



## OPEN A DNA base-specific sequence interposed between CRX and NRL contributes to *RHODOPSIN* expression

Rosa Maritato<sup>1</sup>, Alessia Medugno<sup>1</sup>, Emanuela D'Andretta<sup>1</sup>, Giulia De Riso<sup>2,3</sup>, Mariangela Lupo<sup>4</sup>, Salvatore Botta<sup>6</sup>, Elena Marrocco<sup>4</sup>, Mario Renda<sup>4</sup>, Martina Sofia<sup>4</sup>, Claudio Mussolino<sup>7</sup>, Maria Laura Bacci<sup>5</sup> & Enrico Maria Surace<sup>1</sup>✉

Gene expression emerges from DNA sequences through the interaction of transcription factors (TFs) with DNA *cis*-regulatory sequences. In eukaryotes, TFs bind to transcription factor binding sites (TFBSs) with differential affinities, enabling cell-specific gene expression. In this view, DNA enables TF binding along a continuum ranging from low to high affinity depending on its sequence composition; however, it is not known whether evolution has entailed a further level of entanglement between DNA-protein interaction. Here we found that the composition and length (22 bp) of the DNA sequence interposed between the CRX and NRL retinal TFs in the proximal promoter of *RHODOPSIN* (*RHO*) largely controls the expression levels of *RHO*. Mutagenesis of CRX-NRL DNA linking sequences (here termed "DNA-linker") results in uncorrelated gene expression variation. In contrast, mutual exchange of naturally occurring divergent human and mouse Rho *cis*-regulatory elements conferred similar yet species-specific Rho expression levels. Two orthogonal DNA-binding proteins targeted to the DNA-linker either activate or repress the expression of Rho depending on the DNA-linker orientation relative to the CRX and NRL binding sites. These results argue that, in this instance, DNA itself contributes to CRX and NRL activities through a code based on specific base sequences of a defined length, ultimately determining optimal *RHO* expression levels.

DNA plays a dual role in gene expression. It encodes for regulators of transcription such as transcription factors (TFs) and allows their positioning in the transcriptional machinery by bearing specific transcription factor binding sites (TFBSs)<sup>1,2</sup>. The fundamental interaction that generates DNA-protein binding specificity and eventually gene expression, is generated by the specific affinity of TF protein contacts with specific DNA bases within TFBSs<sup>3-5</sup>. However, in metazoans, a tight control of the level, pattern, and timing of gene expression of different cell types has arisen by evolving TFs (protein) and their cognate TFBSs (DNA) interaction<sup>6</sup>. Each gene is regulated by multiple TFs-TFBS interactions<sup>7</sup>. TFBSs contain DNA sequence variation that exceeds that present in the evolutionarily more constrained coding sequence of TFs, in a model of co-evolution in which TFBSs evolve faster than their binders, TFs<sup>8-10</sup>. TFBSs are short degenerate DNA sequences (also known as motifs) that occur as monomeric<sup>11,12</sup>, dimeric, and overlapping sequences embedded in *cis*-regulatory elements, so that the DNA context surrounding them can also influence how they are bound by TFs<sup>13-18</sup>. On the other hand, TFs independently<sup>11,12,19</sup>, or coupling TF-TF interactions (TF-TFs) cooperatively<sup>20-22,18</sup>, and combinatorially<sup>23,24</sup> recognize and bind TFBSs and their flanks through protein<sup>23,25</sup> or DNA-protein-driven mechanisms<sup>15,26-29</sup>. On a larger scale, TFs and TFBSs combine their binding modes functionally to determine gene expression. The number, position, helical DNA face and binding site (strand) of TFBSs and are recognized by multiple TFs, resulting in qualitative and quantitative binding at the promoters and enhancers<sup>30-32</sup>, eventually leading to transcription, propelled by protein-proteins interaction<sup>33</sup>. These different modes originate and resolve in one recognized basic mechanism, TF-DNA binding affinity<sup>34,35</sup>. The dynamic interactions that regulate

<sup>1</sup>Department of Translational Medicine, University of Naples Federico II, Naples, Italy. <sup>2</sup>Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Naples, Italy. <sup>3</sup>AOU Federico II, Naples, Italy. <sup>4</sup>Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy. <sup>5</sup>Department of Veterinary Medical Sciences, University of Bologna, Bologna, Italy. <sup>6</sup>Department of Translational Medical Science, University of Campania Luigi Vanvitelli, Naples, Italy. <sup>7</sup> Medical Center, University of Freiburg, Freiburg, Germany. ✉email: enricomaria.surace@unina.it

DNA-protein recognition passing through chromatin status and interaction of TFs with nucleosomes<sup>36,37</sup>, eventually lead to rates of association and dissociation of DNA-binding proteins to and from DNA, which in turn determine cell-specific transcriptional activity<sup>38–40</sup>. Therefore, DNA-protein interactions ranging from low to high affinity<sup>41</sup> (the number and types of TFBS and concentrations of TFs<sup>42–44</sup> and types) account for gene expression from noise<sup>45</sup> to gene expression programs<sup>46–48</sup>.

The human *RHODOPSIN* (*RHO*) gene encodes for a G-protein-coupled receptor expressed exclusively in retinal rod photoreceptors and located in their outer segment (OS)<sup>49</sup>. *RHO* in the OS intercepts and converts light photons into electrical signals through the phototransduction cascade and, when mutated, causes retinitis pigmentosa (RP) and other blinding disorders<sup>50,51</sup>. The *Rho* gene is also tightly regulated, as *Rho* gene expression levels determines the size of the outer segment of the rods<sup>52–54</sup>, and both low and high expression levels can lead to photoreceptor degeneration<sup>55,56</sup>. *RHO* expression during development increases linearly and after birth becomes highly and constantly transcribed in rod photoreceptors<sup>57</sup>. Two conserved enhancers (*Rho Enhancer Region* (*RER*)<sup>58</sup> and *CRX-Bound Region 1* (*CBR*)<sup>59</sup>) and the *Rho* promoter contribute differently along pre-natal and post-natal development of rod progenitors<sup>60</sup>. However, genetic studies on murine suggest that in the adult retina *Rho* expression maintenance is strongly controlled by the adult-specific accessible regions at the proximal promoter<sup>60,61</sup>, accounting for 50% of its expression<sup>60</sup>.

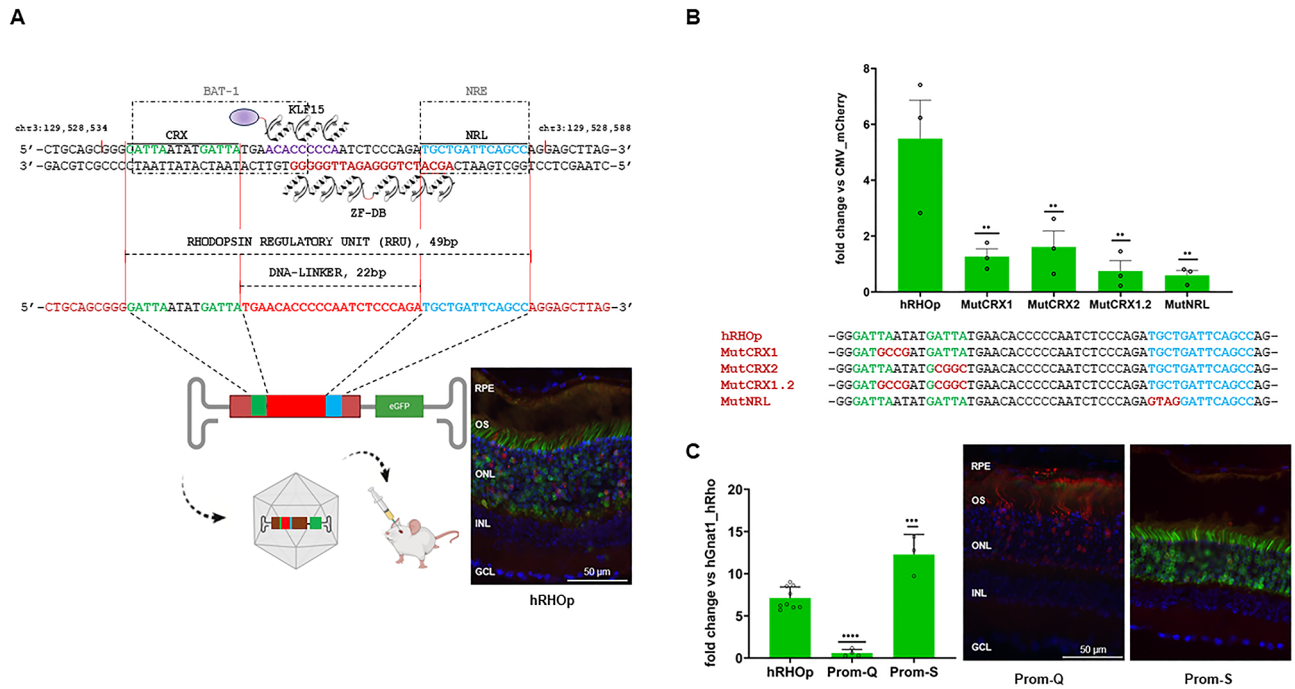
We previously have shown that, both a synthetic zinc finger protein (ZF-DB) and an ectopically expressed transcription factor (KLF15) block *Rhodopsin* (*Rho*) expression when delivered in vivo to rod photoreceptors of adult porcine retina by Adeno-associated virus (AAV) vectors<sup>62,63</sup>. ZF-DB and KLF15 bind 20- and 9-bp DNA sequences, partially overlapping (6 bp), between two key TFs for *RHO* expression, the Cone Rod Homeobox protein (CRX)<sup>64</sup> and the Neural Retina-specific Leucine zipper protein (NRL)<sup>65</sup> in the proximal promoter region of *RHO* (-82 to -62 and -84 to -75 from the transcription start site, TSS, respectively; (Fig. 1A)). ZF-DB and KLF15 repression of *Rho* caught our attention for two main aspects: (i) ZF-DB and KLF15 bind sensitively a DNase-unprotected<sup>66,67</sup> region at 20-fold lower concentrations than those of CRX and NRL, and selectively, as low number of transcriptional off targets are induced<sup>62,63</sup>; (ii) ZF-DB and KLF15 resulted in *Rho* repression despite their biochemical differences, ZF-DB possesses exclusively DNA-binding properties (without effector domains), KLF15 is a fully competent TF<sup>68</sup>, overall suggesting that a very compact DNA *cis*-regulatory sequence near to TSS controls *Rho* expression in terminally differentiated rods. Thus, ZF-DB and KLF15 within *Rho* native chromosomal context in vivo, upon binding to an accessible DNA sequence between CRX and NRL, are sufficient to overcome the transactivation effects of CRX and NRL and in turn block *Rho* expression. These observations led us to hypothesize that CRX and NRL TFs pair additionally to binding, interact with the specific DNA sequence interposed between them by a novel mechanism, which ultimately contributes to determining *Rho* expression levels.

## Results

### Forty-nine bp containing CRX and NRL BSs (*RHO* regulatory unit, RRU) control *RHO* expression in adult retina

ZF-DB and KLF15 target a sequence between two extensively validated *cis*-regulatory elements, BAT-1 and NRE<sup>59,66,69</sup>, which are bound by the retinal-specific TF, CRX<sup>64</sup> and by the rod photoreceptor-specific NRL<sup>65</sup>, respectively. CRX-NRL pair operate synergistically to transactivate numerous rod-specific genes, including *RHO*<sup>59,69</sup>. On this basis, we assessed whether the 49 bp long *cis*-regulatory elements encompassing the CRX and NRL binding sites and the interposed linker sequence bound by ZF-DB and KLF15 may represent a *cis*-regulatory unit (here named *RHO* regulatory unit, RRU) controlling *RHO* expression (Fig. 1A). To this end, we generated constructs for an in vivo Adeno-associated virus (AAV) vectors reporter assay (Fig. 1A). In vivo reporter assay<sup>70</sup> links variation of a DNA *cis*-regulatory sequence to the activity of a reporter gene. Here, gene expression variation output of the enhanced Green Fluorescent Protein reporter gene, eGFP, driven by *cis*-regulatory sequence (human *RHO* promoter fragment and multiple variants, see below) delivered by AAV was assessed into post-mitotic adult mouse retina<sup>71</sup> (Fig. 1A). To dissect the sequence-specific-function relationships derived by variation of human *RHO* promoter fragment in object, we selected conventional, one-at-a-time, one AAV-promoter variant per eye<sup>72</sup>. Relevant for the present study, and diversely from in vivo electroporation<sup>70,73</sup>, AAV vectors efficiently transduce post-mitotic adult photoreceptors<sup>71</sup>. AAV vectors delivered DNA to the target tissue, including in the retina<sup>71</sup>, remains almost exclusively episomal<sup>71,74,75</sup>, thereby isolating *cis*-regulatory activity from the genome context<sup>76,77</sup>. We cloned a human *RHO* proximal promoter fragment of 259 bp (hRHOP: -164 bp from TSS and 95 bp of 5'UTR), containing the 49 bp of the RRU, to drive the expression of the eGFP (Fig. 1A). To transduce photoreceptors, we used the extensively validated subretinal delivery of AAV serotype 8<sup>72</sup> at a dose of  $1 \times 10^9$  (AAV8 genome copies, GC) in adult C57BL/6 mice (4 weeks old). To minimize the potential variability in the dose of the vector delivered following subretinal administration, we co-injected with the test AAV8 vectors a low dose of  $1 \times 10^8$  GC of an AAV8 vector encoding for the red fluorophore mCherry and AAV8 vector encoding for the human *RHO* (hRHOP) under the control of the photoreceptor specific promoter fragment of G protein subunit alpha transducin 1 gene (mCherry; AAV8-hGNAT1-mCherry; GNAT1; AAV8-hGNAT1-hRHOP;  $1 \times 10^8$  GC). Fifteen days after subretinal administration, eyes were harvested, and retinas were processed to measure eGFP levels by RT-PCR or immunofluorescence analysis to assess eGFP expression pattern (Fig. 1). The relative abundance (ratio between transcript levels of eGFP and mCherry or hRHOP) of eGFP was evident and yielded similar values between mCherry or hRHOP (data not shown; we chose to use AAV8-hGNAT1-hRHOP as the sole normalizer in the following experiments, paragraphs below).

To dissect the contribution of DNA sequence determinants of RRU, we first mutated the CRX and NRL binding sites. AAV8-hRHOP-eGFP carrying individual mutations in mCRX1, mCRX2 or NRL binding sites showed that each binding site is required to sustained eGFP reporter expression (Fig. 1B). Next to assess whether RRU is necessary for *RHO* expression, we either deleted (Prom-Q) or duplicated (Prom-S) the RRU (Fig. 1C).

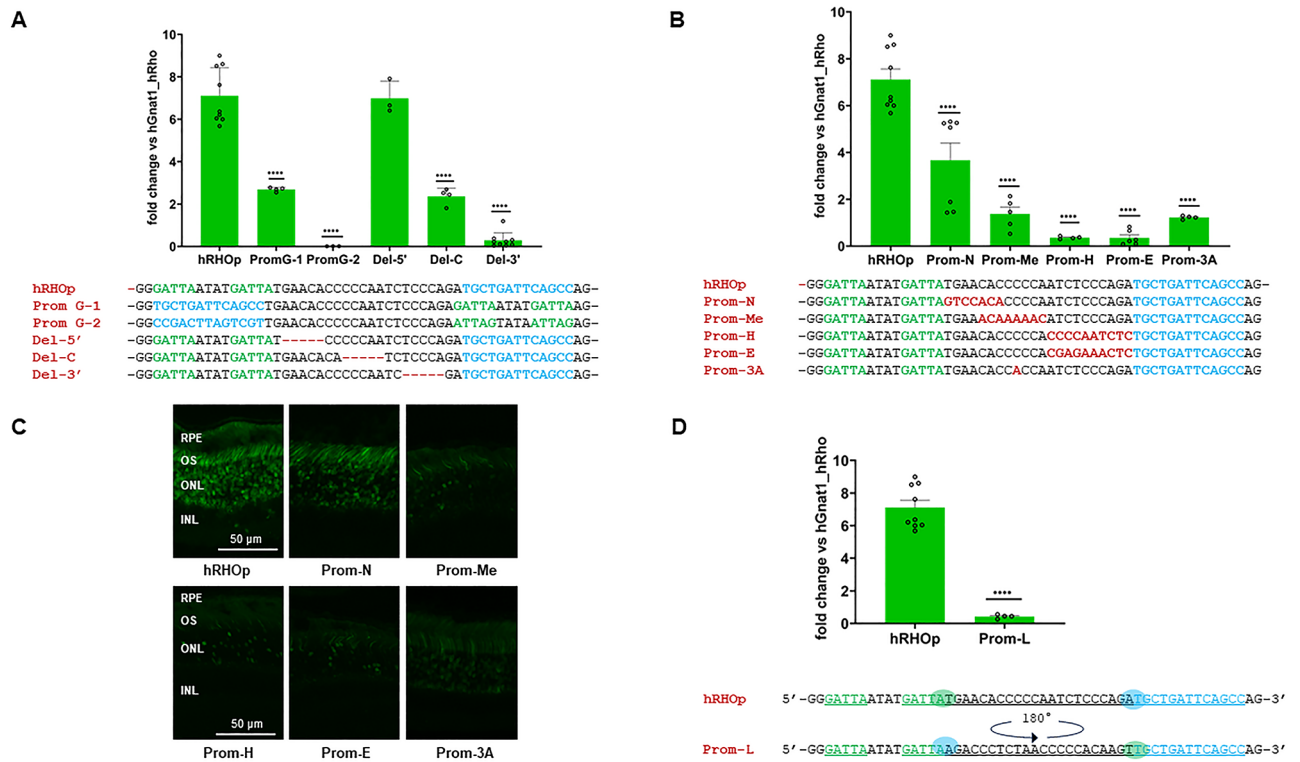


**Fig. 1.** Forty-nine bp containing CRX and NRL BSs control RHO expression in adult retina. **(A)** Proximal promoter region of the human *RHO* containing the Rho Regulatory Unit (hRRU, chr3:129,528,537 – 129,528,585). hRRU is composed of the BAT-1 and NRE elements (dashed grey boxes), CRX and NRL TFs binding sites (in green and light blue, respectively) and DNA-linker (in red). The “footprints” onto hRHOp of the synthetic zinc finger protein (drawing of the six zinc-finger array of ZF-DB below the negative strand, amaranth) and the TF KLF15 (drawing of KLF15 including the effector domain in violet above the positive strand). Below, schematic representation of gene reporter assay with Adeno-associated virus (AAV8) vector carrying eGFP under the control of the RHO proximal promoter (hRHOp) injected sub-retina in adult mice (4 weeks old), on the right-handed panel immunofluorescence analysis of a cross section of a retina transduced with AAV8-hRHOp-eGFP ( $3 \times 10^9$  genome copies, GC), co-injected with AAV8-hGNAT1-mCherry ( $3 \times 10^8$  GC) 15 days post-injection. **(B)** Histogram, quantitative PCR analysis (qPCR), eGFP expression levels 15 days after subretinal administration of AAV8-eGFP ( $3 \times 10^9$  GC) co-delivered with AAV8-hGNAT1-mCherry ( $3 \times 10^8$  GC) to account variability associated with subretinal AAV8 vector delivery (eGFP expression levels were normalised to those of mCherry, Methods; one promoter variant per eye, three replicates). **(C)** qPCR upon removal of hRRU (Prom-Q) and doubling the hRRU (Prom-S), and immunofluorescence analysis on Prom-Q, and Prom-S co-injected with a AAV8 vector encoding for mCherry (eGFP expression levels were normalised to those of mCherry, Methods; three individual retinal replicates for each construct). RPE, retinal pigment epithelium; OS, outer segment; ONL, outer nuclear layer; INL, inner nuclear layer; GCL, ganglion cell layer. Statistical significance was computed using the One-Way Anova test ( $p$ -values  $\leq 0.05$ ), and post-hoc Dunnett’s multiple comparisons test ( $p$ -values  $\leq 0.05$ ).

AAV8-Prom-Q abolished reporter expression levels, whereas AAV8-Prom-S duplication almost doubled eGFP expression. These results suggest that RRU (a 49 bp long RHO compact *cis*-regulatory element containing CRX and NRL binding sites and the 22 bp long interposed linker sequence) contains key features necessary to control both levels and expression pattern restricted to rod photoreceptors expression in adult mice.

### Changes of the DNA sequence interposed between CRX and NRL lead to differential reporter gene expression

To dissect the DNA sequence determinants within RRU, we restricted the analysis to the 22 bp long DNA sequence intervening between CRX and NRL, from now on termed the DNA-linker (Fig. 1A), and its relationship with the flanking CRX and NRL binding sites. To do so, we generated constructs carrying changes in CRX and NRL binding sites order, orientation, and the spacing between them (the length of the DNA-linker), referred to as the *cis*-regulatory syntax<sup>39</sup>, and tested them *in vivo* with AAV reporter assay. Mutual changes of CRX and NRL binding sites order (AAV8-Prom G-1), and concurrent change of order and orientation (AAV8-Prom G-2) resulted in less than half, and abolished reporter expression levels, respectively, relative to control (Fig. 2A). To manipulate the spacing between CRX and NRL binding sites, we generated three distinct non overlapping deletions of 5 bp, Del-5', Del-C and Del-3', in their respective AAV-reporter vectors, within CRX and NRL DNA-linker. These deletions entail a half DNA helical turn (10.5 bp represents an entire DNA turn) of the CRX and NRL binding sites, arranging them on the opposite sides relative to native configuration. *In vivo* reporter gene expression varied in these three constructs from null levels (AAV8-Del-3') to levels equivalent (AAV8-Del-5') to



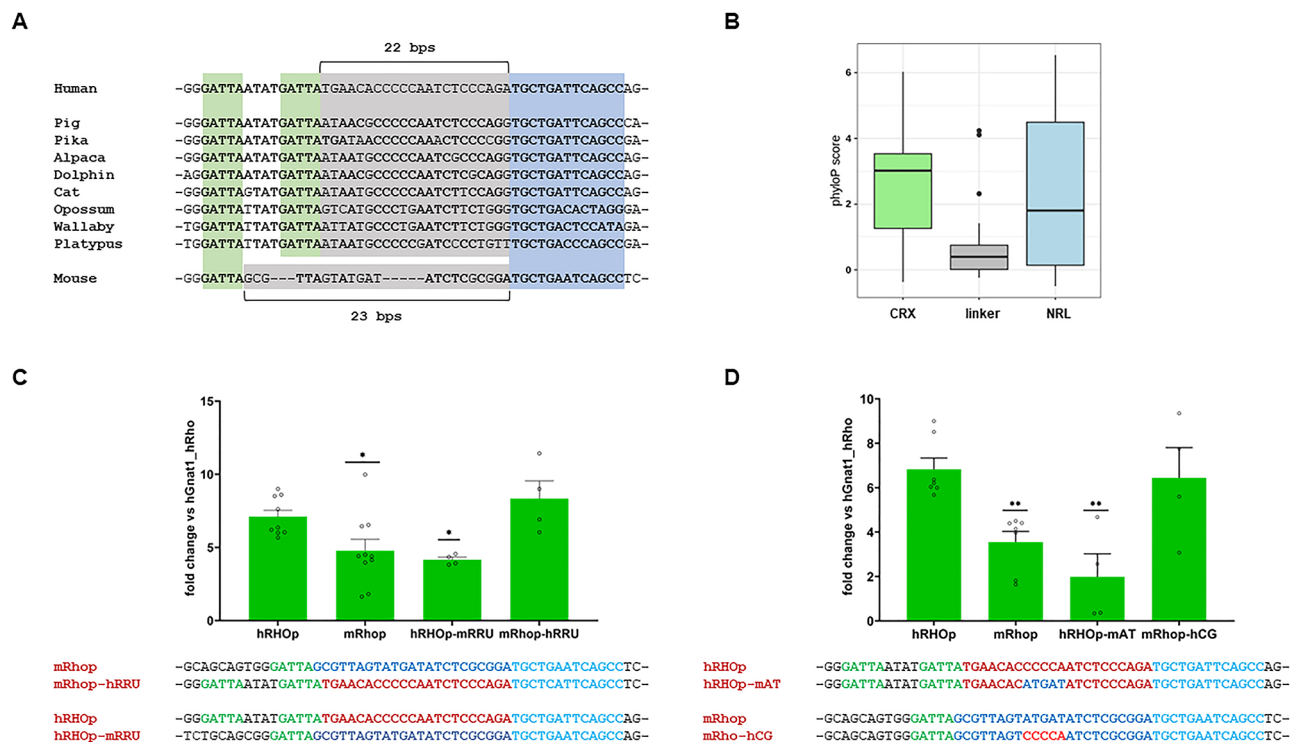
**Fig. 2.** Changes of the DNA sequence interposed between CRX and NRL lead to uncorrelated reporter gene expression. (A), Sequences, and histogram, qPCR analysis, eGFP expression levels 15 days after AAV8-eGFP in vivo retinal delivery in adult mice (Methods), showing the impact of change of the order, orientation, and spacing, of CRX and NRL BSs. (B), Sequences, and qPCR of nucleotide substitutions in the DNA-linker (C), Immunofluorescence analysis on AAV8-hRHOp, AAV8-Prom-N, AAV8-Prom-Me, AAV8-Prom-H, AAV8-Prom-E and AAV8-Prom-3A. RPE, retinal pigment epithelium; OS, outer segment; ONL, outer nuclear layer; INL, inner nuclear layer; GCL, ganglion cell layer. (D), Quantitative effects (qPCR) of DNA-linker sequence cloned in the 5' and 3' orientation (sequence, AAV8-Prom-L). Statistical significance was computed using the One-Way Anova test ( $p$ -values  $\leq 0.05$ ), and post-hoc Dunnett's multiple comparisons test ( $p$ -values  $\leq 0.05$ ). Only in Fig. 2D we performed t-student test ( $p$ -values  $\leq 0.0001$ ).

the native sequence (Fig. 2A). The fact that the same reduction of spacing resulted in differential eGFP expression levels, with an apparent decrease in eGFP expression levels in the 5' to 3' direction of the DNA-linker, points that the source of the variation is contained into CRX and NRL linking sequence; thus, we next mutated its DNA base sequence. We generated both contiguous (7 to 10 bp) and a single base substitution. These changes resulted in diversified levels of reporter gene expression (Fig. 2B). Compared with deletions, in this case only the first 5' mutagenesis of AAV8-Prom-N retains some of the activity, while a few bases away from the 3' inner flank of the DNA-linker (AAV8-Prom-Me and AAV8-Prom-3a) reporter expression begins to be greatly reduced. Thus, deletions and base substitutions within the DNA-linker lead to highly differentiated gene expression (Fig. 2B, C). This suggest that DNA base content and DNA-linker length are key features for the function of RRU. However, these effects may be due to the generation of *de novo* binding sites for either activators or repressors. To rule out this possibility we scanned each sequence for the respective BSs from the *HOCOMOCO* human collection of TFs (see Methods)<sup>78</sup>. The NRL and CRX BSs were consistently retained in the flanks of the DNA-linker in all hRHOp artificial promoter sequences tested, thus, compatible with previously described occupancy on the human RHO promoter<sup>67</sup>. On the contrary, in the DNA-linkers of both wild-type and synthetic promoters we were not able to identify any match with TFs expressed in the human retina (Supplementary Table 2). Thus, under these experimental conditions, equal spacing between CRX and NRL corresponds to a transcriptional output of RHO dictated by the DNA spacer sequence composition between them. To further rule out the possibility of generating novel BSs we inverted the 5' and 3' orientation (Prom-L; Fig. 2D). This arrangement affects only the 3' and the 5' flanks of CRX and NRL binding sites, respectively (Fig. 2D), while ensuring both the conservation of the DNA-linker sequence and its plus-minus DNA strands configuration relative to CRX and NRL BSs. Notably, in this arrangement the reporter expression levels were strongly reduced (Fig. 2D). Finally, immunofluorescence analysis showed that these changes of levels of reporter expression across different mutants corresponded to a variation of the intensity of eGFP expression and maintenance of the same eGFP pattern, restricted to rod photoreceptors (Figs. 1A and C and 2C). Collectively, these results suggest that without the apparent intervening action of other retina-specific TFs, the DNA linker connecting CRX and NRL controls the levels but not the pattern of RHO expression. Thus, the source of variation in RHO reporter expression depends on the syntax of

CRX and NRL BSs (order, orientation, and distance of TFBSs), and in addition on the base composition of the DNA-linker sequence interposed between CRX and NRL (Prom-L).

### Human and murine RRUs diverge but lead to correlated reporter gene expression

Next, we compared the DNA sequence features of RRU along the mammalian phylogeny. The alignment of 100 mammalian genomes shows that, apart from rodents, the two CRX and the NRL binding sites are conserved (Fig. 3A, B and Supplementary Table 2). Another noteworthy feature is the conserved spacing of 22 nucleotides of the DNA-linker. However, alternative nucleotide changes are present in several mammals, although not completely evenly distributed (the central sequence is more conserved than the 5' and 3' regions), including the internal bases of the flanking core CRX and NRL binding sites (Fig. 3A). Therefore, given conserved spacing, nucleotide sequence varies across mammalian phylogeny (Supplementary Fig. 3). Unlike other mammals, in rodents, and particularly in mouse and rat, RRU has one CRX binding site, and the DNA-linker distance has one more nucleotide (23 bp), while nucleotide sequence of the DNA-linker diverges extensively (Fig. 3A). Thus, we compared the human and murine homolog sequence. To study murine RRU (mRRU), we generated AAV reporter constructs with the proximal murine rhodopsin promoter (mRhop, -164 bp from the TSS and -78 bp from the 5'UTR). The murine promoter resulted in slightly lower reporter gene expression than the human counterpart (Fig. 3C). However, reciprocal replacement of human RRU with murine RRU in human RHO promoter (hRHOp) and of murine RRU with human RRU in mouseRho promoter restored the species-specific reporter expression levels. This suggests that despite the divergences between the DNA-linker, human and murine RRUs are functionally equivalent and relatively independent from both the embedded motif and the background sequences<sup>17</sup>. Furthermore, between humans and mice, nucleotide divergence in the DNA-linker corresponded to similar levels of reporter expression, supporting that, unlike synthetic constructs, a functional correlation may be obtained via an evolution-driven nucleotide arrangement. To further explore and isolate the properties of the human and murine DNA-linkers we swapped the 5 bases in the middle of them. These sequences are highly divergent, GC-rich in human and AT-rich in mouse (Fig. 3D). In addition, these bases are centered in the DNase-unprotected sequence<sup>66,67</sup>. AAV constructs containing the humanRHO promoter



**Fig. 3.** Human and murine DNA-linkers diverge but lead to species-specific RHO reporter expression. **(A)**, Multiple sequence alignment for representative mammalian species from UCSC track. The CRX and NRL binding sites are highlighted in green and blue, respectively. The DNA-linker is highlighted in grey. Conserved bases are indicated in bold. **(B)**, Distribution of PhyloP score of DNA bases in the NRL (green box) and CRX binding sites (blue box) and in the DNA-linker of the human RRU. PhyloP score indicates the degree of conservation of a DNA base in vertebrates. **(C)**, Sequences and qPCR analysis of mutual changes of hRRU (AAV8-mRhop-hRRU) and mRRU (AAV8-hRHOp-mRRU) insertion in murine and human Rho/RHO promoters, respectively. **(D)**, Sequence and qPCR analysis of the human RHO promoter with 5 bp of central murine DNA-linker (ATGAT, AAV8-hRHOp-mAT) and mouse Rho promoter with 5 bp of central human DNA-linker (CCCCA, AAV8-mRhop-hCG). Statistical significance was computed using the One-Way Anova test ( $p$ -values  $\leq 0.05$ ), and post-hoc Dunnett's multiple comparisons test ( $p$ -values  $\leq 0.05$ ).

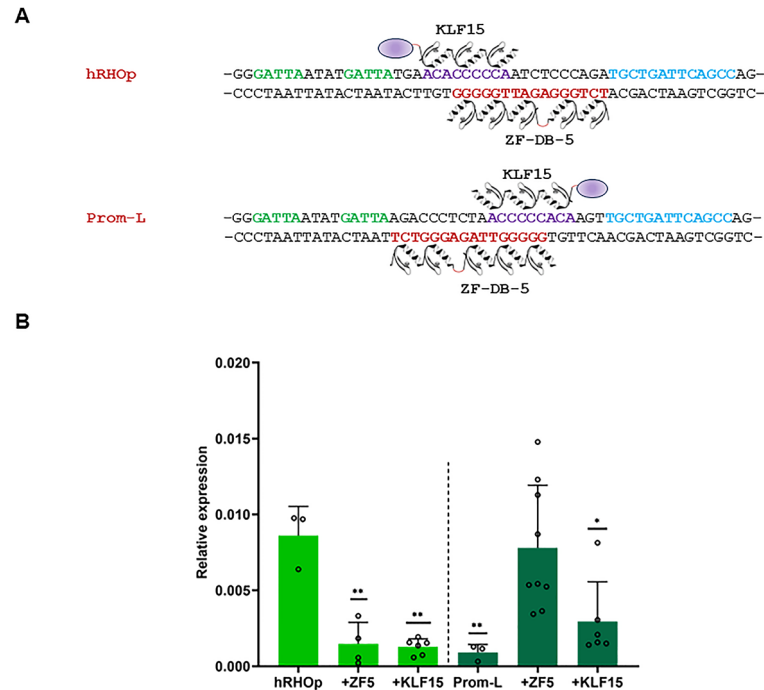
carrying the mouse 5 bp element (ATGAT, AAV8-hRHOp-mAT) and the mouseRho promoter carrying the human 5 bp element (CCCCA, AAV8-mRhop-hCG), showed a precise inversion of the expression levels. Indeed, the human RHOpromoter with the mouse sequence (hRHOp-mAT) equals mouseRho promoter levels, whereas the murine Rhopromoter (mRhop-hCG) equals the human RHOpromoter. This suggests that an even shorter sequence within the DNA-linkers is responsible for the species-specific *Rho* expression levels.

### Orthogonal DNA-linker binding proteins perturb reporter gene expression

To determine whether these *cis*-regulatory sequence properties are retained in isolation, in an unrelated *trans*-regulatory system in which retina-specific genes are not expressed, including CRX and NRL, we co-transfected HEK293 cell line with the same plasmids used to generate AAV vectors and those encoding for CRX and NRL (Methods). hRHOp showed synergistic activation upon co-transfection with CRX and NRL and similar gene expression variation as observed *in vivo* on several constructs (Supplementary Fig. 1B and Fig. 1C and Supplementary Table 1). Next, we interfered the DNA-linker activity with the DNA-linker binding proteins which we have previously used to repress *RHO* expression *in vivo*<sup>62,63</sup>. Since the first of the six fingers of the ZF-DB protein intrudes into the NRL-binding site (Fig. 1A), we removed it obtaining ZF-DB-5 (Fig. 4A). This ZF-DB-5 mutant, which has shown similar *RHO* repression properties as those of the ZF-DB, in *RHO*-P347S mice retina *in vivo* (Supplementary Fig. 2)<sup>62</sup>, when transfected with hRHOp *in vitro*, repressed *RHO* reporter expression (Fig. 4). Next, to get insights into the potential mechanisms underlying the action of the DNA-linker, we speculated that by interfering DNA-linker properties with the orthogonal proteins ZF-DB-5 and KLF15 we could affect the presumptive activity passing through the DNA-linker. To this end we tested whether the DNA-linker (Prom-L) which carries the DNA-linker sequence cloned in the 3'-5' direction, thus maintaining the same binding sites for ZF-DB-5 and KLF15 (Fig. 4A) could interfere its gene expression output. Surprisingly, Prom-L, which has shown to strongly reduce *RHO* reporter activity when co-transfected with ZF-DB-5 and KLF15 (Fig. 4B light green), promoted a transition to transactivation (Fig. 4B dark green). Notably, ZF-DB-5 without a transactivation domain increased the expression of the reporter more than KLF15, which possesses a catalytic domain<sup>68</sup>. These results suggest that interfering the DNA-linker sequence with orthogonal proteins results in down- or up-regulation of gene expression, depending on the orientation of the DNA-linker sequence between CRX and NRL binding sites.

### Discussion

This study shows that DNA is itself a source of contribution to variation of gene expression levels of *RHO*. We found that a 49-bp-long sequence of the human *RHO* proximal promoter, containing binding sites (BSs) for CRX



**Fig. 4.** Orthogonal DNA-linker binding proteins perturb reporter gene expression *in vitro*. (**A**), Drawing of hRHOp and Prom-L promoters with the orthogonal protein ZF-DB-5 and KLF15 bound to the DNA-linker in the opposite direction while maintaining the same double strand side. (**B**), Quantitative effects (qPCR) of transfecting ZF-DB-5 or KLF15<sup>62,63</sup>, and Fig. 4A, in HEK-293 cell line with hRHOp (light green, left hand side, and panel B) or Prom-L (DNA-linker sequence cloned in the 3'-5' direction, dark green, right hand side, and panel B). Statistical significance was computed using the One-Way Anova test ( $p$ -values  $\leq 0.05$ ), and post-hoc Dunnett's multiple comparisons test ( $p$ -values  $\leq 0.05$ ).

and NRL, is a functional discrete *cis*-regulatory unit (*RHO* regulatory unit, RRU) that controls *Rho* expression levels in the adult rod photoreceptors of the retina. This result is in accordance with previous observations which have shown that short proximal sequences of *RHO* promoter enable *Rho* expression *in vivo* in adult mouse rod photoreceptors<sup>60,61,79</sup>. Whereas, in adult retina, *Rho* enhancers account for a relative minor role in maintenance of *Rho* expression compared to *Rho* proximal promoter, as recently genetically determined<sup>60</sup>. These results complement those obtained *in vivo* with two transcriptional repressors, ZF-DB or KLF15 in both humanized P347S mouse (carrying 4 kb of the promoter of human *RHO*, including the DNA-linker element)<sup>62,63</sup> and pig post-mitotic retina<sup>62,63</sup> and along murine retinal development upon their germ-line or somatic expression<sup>82</sup>. Expression in post-mitotic rod photoreceptors of these two trans-acting proteins (in both absence or presence of an effector domain<sup>68</sup> ZF-DB, KLF15, respectively) upon binding, independently, to a partially overlapping 23 bp of the DNA-linker result in robust *RHO*-specific repressive action (~50%)<sup>62,63</sup>. Strongly supporting that: (i) the proximal promoter elements of *Rho* play a central role in *Rho* expression in the adult retina, (ii) more specifically these key elements are contained in the compact RRU (Fig. 1A).

This “strong self-functioning promoter”<sup>60</sup> allowed us to intercept a potential additional property of DNA-protein interaction. Indeed, here we found that within RRU, the base specific composition and length (22 bp) of the DNA-linker, contribute to the *RHO* expression levels. We first found that the shortening (5 bp) of three different consecutive regions of the DNA-linker, which leads to a different DNA sequence between the CRX and NRL binding sites (Fig. 2A), caused a differential eGFP expression in the retina, *in vivo* (Supplementary Fig. 1A and Supplementary Table 1). Similarly, mutagenesis of the DNA-linker resulted in highly differentiated eGFP expression, consistent with that observed by others in the homologous sequence of the murine *Rho* promoter<sup>83</sup>. Notably, the sequence derived by both these shortened and mutagenized DNA-linkers did not generate novel binding sites for TFs expressed in the human retina. These results were also confirmed by immunofluorescence analysis, which showed that the expression of eGFP was found invariably restricted to photoreceptors, as the native sequence, albeit with different intensities, thus suggesting that the mutated linkers do not allow binding of additional TFs. Complementarily, when guided by evolution, we found that the relative optimal expression levels of the human and murine *RHO* proximal promoters were determined by the specific composition and length of the DNA sequence of their respective DNA-linkers. Human and murine *RHO* proximal promoters show a slightly different reporter expression levels in favour of the human promoter, in our experimental setting. Notably, swapping their respective RRUs translated functionally in an inversion of promoter strengths. Furthermore, we showed that the exchange of the 5 bp divergent sequences between the human and mouse, in the middle of their DNA-linkers, enables an exact species-specific segregation of their respective *RHO* promoter strength (Fig. 3D). Therefore, in contrast to shortened and mutagenized DNA-linkers, these naturally occurring divergent human and murine DNA-linkers promoted *Rho* gene expression in a correlated manner. These results suggest that a DNA sequence-dependent feature impacts *Rho* expression, and that a precise base DNA sequence content and length (22 in humans and 23 bp in mice) ensures adequate quantitative gene expression levels of human and mouse *Rho*. Considering that the size of the DNA sequence space grows exponentially with sequence length ( $4^l$  for DNA, where  $l$  is the sequence length), the possible combinations for 22 bp are exceptionally high. However, the phylogenetic analysis has shown that human and mouse share in their respective DNA-linkers both sparse bases and almost complete conservation of the length of the DNA-linkers (Fig. 3 and Supplementary Fig. 3). These aspects suggest that the DNA-linker code is evolutionarily constrained, and that evolution may have created the respective DNA-linkers after separation from the common ancestor to meet physiological ranges of *Rho* expression relevant for the two species. Interestingly, it has been suggested that *Rho* gene expression levels determine the size of the outer segment of photoreceptors<sup>53,54</sup>. What we observed suggests that the human DNA-linker provides higher transcriptional output than the murine DNA-linker. Considering that human rod photoreceptors are larger than murine<sup>84</sup>, it will be interesting to determine whether these species-specific size differences are due to the contribution of the two diversified DNA-linkers.

We additionally showed that DNA-linker-derived variation is interconnected with the binding sites of CRX and NRL. Single BS mutations of CRX and NRL abolish *RHO* promoter activity (Fig. 1B). In addition, changing their order and orientation led to a profound change in reporter expression activity (Fig. 2A). Although the BSs of CRX and NRL (as well as the DNA-linker itself) are not palindromic and can have four orientations relative to each other, they show complete conservation of both the sequence of the BSs and their orientation, supporting that these features are also evolutionarily constrained. Similarly, although the linker DNA sequence conservation in mammals is poor, it is nevertheless evident that the DNA-linker orientation is also constrained. In fact, changing its orientation with respect to the CRX and NRL BSs, by reversing the 5' – 3' axis (Prom-L), greatly reduced the expression of the *RHO* reporter. However, this configuration modifies the 3' and 5' flanks of CRX and NRL, respectively, which, in turn may affect CRX and NRL binding. Indeed, the sequences flanking TFBSs are important for cooperative binding of TFs<sup>14,16,26</sup>. Nonetheless, (i), the binding preferences of TF pairs are typically limited to a few bases flanking the TFBSs; (ii), the 3' and 5' flanks of the DNA-linker are variable along the mammalian phylogeny (Fig. 3A), and (iii), we observed that changes along the entire 22 bp length of the DNA-linker induce a wide variation in reporter gene expression. Thus, these aspects seem to be too broad both qualitatively and quantitatively to justify that what is observed here might be determined by the binding preferences of TF pairs alone, even considering mechanisms concerning the mutual influence of TF pairs that are exerted at greater distance<sup>85</sup>. As opposed to short-mid range binding effects, the DNA-linker length appears too short to allow loop formation between CRX and NRL. Thus, we favour a model in which the DNA-linker is coupled in a continuum to its flanking CRX and NRL BSs.

For there to exist a mechanism of gene expression due to the contribution of the DNA sequence contained between two TFs, it should be possible to admit (and observe) the existence of an additional component, independent of the DNA-protein interaction by binding affinity, which is the accepted core mechanism from which gene expression originates and is implemented<sup>41</sup>. The ability of both ZF-DB and KLF15 to block *Rho* expression

in vivo at 20-fold lower concentrations than CRX and NRL<sup>62,63</sup> and the unprotection from DNase of the DNA-linker sequence<sup>66,67</sup> open the possibility that the DNA-linker, being not bound, may act subsequently of TF binding, and thus, by a DNA binding affinity-independent mechanism. Thus, in this view, our model admits that to the initial TFs binding of CRX and NRL occurring in affinity-dependent manner, a second event intervenes consequently (post-binding) as a DNA sequence-dependent intramolecular mechanism, acting as an additional component to the activity of the two TFs CRX and NRL. Furthermore, the same orthogonal *trans*-acting factors, ZF-DB-5 and KLF15, which block *Rho* expression (above, Fig. 1A)<sup>62,63</sup> when bind the DNA-linker sequence arranged in the opposite direction (orientation 3'-5', Prom-L, Fig. 4), which severely impairs reporter expression, surprisingly resulted in reporter gene activation. Strikingly, this activation is more pronounced when the DNA-linker is bound by ZF-DB-5, which lacks catalytic domains<sup>68</sup>. Thus, rather than a protein-protein steric interference, or a direct protein-protein interaction with CRX and NRL, which appears unlikely because of their biochemically distinct properties and their binding arrangement (Fig. 4), may further support that ZF-DB-5 and KLF15 modify an activity passing through an exact DNA bases composition and length coupled to the action of two TFs.

The results obtained with the AAV reporter assay decouple *cis*-regulatory DNA sequences from their native context by being episomal, thus allowing their function to be interrogated in isolation<sup>76,77</sup>. On the other hand, however, this method cannot fully account for the effects of DNA-linker sequence variation at the locus, including considering the impact of copy number variations of AAV episomes, their unequal distribution within photoreceptors, and the potential interference effects resulting from the use of the eGFP reporter. It will therefore be necessary to design species-specific exchanges by knocking-in the DNA-linker sequence at the promoter locus and to search for possible inter- and intra-specific variants to establish the net contribution of DNA-linker sequence composition and length in *Rho* gene expression in health and disease. To uncover the basis of the biophysical nature of the DNA-TF interaction, beyond CRX-NRL DNA binding interactions<sup>86,87</sup>, by including and resolving the ternary CRX-NRL-DNA-linker complexes structure, determine the dynamics of TF-binding events<sup>87,88</sup>, resolve at single-nucleotide resolution the fraction of the DNA-linker effectively bound/unbound by proteins<sup>89</sup> and its chromatin profiling<sup>90</sup>, to finally decouple the relative contribution of the canonical binding-dependent mechanisms (binding affinity) from the binding-independent mechanism proposed here that explains *Rho* gene expression.

In conclusion, beyond pattern and timing, we propose that levels of gene expression are contributed by DNA itself via a binding-independent mechanism which adds to the TFs binding-dependent mechanisms (Supplementary Fig. 4). Once TFs are bound to TFBSs, by the specific DNA base sequence interposed between them allows the activity of TF-TFs to be integrated; as a result, the DNA base sequence and length constitute a new code that operates within a TF-DNA-TF complex, which thus represents a unit of activity, determining the appropriate final levels of gene expression. In this view, the role of DNA is thus threefold: it contains information that encodes for proteins, regulates their interactions (with a first feedback, binding to DNA), and participates in their own end action (with a second feedback activity) by becoming a phenotype.

The control of gene expression proposed here for RHO may have evolved, and function more generally, as a still-hidden property, adding to the DNA-binding-dependent mechanisms determinant of gene expression.

## Materials and methods

### Plasmid construction

The human rhodopsin proximal promoters (hRHOp, 164 bp from the transcription starting site (TSS) and 95bp of 5'UTR) and the mouse proximal promoter (mRHOp, 164bp from the transcriptional starting site (TSS) and 78bp of 5'UTR) were generated by gene synthesis of Eurofins MWG<sup>®</sup> and cloned in pAAV2.1-eGFP using NheI and NotI restriction enzymes. All variant promoters of hRHOp (Prom-3A, Prom-H, Prom-Me, Del-3', Del-C, Prom-E, Prom-Q, Prom-N, Del-5', Prom-S, hRHOp-mRRU, mRHOp-hRRU, PromG-1, PromG-2, Prom-L, MutCRX1, MutCRX2, MutCRX1.2, MutNRL) were generated by gene synthesis of Eurofins MWG<sup>®</sup> and cloned in pAAV2.1-eGFP using NheI and NotI restriction enzymes.

### AAV vector preparation and purification

AAV vectors were produced by the TIGEM AAV Vector Core by triple transfection of HEK293 cells followed by two rounds of CsCl<sub>2</sub> purification. For each viral preparation, physical titers [genome copies (GC)/mL] were determined by averaging the titer achieved by dot-blot analysis and by PCR quantification using TaqMan (Applied Biosystems, Carlsbad, CA, USA)<sup>91</sup>.

### Mice

All studies on mice were performed in accordance with the institutional guidelines for animal research (ARRIVE guidelines) and approved by the Italian Ministry of Health; Department of Public Health, Animal Health, Nutrition and Food Safety in accordance with the law on animal experimentation (article 7; D.L. 116/92; protocol number: 114/2015-PR). C57BL/6J mice (Charles Rivers Laboratories, Calco, Italy) and P347S+/+ animals<sup>80</sup>, were bred as follow: P347S-/- mice were crossed with C57BL/6J mice (Charles Rivers Laboratories) to obtain the P347S+/- mice<sup>80,81,62,92</sup>. These mouse lines were tested free of rd8 mutations<sup>92</sup>.

### Vector administration

Intraperitoneal injection of ketamine and medetomidine (100mg/kg and 0.25mg/kg respectively), then AAV vectors were delivered sub-retinal via a trans-scleral transchoroidal approach. Once anesthetized, the pupils are dilated with 0.5% tropicamide. A conjunctival incision is made and used to grasp and rotate the ocular globe. Subsequently, a sclerotomy is made by penetrating the sclera with a 27-gauge needle. A 33-gauge needle connected to a Hamilton syringe is inserted through the sclerotomy 2 mm posterior to the temporal limbus. The



cannula is then advanced tangential to the curvature of the globe to the subretinal space in the posterior pole. Each eye was co-injected with two AAV8 vectors: AAV8 vector carrying AAV8-eGFP driven by human and murine proximal promoter variants (dose  $1 \times 10^9$  GC/ $\mu$ l) and an AAV8 control vector encoding for human *RHO* gene driven by the GNAT1 proximal promoter, or AAV8 control vector encoding for mCherry driven by the GNAT1 proximal promoter (dose  $1 \times 10^8$  GC/ $\mu$ l).

### qReal time PCR

RNAs from tissues were isolated using RNeasy Mini Kit (Qiagen), according to the manufacturer protocol. cDNA was amplified from 1  $\mu$ g isolated RNA using QuantiTect Reverse Transcription Kit (Qiagen), as indicated in the manufacturer instructions. The PCRs with cDNA were carried out in a total volume of 20  $\mu$ l, using 10  $\mu$ l LightCycler 480 SYBR Green I Master Mix (Roche) and 400nM primers under the following conditions: pre-Incubation, 50°C for 5 min, cycling: 45 cycles of 95°C for 10 s, 60°C for 20 s and 72°C for 20 s. Each sample was analyzed at least in triplicate. Transcript levels of eGFP, hRHO or mCherry were normalized against murine *Gapdh* and *Act $\beta$*  (DcT), and then the eGFP expression shown is obtained as fold change over control (co-injected vector: AAV8-hGNAT1-hRHO or AAV8-hGNAT1-mCherry). We used the following primers: *Act\_forward* (CAAGATCATTTGCTCCTCCTGA) and *Act\_reverse* (CATCGTACTCCTGCTTGCTGA), *Gapdh\_forward* (GTCCGTGTGAACGGATTTG) and *Gapdh\_reverse* (CAATGAAGGGGTCGTTGATG); *eGFP\_forward* (ACGTAAA CCGCCACAAGTTC) and *eGFP\_reverse* (AAGTCGTGCTGCTTCATGTG); *hRHO\_forward* (TCATGGTCC TAGGTGGCTTC) and *hRHO\_reverse* (GGAAGTTGCTCATGGGCTTA); *mCherry\_forward* (CACTACGAC GCTGAGGTCAA) and *mCherry\_reverse* (GTGGGAGGTGATGTCCAAC).

### Histological analysis

For morphological studies, after the mice were killed by cervical dislocation, the eyecups were harvested, fixed by immersion in 4% paraformaldehyde, and then embedded in OCT (KaltK). For each eye, 150 to 200 serial sections (10- $\mu$ m thick) were cut along the horizontal plane; the sections were progressively distributed on 10 glass slides so that each slide contained 15 to 20 sections representative of the whole eye at different levels. Slides were coverslipped with Vectashield containing DAPI (4',6-diamidino-2-phenylindole; Vector laboratories, Burlingame, CA, USA) to stain cells nuclei and retinal histology was analyzed a Leica Fluorescence Microscope System (Leica Microsystems GmbH, Wetzlar, Germany).

### Transient transfection and analysis of promoter activity by qReal Time PCR

HEK293 cells were plated in six-well plates at a density of  $1 \times 10^6$  cells/well in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin and grown for 24h at 37°C. The cells were transfected at 80% confluency using a TransIT-X2 Dynamic Delivery System (MIRUS), as indicated in the manufacturer instructions. We co-transfected the following constructs: 50 ng reported plasmid, 250ng of a plasmid encoding for human CRX, 125 ng of a plasmid expressing human NRL and 125ng CMV-ZF5 or CMV-KLF15 plasmid. The DNA amount was kept at 550 ng/well with pUC19. Cells were grown and harvested 48h post transfection in TRIzol™ Reagent (Life Technologies). The cDNA was amplified from 1  $\mu$ g isolated RNA using QuantiTect Reverse Transcription Kit (Qiagen), as indicated in the manufacturer instructions. The PCRs with cDNA were carried out in a total volume of 20  $\mu$ l, using 10  $\mu$ l LightCycler 480 SYBR Green I Master Mix (Roche), 400nM primers under the following conditions: pre-Incubation, 50°C for 5 min, cycling: 45 cycles of 95°C for 10 s, 60°C for 20 s and 72°C for 20 s. Each sample was analysed in triplicate. Transcript levels of cells were measured by real-time PCR using the LightCycler 480 (Roche) and the following primers: *PCR eGFP\_forward* (ACGTA AACGGCCACAAGTTC) and *eGFP\_reverse* (AAGTCGTGCTGCTTCATGTG), *GAPDH\_forward* (GAAGGT GAAGGTCCGAGT) and *GAPDH\_reverse* (GAAGATGGTGTGATGGGATTTTC). The eGFP expression shown is obtained normalizing to GAPDH.

### Statistical analyses

Data are presented as mean  $\pm$  Error bars indicate standard error mean (SEM). Statistical significance was computed using the One-Way Anova test (p-values  $\leq 0.05$ ), and post-hoc Dunnett's multiple comparisons test (p-values  $\leq 0.05$ ). Only in Fig. 2D we performed t-student test (p-values  $\leq 0.0001$ ).

### Motif Scan

TF human motifs were retrieved from HOCOMOCO (version 11)<sup>78</sup> and scanned on target sequences using FIMO (meme-suite 5.5) (Grant CE,2011), with background model derived from target sequence composition and parameters --norc and -thresh 0.01. Matches with positive scores and p-value $<0.01$  were considered. Alternative matches for a given TF on the same sequence were disambiguated by considering the match with the highest score.

Matches with positive scores and p-value $<0.01$  were considered. Alternative matches for a given TF on the same sequence were disambiguated by considering the match with the highest score.

### TF expression in retina

Transcript and protein expression data were retrieved from the Human Protein Atlas version 23.0 (<https://www.proteinatlas.org/>). We considered as positively expressed transcripts with normalized TPM  $>1$  in human retina (n=13.114), and proteins with positive detection in photoreceptor cells (n=79).

### Electrophysiological testing

The method is as described<sup>93</sup>. Briefly, mice were dark reared for 3 hr and anesthetized. Flash electroretinograms (ERGs) were evoked by light flashes generated through a Ganzfeld stimulator (CSO, Costruzione Strumenti

Oftalmici, Florence, Italy). ERG analysis in scotopic conditions were evoked by 11 stimuli (from  $-4$  to  $+1.3$  log  $\text{cd}^*/\text{m}^2$ ) with an interval of 0.6 log unit. and registered as previously described.

## Data availability

“Sequence data that support the findings of this study have been deposited at GenBank with the primary accession codes: BankIt2849745 hRHOp PQ037961BankIt2849745 Prom\_G-1 PQ037962BankIt2849745 Prom\_G-2 PQ037963BankIt2849745 Del-5' PQ037964BankIt2849745 Del-C PQ037965BankIt2849745 Del-3' PQ037966BankIt2849745 Prom\_L PQ037967BankIt2849745 Prom\_N PQ037968BankIt2849745 Prom-Me PQ037969BankIt2849745 PromH PQ037970BankIt2849745 Prom\_E PQ037971BankIt2849745 Prom-3A PQ-037972BankIt2849745 Prom\_Q PQ037973BankIt2849745 Prom\_S PQ037974BankIt2849745 hRHOp\_mRRU PQ037975BankIt2849745 MutCRX1 PQ037976BankIt2849745 MutCRX2 PQ037977BankIt2849745 MutCRX1.2 PQ037978BankIt2849745 MutNRL PQ037979BankIt2849745 mRhop PQ-037980BankIt2849745 mRhop\_hRRU PQ037981BankIt2849745 hRHO-mAT PQ037982BankIt2849745 mRho-hCG PQ037983.

Received: 5 July 2024; Accepted: 15 October 2024

Published online: 01 November 2024

## References

- Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
- Gilbert, W. Müller-Hill, B. The lac operator is DNA. *Proc. Natl. Acad. Sci. U S A.* **58**, 2415–2421 (1967).
- Klug, A. Gene regulatory proteins and their interaction with DNA. *Ann. N Y Acad. Sci.* **758**, 143–160 (1995).
- Pabo, C. O. & Sauer, R. T. Transcription factors: Structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095 (1992).
- Rohs, R. et al. Origins of specificity in Protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269 (2010).
- Slatery, M. et al. Absence of a simple code: How transcription factors read the genome. *Trends Biochem. Sci.* **39**, 381–399 (2014).
- Garber, M. et al. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell.* **47**, 810–822 (2012).
- Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature.* **430**, 85–88 (2004).
- Carroll, S. B. *Endless Forms Cell.* **101**, 577–580 (2000).
- Signor, S. A. & Nuzhdin, S. V. The evolution of Gene expression in cis and trans. *Trends Genet.* **34**, 532–544 (2018).
- Vierstra, J. et al. Global reference mapping of human transcription factor footprints. *Nature.* **583**, 729–736 (2020).
- Sahu, B. et al. Sequence determinants of human gene regulatory elements. *Nat. Genet.* **54**, 283–294 (2022).
- Gordán, R., Hartemink, A. J. & Bulyk, M. L. Distinguishing direct versus indirect transcription factor–DNA interactions. *Genome Res.* **19**, 2090–2100 (2009).
- Horton, C. A. et al. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science.* **381**, eadd1250 (2023).
- Jolma, A. et al. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature.* **527**, 384–388 (2015).
- Kim, S. et al. DNA-guided transcription factor cooperativity shapes face and limb mesenchyme. *Cell.* **187**, 692–711e26 (2024).
- Sharon, E. et al. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30**, 521–530 (2012).
- Wilson, D. S., Guenther, B., Desplan, C. & Kuriyan, J. High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell.* **82**, 709–719 (1995).
- Kulkarni, M. M. & Arnosti, D. N. Information display by transcriptional enhancers. *Development.* **130**, 6569–6575 (2003).
- Arnosti, D. N. & Kulkarni, M. M. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.* **94**, 890–898 (2005).
- Rodda, D. J. et al. Transcriptional regulation of nanog by OCT4 and SOX2. *J. Biol. Chem.* **280**, 24731–24737 (2005).
- Thanos, D. & Maniatis, T. Virus induction of human IFN $\beta$  gene expression requires the assembly of an enhanceosome. *Cell.* **83**, 1091–1100 (1995).
- Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Curr. Opin. Genet. Dev.* **43**, 73–81 (2017).
- Giniger, E. & Ptashne, M. Cooperative DNA binding of the yeast transcriptional activator GAL4. *Proc. Natl. Acad. Sci. U S A.* **85**, 382–386 (1988).
- Mann, R. S. & Affolter, M. Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* **8**, 423–429 (1998).
- Gordán, R. et al. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell. Rep.* **3**, 1093–1104 (2013).
- Morgunova, E. & Taipale, J. Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.* **47**, 1–8 (2017).
- Bintu, L. et al. Transcriptional regulation by the numbers: Models. *Curr. Opin. Genet. Dev.* **15**, 116–124 (2005).
- Ptashne, M. Gene regulation by proteins acting nearby and at a distance. *Nature.* **322**, 697–701 (1986).
- Li, Q., Peterson, K. R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood.* **100**, 3077–3086 (2002).
- Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell. Biol.* **19**, 621–637 (2018).
- Hafner, A. & Boettiger, A. The spatial organization of transcriptional control. *Nat. Rev. Genet.* **24**, 53–68 (2023).
- Cramer, P. Organization and regulation of gene transcription. *Nature.* **573**, 45–54 (2019).
- Milo, R. & Phillips, R. Cell Biology by the numbers. *Garland Sci.* <https://doi.org/10.1201/9780429258770> (2015).
- Marklund, E. et al. Sequence specificity in DNA binding is mainly governed by association. *Science.* **375**, 442–445 (2022).
- Svaren, J., Klebanow, E., Sealy, L. & Chalkley, R. Analysis of the competition between nucleosome formation and transcription factor binding. *J. Biol. Chem.* **269**, 9335–9344 (1994).
- Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U S A.* **107**, 22534–22539 (2010).
- Crocker, J. et al. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell.* **160**, 191–203 (2015).
- Farley, E. K. et al. Suboptimization of developmental enhancers. *Science.* **350**, 325–328 (2015).
- Segal, E. & Widom, J. From DNA sequence to transcriptional behaviour: A quantitative approach. *Nat. Rev. Genet.* **10**, 443–456 (2009).
- de Boer, C. G. The continuum of transcription factor affinities. *Nat. Rev. Genet.* **25**, 378 (2024).

42. Kim, H. D. & O'Shea, E. K. A quantitative model of transcription factor-activated gene expression. *Nat. Struct. Mol. Biol.* **15**, 1192–1198 (2008).
43. Hammar, P. et al. The lac repressor displays facilitated diffusion in living cells. *Science*. **336**, 1595–1598 (2012).
44. Naqvi, S. et al. Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nat. Genet.* **55**, 841–851 (2023).
45. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science*. **304**, 1811–1814 (2004).
46. Carroll, S. B. Evo-Devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell*. **134**, 25–36 (2008).
47. Sorrells, T. R. & Johnson, A. D. Making sense of transcription networks. *Cell*. **161**, 714–723 (2015).
48. Crocker, J., Preger-Ben Noon, E. & Stern, D. L. The soft touch. in *Current Topics in Developmental Biology* vol. 117 455–469 (Elsevier, (2016).
49. Nathans, J. & Hogness, D. S. Isolation and nucleotide sequence of the gene encoding human rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4851–4855 (1984).
50. Dryja, T. & Li, T. Molecular genetics of retinitis pigmentosa. *Hum. Mol. Genet.* **4**, 1739–1743 (1995).
51. Athanasiou, D. et al. The molecular and cellular basis of rhodopsin retinitis pigmentosa reveals potential strategies for therapy. *Prog. Retin. Eye Res.* **62**, 1–23 (2018).
52. Olsson, J. E. et al. Transgenic mice with a rhodopsin mutation (Pro23His): A mouse model of autosomal dominant retinitis pigmentosa. *Neuron*. **9**, 815–830 (1992).
53. Price, B. A. et al. Rhodopsin gene expression determines rod outer segment size and Rod Cell Resistance to a Dominant-negative neurodegeneration mutant. *PLoS One*. **7**, e49889 (2012).
54. Makino, C. L. et al. Rhodopsin expression level affects rod outer segment morphology and photoresponse kinetics. *PLoS One*. **7**, e37832 (2012).
55. Tan, E. et al. The relationship between opsin overexpression and photoreceptor degeneration. *Invest. Ophthalmol. Vis. Sci.* **42**, 589–600 (2001).
56. Liang, Y. et al. Rhodopsin Signaling and Organization in heterozygote rhodopsin knockout mice. *J. Biol. Chem.* **279**, 48189–48196 (2004).
57. Treisman, J. E., Morabito, M. A. & Barnstable, C. J. Opsin expression in the rat retina is developmentally regulated by transcriptional activation. *Mol. Cell. Biol.* **8**, 1570–1579 (1988).
58. Nie, Z., Chen, S., Kumar, R. & Zack, D. J. RER, an evolutionarily conserved sequence upstream of the rhodopsin gene, has enhancer activity. *J. Biol. Chem.* **271**, 2667–2675 (1996).
59. Hsiau, T. H. C. et al. The cis-regulatory logic of the mammalian photoreceptor transcriptional network. *PLoS One*. **2**, e643 (2007).
60. Sun, C., Ruzycski, P. A. & Chen, S. Rho enhancers play unexpectedly minor roles in rhodopsin transcription and rod cell integrity. *Sci. Rep.* **13**, 12899 (2023).
61. Zack, D. J. et al. Unusual topography of bovine rhodopsin promoter-lacZ fusion gene expression in transgenic mouse retinas. *Neuron*. **6**, 187–199 (1991).
62. Botta, S. et al. Rhodopsin targeted transcriptional silencing by DNA-binding. *eLife*. **5**, e12242 (2016).
63. Botta, S. et al. Targeting and silencing of rhodopsin by ectopic expression of the transcription factor KLF15. *JCI Insight*. **2**, e96560 (2017).
64. Chen, S. et al. Crx, a novel otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes. *Neuron*. **19**, 1017–1030 (1997).
65. Rehemtulla, A. et al. The basic motif-leucine zipper transcription factor nrl can positively regulate rhodopsin gene expression. *Proc. Natl. Acad. Sci. U S A*. **93**, 191–195 (1996).
66. Desjardin, L. E. & Hauswirth, W. W. Developmentally important DNA elements within the bovine opsin upstream region. *Invest. Ophthalmol. Vis. Sci.* **37**, 154–165 (1996).
67. Cherry, T. J. et al. Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9001–9012 (2020).
68. Mas, C. et al. Structural and functional characterization of an atypical activation domain in erythroid Krüppel-like factor (EKLF). *Proc. Natl. Acad. Sci. U S A*. **108**, 10484–10489 (2011).
69. Mitton, K. P. et al. The leucine Zipper of NRL interacts with the CRX homeodomain. *J. Biol. Chem.* **275**, 29794–29799 (2000).
70. Matsuda, T. & Cepko, C. L. Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proc. Natl. Acad. Sci. U S A*. **101**, 16–22 (2004).
71. Surace, E. M. et al. Delivery of adeno-associated virus vectors to the fetal retina: Impact of viral capsid proteins on retinal neuronal progenitor transduction. *J. Virol.* **77**, 7957–7963 (2003).
72. Allocca, M. et al. Novel adeno-associated virus serotypes efficiently transduce murine photoreceptors. *J. Virol.* **81**, 11372–11380 (2007).
73. Montana, C. L., Myers, C. A. & Corbo, J. C. Quantifying the activity of cis-regulatory elements in the mouse retina by explant electroporation. *J. Vis. Exp.* **2821** <https://doi.org/10.3791/2821> (2011).
74. McCarty, D. M., Young, S. M. & Samulski, R. J. Integration of adeno-associated virus (AAV) and recombinant AAV vectors. *Annu. Rev. Genet.* **38**, 819–845 (2004).
75. Penaud-Budloo, M. et al. Adeno-associated virus vector genomes persist as episomal chromatin in primate muscle. *J. Virol.* **82**, 7875–7885 (2008).
76. Klein, J. C. et al. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat. Methods*. **17**, 1083–1091 (2020).
77. Maricque, B. B., Chaudhari, H. G. & Cohen, B. A. A massively parallel reporter assay dissects the influence of chromatin structure on cis-regulatory activity. *Nat. Biotechnol.* **37**, 90–95 (2019).
78. Kulakovskiy, I. V. et al. HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).
79. Flannery, J. G. et al. Efficient photoreceptor-targeted gene expression in vivo by recombinant adeno-associated virus. *Proc. Natl. Acad. Sci. U S A*. **94**, 6916–6921 (1997).
80. Li, T., Snyder, W. K., Olsson, J. E. & Dryja, T. P. Transgenic mice carrying the dominant rhodopsin mutation P347S: Evidence for defective vectorial transport of rhodopsin to the outer segments. *Proc. Natl. Acad. Sci. U S A*. **93**, 14176–14181 (1996).
81. Chadderton, N. et al. Improved retinal function in a mouse model of dominant retinitis pigmentosa following AAV-delivered gene therapy. *Mol. Ther.* **17**, 593–599 (2009).
82. Marrocco, E., Maritato, R., Botta, S., Esposito, M. & Surace, E. M. Challenging Safety and Efficacy of Retinal Gene therapies by Retinogenesis. *IJMS*. **22**, 5767 (2021).
83. Kwansieski, J. C., Mogno, I., Myers, C. A., Corbo, J. C. & Cohen, B. A. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl. Acad. Sci. U S A*. **109**, 19498–19503 (2012).
84. Fu, Y. & Yau, K. W. Phototransduction in mouse rods and cones. *Pflugers Arch. - Eur. J. Physiol.* **454**, 805–819 (2007).
85. Kim, S. et al. Probing Allostery through DNA Sci. **339**, 816–819 (2013).
86. Clanor, P. B. et al. Structural and functional analysis of the human cone-rod homeobox transcription factor. *Proteins*. **90**, 1584–1593 (2022).

87. Srivastava, D. et al. Molecular basis of CRX/DNA recognition and stoichiometry at the Ret4 response element. *Structure*. **S0969-2126(24)00242-9**<https://doi.org/10.1016/j.str.2024.07.004> (2024).
88. Zhuo, X. & Knox, B. E. Interaction of human CRX and NRL in live HEK293T cells measured using fluorescence resonance energy transfer (FRET). *Sci. Rep.* **12**, 6937 (2022).
89. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*. **147**, 1408–1419 (2011).
90. Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. **356**, eaaj2239 (2017).
91. Doria, M., Ferrara, A. & Auricchio, A. AAV2/8 vectors purified from culture medium with a simple and rapid protocol transduce murine liver, muscle, and retina efficiently. *Hum. Gene Ther. Methods*. **24**, 392–398 (2013).
92. Carrella, S. et al. miR-181a/b downregulation: A mutation-independent therapeutic approach for inherited retinal diseases. *EMBO Mol. Med.* **14**, e15941 (2022).
93. Surace, E. M. et al. Amelioration of both functional and morphological abnormalities in the retina of a mouse model of ocular albinism following AAV-mediated gene transfer. *Mol. Ther.* **12**, 652–658 (2005).

## Acknowledgements

We thank Fondazione Telethon and the Telethon Institute of Genetics and Medicine (TIGEM) and the TIGEM Vector Core for vector production. We thank Irene Cantone, Pasqualina Colella and Paolo Sordino for helpful discussions.

## Author contributions

EMS conceived the study. RM, MP, SB, EMS, designed the experiments and analyzed data. MS, EM, performed histological analysis. GD performed bioinformatic analysis. EM, EMS, MLB performed in vivo studies. RM, AM, ED, CM and EMS draft the manuscript. RM, AM, ED, MP, SB, EM, MR, performed the molecular biology experiments. EMS wrote the manuscript and supervised the study.

## Funding

This work was supported by the European Research Council/ERC grant 311682 “AlleleChoker”. (to EMS); ERC proof-of concept (ERC-POC). “Insight” 813223 (to EMS); Missione 4 - Center. for Gene Therapy and Drugs based on RNA Technology- WP3.5 Stroke and other brain diseases, WP 3.5.7 (to EMS); Prin MUR: 2022LB-ZSTR (to EMS).

## Competing interests

E.M.S., S.B. and E.M. are inventors on patents, “Artificial DNA-binding proteins and uses thereof”, WO2015075154A3, US20160289284A1. E.M.S., S.B. and E.M. are inventors on a pending patent “Ectopically expressed transcription factors and uses thereof”, PCT/EP2018/086782, Department of Translational Medical Sciences (DiSMET) University of Naples “Federico II”, Italy, which may encompass the findings. The remaining authors declare no competing interests.

## Ethical approval

All methods were carried out in accordance with relevant guidelines and regulations.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-76664-8>.

**Correspondence** and requests for materials should be addressed to E.M.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024