derived cells, flanked by photoreceptor (PR) cells and the choroid complex. The significance of the RPE in the ocular system is exemplified by the major association of these pigmented cells in genetically determined retinal diseases such as age-related macular degeneration and retinitis pigmentosa. Accordingly, various *in vitro* human RPE models have been established as a convenient platform to study RPE functions, where the two most commonly used models are primary human fetal RPE cells and immortalized cell lines such as ARPE-19 and hRPE7. Nonetheless, studies of their morphologic and functional characteristics have produced contradictory results (Abлонczy et al., 2011) demonstrating the limitation of immortalized cells in studying native human RPE function. The more recent discovery of induced pluripotent stem (iPS) cells has also yielded iPS-derived RPE cells that closely mimic the gene expression, polarity, and physiology of native human RPE cells (Kokkinaki et al., 2011; Liao et al., 2010). However, despite a significant amount of research on global gene expression profiling of stem-cell-derived/primary RPE cells and native tissue, there are no methods to systematically link transcriptomic data sets with genomic data to identify cis-regulatory elements that would provide inherently RPE-specific expression of recombinant genes.

The recent approval of voretigene neparvovec (Luxturna®) to treat retinal degeneration highlights that gene therapies to disease-causing genetic mutations are possible. This achievement is partly made possible by the use of adeno-associated viral (AAV) vectors that can transduce and maintain therapeutic gene expression in non-dividing cells (including the retina) with minimal immune responses (Naso et al., 2017). Ubiquitous promoters such as that derived from human cytomegalovirus (CMV) are often used to drive high transgene expression despite potentially undesirable attributes such as promoter silencing and lack of cell-type specificity. Endogenous promoters, on the other hand, often have lower activity compared to viral-derived promoters and are a relatively large size, thus limiting their use in viral vectors. For example, the ~1.6 kb-long RPE65 promoter displayed only 10% of CMV activity when used to induce targeted expression of the RPE65 gene in RPE65-deficient canines, and was inactive in older animals (Le Meur et al., 2007). The latter illustrates a further possible constraint in the use of endogenous cell-specific promoters in which their expression may be downregulated for target tissues that are already in a state of disease. Similarly, Komáromy et al. (2010) also reported (endogenous) promoter length and age dependency, where the long version of the red cone opsin promoter (~2.1 kb) in younger animals led to a more stable therapeutic effect for achromatopsia. In this context, bespoke synthetic promoters are an attractive alternative as they can be custom-designed to control recombinant gene expression predictably in a specific cellular context.

A number of studies report engineering of natural promoters for improved activity in a therapeutic context. Examples include the creation of hybrid promoters to selectively kill cancerous tissues via

incorporation of the prostate-specific probasin promoter into the retroviral LTR to target prostate cancer cells (Logg et al., 2002) or by coupling the endothelin enhancer element with the Cdc6 promoter to target dividing tumor endothelial cells (Szymanski et al., 2006). More advanced attempts to create promoters with increased tissue-specificity are exemplified via *de novo* design of synthetic promoters that could specifically mediate gene expression in muscle cells (Li et al., 1999), colorectal cancer cells (Roberts et al., 2017) or liver cells with responsiveness to glucose (Han et al., 2011). However, these studies involved screening hundreds to thousands of synthetic promoters, which is unfeasible for primary and iPS-derived cells, such as RPE, with a limited capacity for expansion and which exhibit a particular differentiated morphological state. Further, there remains no information on how it is possible to utilize RPE model cells *in vitro* to characterize the function of individual genetic components to eliminate uncontrollable and functionally ill-defined parts of endogenous promoter assemblies.

In this study, we test the hypothesis that RPE genomic information can be mined to identify active transcription factor regulatory elements (TFREs) that could be utilized to design compact, space-efficient synthetic promoter assemblies that exhibit both a high degree of cell type specificity and transcriptional activity. Through systematic bioinformatic analysis of 'omic data streams coupled with *in vitro* screening we identified endogenous human TFRE sequences that are potentially active in the different eye components (i.e., RPE vs. PR) as well as those that function as transcriptional repressors. We further identified highly active TFREs present in the human CMV promoter that enable active and space-efficient synthetic promoter constructs that recruit RPE cell intrinsic transcriptional capacity. Based on these data, we designed RPE-active promoter/TFRE assemblies *de novo* with different objective functions, either high RPE specificity (no, or low transcriptional activity in PR cells) and/or high RPE transcriptional activity. We also compared the *de novo* (bottom-up) synthetic promoter design strategy to (top-down) targeted re-engineering of the RPE-specific bestrophin-1 (BEST1) promoter for improved activity in RPE cells. While this study demonstrates effective construction of promoters for RPE cells, similar approaches could be used to design promoters for applications requiring specific and/or high expression of recombinant genes in other cell types.

2 | MATERIALS AND METHODS

2.1 | iPS-derived and primary human RPE cell cultures

Cryopreserved iPS-derived human RPE cells (iCell® RPE) were obtained from FujiFilm Cellular Dynamics and cultured according to the manufacturer's instructions using MEM α (Thermo Fisher Scientific) supplemented with KnockOut SR (Thermo Fisher Scientific), N-2 supplement (Thermo Fisher Scientific), hydrocortisone (Sigma), taurine (Sigma), triiodo-L-thyronine (Sigma) and

gentamicin (Thermo Fisher Scientific). Cryopreserved human fetal RPE cells (Clonetics® RPE) were obtained from Lonza and cultured according to the manufacturer's instructions using RtEGM RPE Cell Growth Medium BulletKit (Lonza) and ReagentPack sub-culture reagents (Lonza). iPS-derived cells were maintained in vitronectin-coated vessels and primary cells were maintained in tissue culture-treated vessels at 37°C under 5% CO₂ in humidified incubator. Cell concentration and viability were measured using a Vi-CELL XR (Beckman Coulter).

2.2 | In silico analysis of TFREs

RPE and PR microarray data were obtained from the literature (Booij et al., 2009, 2010; Table S1). For each gene, 2000 bp upstream of the start codon was extracted from human genome database GRCh38/hg38. Transcriptional start sites (TSSs) were determined based on the literature, Eukaryotic Promoter Database (EPD; epd.epfl.ch) or annotated 5'-UTRs. Genomatix Gene Regulation software (MatInspector Release 8.2 and MatBase Version 9.4; Genomatix) was used to analyze the region -1000 to +200 relative to the TSS (or up to the start codon) to find putative TFREs. Overrepresented TFREs were identified by analyzing the promoters against Genomatix-defined human promoter background followed by selection of the TFREs with Z-score > 2.5, whereas common TFREs were identified using core similarity of 1.0 and optimized matrix similarity of +0.01 followed by selection of the TFREs with $p > .2$ against Genomatix-defined randomly drawn promoter samples. Identification of TFREs in hCMV-IE promoter using MatInspector was performed using core similarity of 0.75 and optimized matrix similarity.

2.3 | TFRE-reporter vector construction

pmaxGFP vector (Lonza) was utilized as a backbone. The CMV promoter and chimeric intron of pmaxGFP were deleted by digestion with BsrGI and KpnI, and replaced with a short DNA fragment containing a HindIII site. A minimal CMV core promoter from the human CMV was synthesized (Eurofins Genomics), PCR amplified (Q5 high-fidelity 2× master mix; NEB), and purified (QIAquick PCR Purification Kit; Qiagen). The PCR products were then digested with HindIII and KpnI enzymes (NEB), gel extracted (QIAquick Gel Extraction Kit; Qiagen) and inserted directly upstream of the green fluorescent protein (GFP) open reading frame (ORF) of the promoterless pmaxGFP vector. The CMV core promoter sequence used was as follows: 5'-AGGCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAGATCGCCTGGAGACGCATCCACGCTGTTTTGACCTCCATAGAAGAC-3'. To create TFRE reporter plasmids, synthetic oligonucleotides containing 7× repeat copies of the TFRE sequences in Tables 1 and 2 were synthesized, PCR amplified, and inserted into BsrGI and HindIII sites upstream of the CMV core promoter. Clonally derived plasmids were purified using a QIAGEN Plasmid Plus kit (Qiagen). The sequence of all plasmid constructs was confirmed by DNA sequencing.

2.4 | Synthetic promoter vector construction

To create synthetic promoters, synthetic genes containing combinations of specific TFREs were designed in silico (Tables 3 and 4). The positions of the TFRE blocks within the promoters were randomly arranged using R software in forward orientation of 5' DNA strand. Synthetic genes were synthesized (Eurofins Genomics) and inserted into BsrGI and HindIII sites upstream of the CMV core promoter. A full length CMV promoter containing the same CMV core (-600 to +50 relative to the TSS) and endogenous BEST1 promoter (-585 to +76 relative to the TSS) were also synthesized and inserted separately into BsrGI and KpnI sites upstream of the GFP ORF. The sequence of all plasmid constructs was confirmed by DNA sequencing. The estimated RPE/PR specificity ratio of a promoter was calculated as follows:

$$\text{Estimated RPE/PR specificity ratio} = \frac{\text{RPE activity}}{\text{PR activity}} = \frac{\sum T_{a_i} \times N_i \times \text{mRNA}_{\text{RPE/PR}_i}}{\sum T_{a_i} \times N_i \times 1}, \quad (1)$$

where T_a is the transcriptional activity of a specific TFRE, i is a specific TFRE type, N is the copy number of a specific TFRE in the promoter, and $\text{mRNA}_{\text{RPE/PR}}$ is the cognate TF mRNA expression fold-change in RPE over PR (derived from Booij et al., 2010; Table S2).

2.5 | Transient transfection of iPS-derived and primary RPE cells

iPS-derived RPE cells were transfected using Lipofectamine Stem (Thermo Fisher Scientific) at Day 3 post-seeding according to the manufacturer's instructions. For each transfection in a 96-well, 0.4 μg plasmid was diluted in 5 μl Opti-MEM I reduced serum medium (Thermo Fisher Scientific), and then combined with 0.9 μl Lipofectamine pre-diluted in 5 μL Opti-MEM medium. The Lipofectamine/DNA mixture was allowed to stand at room temperature for 10 min before being added to the culture well. Cells were maintained at 37°C under 5% CO₂ with transfection complexes removed 24 h post-transfection.

Primary RPE cells were transfected using a P3 Primary Cell 96-well Nucleofector system (Lonza) at ~90% culture confluency on the second passage according to the manufacturer's instructions. Briefly, 3×10^5 cells were electroporated with 0.72 μg DNA using program EA-104 and immediately diluted with 80 μl serum-supplemented culture media. 25 μl of the diluted reaction was transferred to a 96-well plate containing 175 μl pre-warmed serum-supplemented culture media. Transfected cells were cultured at 37°C under 5% CO₂ for 24 h followed by maintenance in serum-free media.

2.6 | Measurement of recombinant GFP expression in vitro

GFP expression in differentiated iPS-derived RPE cells (Day 28) and primary RPE cells (Day 14) was quantified using a SpectraMax iD5

TFRE	In BEST1	Endogenous sequence	Consensus sequence
Overrepresented and common in at least 25%			
AP-2 ϵ	Y	ACCCCTGAGGCCT	<u>TTGCCTGAGGCGA</u>
ZNF300	N	CGCCCCAG	(endogenous)
MafA	N	GGCGGGGACAGCA	<u>GGCGGAGTCAGCA</u>
ZNF35	N	GCCGGGAAGACC	(endogenous)
ChREBP:Mix	Y ^a	CACGTGGTCCCCAGGTG	<u>CACGTGGCAAGCACGTG</u>
Overrepresented			
SPZ1	Y	TGGAGGGTGTT	(endogenous)
EKLF	Y	CAGGTGGGGTT	<u>CCGGTAGGGTG</u>
ZBTB7	Y	CAGCCCCCTAACC	<u>AGCCCCCAAAA</u>
NF- κ B	Y	TGGGAATTTTCAT	<u>CGGGACTTTCCA</u>
Zic2	Y	CTCAGCATGTG	<u>CACAGCAGGAG</u>
HRE	Y ^a	GGACGTGCC	(endogenous)
c-Myb	Y	CAACAGTCCT	<u>CAACCGCCAT</u>
MYRF	Y	CTGTGCCAGGAA	(endogenous)
SALL2	N	TCGGGTGGGTT	(endogenous)
HELT	Y	GCCCACGTGAGT	<u>GGGCACGTGACC</u>
Common in at least 25%			
AML3	Y	AGCAGTGGTTCTTG	<u>AGCTGTGGTTTGTG</u>
DICE	Y	TGCTCTTTCATG	<u>TGTCTCTCCACAG</u>
AREB6	Y	AGGTTTCAG	(endogenous)
IR1 nGRE	Y	CCTTCCTGGAGAGT	<u>GCCTCCTGGAGAGG</u>
MafF	Y	AGTGCTGAGCCCGTC	<u>ACTGCTGAGTCAGCAAT</u>
MGA	Y	CGGGGTCACCACACACA	<u>AGGTGTGACTTCACACC</u>
NF- κ B (p50)	Y	AGGGAGTCCC	<u>GGGGATTCCC</u>
C/EBP ϵ :ATF4	Y	TGAAGCAA	(endogenous)
AP-4	Y	CACCAGCTGCC	(endogenous)
Lf	Y	GGCACTGGC	<u>GGCACTTGC</u>
Literature			
Otx2	Y	TCCCTAAGCCAGGA	<u>CAATTAATCCCTAC</u>
SIX3	Y	TCTTGTAATCTGCTCAGAA	<u>ATGTGTAATGACTTCACTC</u>
PAX6	Y	TAAATTCAGCCCTG	<u>TTAGTTCAGGTCAG</u>
MITF	Y ^a	(HELT)	<u>GTACAGTGAC</u>

Abbreviations: AML3, runt-related transcription factor 2/CBFA1; AP-2 ϵ , activator protein 2 epsilon; AP-4, activating enhancer binding protein 4; AREB6, Atp1a1 regulatory element binding factor 6; C/EBP ϵ :ATF4, heterodimer of C/EBP epsilon and ATF4; ChREBP:Mix, heterodimer of ChREBP and Mix; c-Myb, cellular myoblastosis virus oncogene v-Myb; DICE, downstream immunoglobulin control element; EKLF, erythroid krueppel like factor; HELT, hey-like bHLH-transcription factor; HRE, hypoxia-response element; IR1 nGRE, binding site for glucocorticoid receptor; Lf, lactotransferrin and delta-lactoferrin; MafA, lens-specific Maf/MafA-sites; MafF, transcription factor MafF; MGA, MAX gene associated; MITF, microphthalmia transcription factor; MYRF, myelin regulatory factor; NF- κ B (p50), nuclear factor- κ B p50 subunit; NF- κ B, nuclear factor κ B; Otx2, orthodenticle homeobox 2; PAX6, PAX6 paired domain and homeodomain; SALL2, zinc finger protein Spalt-2, sal-like 2; SIX3, SIX3/SIX domain and homeodomain; SPZ1, spermatogenic Zip 1; ZBTB7, zinc finger and BTB domain containing 7; Zic2, Zic family member 2; ZNF300, KRAB-containing zinc finger protein 300; ZNF35, zinc finger protein ZNF35.

^aOverlap with HELT in BEST1 promoter (endogenous sequence from another promoter was used if available).

TABLE 1 DNA sequences of RPE-specific transcription factor regulatory elements (TFREs) identified by bioinformatic survey of retinal endogenous promoters. Differing nucleotides between endogenous and consensus sequence are underlined. Measurement of the TFRE relative ability to activate transcription of recombinant GFP genes in RPE cells is shown in Figure 4. A BEST1 promoter sequence map displaying the TFREs is shown in Figure S1A

microplate reader (Molecular Devices). Cells were rinsed and culture media was replaced with Dulbecco's phosphate-buffered saline (DPBS; Sigma) before fluorescence read (excitation: 485 nm, emission: 535 nm). GFP was visualized by fluorescence microscopy using

Olympus IX73 microscope (Olympus). Culture media was replaced with DPBS before fluorescence imaging. To measure transfection efficiency, cells (scaled-up to a 48-well plate) were trypsinized 48 h posttransfection, rinsed with DPBS and analyzed using Attune

TABLE 2 Transcription factor regulatory elements (TFREs) identified by bioinformatic survey of cytomegalovirus (CMV) promoter and their overrepresentation in the retinal pigment epithelium (RPE)-specific, photoreceptor (PR)-specific, and nonspecific groups. Known transcriptional repressors (YY1 and Gfi1) are excluded from analysis. Measurement of the TFRE relative ability to activate transcription of recombinant photoreceptor (GFP) genes in RPE cells is shown in Figure 4b. A CMV promoter sequence map displaying the TFREs is shown in Figure S1B

TFRE	No. of copies	Viral sequence	In BEST1	In RPE-specific	In PR-specific	In nonspecific
Sp1	6	TGGGGCGGAGT	N	Y	Y	Y
CREB	6	ATTGACGTCAATG	Y	N	Y	Y
NF- κ B	4	(consensus)	Y	Y	N	N
NF1	4	TTGGCAGTACATCAA	Y	N	N	Y
TLX1	3	CGGTAATGG	N	N	N	Y
RAR	3	TGCCAGTACATGACCT	N	N	N	N
C/EBP ϵ	2	TGTCGTAAC (C/EBP ϵ :ATF4)	Y	Y	N	N
AhR/ARNT	2	TGGGCGTGGATA	N	Y	N	Y
AP-1	2	CGTGAGTCAAA	N	N	N	N
SRF	1	CCATATATGGA	Y	N	N	Y

Abbreviations: AhR/ARNT, aryl hydrocarbon receptor and nuclear translocator heterodimer; AP-1, activator protein 1; C/EBP ϵ , CCAAT/enhancer binding protein epsilon; CREB, cAMP-responsive element binding protein; NF1, nuclear factor 1; NF- κ B, nuclear factor- κ B; RAR, retinoic acid receptor; Sp1, stimulating protein 1; SRF, serum response factor; TLX1, T-cell leukemia homeobox 1.

Acoustic Focusing Cytometer (Thermo Fisher Scientific). Background fluorescence/absorbance was determined in cells transfected with a promoterless vector.

3 | RESULTS

3.1 | Bioinformatic identification of TFREs in endogenous RPE gene promoters

We previously devised a methodical, comprehensive approach for bioprocess-directed design of synthetic promoters based on genomic sequence information (Johari et al., 2019). The work flow enables identification of TFREs associated with endogenous gene promoters with specific characteristics, thus allowing de novo design of synthetic promoters with relevant functional features. In this study, we hypothesized that it was feasible to construct RPE-specific promoters using assemblies of the corresponding RPE-specific TFREs (summarized in Figure 1). To profile endogenous gene expression in RPE and PR cells, we utilized transcriptomic datasets from Booi et al. (2009, 2010). A subgroup of 35 highly expressed RPE-specific genes was created by selecting genes with a microarray expression level > 6,000 units in RPE and at least four-fold higher expression levels in RPE than in PR cells (as well as choroids). As expected, the BEST1 (VMD2) gene was highly and preferentially expressed in RPE cells, confirming the potential use of its promoter to drive RPE-specific expression in vitro (Esumi et al., 2004) and in vivo (Guziewicz et al., 2013; Kachi et al., 2006). A corresponding PR-specific group was created by selecting 35 genes with expression level < 6,000 units in RPE and RPE/PR expression fold change of < 0.5, whereas a non-specific group was created by selecting 35 genes with expression level > 6,000 units and RPE/PR expression fold change of between 0.97 and 1.03. Table S1 lists the selected genes in each group.

For each selected gene, a transcriptional start site (TSS) was obtained from the literature or the eukaryotic promoter database (EPD; epd.epfl.ch), or if not available, was estimated based on annotated 5'UTR sequences. Relative to the TSS a -1000 to +200 bp (or the start codon) segment was extracted from the human genomic sequences for putative TFRE analysis. TFRE identification was performed using Genomatix Gene Regulation software using two different analyses, (i) *overrepresented* TFREs in each group with promoter Z-score > 2.5 (against a human promoter background pre-defined by Genomatix), and (ii) *common* TFREs, that is, frequently occurring elements in each group, set at $\geq 25\%$ (9/35) of the genes. The results revealed 132 and 383 discrete TFREs in RPE-specific groups using the overrepresented and common methods respectively. To minimize false positives in the RPE-specific group, we filtered out TFREs that also occurred in the PR/nonspecific group (Figure 2) as well as common TFREs that did not occur in BEST1 promoter. To increase the TFRE pool complexity, we selected only up to 2 TFREs with the highest Z-scores from each TF family. 4 TFREs namely Otx2, SIX3, PAX6, and MITF that are required to induce iPSCs to RPE (D'Alessio et al., 2015; Kokkinaki et al., 2011; not selected by the bioinformatic analysis due to the high selection stringency) and present in BEST1 promoter (see Figure S1A) were also included. We note that in our analysis MITF was identified only in BEST1 promoter and the primary site overlapped HELT (Esumi et al., 2004, 2007). Table 1 lists the final set of 29 TFREs incorporated into the functional screen and their endogenous and consensus sequences (derived from the endogenous promoters and Genomatix software, respectively). The TF matrix/family, frequency and Z-score/p-value of the selected TFREs are detailed in Table S3.

To further analyze the "transcriptional landscape" of RPE cells and identify active TFs contained within, we surveyed putative TFREs in hCMV-IE1. Using Genomatix MatInspector tool, 70 discrete TF families were identified in the CMV promoter. A subset of 10 TFs

TFRE	1/01	1/02	1/03	1/04	1/05	1/06	1/07	1/08	1/09	1/10	1/11
Endogenous											
AP-2 ϵ					4					4	2
ZNF300							4			4	
MafA								4	4		
ZNF35			4		4						
ChREBP:Mix	7	4			4						2
SPZ1		4							4		2
EKLF				4		4					
ZBTB7				4						4	2
NF- κ B			4	4							2
Zic2					4			4			2
HRE	7	4		4							2
c-Myb						4	4				2
MYRF		4								4	2
SALL2			4					4			
HELT/MITF	7	6	6	6	6	6	6	6	6	6	2
AML3			4				4				
DICE				4	4						
AREB6		4	4								2
IR1 nGRE						4		4			2
MafF						4			4		2
MGA							4	4			2
NF- κ B (p50)									4	4	
C/EBP ϵ :ATF4			4			4					2
AP-4								4		4	2
Lf	7	4				4					2
Otx2					4		4				2
SIX3							4		4		2
PAX6				4					4		2
Total copies	28	30	30	30	30	30	30	30	30	30	40

TABLE 3 Composition of specific transcription factor regulatory element (TFRE) copies in first-generation synthetic promoters

Abbreviations: AML3, runt-related transcription factor 2/CBFA1; AP-2 ϵ , activator protein 2 epsilon; AP-4, activating enhancer binding protein 4; AREB6, Atp1a1 regulatory element binding factor 6; C/EBP ϵ :ATF4, heterodimer of C/EBP epsilon and ATF4; ChREBP:Mix, heterodimer of ChREBP and Mix; c-Myb, cellular myoblastosis virus oncogene v-Myb; DICE, downstream immunoglobulin control element; EKLF, erythroid krueppel like factor; HELT, hey-like bHLH-transcription factor; HRE, hypoxia-response element; IR1 nGRE, binding site for glucocorticoid receptor; Lf, lactotransferrin and delta-lactoferrin; MafA, lens-specific Maf/MafA-sites; MafF, transcription factor MafF; MGA, MAX gene associated; MITF, microphthalmia transcription factor; MYRF, myelin regulatory factor; NF- κ B, nuclear factor κ B; NF- κ B (p50), nuclear factor- κ B p50 subunit; Otx2, orthodenticle homeobox 2; PAX6, PAX6 paired domain and homeodomain; SALL2, zinc finger protein Spalt-2, sal-like 2; SIX3, SIX3/SIX domain and homeodomain; SPZ1, spermatogenic Zip 1; ZBTB7, zinc finger and BTB domain containing 7; Zic2, Zic family member 2; ZNF300, KRAB-containing zinc finger protein 300; ZNF35, zinc finger protein ZNF35.

that are known to be positive regulators of CMV activity in different cell types (e.g., Brown et al., 2015; Ghazal et al., 1992; Lashmit et al., 2009) were selected for screening (Figure S1B). To further minimize this pool (design space) as well as false positives, we selected TFRE sequences with the highest Genomatix matrix similarity from each TF family as summarized in Table 2. The TF matrix/family, frequency and matrix similarity of the selected TFREs are detailed in Table S4. We note that viral-derived NF- κ B sequence is identical to the consensus sequence (Table 1).

3.2 | Determination of TFRE activity in RPE cells

Previous studies showed that iPS-derived RPE cells exhibit membrane potential, ion transport, polarized vascular endothelial growth factor secretion, and gene expression profile that closely resemble to those of native human RPE (Kokkinaki et al., 2011; Liao et al., 2010). To compare the utility of different RPE models in constructing and evaluating RPE promoters, we utilized iPS-derived human RPE cells (iCell® RPE, FujiFilm Cellular Dynamics) and primary human fetal

TABLE 4 Composition of specific TFRE copies in second-generation synthetic promoters

TFRE	2/01	2/02	2/03	2/04	2/05	2/06	2/07	2/08	2/09	2/10	2/11	2/12	2/13
Endogenous													
AP-2ε		1	2	2	2			1	2	1			
ChREBP:Mix		1	2	4	2		2	1	2	2	1	2	1
SPZ1		1	2	2	2			1	2	1			
Zic2		1	2	2	2			1	2	1			
HRE	7	7	2	4	2	7	6	6	2	4	6	6	6
c-Myb		1	2	2	2			1	2	1			
MYRF		1	2	2	2			1	2	1			
HELT/MITF	7	7	2	4	2	7	6	6	2	4	6	6	6
IR1 nGRE		1	2	2	2			1	2	1			
MGA		1	2	2	2			1	2	1			
C/EBPε:ATF4		1	2	2									
Lf	7	7	2	4	2	7	4	6	2	4	5	5	5
Otx2		1	2	2	2			1	2	1			
Consensus													
MafF		1	2	2	2		3	1	2	1			
NF-κB					2	7	7	6	2	4	7	7	7
PAX6													2
Viral													
Sp1									2	2	2	3	3
CREB									2	4	7	7	7
C/EBPε:ATF4					2			1	2	2	2	2	2
AP-1												3	2
Total copies	21	32	28	36	30	28	28	35	34	35	36	41	41
Size (bp)	323	478	468	570	498	414	430	525	548	538	526	602	552

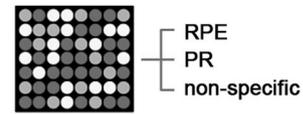
Abbreviations: AP-1, activator protein 1; AP-2ε, activator protein 2 epsilon; C/EBPε:ATF4, heterodimer of C/EBP epsilon and ATF4; ChREBP:Mix, heterodimer of ChREBP and Mix; c-Myb, cellular myoblastosis virus oncogene v-Myb; CREB, cAMP-responsive element binding protein; HELT, hey-like bHLH-transcription factor; HRE, hypoxia-response element; IR1 nGRE, binding site for glucocorticoid receptor; Lf, lactotransferrin and delta-lactoferrin; MafA, lens-specific Maf/MafA-sites; MafF, transcription factor MafF; MGA, MAX gene associated; MITF, microphthalmia transcription factor; MYRF, myelin regulatory factor; Otx2, orthodenticle homeobox 2; PAX6, PAX6 paired domain and homeodomain; Sp1, stimulating protein 1; SPZ1, spermatogenic Zip 1; Zic2, Zic family member 2.

RPE cells (Clonetics® RPE, Lonza). Both culture models developed into monolayers with tight junctions with iPS-derived RPE cells exhibited obvious pigmentation as well as more uniform in size and shape (Figure 3). We note that there were no established cells representative of PRs that could be utilized as control cells (McDougald et al., 2019). The relative transcriptional activity of each TFRE (Tables 1 and 2) in RPE cells was determined as previously described (Johari et al., 2019) using a GFP reporter construct that contained seven repeat copies of a specific TFRE in series, upstream of a minimal mammalian core promoter (hCMV-EI core containing a TATA box and an Inr motif, -34 to +50 relative to the TSS). A control CMV promoter-reporter plasmid was constructed using the hCMV-IE1 promoter (-600 to +50 relative to the TSS, i.e., the complete hCMV-IE1 enhancer containing the distal, proximal and core promoter regions, henceforth referred to as CMV) upstream of the GFP ORF.

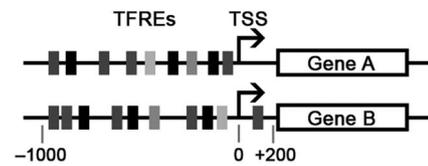
Optimized transient transfection of plasmid DNA into iPS-derived RPE by lipofection and primary RPE by nucleofection yielded

transfection efficiencies of ~71% and ~85%, respectively at 48 h posttransfection (measured using a vector harboring a full-length CMV promoter, flow cytometry data not shown). Measurement of GFP expression deriving from transfection of each endogenous sequence TFRE-reporter plasmid in mature RPE cells is shown in Figure 4a (normalized to expression derived from CMV). This analysis identified four TFREs that significantly increased expression over basal expression from the minimal core promoter (2.6%–10.9% CMV, $p < 0.001$), that is, ChREBP:Mix, HRE, HELT/MITF, and Lf. Two elements (ChREBP:Mix and HELT/MITF) mediated different levels of GFP expression in the iPS-derived and primary cells, likely due to differences in TF relative abundance. As the relative level of reporter expression is also a function of affinity of the TF for its cognate TFRE we tested the consensus binding sequence of the TFREs (Table 1), as well as viral-derived TFREs (Table 2) to further identify elements that can independently mediate activation of gene transcription using available TFs in RPE cells. This analysis (Figure 4b) showed that NF-κB consensus sequence and viral-derived CREB were highly

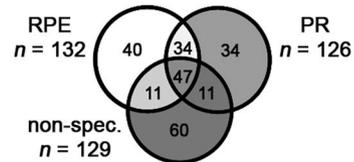
Step 1. Collation of transcriptomic datasets available in the literature and selection of genes from different groups, i.e. RPE-specific, PR-specific and non-specific genes.



Step 2. Identification of putative transcription factor regulatory elements (TFREs) in the promoter region (-1000bp to +200bp) of genes from each group using TFRE-prediction tools.



Step 3. Elimination of TFREs that were also present in non-desirable groups (i.e. PR- and non-specific genes) to minimize false positives, as well as similar TFREs to further reduce the pool.



Step 4. Screening of active TFREs using GFP reporter vectors transfected into iPS-derived and primary RPE cells.



Step 5. Characterization of heterotypic promoters to identify repressor elements (1st generation) and evaluate functionality in RPE cells (2nd generation).

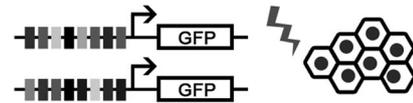


FIGURE 1 Summary of bioinformatic analysis (Steps 1–3) followed by *in vitro* screening of transcription factor regulatory elements (Step 4) and characterization of heterotypic constructs (Step 5) for *de novo* design of retinal pigment epithelial (RPE) synthetic promoters or engineering of endogenous RPE promoters

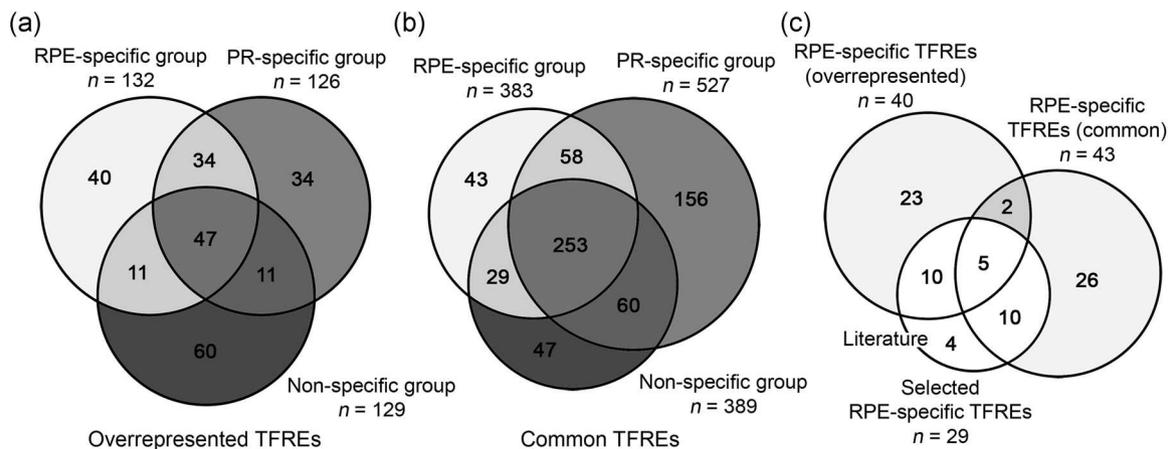
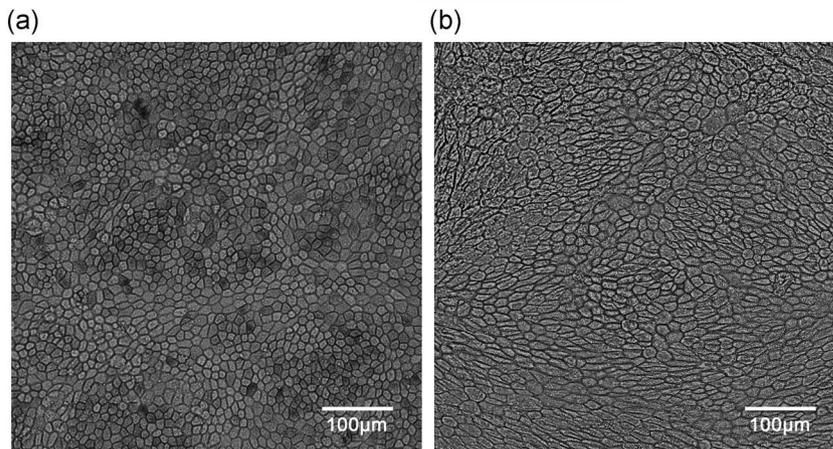


FIGURE 2 Distribution of discrete transcription factor regulatory elements (TFREs) across retinal pigment epithelium (RPE)-specific, photoreceptor (PR)-specific and nonspecific groups. Retinal endogenous promoters mediating high endogenous gene expression in RPE cells, PR cells, and both RPE and PR cells (35 promoters in each group; Table S1) were surveyed for the presence of discrete TFREs using Genomatix Gene Regulation software using (a) overrepresented TFRE method by analyzing the promoters against a pre-defined human promoter background, and (b) common TFRE method by selecting TFREs that present in at least 25% (9/35) of the genes. The region -1000 to +200 relative to putative transcriptional start site (TSS) was analyzed against a human promoter background to find overrepresented TFREs in each group, or for common TFREs that present in at least 25% (9/35) of the promoters. To identify potentially active TFREs in the high activity group, TFREs that also occurred in the PR and/or nonspecific groups were excluded and (c) the remaining TFREs from both methods were narrowed down to 25 as described in text. DNA sequences of the 29 selected TFREs (including 4 from the literature) are listed in Table 1

FIGURE 3 Epithelial morphology in differentiated human retinal pigment epithelial (RPE) cell monolayers. Microscope visualization of (a) induced pluripotent stem (iPS)-derived RPE cells at Day 28 and (b) primary RPE cells at Day 14 post-plating provides evidence of polygonal morphology and pigmentation. Images were taken with a 10 \times objective



active, attaining up to 36% of CMV activity. Consensus NF- κ B (p50), MafF, and PAX6 also displayed substantial activities (2.4%–8.5% CMV, $p < 0.001$) although we suspect the former was contributed by a weak binding of NF- κ B (V\$NFKAPPAB.02 matrix similarity 0.868). Contrarily, an alteration of six nucleotides flanking the E-box (CACGTG) in HELT/MITF endogenous sequence into discrete HELT and MITF consensus sequences resulted in 29%–68% reduction in GFP expression, while the use of an Lf consensus sequence diminished its activity. Furthermore, viral-derived Sp1, C/EBP ϵ :AT4, and AP-1 demonstrated low, but significant activities (>2.0% CMV, $p < 0.01$). Other TFREs showed no observable increase in GFP above core control levels, suggesting discrete mechanisms of TFRE transcriptional activation or absence of their cognate TFs.

As TF abundance may also impact cell specificity, we further analyzed TF expression at the transcript level using the Booi et al. (2009, 2010) datasets. While this does not permit direct quantification of TF activity in PR, it does provide information on general TF expression patterns between RPE and PR cells (mRNA half-lives in two different cell types can be expected to be comparable; see Johari et al., 2019), enabling prediction of cell-type specificity. Furthermore, this method is easily applicable to promoter design for most human cell types, for which transcriptomic datasets are normally accessible. Analysis of the mRNA transcript data (Table S2) indicate that MITF and C/EBP $\alpha,\beta,\delta,\epsilon$ were highly specific to RPE (3.87–10.7-fold change RPE/PR). While the former has been widely reported to be a key TF regulator of RPE cells (e.g., Esumi et al., 2007), this data indicates that C/EBP could be utilized as an RPE-specific element. Lf, AP-1, and MafF (2.31–2.82-fold change), and to a lesser extent C/EBP γ , AhR, NF- κ B, Mix, and PAX6 (1.21–1.75-fold change), were also expressed at relatively higher levels in RPE cells. On the other hand, Sp1 was expressed in RPE and PR cells at the same level (1.10-fold change) whereas CREB was expressed slightly higher in PR (0.82-fold change) — consistent with the bioinformatic analysis of the RPE and PR-specific groups (Table 2). We further note that Sp1 (V\$SP1.01) was the most common element in both RPE and PR-specific groups with 115 binding sites in 31/35 RPE promoters and 115 binding sites in 22/35 PR promoters, respectively (data not shown). In summary, these data

identify a group of transcriptionally active TFRE sequences that can be utilized to construct synthetic promoters for RPE cells. Contrarily, the data indirectly serves as a reference of TFREs that should be avoided for the construction of PR-specific promoters.

3.3 | First generation RPE promoters enable identification of repressor elements in RPE cells

Given the above finding that HELT/MITF could bind to suboptimal E-box sequences (leading to lower activities), we utilized the four positive endogenous TFREs identified in the screening exercise to evaluate their compatibility in a heterotypic construct (promoter 1/01). Importantly, given that a TF can act as a transcriptional activator or repressor (or both), we designed a set of promoters (promoter 1/02–1/10) containing specific combinations of the endogenous TFREs selected in the bioinformatic analysis to assess elements that primarily function as repressors in RPE cells. For each promoter, 6 copies of HELT/MITF were included to provide a promoter basal expression, and the other 27 specific TFREs (4 copies each) were randomly distributed using R software with the following design rules: (i) each specific TFRE occurred twice in different promoters, (ii) no two specific TFREs re-occurred in another promoter, and (iii) the relative order of constituent TFREs was random, separated by minimal spacers (Brown et al., 2017; Johari et al., 2019). Additionally, to test whether the selected endogenous elements from the bioinformatics analysis could generally act as positive effectors in heterotypic constructs, we constructed a promoter (1/11) containing two repeat copies of each element that present in upstream of BEST1 promoter (–900 to –1 relative to the TSS). The synthetic promoter constructs were chemically synthesized and inserted upstream of the minimal CMV core promoter in GFP reporter plasmids. An endogenous BEST1 promoter (–585 to +76 relative to the TSS; Esumi et al., 2004) was used as an RPE-specific control, which in our study exhibited ~1.5%–2% relative activity compared to the ubiquitous CMV promoter (representative GFP fluorescence images are shown in Figure S2). We note that we observed insignificant GFP expression ($\leq 0.5\%$ CMV) using an endogenous RPE65 promoter

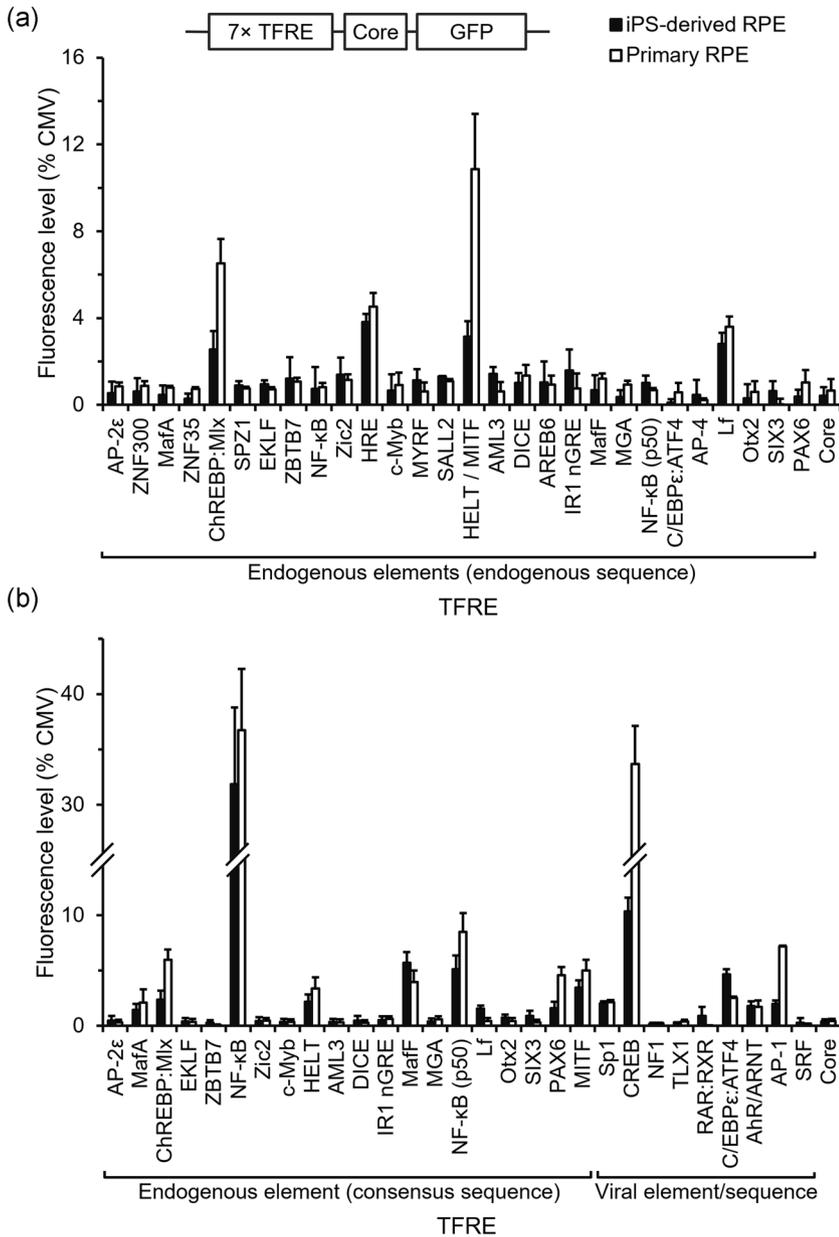


FIGURE 4 Screening discrete transcription factor regulatory element (TFRE) activity in retinal pigment epithelial (RPE) cell models. Seven copies of each TFRE (as described in Tables 1 and 2) comprising (a) endogenous sequences and (b) consensus/viral sequences were cloned in series upstream of a minimal cytomegalovirus (CMV) core promoter in reporter vectors encoding green fluorescent protein (GFP). 5×10^4 iPS-derived RPE cells were plated and $0.4 \mu\text{g}$ of plasmid was transfected into the cells by lipofection at Day 3 post-plating. $0.72 \mu\text{g}$ of plasmid was transfected into 3×10^5 primary RPE cells by nucleofection followed by plating. GFP fluorescence level was measured in the fully differentiated iPS-derived and primary RPE cells at Day 28 and Day 14 post-plating, respectively. Data are expressed as a percentage with respect to the GFP expression of a vector containing CMV-IE promoter. Data shown are the mean value \pm standard deviation of three independent experiments each performed in triplicate

(-655 to +52 relative to the TSS; Boulanger et al., 2000; Kachi et al., 2006).

The GFP expression for the first library of synthetic promoters (Table 3) are shown in Figure 5a. Promoter 1/01, which contained all four active TFREs identified in the screening exercise, did not display improved activity (<10% CMV) compared to testing of independent TFREs in homotypic constructs (Figure 4a). We surmise that HELT/MITF TF was constitutively bound to ChREBP:Mix E-box sequence with lower transcriptional activation as observed in the HELT and MITF consensus sequences (Figure 4b), thus lowering the promoter's activity. Further, the data in Figure 5a demonstrated that active HELT/MITF in promoters 1/02-1/10 were counteracted by repressor elements to produce weak or nonfunctional promoters. To identify these repressor elements, the synthetic promoters' activities were analyzed against their individual components. This breakdown

(Figure 5b) suggests that seven elements, that is, ZNF300, ZBTB7, NF- κ B, AML3, NF- κ B (p50), SIX3, PAX6 had repressive effects on transcriptional activity, where the promoters exhibited < 2% fluorescence level in both iPS-derived and primary RPE cells. With the exception of NF- κ B and PAX6, this finding is consistent with the literature indicating that these TFREs function as transcriptional repressors in mammalian cells (Costoya, 2007; Gou et al., 2004; Guan et al., 2005; Isemann et al., 2007; Liu et al., 2010).

Further bioinformatic examination on the endogenous NF- κ B sequence (V\$NFKAPPAB.02, matrix similarity 0.869) indicated that this particular sequence overlapped with transcriptional repressor RBP-J κ binding sequence (V\$RBPJK.02 matrix similarity 0.958; Figure S1A). This constitutive binding by RBP-J κ (Lee et al., 2000) expounds the repression observed in promoters 1/03 and 1/04 (Figure 5) as well as the inactivity of endogenous NF- κ B sequence

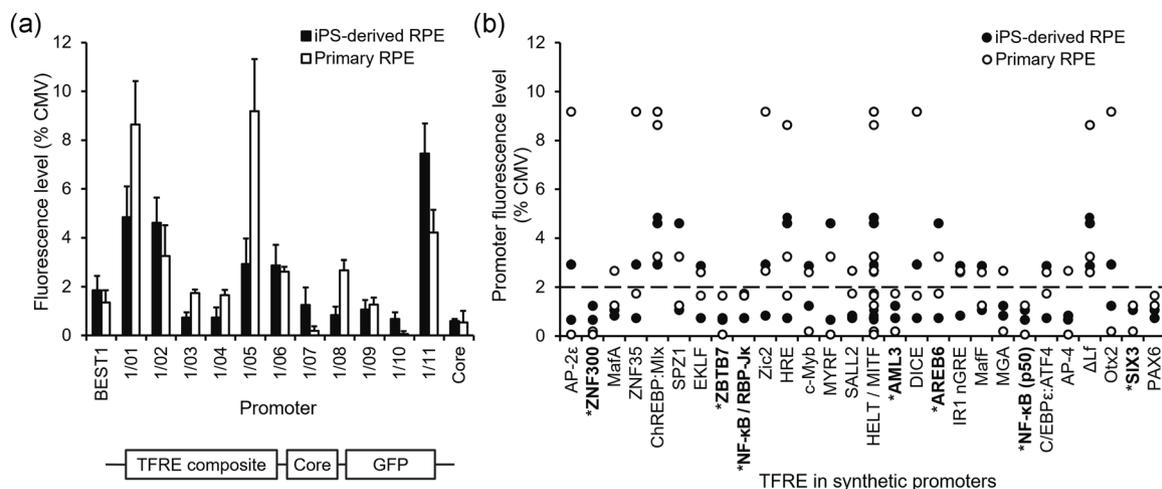


FIGURE 5 Measurement of first-generation synthetic promoter activity in retinal pigment epithelium (RPE) cells. (a) 11 synthetic promoters comprising Library 1 (transcription factor regulatory element [TFRE] compositions described in Table 3) were transfected into induced pluripotent stem (iPS)-derived and primary RPE cells. Intracellular green fluorescent protein (GFP) level was analyzed in differentiated RPE cells. Data are expressed as a percentage with respect to the expression level exhibited by the control cytomegalovirus (CMV) promoter. GFP expression driven by the BEST1 promoter was also tested as RPE-specific promoter control. Data shown are the mean \pm standard deviation of three independent experiments each performed in triplicate. (b) The GFP levels exhibited by the synthetic promoters in (a) were plotted against the promoters' specific TFRE components to identify potential repressor elements (marked by an asterisk [*]) by setting the minimum expression threshold to 2% CMV and further analyzed as described in the text

compared to the consensus sequence (Figure 4). In contrast, PAX6 has been reported to act as a transactivator in RPE cells (Raviv et al., 2014) and its consensus sequence displayed significant activity (Figure 4b). Thus, PAX6 may be excluded as a repressor in which the repression observed in promoters 1/04 and 1/09 could be attributed to RBP-J κ and NF- κ B (p50) elements respectively. Further, as promoter 1/02 exhibited lower GFP expression in primary cells compared to promoters 1/01 and 1/05 and did not contain any of above repressor elements, we deduce that the AREB6 secondary binding sequence employed repressed transcription as reported by Ikeda and Kawakami (1995) with E-box motif in HELT/MITF sequence acted as the main binding site of multi-domain AREB6 protein (see Figure S1A). Promoter 1/11, which contained two copies of selected elements from the BEST1 promoter demonstrated \sim 3–4-fold higher expression than the BEST1 promoter itself, confirming the utility of the RPE cell synthetic promoter design approach.

3.4 | Second generation RPE promoters exhibit an inverse correlation between activity and specificity

Based on the observations from the first library, we created a second library of functional synthetic promoters (Table 4) by omitting probable repressor elements with two objectives: (i) high RPE cell-specificity, attainable by limiting the constructs to endogenous elements (promoters 2/01–2/07), and (ii) high RPE transcriptional activity, attainable by including active viral-derived elements (promoters 2/08–2/13). The former is highly desirable for targeted AAV-mediated transduction in vivo, whereas the latter can be efficiently

applied for recombinant protein expression in vitro to investigate the effects of protein overexpression in retinal diseases (e.g., HtrA1 enrichment; Melo et al., 2018) as well as to reprogram RPE cells to an altered lineage (e.g., neuronal cells; Hu et al., 2014). Additionally, we created an engineered BEST1 promoter by mutating a total of 25 nucleotides to remove repressors and/or introduce active TFREs (Figure S3).

Measurement of GFP expression after transfection into RPE cells is shown in Figure 6. These data show that promoter activities vary by \sim 30-fold, where the most active promoters exceeded the transcriptional activity of CMV in primary cells. The engineered BEST1 promoter exhibited a 30%–32% improvement in expression compared to its native counterpart, although this was largely insignificant ($p = 0.10$ – 0.14). Similarly, promoter 2/03, devoid of repressor elements, displayed a 2.1-fold increase in expression compared to promoter 1/11 in primary cells, although biasing the elements towards active elements (promoter 2/04) appeared to have a negative effect – likely due to suboptimal TFRE stoichiometry (Martinelli & De Simone, 2005). Anticipated to be RPE-specific, the in vitro expression levels obtained from promoters 2/01–2/05 were up to \sim 8-fold higher than that of BEST promoter but significantly lower compared to CMV ($\leq 17\%$). While NF- κ B in promoters 2/06 and 2/07 significantly improved expression with the latter attaining 62% CMV activity in primary cells, we conjecture the use of high copy number of NF- κ B would increase promoter activity in PR (and other) cells, where the cognate TF is also present (Table S2). Expectedly, the data in Figure 6 also shows that strong RPE promoters can be constructed by incorporating viral-derived CREB, Sp1 and AP-1, and biasing the TFRE copy number towards highly active elements resulted in

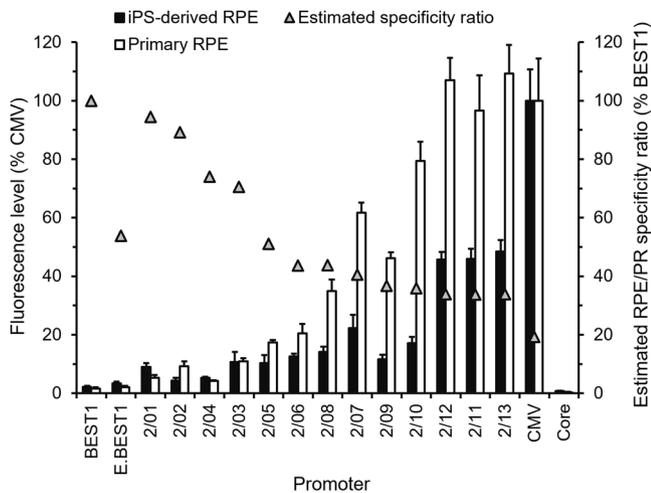


FIGURE 6 Measurement of second-generation synthetic promoter activity and predicted retinal pigment epithelium (RPE) cell specificity. 13 synthetic promoters comprising Library 2 (transcription factor regulatory element [TFRE] compositions described in Table 4) were transfected into induced pluripotent stem (iPS)-derived and primary RPE cells. An engineered BEST1 promoter (E.BEST1) was constructed by mutating a total of 25 nucleotides to remove and/or introduce specific TFREs (Figure S3). Intracellular green fluorescent protein (GFP) level was analyzed in differentiated RPE cells. Data are expressed as a percentage with respect to the expression level exhibited by the control cytomegalovirus (CMV) promoter. Predictive RPE/photoreceptor (PR) specificity ratio of each promoter was calculated using Equation 1 based on the transcriptional activity of a specific TFRE, the copy number of a specific TFRE in the promoter, and the cognate transcription factor (TF) mRNA expression fold-change in RPE/PR. Data are expressed as a percentage with respect to the specificity ratio of the control bestrophin-1 (BEST1) promoter

substantial increase in promoter strength up to 109% of CMV in primary cells. However, reporter expression in iPS-derived cells was relatively lower compared to that observed for the primary cells (achieving only ~50% of CMV activity), indicative of differences in transcriptional landscape. We deduce that iPS-derived cells, reprogrammed from somatic (skin or blood) cells, contained untested TFs that conferred relatively higher CMV expression.

To further illustrate the promoters' (predictive) capability in conferring specific and exclusive cellular tropism for RPE gene therapy, we calculated an "estimated RPE/PR specificity ratio" for each promoter as a function of (i) the transcriptional activity of a specific TFRE (Figure 4), (ii) the copy number of a specific TFRE within the promoter (Table 4), and (iii) the cognate TF mRNA expression fold-change in RPE over PR (Table S2; Equation 1). As shown in Figure 6, this analysis indicates that promoters designed with endogenous elements would drive targeted transgene expression to the RPE cells *in vivo*, whereas promoters with viral-derived elements would have significantly reduced specificity. While not directly relevant, our data from a separate study showed that these promoters exhibited no or very low activity ($\leq 20\%$ CMV) in a human kidney cell line (Figure S4), illustrating that cell-specific control of

recombinant gene transcriptional activity is feasible. Furthermore, one major limitation of AAVs as vectors is that AAV packaging capacity is fundamentally restricted to 5 kb where packaged vector genomes derived from plasmid-encoded vectors exceeding 5 kb are truncated on the 5' end and heterogeneous in length, as well as result in a considerable reduction in viral production yields (Wu et al., 2010). The synthetic promoters in this study were relatively small in size compared the control BEST1 and CMV promoters (Table 4), making them advantageous for AAV-mediated gene delivery. For example, promoter 2/03 achieved a 4.9–6.8-fold increase in transcriptional activity over the BEST1 promoter while being 29% shorter in length at 468 bp. We anticipate that combinatorial, context-dependent empirical modeling (Johari et al., 2019) will further assist the construction of promoters with optimal TFRE stoichiometry for RPE gene therapy applications.

4 | DISCUSSION

In this study we have characterized various regulatory elements in RPE cells derived from endogenous and viral promoters. Using a combined *in silico* and *in vitro* screening approach, we successfully identified active TFRE candidates (sequences) that could be utilized to construct synthetic promoter assemblies with strong and/or RPE specific expression. The data in this study also indirectly serves as a reference for TFREs that should be avoided in constructing PR-specific promoters. Furthermore, although our specific objective was to identify active RPE-associated regulatory elements, it should be recognized that this approach may be universally applicable and adapted to accommodate screening of TFREs associated with other cell/tissue types. Indeed, cell function is largely controlled by the action of TFs that recognize and bind particular sequence motifs in the genome and regulate gene expression. While hundreds of TFs are expressed in any one tissue type (Vaquerizas et al., 2009), only a small range of core TFs are responsible in programming the gene expression that define individual cell identity (D'Alessio et al., 2015). This is evident, for example, where our study indicated that SRF (serum response factor) element had no transcriptional activity in RPE cells (Figure 4b) yet the identical sequence formed the primary building block of synthetic promoters that conferred muscle-specific expression (Li et al., 1999).

We further identified suboptimal TF binding sequences and endogenous TFREs (25%) that acted as transcriptional repressors. These were not entirely unexpected as the promoters in RPE have evolved to function in a complex spatiotemporal gene regulation of the retina including pigment biogenesis, ion transport, and growth factor secretion (Booij et al., 2010). However, it is highly unlikely that they will be optimal for use in a more specific context such as AAV-mediated gene therapy. For example, many of the transcriptional repressors identified in this study are present in BEST1 promoter suggesting that a large proportion of the BEST1 sequence (and that of other RPE promoters) may be functionally redundant for recombinant gene expression. Thus, identification of such repressor

elements, as well as optimization of active TF binding sequences, would permit engineering of endogenous promoters for enhanced performance. With regard to the latter, tens of TFRE motif sequence variants can be characterized simultaneously through *in vitro* use of parallel high-throughput screening techniques, allowing determination of their optimal binding affinity. With regard to the former, the functional impact of repressor elements can be identified and accurately quantitated using the TFRE-specific decoy technology previously developed in this laboratory (Brown et al., 2013, 2015).

Even though we have not tested our promoters against PR (due to lack of reliable model cells, see below), and therefore cannot definitively claim that they will exhibit restricted gene expression in RPE cells, the methodology presented allows creation of promoters using binding sequences that are exclusive to RPE-specific promoters and correspond to relatively high expression of their cognate TFs in RPE cells, thus enabling confident prediction of their functionality. On the other hand, TATA box (present in the minimal core promoter) is a known modular component in that the strength of the TATA-RNA polymerase complex and the ensuing transcription that it mediates has very little noise to promoter activity—it simply acts as a linear amplifier without influencing specificity of gene expression controlled by upstream cis-regulatory elements (Mogno et al., 2010). Indeed, muscle (Li et al., 1999), deregulated β -catenin (Lipinski et al., 2004), liver (Han et al., 2011), and colorectal cancer cell-specific (Roberts et al., 2017) synthetic promoters all contained a TATA box. Furthermore, our bioinformatic analysis indicated that Sp1 is highly prevalent in all three groups (RPE, PR, and nonspecific) analyzed—in agreement with the notion that Sp1 is essential for maintaining basal transcription of genes and protection of CpG islands from *de novo* methylation (Samson & Wong, 2002). Accordingly, we conjecture that Sp1 does not influence the specificity of a promoter that is designed to mediate cell type specific expression. We further acknowledge that the RPE cell models we employed may contain small subsets cells in a variably differentiated state that have a transcriptional landscape deviating from that of the fully differentiated population. To characterize this further experiments utilizing multiplex flow cytometry to directly link markers of RPE differentiation (e.g., Plaza Reyes et al., 2020) to synthetic promoter mediated reporter gene expression would be possible.

Lastly, *in vitro* model cells may have an altered transcriptional landscape compared to cells *in vivo* and may vary from one model to another. Our study demonstrated that iPSC-derived and primary RPE cells transactivated the same TFREs (albeit at different levels), concurring with previous reports that these cells displayed appropriate levels of RPE gene expression compared to native tissue (Ablonczy et al., 2011; Liao et al., 2010). In contrast, there are no established model cells representative of pure human retinal PRs (McDougald et al., 2019). Isolated primary PR cells, deprived of their extracellular matrix and cellular contacts (RPE, retinal and choroidal blood supply, etc), display rapid kinetics of degeneration (Fernandez-Bueno et al., 2012). While iPSC cells can be dependably differentiated into PR cells following a 60-day induction regimen, rapid loss of cells committed to a PR fate (rhodopsin, opsin) was observed at

Day 45–60 (Mellough et al., 2012). On the other hand, 661W PR cell line, derived from a mouse retinal tumor, expresses several markers of cone PR cells but not of rod cells (Tan et al., 2004) and was also reported to exhibit the properties of retinal ganglion cells (Sayyad et al., 2017). Nevertheless, where reliable model cells are not available, we demonstrated that it may possible to design cell-specific promoters *in silico* for *in vivo* applications. As the quality and volume of 'omics data continues to increase, and, given the progressive development of TFRE database and informatic tools (e.g., Wu et al., 2019), we envisage that synthetic promoters will facilitate advancement of the current revolution in AAV-mediated gene therapy.

ACKNOWLEDGMENTS

This study was supported by REGENXBIO, USA. The authors thank Stephen Jaffé and William Morgan-Evans (University of Sheffield) for assistance in cell culture and fluorescence microscopy, respectively, and Laura Moran (REGENXBIO) for project management.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Yusuf B. Johari: *conceptualization, methodology, investigation, formal analysis, validation, data curation, visualization, project administration, writing—original draft.* Andrew C. Mercer: *conceptualization, methodology, validation, funding acquisition, project administration.* Ye Liu: *conceptualization, methodology, validation.* Adam J. Brown: *conceptualization, methodology, validation, funding acquisition.* David C. James: *conceptualization, methodology, validation, funding acquisition, supervision, writing—review and editing.*

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Yusuf B. Johari  <http://orcid.org/0000-0001-9933-5764>

Adam J. Brown  <http://orcid.org/0000-0002-3290-4560>

David C. James  <http://orcid.org/0000-0002-1697-151X>

REFERENCES

- Ablonczy, Z., Dahrouj, M., Tang, P. H., Liu, Y., Sambamurti, K., Marmorstein, A. D., & Crosson, C. E. (2011). Human retinal pigment epithelium cells as functional models for the RPE *in vivo*. *Investigative Ophthalmology & Visual Science*, 52, 8614–8620. <https://doi.org/10.1167/iovs.11-8021>
- Booij, J. C., tenBrink, J. B., Swagemakers, S. M., Verkerk, A. J., Essing, A. H., van der Spek, P. J., & Bergen, A. A. (2010). A new strategy to identify and annotate human RPE-specific gene expression. *PLOS One*, 5, e9341. <https://doi.org/10.1371/journal.pone.0009341>
- Booij, J. C., van Soest, S., Swagemakers, S. M., Essing, A. H., Verkerk, A. J., van der Spek, P. J., Gorgels, T. G., & Bergen, A. A. (2009). Functional annotation of the human retinal pigment epithelium transcriptome. *BMC Genomics*, 10, 164. <https://doi.org/10.1186/1471-2164-10-164>

- Boulanger, A., Liu, S., Henningsgaard, A. A., Yu, S., & Redmond, T. M. (2000). The upstream region of the Rpe65 gene confers retinal pigment epithelium-specific expression in vivo and in vitro and contains critical octamer and E-box binding sites. *The Journal of Biological Chemistry*, 275, 31274–31282. <https://doi.org/10.1074/jbc.M003441200>
- Brown, A. J., Gibson, S. J., Hatton, D., & James, D. C. (2017). In silico design of context-responsive mammalian promoters with user-defined functionality. *Nucleic Acids Research*, 45, 10906–10919. <https://doi.org/10.1093/nar/gkx768>
- Brown, A. J., Mainwaring, D. O., Sweeney, B., & James, D. C. (2013). Block decoys: Transcription-factor decoys designed for in vitro gene regulation studies. *Analytical Biochemistry*, 443, 205–210. <https://doi.org/10.1016/j.ab.2013.09.003>
- Brown, A. J., Sweeney, B., Mainwaring, D. O., & James, D. C. (2015). NF- κ B, CRE and YY1 elements are key functional regulators of CMV promoter-driven transient gene expression in CHO cells. *Biotechnology Journal*, 10, 1019–1028. <https://doi.org/10.1002/biot.201400744>
- Costoya, J. A. (2007). Functional analysis of the role of POK transcriptional repressors. *Briefing in Functional Genomics*, 6, 8–18. <https://doi.org/10.1093/bfpg/elm002>
- D'Alessio, A. C., Fan, Z. P., Wert, K. J., Baranov, P., Cohen, M. A., Saini, J. S., Cohick, E., Charniga, C., Dadon, D., Hannett, N. M., Young, M. J., Temple, S., Jaenisch, R., Lee, T. I., & Young, R. A. (2015). A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Reports*, 5, 763–775. <https://doi.org/10.1016/j.stemcr.2015.09.016>
- Esumi, N., Kachi, S., Campochiaro, P. A., & Zack, D. J. (2007). VMD2 promoter requires two proximal E-box sites for its activity in vivo and is regulated by the MITF-TFE family. *Journal of Biological Chemistry*, 282, 1838–1850. <https://doi.org/10.1074/jbc.M609517200>
- Esumi, N., Oshima, Y., Li, Y., Campochiaro, P. A., & Zack, D. J. (2004). Analysis of the VMD2 promoter and implication of E-box binding factors in its regulation. *Journal of Biological Chemistry*, 279, 19064–19073. <https://doi.org/10.1074/jbc.M309881200>
- Fernandez-Bueno, I., Fernández-Sánchez, L., Gayoso, M. J., García-Gutierrez, M. T., Pastor, J. C., & Cuenca, N. (2012). Time course modifications in organotypic culture of human neuroretina. *Experimental Eye Research*, 104, 26–38. <https://doi.org/10.1016/j.exer.2012.08.012>
- Ghazal, P., DeMattei, C., Giulietti, E., Kliewer, S. A., Umesono, K., & Evans, R. M. (1992). Retinoic acid receptors initiate induction of the cytomegalovirus enhancer in embryonal cells. *Proceedings of the National Academy of Sciences of the U.S.A.*, 89, 7630–7634. <https://doi.org/10.1073/pnas.89.16.7630>
- Gou, D., Wang, J., Gao, L., Sun, Y., Peng, X., Huang, J., & Li, W. (2004). Identification and functional analysis of a novel human KRAB/C₂H₂ zinc finger gene ZNF300. *Biochimica et Biophysica Acta*, 1676, 203–209. <https://doi.org/10.1016/j.bbexp.2003.11.011>
- Guan, H., Hou, S., & Ricciardi, R. P. (2005). DNA binding of repressor nuclear factor- κ B p50/p50 depends on phosphorylation of Ser³³⁷ by the protein kinase A catalytic subunit. *Journal of Biological Chemistry*, 280, 9957–9962. <https://doi.org/10.1074/jbc.M412180200>
- Guziewicz, K. E., Zangerl, B., Komáromy, A. M., Iwabe, S., Chiodo, V. A., Boye, S. L., Hauswirth, W. W., Beltran, W. A., & Aguirre, G. D. (2013). Recombinant AAV-mediated BEST1 transfer to the retinal pigment epithelium: Analysis of serotype-dependent retinal effects. *PLOS One*, 8, e75666. <https://doi.org/10.1371/journal.pone.0075666>
- Han, J., McLane, B., Kim, E.-H., Yoon, J.-W., & Jun, H.-S. (2011). Remission of diabetes by insulin gene therapy using a hepatocyte-specific and glucose-responsive synthetic promoter. *Molecular Therapy*, 19, 470–478. <https://doi.org/10.1038/mt.2010.255>
- Hu, Q., Chen, R., Teesalu, T., Ruoslahti, E., & Clegg, D. O. (2014). Reprogramming human retinal pigmented epithelial cells to neurons using recombinant proteins. *Stem Cells Translational Medicine*, 3, 1526–1534. <https://doi.org/10.5966/sctm.2014-0038>
- Ikeda, K., & Kawakami, K. (1995). DNA binding through distinct domains of zinc-finger-homeodomain protein AREB6 has different effects on gene transcription. *European Journal of Biochemistry*, 233, 73–82. https://doi.org/10.1111/j.1432-1033.1995.073_1.x
- Iseman, S., Cakouros, D., Zannettino, A., Shi, S., & Gronthos, S. (2007). hTERT transcription is repressed by Cbfa1 in human mesenchymal stem cell populations. *Journal of Bone and Mineral Research*, 22, 897–906. <https://doi.org/10.1359/jbmr.070308>
- Johari, Y. B., Brown, A. J., Alves, C. S., Zhou, Y., Wright, C. M., Estes, S. D., Kshirsagar, R., & James, D. C. (2019). CHO genome mining for synthetic promoter design. *Journal of Biotechnology*, 294, 1–13. <https://doi.org/10.1016/j.jbiotec.2019.01.015>
- Kachi, S., Esumi, N., Zack, D. J., & Campochiaro, P. A. (2006). Sustained expression after nonviral ocular gene transfer using mammalian promoters. *Gene Therapy*, 13, 798–804. <https://doi.org/10.1038/sj.gt.3302700>
- Kokkinaki, M., Sahibzada, N., & Golestaneh, N. (2011). Human induced pluripotent stem-derived retinal pigment epithelium (RPE) cells exhibition transport, membrane potential, polarized vascular endothelial growth factor secretion, and gene expression pattern similar to native RPE. *Stem Cells*, 29, 825–835. <https://doi.org/10.1002/stem.635>
- Komáromy, A. M., Alexander, J. J., Rowlan, J. S., Garcia, M. M., Chiodo, V. A., Kaya, A., Tanaka, J. C., Acland, G. M., Hauswirth, W. W., & Aguirre, G. D. (2010). Gene therapy rescues cone function in congenital achromatopsia. *Human Molecular Genetics*, 19, 581–2593. <https://doi.org/10.1093/hmg/ddq136>
- Lashmit, P., Wang, S., Li, H., Isomura, H., & Stinski, M. F. (2009). The CREB site in the proximal enhancer is critical for cooperative interaction with the other transcription factor binding sites to enhance transcription of the major intermediate-early genes in human cytomegalovirus-infected cells. *Journal of Virology*, 83, 8893–8904. <https://doi.org/10.1128/JVI.02239-08>
- Lee, S. H., Wang, X., & DeJong, J. (2000). Functional interactions between an atypical NF- κ B site from the rat CYP2B1 promoter and the transcriptional repressor RBP-J κ /CBF1. *Nucleic Acids Research*, 28, 2091–2098. <https://doi.org/10.1093/nar/28.10.2091>
- Li, X., Eastman, E. M., Schwartz, R. J., & Draghia-Akli, R. (1999). Synthetic muscle promoters: Activities exceeding naturally occurring regulatory sequences. *Nature Biotechnology*, 17, 241–245. <https://doi.org/10.1038/6981>
- Liao, J. L., Yu, J., Huang, K., Hu, J., Diemer, T., Ma, Z., Dvash, T., Yang, X. J., Travis, G. H., Williams, D. S., Bok, D., & Fan, G. (2010). Molecular signature of primary retinal pigment epithelium and stem-cell-derived RPE cells. *Human Molecular Genetics*, 19, 4229–4238. <https://doi.org/10.1093/hmg/ddq341>
- Lipinski, K. S., Djeha, H. A., Gawn, J., Cliffe, S., Maitland, N. J., Palmer, D. H., Mountain, A., Irvine, A. S., & Wrighton, C. J. (2004). Optimization of a synthetic β -catenin-dependent promoter for tumor-specific cancer gene therapy. *Molecular Therapy*, 10, 150–161. <https://doi.org/10.1016/j.ymthe.2004.03.021>
- Liu, W., Lagutin, O., Swindell, E., Jamrich, M., & Oliver, G. (2010). Neuroretina specification in mouse embryos requires Six3-mediated suppression of Wnt8b in the anterior neural plate. *The Journal of Clinical Investigation*, 120, 3568–3577. <https://doi.org/10.1172/JCI43219>
- Logg, C. R., Logg, A., Matusik, R. J., Bochner, B. H., & Kasahara, N. (2002). Tissue-specific transcriptional targeting of a replication-competent retroviral vector. *Journal of Virology*, 76, 12783–12791. <https://doi.org/10.1128/jvi.76.24.12783-12791.2002>

- Martinelli, R., & De Simone, V. (2005). Short and highly efficient synthetic promoters for melanoma-specific gene expression. *FEBS Letter*, 579, 153–156. <https://doi.org/10.1016/j.febslet.2004.11.068>
- McDougald, D. S., Duong, T. T., Palozola, K. C., Marsh, A., Papp, T. E., Mills, J. A., Zhou, S., & Bennett, J. (2019). CRISPR activation enhances in vitro potency of AAV vectors driven by tissue-specific promoters. *Molecular Therapy: Methods & Clinical Development*, 13, 380–389. <https://doi.org/10.1016/j.omtm.2019.03.004>
- Mellough, C. B., Sernagor, E., Moreno-Gimeno, I., Steel, D. H. W., & Lako, M. (2012). Efficient stage-specific differentiation of human pluripotent stem cells toward retinal photoreceptor cells. *Stem Cells*, 30, 673–686. <https://doi.org/10.1002/stem.1037>
- Melo, E., Oertle, P., Trepp, C., Meistermann, H., Burgoyne, T., Sborgi, L., Cabrera, A. C., Chen, C.-Y., Hoflack, J.-C., Kam-Thong, T., Schmucki, R., Badi, L., Flint, N., Ghiani, Z. E., Delobel, F., Stucki, C., Gromo, G., Einhaus, A., Hornsperger, B., ... Iacone, R. (2018). HtrA1 mediated intracellular effects on tubulin using a polarized RPE disease model. *EBioMedicine*, 27, 258–274. <https://doi.org/10.1016/j.ebiom.2017.12.011>
- Le Meur, G., Stieger, K., Smith, A. J., Weber, M., Deschamps, J. Y., Nivard, D., Mendes-Madeira, A., Provost, N., Péréon, Y., Cherel, Y., Ali, R. R., Hamel, C., Moullier, P., & Rolling, F. (2007). Restoration of vision in RPE65-deficient Briard dogs using an AAV serotype 4 vector that specifically targets the retinal pigmented epithelium. *Gene Therapy*, 14, 292–303. <https://doi.org/10.1038/sj.gt.3302861>
- Mogno, I., Vallania, F., Mitra, R. D., & Cohen, B. A. (2010). TATA is a modular component of synthetic promoters. *Genome Research*, 20, 1391–1397. <https://doi.org/10.1101/gr.106732.110>
- Naso, M. F., Tomkowicz, B., Perry, W. L., III, & Strohl, W. R. (2017). Adeno-associated virus (AAV) as a vector for gene therapy. *BioDrugs*, 31, 317–334. <https://doi.org/10.1007/s40259-017-0234-5>
- Plaza Reyes, A., Petrus-Reurer, S., Padrell Sánchez, S., Kumar, P., Douagi, I., Bartuma, H., Aronsson, M., Westman, S., Lardner, E., André, H., Falk, A., Nandrot, E. F., Kvant, A., & Lanner, F. (2020). Identification of cell surface markers and establishment of monolayer differentiation to retinal pigment epithelial cells. *Nature Communications*, 11, 1609. <https://doi.org/10.1038/s41467-020-15326-5>
- Raviv, S., Bharti, K., Rencus-Lazar, S., Cohen-Tayar, Y., Schyr, R., Evantal, N., Meshorer, E., Zilberberg, A., Idelson, M., Reubinoff, B., Grebe, R., Rosin-Arbesfeld, R., Lauderdale, J., Luty, G., Arnheiter, H., & Ashery-Padan, R. (2014). PAX6 regulates melanogenesis in the retinal pigmented epithelium through feed-forward regulatory interactions with MITF. *PLOS Genetics*, 10, e1004360. <https://doi.org/10.1371/journal.pgen.1004360>
- Roberts, M. L., Katsoupi, P., Tseveleki, V., & Taoufik, E. (2017). Bioinformatically informed design of synthetic mammalian promoters. *Methods in Molecular Biology*, 1651, 93–112. https://doi.org/10.1007/978-1-4939-7223-4_8
- Samson, S. L.-A., & Wong, N. C. W. (2002). Role of Sp1 in insulin regulation of gene expression. *Journal of Molecular Endocrinology*, 29, 265–279. <https://doi.org/10.1677/jme.0.0290265>
- Sayyad, Z., Sirohi, K., Radha, V., & Swarup, G. (2017). 661W is a retinal ganglion precursor-like cell line in which glaucoma-associated optineurin mutants induce cell death selectively. *Science Reports*, 7, 16855. <https://doi.org/10.1038/s41598-017-17241-0>
- Szymanski, P., Anwer, K., & Sullivan, S. M. (2006). Development and characterization of a synthetic promoter for selective expression in proliferating endothelial cells. *The Journal of Gene Medicine*, 8, 514–523. <https://doi.org/10.1002/jgm.875>
- Tan, E., Ding, X.-Q., Saadi, A., Agarwal, N., Naash, M. I., & Al-Ubaidi, M. R. (2004). Expression of cone-photoreceptor-specific antigens in a cell line derived from retinal tumors in transgenic mice. *Investigative Ophthalmology & Visual Science*, 45, 764–768. <https://doi.org/10.1167/iovs.03-1114>
- Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., & Luscombe, N. M. (2009). A census of human transcription factors: Function, expression and evolution. *Nature Reviews Genetics*, 10, 252–263. <https://doi.org/10.1038/nrg2538>
- Wu, M. R., Nissim, L., Stupp, D., Pery, E., Binder-Nissim, A., Weisinger, K., Enghuus, C., Palacios, S. R., Humphrey, M., Zhang, Z., Maria Novoa, E., Kellis, M., Weiss, R., Rabkin, S. D., Tabach, Y., & Lu, T. K. (2019). A high-throughput screening and computation platform for identifying synthetic promoters with enhanced cell-state specificity (SPECS). *Nature Communications*, 10, 2880. <https://doi.org/10.1038/s41467-019-10912-8>
- Wu, Z., Yang, H., & Colosi, P. (2010). Effect of genome size on AAV vector packaging. *Molecular Therapy*, 18, 80–86. <https://doi.org/10.1038/mt.2009.255>

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Johari, Y. B., Mercer, A. C., Liu, Y., Brown, A. J., & James, D. C. (2021). Design of synthetic promoters for controlled expression of therapeutic genes in retinal pigment epithelial cells. *Biotechnology and Bioengineering*, 1–15. <https://doi.org/10.1002/bit.27713>