Gene Therapy (2010), 1–10

www.nature.com/gt

ORIGINAL ARTICLE

Quantitative fine-tuning of photoreceptor cis-regulatory elements through affinity modulation of transcription factor binding sites

J Lee, CA Myers, N Williams, M Abdelaziz and JC Corbo
Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO, USA

Given the remarkable recent progress in gene therapy-based treatments for retinal disease, there is an urgent need for the development of new approaches to quantitative design and analysis of photoreceptor-specific promoters. In this study, we determined the relative binding affinity of all single-nucleotide variants of the consensus binding site of the mammalian photoreceptor transcription factor, Crx. We then showed that it is possible to use these data to accurately predict the relative binding affinity of Crx for all possible 8 bp sequences. By rationally adjusting the binding affinity of three Crx sites, we were able to fine-tune the expression of the rod-specific Rhodopsin *promoter over a 225-fold range in living retinas. In addition, we showed that it is possible to fine-tune the activity of the rod-specific* Gnat1 *promoter over ~275-fold range by modulating the affinity of a single Crx-binding site. We found that the action of individual binding sites depends on the precise promoter context of the site and that increasing binding affinity does not always equate with increased promoter output. Despite these caveats, this tuning approach permits quantitative engineering of photoreceptor-specific* cis*-regulatory elements, which can be used as drivers in gene therapy vectors for the treatment of blindness.*
Gene Therapy advance online publication, 13 May 2010; doi:10.1038/gt.2010.77

Keywords: *photoreceptor;* cis*-regulatory elements; promoters; affinity; retina; Crx*

Introduction

Cis-regulatory elements (CREs; that is, enhancers or promoters) consist of short regions of noncoding DNA that contain binding sites for transcriptional activators and repressors, which can act in a combinatorial manner to dictate the spatial, temporal and quantitative levels of the gene whose expression they control.[1] Given the central role of CREs in organismal development and physiology,[2] there has been a great deal of effort aimed at elucidating the mechanisms whereby CREs establish spatiotemporal patterns of gene expression. Considerably less is known, however, about the mechanism by which CREs control quantitative levels of gene expression, especially in living mammalian tissues. Transgenic mice carrying CRE-reporter fusions remain the gold standard in mammalian *cis*-regulatory analysis, but given the wide-ranging effects of insertion site on CRE activity, such transgenes are generally not suitable for quantitative analysis. Thus, although transgenic mice have permitted analysis of the effects of specific truncations, deletions and point mutations on the spatiotemporal distribution of transgene expression, the results are largely nonquantitative. New, rapid approaches to *cis*-regulatory analysis are needed to dissect the quantitative aspects of mammalian CRE function in living tissues.

There is a clinical urgency to understand the mechanisms of cell type-specific gene regulation in photoreceptors as almost half of nearly 200 different human blindness genes seem to be enriched in this cell type.[3,4] Promising gene therapy approaches are being developed for a number of blinding diseases, and the viral delivery vectors used in these studies typically carry a cell type-specific promoter driving the expression of a rescue transgene.[5–7] Adeno-associated virus is by far the most commonly used vector for targeting therapy to photoreceptors.[5] Unfortunately, the single most significant limitation of this virus is its small cloning capacity of only 4.7 kb.[8] As the coding sequences of many disease genes are of this size or larger, using this vector to target such diseases is not easily achieved. Even when targeting genes whose coding sequence is <4.7 kb, relatively little space is left for inclusion of an appropriate promoter to drive transgene expression. Overall, the repertoire of compact promoters available for retinal gene therapy is limited, and their activities *in vivo* have not been quantitatively characterized on account of the lack of an appropriate assay system in living retinas.[9] For these reasons, the development of quantitative approaches to analyzing promoter activity in living photoreceptors would greatly facilitate the engineering of therapeutic vectors for the treatment of specific retinal diseases.

Numerous studies have elucidated the structure of the photoreceptor transcriptional network. Three transcription

Correspondence: Professor JC Corbo, Department of Pathology and Immunology, Washington University School of Medicine, 660 South Euclid Avenue, Campus Box 8118, St Louis, MO 63110, USA.
E-mail: jcorbo@pathology.wustl.edu
Received 30 November 2009; revised 15 April 2010; accepted 15 April 2010

factors in particular, Crx, Nrl and Nr2e3, have a central role in the control of photoreceptor gene expression.[3,10–15] Crx is a paired-type homeobox transcription factor, which is expressed in both rods and cones in the retina.[16,17] Mutations in human *CRX* are associated with several retinal diseases,[18–21] and *Crx* mutant mice lack detectable photoreceptor function.[22] Crx functions predominantly as an activator of both rod and cone gene expression and is critical for normal photoreceptor differentiation.[3,4,15] Nrl is a leucine zipper transcription factor, which is expressed in rod photoreceptors.[23–25] Mutations in human *NRL* cause autosomal-dominant retinitis pigmentosa, and mice mutant for *Nrl* show *en masse* conversion of presumptive rod photoreceptors into cones.[26–29] Microarray studies have identified a variety of putative *Nrl* target genes.[10,11,30] One such target of *Nrl* is the nuclear receptor superfamily member, *Nr2e3*, which, when mutated in humans, results in the Enhanced S-cone Syndrome.[31–35] Microarray and *in situ* hybridization analysis of the target genes controlled by *Crx*, *Nrl* and *Nr2e3*[14] have permitted the creation of a comprehensive model of the mammalian photoreceptor transcriptional network, which contains over 600 genes including most retinal disease genes expressed in photoreceptors.[4]

Given the central role of Crx in the control of photoreceptor gene expression, we have carried out a detailed quantitative analysis of this transcription factor's DNA-binding preferences and have analyzed the effects of modulating the affinity of identified Crx-binding sites on overall promoter activity in living explanted retinas. For this purpose, we devised a method for the rapid quantitation of promoter activity in living retinas using measurements of the light emitted by fluorescent reporters. With this assay system, we were able to fine-tune the activity of rod-specific promoters over a >200-fold range. This approach should prove useful for the rational, quantitative design of CREs for use in retinal gene therapy vectors.

Results

Determining the relative affinity of all single-nucleotide variants of the Crx-binding site

To determine the relative affinity of the Crx homeodomain for all single-nucleotide variants of its consensus sequence, we used the quantitative multiple fluorescence relative affinity assay.[36,37] On the basis of previous studies,[16,38] we chose to use the eight-nucleotide consensus sequence, CTAATCCC, as a starting point for our analysis of Crx's binding preferences. In our initial analysis, we quantified the binding affinity of all 24 single-nucleotide variants of the Crx consensus binding site (Figure 1 and Supplementary Table S1). The measurements were made in triplicate and were highly reproducible. One of the variants we assayed, '1T' ('1T' indicates that this variant differs from the consensus sequence in having a 'T' in position '1'), actually has a higher affinity than the so-called 'consensus' site. Furthermore, even though earlier studies defined a seven-nucleotide consensus (C/TTAATCC) for Crx,[16] our analysis demonstrates that Crx shows a distinct preference for a 'C' in position '8' (Figure 1b). Any other nucleotide in this position results in a moderate (~20–40%) reduction in affinity. Crx's strongest preference is for an 'A' in position 4, which is almost absolute. All three variants at this position (4G, 4C and 4T) show almost no binding (Figure 1b). In considering the 'TAAT' core further, one can see that Crx does tolerate some variation in these nucleotides. The three variants within the 'TAAT' core with the highest relative affinity are 2A, 3G and 5G (see Figure 1b). All these sequences have relative affinities ~20% of consensus. Crx's weakest nucleotide preferences occur at the ends of the motif (positions 1 and 8).

To assess whether Crx has preferences beyond positions 1 and 8, we examined all four variants in position '0' and position '9' (Figure 1b). Crx bound with very similar affinity to sequences containing any one of the four nucleotides in position 0, indicating little or no role for this position in determining Crx binding. Crx showed only a weak preference for 'C' and 'T' over 'A' and 'G' in the position 9. Although, as far as we know, there is no structural data to suggest that homeodomains directly bind with the nucleotide in position 9, it is possible that nucleotides in this position (and perhaps even further removed) can subtly affect binding through structural changes in the DNA that are transmitted to adjacent nucleotides, an effect known as 'indirect readout'.[39,40]

Predicting Crx's affinity for all possible eight-nucleotide sequences

Next, we set out to determine whether the affinity of binding sites that differ from the Crx consensus sequence by more than one nucleotide (for example, CTGATGCC, ATAATAAC and so on.) can be predicted from the single-nucleotide data alone. As there are 65 536 ($=4^8$) possible 8-bp sequence variants, it is not practical to determine Crx's relative affinity for all these variants using quantitative multiple fluorescence relative affinity assay. However, a recent report concluded that although multiple nucleotide changes within a site are not necessarily independent (that is, some combinations of changes show nonadditive effects on affinity), very good estimates of relative affinity can be obtained under the assumption of strict additivity.[41] To test this assumption for Crx, we made 22 dinucleotide variants from the consensus binding site and measured their relative affinity in triplicate using quantitative multiple fluorescence relative affinity assay. We then compared these measurements with the relative affinities predicted from the single-nucleotide variant data (Figure 2a and Supplementary Table S1). This analysis showed a strong linear correlation ($r=0.97$) between the measured affinities and those predicted from the single-nucleotide variant data under the assumption of additivity (Figure 2b). We did not find any systematic differences between the measured and predicted affinities of the dinucleotide variants. However, there is a mild tendency for those dinucleotide variants containing '5G' to have a somewhat higher measured affinity than predicted (Figure 2a). Given the excellent match between measured and predicted affinity for these selected dinucleotide variants, we predicted Crx's affinity for all possible eight-nucleotide sites using the single-nucleotide data alone (Supplementary Table S2). These results therefore permit quantitative prediction of the affinity of any possible Crx-binding site. The sequence logo for Crx derived from the single-nucleotide variant data is shown in Figure 2c. The present findings agree well with a recent study, which used protein-binding microarrays to determine

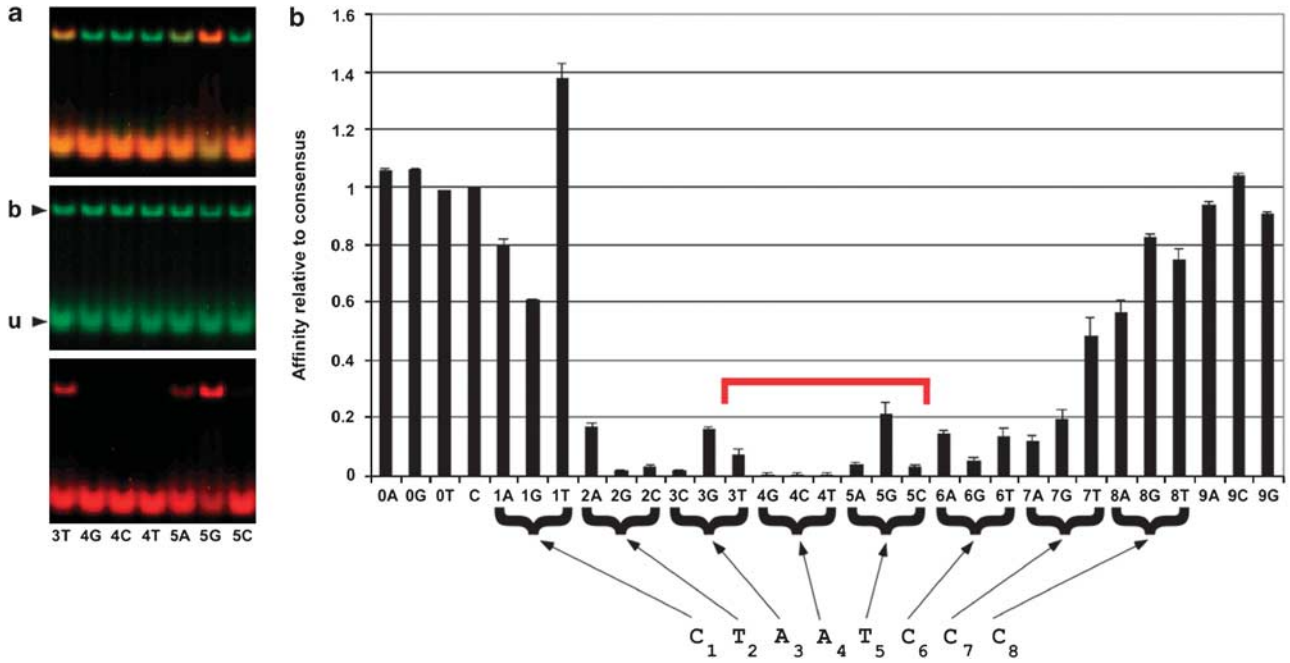

Figure 1 The relative affinity of all single-nucleotide variants of the Crx-binding site. (**a**) Example of a gel-shift assay used to quantify the relative affinity of variant Crx-binding sites. Two different binding sites, a 'test' and a 'control' were tagged with red and green fluorophores respectively, and then allowed to compete for binding to Crx protein in the same reaction mix. The upper and lower fluorescent bands represent the bound (b) and unbound (u) fractions, respectively. '3T' indicates that this oligonucleotide differs from the consensus sequence, CTAATCCC, in having 'T' in nucleotide position '3'. (**b**) Affinity of all single-nucleotide variants of the Crx consensus site (mean ± s.d.). The affinity of the consensus site (indicated by 'C') is normalized to '1'. The horizontal red bracket covers the seven variants shown in (**a**).

the binding preference of 168 mouse homeodomains including that of Crx.[42]

Fine-tuning the activity of rod-specific CREs in living retinas

To determine the effects on promoter activity of quantitatively modulating the affinity of identified Crx-binding sites within endogenous CREs, we developed a novel dual fluorescence assay (see Materials and methods for details). Our initial analysis focused on the rod-specific promoter from the *Rhodopsin* (*Rho*) gene. Fewer than 200 bp upstream of the transcription start site (TSS) of this gene are sufficient to drive photoreceptor-specific expression in transgenic mice.[43,44] We quantified the promoter activity of the endogenous *Rho* promoter, as well as 11 variant promoters carrying mutations in one or more of the Crx-binding sites (Figure 3). We examined a total of three affinity variants of Crx sites 1 and 2 in addition to the consensus sites. We also evaluated four variants of Crx site 3 in addition to the consensus site, all on a background in which Crx 1 had been made into a consensus site (Figure 3b,c). We found a strong linear correlation between Crx-binding site affinity and promoter activity for both Crx 2 (correlation coefficient, $r = 0.95$) and Crx 3 ($r = 0.95$; Figure 3b). In contrast, there was a poor correlation between binding site affinity and promoter activity at Crx site 1 ($r = 0.47$; Figure 3b). The reason for this poor correlation is not clear but one possibility is that certain binding sites, even though they have relatively high affinity for Crx, might disrupt binding of a second unknown factor whose recognition site partially overlaps with Crx 1. Thus, it seems that the tuning function of individual Crx-binding sites within the *Rho* promoter are dependent on the precise sequence context of the site. All variant promoters we examined, maintained a photoreceptor-specific expression in the retina (Figure 3e and Supplementary Figure S1). Overall, by tuning the affinity of these three Crx-binding sites within the context of a short (165 bp) promoter region, it was possible to generate a series of promoters with expression strengths varying over a 225-fold range (Figures 3c and d).

Next, we wished to determine whether this same tuning approach could be applied to a second rod-specific promoter with only a single identified Crx site. For this purpose, we chose the promoter of *Gnat1*, which encodes the α-subunit of rod transducin, a component of the rod phototransduction cascade.[45] As in the case with *Rho*, expression of *Gnat1* is severely compromised in the *Crx* mutant mouse.[4,22] We showed previously that a construct containing the first 541 bp upstream of the TSS of *Gnat1* drives strong photoreceptor-specific expression of a fluorescent transgene when introduced into the retina by electroporation.[4] This promoter region contains only a single high-affinity Crx-binding site at −158 bp relative to the TSS (TTAATCTG; relative affinity = 54.7% of consensus). We engineered promoters containing nine affinity variants of this site and quantified their promoter activity along with that of the endogenous promoter (Figure 4). We found that it was possible to tune *Gnat1* promoter activity over a 277-fold range by modulating the affinity of this single Crx site. As was the case with Crx site 1 in the *Rho* promoter, there was a relatively poor linear correlation between *Gnat1* promoter activity and Crx site affinity ($r = 0.64$). As a control, we engineered an inactivating mutation in the single high-affinity Nrl site within this promoter

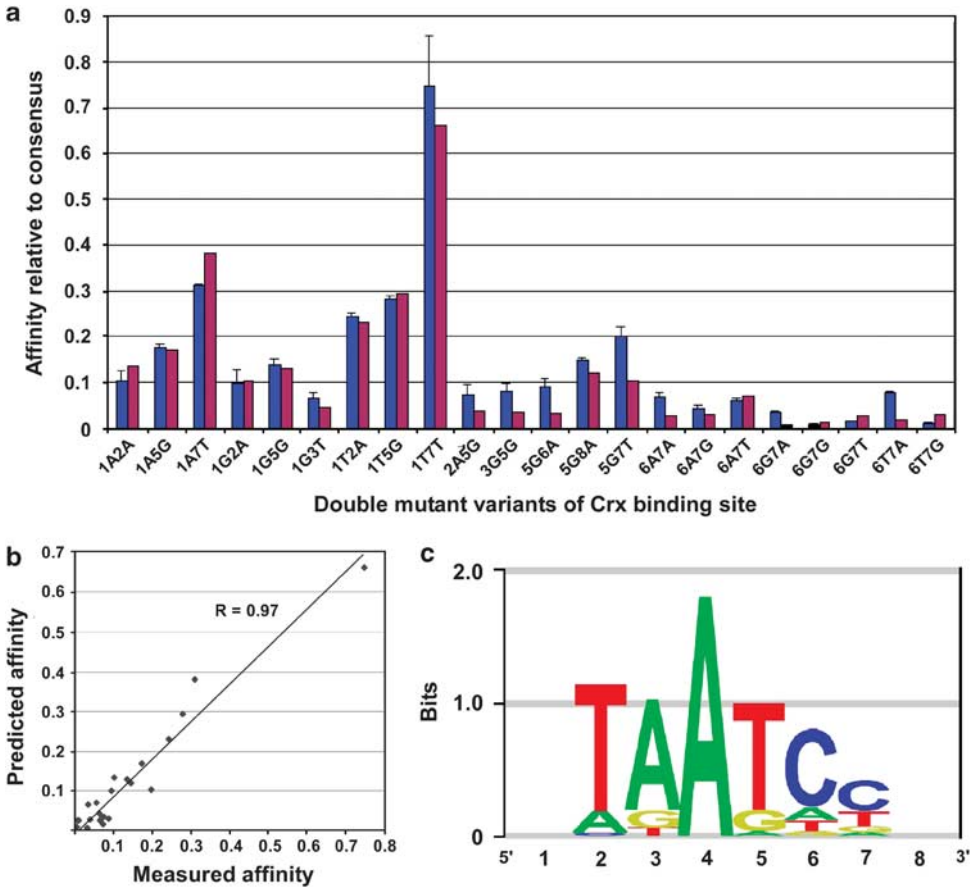

Figure 2 Individual nucleotide changes in the Crx-binding site have additive effects on binding. (**a**) Affinity measurements of 22 double-nucleotide variants of the Crx consensus site determined in three independent assays (mean ± s.d. in purple) compared with the affinity predicted from single-nucleotide variants (magenta). (**b**) Correlation between the measured affinity of individual double-mutant variants of the Crx site and the affinity predicted from single-nucleotide variants (correlation coefficient, $r = 0.97$). (**c**) Sequence logo representation of the DNA-binding preferences of Crx.[52]

(Figures 4a and b). This mutation resulted in a severe decrease in promoter activity comparable to that of the worst Crx mutation (Figure 4b). This result accords well with the total loss of *Gnat1* expression in the *Nrl* mutant retina.[4,29] We also evaluated the cellular expression pattern of a range of *Gnat1* promoter variants including those with the highest levels of expression by histological sectioning and found that they all maintain photoreceptor-specific expression (Supplementary Figure S2).

Discussion

In this study, we have determined the binding affinity of Crx for all single-nucleotide variants of its consensus sequence and used this information to predict its affinity for all possible eight-nucleotide binding sites. These data were then used to study the relationship between the promoter activity of two rod-specific CREs and the affinity of individual Crx sites within the promoters. We found that this relationship is strongly site and context dependent. Nevertheless, a close comparison of the tuning function at each of the examined Crx sites permits the establishment of some generalizations. In Figure 5, the promoter activities resulting from the introduction of five different Crx sites with a wide range of affinities at four different positions (three sites in the *Rho* promoter and one in the *Gnat1* promoter) are shown. The promoter activities of all variants at a given position have been normalized to that of a single variant (1T2A), which was assigned a value of ten. When normalized in this manner, a clear trend becomes apparent. As one progresses from a site with modest affinity (1T2A = 23% of consensus) through 7T (48% of consensus) and 1A (79% of consensus) up to a consensus site, there is a progressive increase in promoter activity (Figure 5b). This fact indicates that using even a relatively small number of different affinity variants, it is possible to achieve a significant range of tuning. However, a couple exceptions to this trend are apparent in the data. For example, the 1A variant and the consensus site at position Crx 3 in the *Rho* promoter result in nearly identical promoter activities, even though the consensus site would normally be expected to yield considerably stronger activity than the 1A variant (for example, at Crx 1 and at the site in the *Gnat1* promoter). One possible explanation for this result is that the variants at position Crx 3 were made on the background of a promoter that had already been 'upregulated' by the introduction of a consensus site at position Crx 1. Thus, it is possible that the dynamic range of this short promoter had been exhausted with the introduction of the 1A variant such that no further tuning could be achieved. This observation further underscores the fact that the tuning function

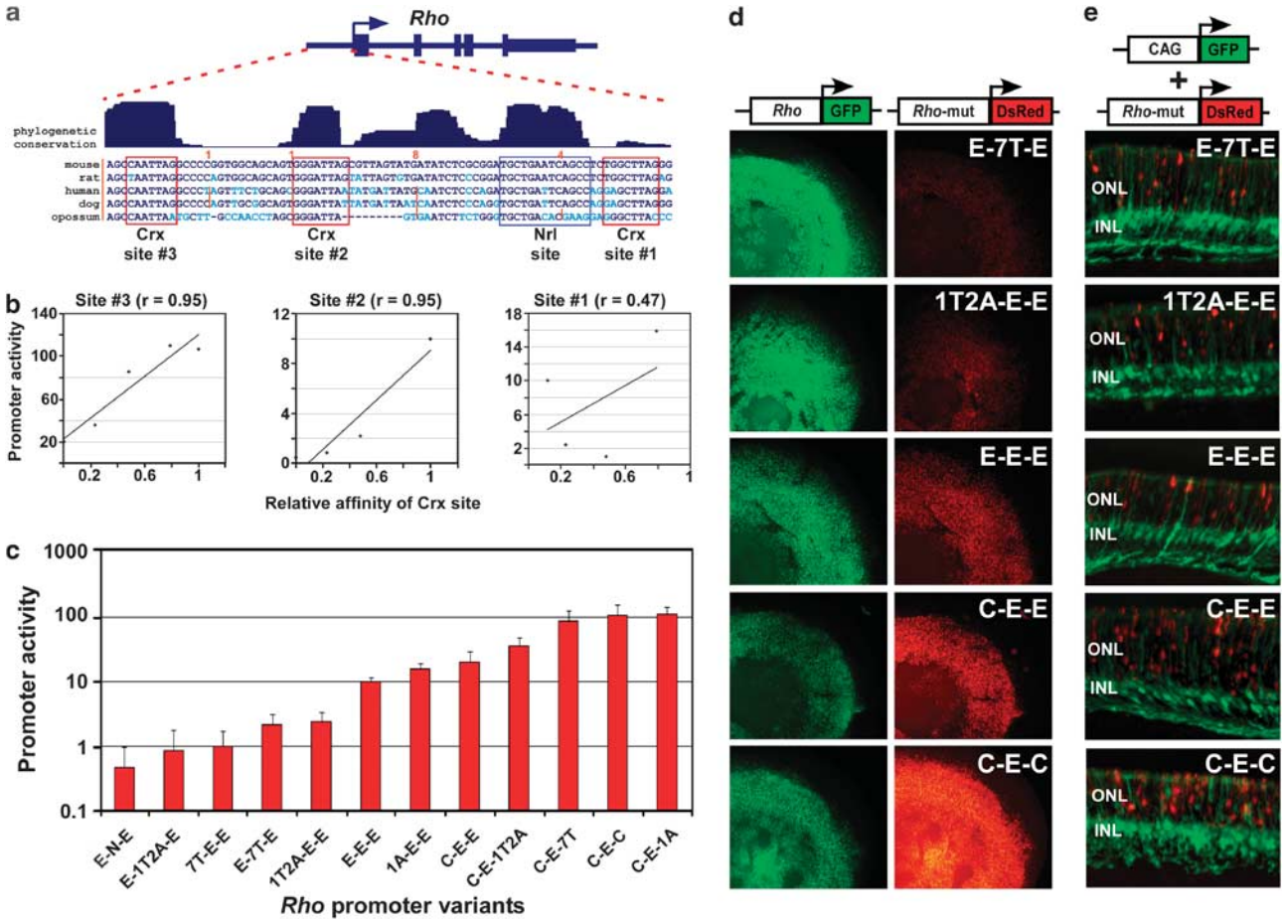

Figure 3 Quantitative fine-tuning of the mouse *Rhodopsin* (*Rho*) promoter in explanted retinas. (**a**) The proximal *Rho* promoter has one Nrl-binding site and three Crx-binding sites, all of which are phylogenetically conserved. The 'conservation' track represents the Phastcons score from the UCSC genome browser.[53,54] (**b**) Correlations between Crx-binding site affinity and promoter activity for the three Crx-binding sites ('r' represents the linear correlation coefficient). (**c**) Quantitation of the promoter activity of the endogenous *Rho* promoter and 11 variant promoters that carry mutations in one (or more) Crx sites as indicated. The activity of the endogenous *Rho* promoter is normalized to 10. Promoter activity is shown on a logarithmic scale. E: endogenous Crx site for that position; N: 'null' site, CGGGGCCC; C: 'consensus'; other variants are as described in Figure 1. For example, 'C-E-1T2A' indicates a variant *Rho* promoters with Crx site 1: consensus; Crx site 2: endogenous; and Crx site 3: 1T2A variant. (**d**) Flatmount images of mouse retinas, which were co-electroporated with the indicated *Rho* promoter variant driving DsRed and a 'loading' control (endogenous *Rho* promoter driving GFP) at P0, cultured in explant and harvested at P8. All flatmount images are taken at the same exposure time. (**e**) Cross-sections of retinas electroporated with a ubiquitously expressing CAG promoter driving GFP and the indicated *Rho* promoter variant driving DsRed. ONL, outer nuclear layer; INL, inner nuclear layer.

of individual Crx sites is highly dependent on the precise promoter context in which they occur.

A second intriguing feature of the tuning data is that some binding sites seem to drive promoter activity out of proportion to their affinity. One striking example is the site 5G8A (Figure 5). Although the measured affinity of this site is only 15% of consensus, it drives expression of the reporter gene at a level comparable to the 1A variant, which has an affinity of 79% of consensus (Figure 5). The reason for such discrepancies between affinity and promoter output is not known, but there are a number of potential explanations. First, as mentioned above, the sequence of a given binding site may differentially affect the binding of other factors whose recognition motif are adjacent to or partially overlap to that of Crx. Second, it is possible that binding to certain sites, but not others, can allosterically alter the structure of the Crx protein outside the DNA-binding domain in such a way as to create more favorable surfaces for activation. Such allostery within transcription factors has been described.[46–48] Finally, it is possible that other homeodomain transcription factors in photoreceptors such as Otx2 or Rx may bind some 'Crx' sites more avidly than others and thereby cause differential effects on transcription, which are not directly related to binding by Crx at all. Given the complexity of *cis*-regulatory architecture and the presence of numerous transacting factors within the nucleus of a cell, it is clear that there may not always exist a simple linear relationship between the affinity of a single factor (as determined *in vitro*) and promoter output in living retinas.

The results of this study have important implications for the design of gene therapy vectors targeting photoreceptors. We have fine-tuned the expression of the *Rho* and *Gnat1* promoters over several orders of magnitude, thus creating a series of compact promoters that can be used for gene therapy. The *Rho* promoter is particularly striking as it is only 165 bp in length, but has been engineered to drive very strong expression in rods (see Figure 3). Furthermore, given the nearly ubiquitous

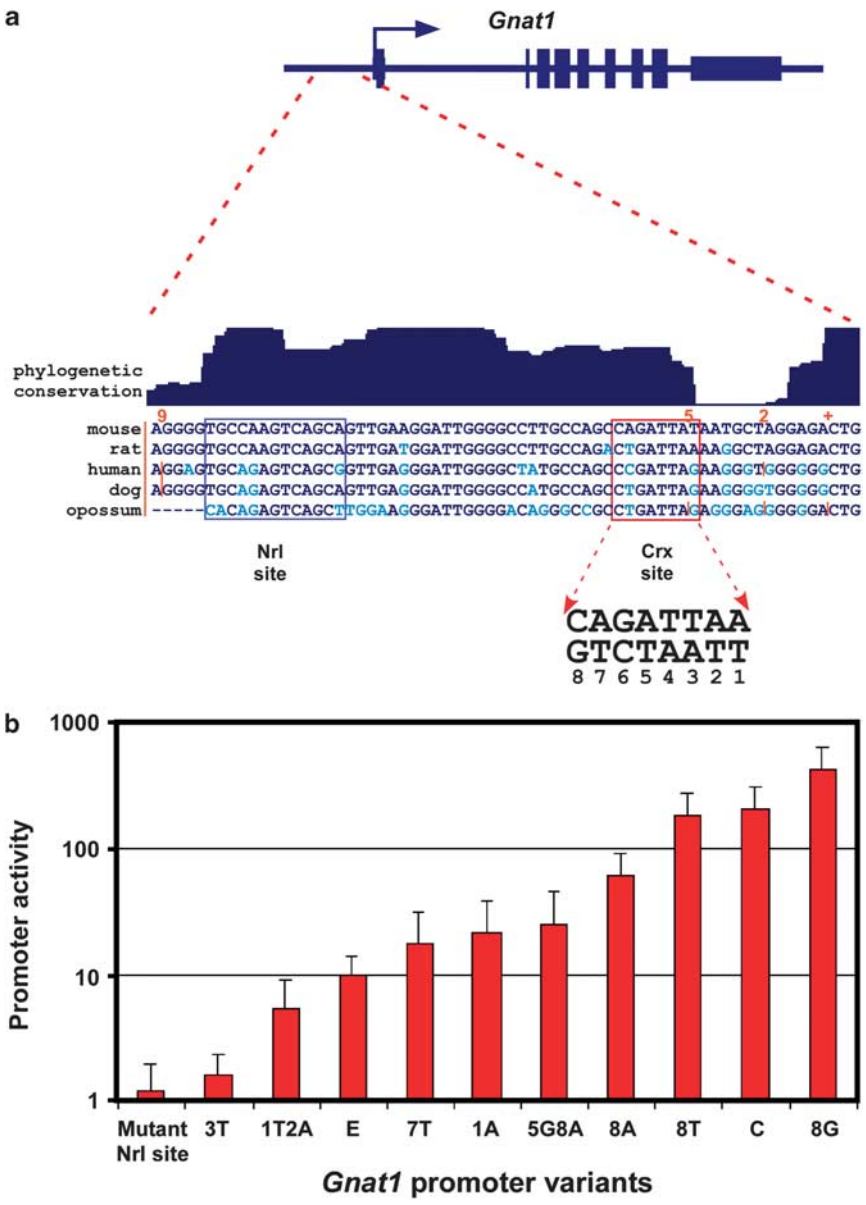

Figure 4 Quantitative fine-tuning of the *Gnat1* promoter using only a single Crx site. (**a**) The *Gnat1* promoter region contains one high-affinity Crx site and one Nrl site, both phylogenetically conserved. The Crx site is in the ''reverse' orientation relative to the transcription start site of *Gnat1*, but the nucleotides of the site are numbered according to the ''forward' strand as shown. (**b**) Quantitation of the promoter activity of the endogenous ('E') *Gnat1* promoter and 10 variant promoters. All the variant promoters carry alterations in the Crx site except 'Mutant Nrl site', which has a mutation that abolishes the Nrl site. Nomenclature is as described in Figure 1.

regulation of photoreceptor-specific promoters by Crx,[4] this same tuning approach can, in principle, be applied to almost any photoreceptor promoter. As endogenous photoreceptor gene expression varies over several orders of magnitude,[4,49] it will be highly desirable to design quantitatively appropriate promoters for driving the expression of specific rescue transgenes. Affinity modulation of Crx-binding sites represents a means of achieving this goal.

The present approach to fine-tuning promoter activity should be applicable to a wide variety of *cis*-elements beyond the retina, provided the relevant transcription factor binding sites are known. The advent of large-scale studies aimed at determining the binding preferences for a wide range of mammalian transcription factors will greatly facilitate such approaches.[42,50,51] However, given the strict dependence of the tuning function on the particular promoter context of the binding site, there will remain a need to test a range of binding site variants in intact tissues to engineer a precise, quantitative level of promoter output.

Materials and methods

Preparation of Crx–glutathione S*-transferase fusion protein*

The 60 amino acid homeodomain of Crx (amino-acid residues 38–98 of mouse Crx) was generated by PCR from a plasmid template using a forward primer, 5′-ccggaattccagcggcgggagcggacc-3′, and a reverse primer, 5′-tcagtcacgatgcggccgctcgctgctgtctgcattt-3′. *Eco*RI and

a

Variant	Affinity	Crx sites				Mean ± SD
		Crx #1	Crx #2	Crx #3	Gnat1	
C	1	205.74	113.68	29.69	387.39	184.12 ± 153.4
1A	0.79	65.47	ND	30.58	40.49	45.52 ± 17.98
7T	0.48	4.18	25.26	23.83	32.88	21.54 ± 12.23
1T2A	0.24	10	10	10	10	10
5G8A	0.15	41.28	ND	ND	47.59	44.43 ± 4.47

b

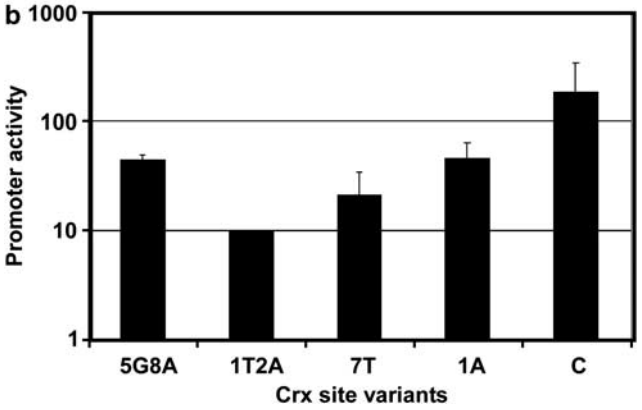

Figure 5 Selected Crx-binding site variants show similar quantitative effects on promoter activity at different positions. (**a**) Table showing the promoter activity driven by five different Crx-binding site variants at four different positions. The promoter activities at a given site are all normalized to that of the 1T2A variant whose value is set to 10. ND, not done. (**b**) Graph of the mean (±s.d.) of the promoter activity for each of the five Crx site variants in (**a**).

*Not*I sites (underlined) were added to 5′ ends of the primers to facilitate cloning of the PCR product into pGEX-5X-1 (Amersham Biosciences/Pharmacia Biotech/GE Healthcare, Piscataway, NJ, USA). The resulting fusion protein includes a 5′-glutathione *S*-transferase tag, a factor X cleavage site, a short linker and the homeodomain of Crx. The Crx–glutathione *S*-transferase fusion construct was transformed into the *Escherichia coli* strain BL21. After 3–6 h of induction with 0.2 mM isopropyl-β-D-thiogalactoside, the bacterial culture was pelleted by centrifugation. The resulting pellet was resuspended in ice-cold 1 × phosphate-buffered saline with the complete protease inhibitor cocktail (Roche Applied Science, Indianapolis, IN, USA). Glutathione Sepharose 4B column (Amersham Biosciences/GE Healthcare) was used to purify the fusion protein. The volume of the various reagents and the time required for each purification step followed the manufacturer's recommendation. Vivaspin6 (Vivascience, Littleton, MA, USA) was used to exchange the buffers of the purified proteins, which were then stored at −70 °C for future use. The composition of the buffer used for the exchange and the storage was as follows: 60 mM KCl, 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.6), 1 mM Dithiothreitol (DTT), 1 mM EDTA, 5% glycerol. The concentration of the prepared fusion protein was determined using the Bio-Rad protein assay kit (Bio-Rad, Hercules, CA, USA) followed by measurement on a Thermomax microplate reader (Molecular Devices, Sunnyvale, CA, USA).

Quantitative multiple fluorescence relative affinity assay

This assay was carried out as previously published with the following modifications.[36,37] DNA probes were prepared by PCR using two primers, KS-1, 5′-tcgaggtcg acggtgtc-3′, and SK-1, 5′-ctggcggccgctctagagcaagtgga-3′, on a template, which included the two primer-binding sites (in lower case) and a central eight-nucleotide Crx-binding site (in upper case): 5′-tcgaggtcgacggtgtcCT AATCCCtccacttgctctagagcggccgccag-3′ (shown sequence is the template for the consensus Crx site). Appropriate variant templates were used to make all variant sites. The KS-1 primers carried a 5′ fluorophore, either 6-carboxy-fluorescein (FAM) or carboxy-x-rhodamine (Rox) (Integrated DNA Technologies, Coralville, IA, USA). After 30 cycles of PCR amplification (melting: 94 °C for 30 s; annealing: 55 °C for 30 s; extension: 68 °C for 30 s), the PCR products were ethanol precipitated and resuspended. DNA-binding reactions were carried out at 4 °C in the presence of various amount of proteins ranging between 0.2 and 2.0 μg, 0.1 μg probes, 2 μg poly(dI-dC), 60 mM KCl, 25 mM HEPES (pH 7.6), 1 mM DTT, 1 mM EDTA, 5% glycerol for 1 h. In any given reaction, an equal amount (0.1 μg) of the FAM-labeled control probe was competed with the Rox-labeled test probe. For high-affinity variants the consensus probe was used as a control and for low-affinity variants either the 5G or 3T variants were used.

Protein–DNA complexes were separated from the free probe by electrophoresis through a 10% TBE native polyacrylamide gel (Invitrogen, Carlsbad, CA, USA) at 100 V for 1 h at room temperature. The quantities of protein–DNA complex and free DNA were determined from the Electrophoretic mobility shift assay (EMSA) gel using a Typhoon variable scanner (Molecular Dynamics, Sunnyvale, CA, USA). The settings on the scanner were as follows: excitation laser at 532 nm, emission filter at 526 nm for the FAM fluorophore and emission filter at 610 nm for the Rox fluorophore. The fluorescent intensities of the bound and unbound fractions in each lane at each emission wavelength were quantified through the volume analysis method of the ImageQuant software (Molecular Dynamics). The background fluorescent intensities were subtracted using the object average method. The binding constant (K_b) of a probe DNA 1 is equal to: $K_b(D_1) = [P \cdot D_1]/[P][D_1]$ where $[P \cdot D_1]$ is the amount of protein–DNA complex and $[P]$ and $[D_1]$ are the amounts of free protein and DNA, respectively. When the FAM-labeled reference probe and Rox-labeled mutant probe bind to the same pool of protein, the relative affinities can be determined without knowing the free protein concentration using the ratio of binding constants for the two DNA species, D_{FAM} and D_{Rox}: $K_b(D_{FAM})/K_b(D_{Rox}) = [P \cdot D_{FAM}][D_{Rox}]/[P \cdot D_{Rox}][D_{FAM}]$.

The predicted affinities of Crx for all possible eight-nucleotide binding sites (considering only the top strand) are given in Supplementary Table S2. These values were calculated from the Crx single-nucleotide variant binding data under an assumption of strict additivity. For example, Crx's predicted relative binding affinity for the site 1G2A8T (GAAATCCT) would be the product of the measured relative affinity of each of the single-nucleotide variants of which it is comprised: 0.605 (1G) × 0.167 (2A) × 0.744 (8T) = ∼0.075.

Promoter–reporter fusion constructs

The mouse *Rho* promoter region (containing the first 165 bp upstream of the TSS) was obtained by PCR from

mouse genomic DNA using the following primers: 5′-tcc ccggaattctgacggctcccacagcgcaac-3′ and 5′-tcccctctagaatg tcaccttggcccctctgc-3′. This product was digested with *Eco*RI and *Xba*I (sites underlined in primers) and subcloned into the 'No-basal' reporter vector, which contains a polycloning site upstream of the coding sequence of Discosoma red fluorescent protein (DsRed) followed by a polyadenylation signal.[4] The *Gnat1*-DsRed reporter used here was previously described[4] and contains the first 541 bp upstream of the TSS. A control construct for the *Rho* reporters was generated by removing the DsRed coding sequence and replacing it with that of green fluorescent protein. Variant promoters containing Crx sites with different affinities were generated by site-directed mutagenesis using the QuikChange II XL kit (Stratagene, Cedar Creek, TX, USA) according to the manufacturer's instructions.

Electroporation and culture of explanted retinas

Electroporations were carried out using two DNA constructs, a loading control (consisting of the 'wild-type' version of the experimental construct driving green fluorescent protein for the *Rho* promoter analysis and a 2.2 kb bovine *Rho* promoter for the *Gnat1* analysis) and the experimental construct driving DsRed, both at a final concentration of 0.5 μg μl^{-1}. The constructs were ethanol precipitated together, washed in 70% ethanol, dried and resuspended in 60 μl of sterile 1 × phosphate-buffered saline. The DNA was then transferred to the wells of a 453 microslide chamber (BTX Harvard Apparatus, Holliston, MA, USA), which had been divided into five equal wells (well volume ~70 μl) using aquarium cement. Retinas from P0 CD-1 mice were dissected free of sclera and retinal pigment epithelium and placed into the microslide chamber with the lens facing the positive electrode. Once the retinas had been placed in the well, they were subjected to five 30 V pulses, 50 ms in length, 950 ms apart using an ECM 830 square-wave electroporator (BTX Harvard Apparatus). Three retinas were electroporated per construct per electroporation and three replicate electroporations were performed per construct. The retinas were then rinsed in media (1:1 ratio of Dulbecco's modified Eagle's medium and F12, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, 0.2 mg ml^{-1} L-glutamine and 5 μg ml^{-1} insulin), once without fetal calf serum (FCS) and once with 10% FCS placed on a circular Nuclepore filter (25 mm, 0.2 μm; Whatman, Piscataway, NJ, USA) floating in 3 ml of media containing 10% FCS, and allowed to incubate at 37 °C. On P8, the explanted retinas were examined using a fluorescent dissecting microscope (Fluo III; Leica, Richmond, IL, USA) in order to subjectively estimate the brightness of the various electroporated constructs. The retinas were then fixed in 4% paraformaldehyde for 15 min followed by two 5 min washes with 1 × phosphate-buffered saline. The retinas were placed on slides, retina side up, along with ~300 μl 1 × phosphate-buffered saline. Fragments of glass coverslips (Fisher Scientific #1.5, Pittsburg, PA, USA; 0.16 mm thick) were used as legs to elevate an intact 24 × 50 mm no. 1.5 coverslip, which was placed on top.

For cross-sectional images in Figure 3 and Supplementary Figure 2, co-electroporations were carried out using a ubiquitously expressing CAG-green fluorescent protein and the indicated *Rho* and *Gnat1* promoter constructs fused to DsRed. The electroporated retinas were harvested after 8 days in culture, fixed, embedded in optimal cutting temperature compound (OCT) (Tissue Tek, CRYO-OCT Compound, Andwin Scientific (Addison, IL, USA) and sectioned on a cryostat. The sections were then counterstained with 4,6-diamidino-2-phenylindole, coverslipped and imaged.

Quantitation of promoter activity in whole-mount retinas

Images were collected from whole-mount retinas either in wide field or as single-optical sections using a DSU spinning disk on an Olympus (Center Valley, PA, USA) BX61WI microscope equipped with a Xenon light source and a monochromatic ORCA-ER camera (Hamamatsu Photonic Systems, Bridgewater, NJ, USA). Quantitation using either optical sections or wide-field images gave quantitatively similar results. Both red and green channel images were captured in gray scale and an optimal exposure time for each color channel was determined using the 'strongest' expressing construct, as previously assessed subjectively using a fluorescent dissecting microscope. An exposure time was chosen such that the average pixel value of this 'strongest' construct was as high as possible without having saturated pixels, so as to maximize the dynamic range of quantitation. Given that the pixel depth was 12 bit (that is, pixel values ranged from 0 to 4095), we ensured that the highest pixel values in a given image remained under ~4000. The exposure time as determined in this manner was then used for all subsequent retinas. SlideBook Software, Intelligent Imaging Innovations, Inc., Denver, CO, USA (version 4.2) was used to collect images, which were then exported for subsequent analysis. Images from both green and red channels were opened using ImageJ (NIH, Bethesda, MD, USA) and a circle of interest was defined. This circle was then used to select five regions of interest from each retina for quantitation. We selected regions that were uniformly electroporated, avoiding the outer edges of the retina and the region of the retina overlying the lens. Three circles were placed outside the retina and were used for background subtraction. The mean pixel value was recorded for each region for both channels and the coordinates of the circles of interest were saved. To normalize the expression levels of the experimental construct to that of the loading control, the mean pixel value of the experimental construct (post-background subtraction) was divided by the mean pixel value of the control construct (post-background subtraction) for each of the regions of interest. To quantitatively compare the results of electroporations carried out on different days, one 'standard' construct was always included in every electroporation set. Relative expression values across experiments could be compared by normalizing to the expression level of this 'standard'.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank G Stormo and colleagues for their advice on the QuMFRA assay and for permitting us to use their

Typhoon scanner. We also thank the other members of the Corbo lab for their comments and thank C Montana for carefully reading the paper. This work was supported by Grant R01 EY018826 from NIH/NEI and by grants from the McDonnell Center for Cellular and Molecular Neurobiology and from the American Health Assistance Foundation.

References

1 Davidson EH. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. Academic Press: London, 2006.
2 Davidson EH. *Genomic Regulatory Systems: Development and Evolution*, 1st edn. Academic Press: San Diego, 2001.
3 Blackshaw S, Fraioli RE, Furukawa T, Cepko CL. Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes. *Cell* 2001; **107**: 579–589.
4 Hsiau TH, Diaconu C, Myers CA, Lee J, Cepko CL, Corbo JC. The *cis*-regulatory logic of the mammalian photoreceptor transcriptional network. *PLoS ONE* 2007; **2**: e643.
5 Dinculescu A, Glushakova L, Min SH, Hauswirth WW. Adeno-associated virus-vectored gene therapy for retinal disease. *Hum Gene Ther* 2005; **16**: 649–663.
6 Bainbridge JW. Gene therapy clinical trials for inherited eye disease. *Exp Rev Ophthalmol* 2007; **2**: 517–519.
7 Montana CL, Corbo JC. Inherited diseases of photoreceptors and prospects for gene therapy. *Pharmacogenomics* 2008; **9**: 335–347.
8 Goverdhana S, Puntel M, Xiong W, Zirger JM, Barcia C, Curtin JF *et al*. Regulatable gene expression systems for gene therapy applications: progress and future challenges. *Mol Ther* 2005; **12**: 189–211.
9 Corbo JC. The role of *cis*-regulatory elements in the design of gene therapy vectors for inherited blindness. *Expert Opin Biol Ther* 2008; **8**: 599–608.
10 Yu J, He S, Friedman JS, Akimoto M, Ghosh D, Mears AJ *et al*. Altered expression of genes of the Bmp/Smad and Wnt/calcium signaling pathways in the cone-only Nrl−/− mouse retina, revealed by gene profiling using custom cDNA microarrays. *J Biol Chem* 2004; **279**: 42211–42220.
11 Yoshida S, Mears AJ, Friedman JS, Carter T, He S, Oh E *et al*. Expression profiling of the developing and mature Nrl−/− mouse retina: identification of retinal disease candidates and transcriptional regulatory targets of Nrl. *Hum Mol Genet* 2004; **13**: 1487–1503.
12 Peng GH, Ahmad O, Ahmad F, Liu J, Chen S. The photoreceptor-specific nuclear receptor Nr2e3 interacts with Crx and exerts opposing effects on the transcription of rod versus cone genes. *Hum Mol Genet* 2005; **14**: 747–764.
13 Chen J, Rattner A, Nathans J. The rod photoreceptor-specific nuclear receptor Nr2e3 represses transcription of multiple cone-specific genes. *J Neurosci* 2005; **25**: 118–129.
14 Corbo JC, Cepko CL. A hybrid photoreceptor expressing both rod and cone genes in a mouse model of enhanced s-cone syndrome. *PLoS Genet* 2005; **1**: e11.
15 Livesey FJ, Furukawa T, Steffen MA, Church GM, Cepko CL. Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx. *Curr Biol* 2000; **10**: 301–310.
16 Chen S, Wang QL, Nie Z, Sun H, Lennon G, Copeland NG *et al*. Crx, a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes. *Neuron* 1997; **19**: 1017–1030.
17 Furukawa T, Morrow EM, Cepko CL. Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell* 1997; **91**: 531–541.
18 Freund CL, Gregory-Evans CY, Furukawa T, Papaioannou M, Looser J, Ploder L *et al*. Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell* 1997; **91**: 543–553.
19 Freund CL, Wang QL, Chen S, Muskat BL, Wiles CD, Sheffield VC *et al*. *De novo* mutations in the CRX homeobox gene associated with Leber congenital amaurosis [letter]. *Nat Genet* 1998; **18**: 311–312.
20 Jacobson SG, Cideciyan AV, Huang Y, Hanna DB, Freund CL, Affatigato LM *et al*. Retinal degenerations with truncation mutations in the cone-rod homeobox (CRX) gene. *Invest Ophthalmol Vis Sci* 1998; **39**: 2417–2426.
21 Sohocki MM, Sullivan LS, Mintz-Hittner HA, Birch D, Heckenlively JR, Freund CL *et al*. A range of clinical phenotypes associated with mutations in CRX, a photoreceptor transcription-factor gene. *Am J Hum Genet* 1998; **63**: 1307–1315.
22 Furukawa T, Morrow EM, Li T, Davis FC, Cepko CL. Retinopathy and attenuated circadian entrainment in Crx-deficient mice. *Nat Genet* 1999; **23**: 466–470.
23 Liu Q, Ji X, Breitman ML, Hitchcock PF, Swaroop A. Expression of the bZIP transcription factor gene Nrl in the developing nervous system. *Oncogene* 1996; **12**: 207–211.
24 Swain PK, Hicks D, Mears AJ, Apel IJ, Smith JE, John SK *et al*. Multiple phosphorylated isoforms of NRL are expressed in rod photoreceptors. *J Biol Chem* 2001; **276**: 36824–36830.
25 Swaroop A, Xu JZ, Pawar H, Jackson A, Skolnick C, Agarwal N. A conserved retina-specific gene encodes a basic motif/leucine zipper domain. *Proc Natl Acad Sci USA* 1992; **89**: 266–270.
26 Bessant DA, Payne AM, Mitton KP, Wang QL, Swain PK, Plant C *et al*. A mutation in NRL is associated with autosomal dominant retinitis pigmentosa. *Nat Genet* 1999; **21**: 355–356.
27 DeAngelis MM, Grimsby JL, Sandberg MA, Berson EL, Dryja TP. Novel mutations in the NRL gene and associated clinical findings in patients with dominant retinitis pigmentosa. *Arch Ophthalmol* 2002; **120**: 369–375.
28 Daniele LL, Lillo C, Lyubarsky AL, Nikonov SS, Philp N, Mears AJ *et al*. Cone-like morphological, molecular, and electrophysiological features of the photoreceptors of the Nrl knockout mouse. *Invest Ophthalmol Vis Sci* 2005; **46**: 2156–2167.
29 Mears AJ, Kondo M, Swain PK, Takada Y, Bush RA, Saunders TL *et al*. Nrl is required for rod photoreceptor development. *Nat Genet* 2001; **29**: 447–452.
30 Corbo JC, Myers CA, Lawrence KA, Jadhav AP, Cepko CL. A typology of photoreceptor gene expression patterns in the mouse. *Proc Natl Acad Sci USA* 2007; **104**: 12069–12074.
31 Akhmedov NB, Piriev NI, Chang B, Rapoport AL, Hawes NL, Nishina PM *et al*. A deletion in a photoreceptor-specific nuclear receptor mRNA causes retinal degeneration in the rd7 mouse. *Proc Natl Acad Sci USA* 2000; **97**: 5551–5556.
32 Haider NB, Jacobson SG, Cideciyan AV, Swiderski R, Streb LM, Searby C *et al*. Mutation of a nuclear receptor gene, NR2E3, causes enhanced S cone syndrome, a disorder of retinal cell fate. *Nat Genet* 2000; **24**: 127–131.
33 Haider NB, Naggert JK, Nishina PM. Excess cone cell proliferation due to lack of a functional NR2E3 causes retinal dysplasia and degeneration in rd7/rd7 mice. *Hum Mol Genet* 2001; **10**: 1619–1626.
34 Kobayashi M, Takezawa S, Hara K, Yu RT, Umesono Y, Agata K *et al*. Identification of a photoreceptor cell-specific nuclear receptor. *Proc Natl Acad Sci USA* 1999; **96**: 4814–4819.
35 Milam AH, Rose L, Cideciyan AV, Barakat MR, Tang WX, Gupta N *et al*. The nuclear receptor NR2E3 plays a role in human retinal photoreceptor differentiation and degeneration. *Proc Natl Acad Sci USA* 2002; **99**: 473–478.
36 Man TK, Yang JS, Stormo GD. Quantitative modeling of DNA-protein interactions: effects of amino acid substitutions on binding specificity of the Mnt repressor. *Nucleic Acids Res* 2004; **32**: 4026–4032.

37 Man TK, Stormo GD. Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. *Nucleic Acids Res* 2001; **29**: 2471–2478.
38 Chatelain G, Fossat N, Brun G, Lamonerie T. Molecular dissection reveals decreased activity and not dominant negative effect in human OTX2 mutants. *J Mol Med* 2006; **84**: 604–615.
39 Morozov AV, Havranek JJ, Baker D, Siggia ED. Protein-DNA binding specificity predictions with structural models. *Nucleic Acids Res* 2005; **33**: 5781–5798.
40 Little EJ, Babic AC, Horton NC. Early interrogation and recognition of DNA sequence by indirect readout. *Structure* 2008; **16**: 1828–1837.
41 Benos PV, Bulyk ML, Stormo GD. Additivity in protein-DNA interactions: how good an approximation is it? *Nucleic Acids Res* 2002; **30**: 4442–4451.
42 Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Pena-Castillo L *et al*. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* 2008; **133**: 1266–1276.
43 Zack DJ, Bennett J, Wang Y, Davenport C, Klaunberg B, Gearhart J *et al*. Unusual topography of rhodopsin promoter-lacZ fusion gene expression in transgenic mouse retinas. *Neuron* 1991; **6**: 187–199.
44 Lem J, Applebury ML, Falk JD, Flannery JG, Simon MI. Tissue-specific and developmental regulation of rodopsin chimeric genes in transgenic mice. *Neuron* 1991; **6**: 201–210.
45 Dryja TP, Hahn LB, Reboul T, Arnaud B. Missense mutation in the gene encoding the alpha subunit of rod transducin in the Nougaret form of congenital stationary night blindness. *Nat Genet* 1996; **13**: 358–360.
46 Scully KM, Jacobson EM, Jepsen K, Lunyak V, Viadiu H, Carriere C *et al*. Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science* 2000; **290**: 1127–1131.
47 Leung TH, Hoffmann A, Baltimore D. One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. *Cell* 2004; **118**: 453–464.
48 Meijsing SH, Pufall MA, So AY, Bates DL, Chen L, Yamamoto KR. DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* 2009; **324**: 407–410.
49 Blackshaw S, Harpavat S, Trimarchi J, Cai L, Huang H, Kuo WP *et al*. Genomic analysis of mouse retinal development. *PLoS Biol* 2004; **2**: E247.
50 Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA *et al*. Diversity and complexity in DNA recognition by transcription factors. *Science* 2009; **324**: 1720–1723.
51 Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* 2008; **133**: 1277–1289.
52 Workman CT, Yin Y, Corcoran DL, Ideker T, Stormo GD, Benos PV. enoLOGOS: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Res* 2005; **33**: W389–W392; Web Server issue.
53 Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K *et al*. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005; **15**: 1034–1050.
54 Kuhn RM, Karolchik D, Zweig AS, Wang T, Smith KE, Rosenbloom KR *et al*. The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res* 2009; **37**: D755–D761; Database issue.

Supplementary Information accompanies the paper on Gene Therapy website (http://www.nature.com/gt)