

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed preprint posted

25 April 2023

Sent for peer review

24 February 2023

Posted to bioRxiv

6 February 2023

Missense mutations in CRX homeodomain cause dominant retinopathies through two distinct mechanisms

Yiqiao Zheng, Chi Sun, Xiaodong Zhang, Philip A. Ruzyczki, Shiming Chen

Molecular Genetic and Genomics Graduate program, Division of Biological and Biomedical Sciences, Washington University in St Louis, Saint Louis, Missouri, USA • Department of Ophthalmology and Visual Sciences, Washington University in St Louis, Saint Louis, Missouri, USA • Department of Genetics, Washington University in St Louis, Saint Louis, Missouri, USA • Department of Developmental Biology, Washington University in St Louis, Saint Louis, Missouri, USA

 (https://en.wikipedia.org/wiki/Open_access)

 (<https://creativecommons.org/licenses/by/4.0/>)

Abstract

Summary

Homeodomain transcription factors (HD TFs) are instrumental to vertebrate development. Mutations in HD TFs have been linked to human diseases, but their pathogenic mechanisms remain elusive. Here we use *Cone-Rod Homeobox (CRX)* as a model to decipher the disease-causing mechanisms of two HD mutations, *p.E80A* and *p.K88N*, that produce severe dominant retinopathies. Through integrated analysis of molecular and functional evidence *in vitro* and in knock-in mouse models, we uncover two novel gain-of-function mechanisms: *p.E80A* increases CRX-mediated transactivation of canonical CRX target genes in developing photoreceptors; *p.K88N* alters CRX DNA-binding specificity resulting in binding at ectopic sites and severe perturbation of CRX target gene expression. Both mechanisms produce novel retinal morphological defects and hinder photoreceptor maturation distinct from loss-of-function models. This study reveals the distinct roles of E80 and K88 residues in CRX HD regulatory functions and emphasizes the importance of transcriptional precision in normal development.

eLife assessment

This manuscript is of interest to readers in the field of neural development. The study is **important** as it examines two mutations within a homeodomain transcription factor called Cone-Rod Homeobox (CRX) that cause retinopathy in humans. The data are **solid**, and the work contributes to our understanding of the underlying pathogenetic mechanisms.

Introduction

Homeodomain transcription factors (HD TFs) play a fundamental role in vertebrate development. Members of the HD TF family are characterized by the presence of a highly conserved 60 amino acid helix-turn-helix DNA-binding domain known as the homeodomain (HD). The HD is one of the most studied eukaryotic DNA-binding motifs since its discovery in *Drosophila* homeotic transformations^{(1),(2)}. Hundreds of HD TFs have subsequently been documented in regulating gene expression programs important for body plan specification, pattern formation and cell fate determination^{(1),(2)}. Mutations in HD TFs have been linked to many human diseases, including neuropsychiatric and neurodegenerative conditions^{(3),(4)}. Although significant progress has been made in understanding HD-DNA interactions, uncovering the pathogenetic mechanisms of disease-causing missense mutations in HD have proven challenging.

The retina has long been used as a model system to study the role of HD TFs in normal central nervous system (CNS) development and in neurological diseases⁽⁵⁾. During retinogenesis, HD TFs play essential roles in the patterning of neuroepithelium, specification of retinal progenitors and differentiation of all retinal cell classes that derive from a common progenitor⁽⁶⁾. Importantly, many HD TFs are shared between the brain and the retina during development and mutations in these TFs can lead to disease manifestation in both tissues^{(7)–(12)}. The accessibility and wealth of available molecular tools makes the retina a valuable tool to decipher the pathogenic mechanisms of HD TF mutations associated with neurological diseases.

Here, we study CRX, a HD TF essential for photoreceptor cells in the retina, as a model to understand how single amino acid substitutions in the HD impact TF functions and cause blinding diseases. Photoreceptors are the most numerous neurons in the retina and are specialized to sense light and initiate vision through a process called phototransduction. Animal studies have demonstrated that *Crx* is first expressed in post-mitotic photoreceptor precursors and maintained throughout life^{(13),(14)}. Loss of CRX results in impaired photoreceptor gene expression, failure of maturation and rapid degeneration of immature, non-functional photoreceptors⁽¹⁵⁾. Protein-coding sequence variants in human *CRX* have been associated with inherited retinal diseases (IRDs) that affect photoreceptors: Leber congenital amaurosis (LCA), cone-rod dystrophy (CoRD), and retinitis pigmentosa (RP) (OMIM:602225). However, the recessive phenotype observed in *Crx* knockout mouse models fails to recapitulate many dominant human *CRX* mutations that arise *de novo*⁽¹⁵⁾.

CRX contains two functional domains – the N-terminal HD and C-terminal activation domain (AD) (Figure 1A), both are required for proper activation of target genes and maintenance of normal *Crx* mRNA transcript abundance^{(16),(17)}. To understand how CRX HD mutations cause diseases, we have previously reported a mutation knock-in mouse model carrying a hypomorphic mutation *p.R90W* (R90W) in CRX HD^{(18),(19)}. We found that R90W mutation produces a recessive loss-of-function phenotype very similar to that of *Crx* knockout mice.

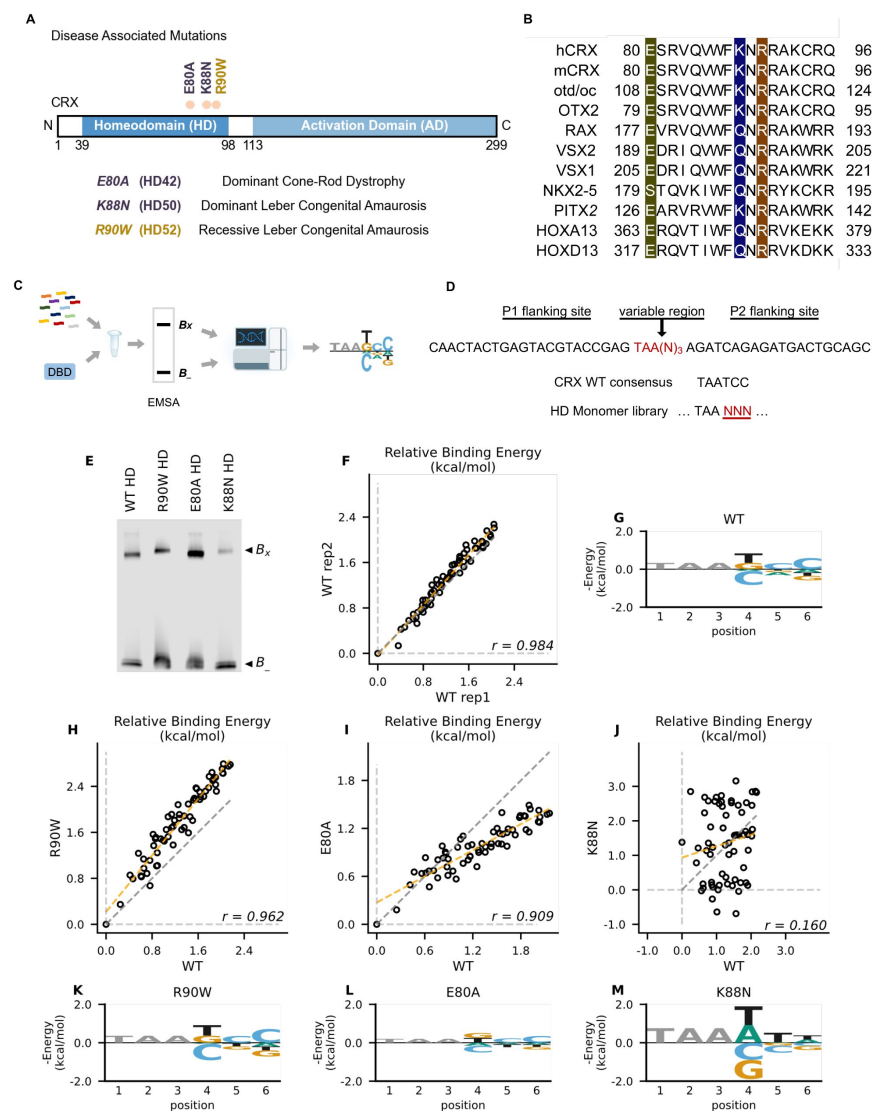


Figure 1.

Disease associated missense mutations altered CRX HD DNA binding specificity.

(A) Diagram of CRX functional domains: Homeodomain (HD) for DNA-binding and Activation Domain (AD) for target gene transactivation. The three missense mutations in this study are located at the C-terminus of CRX HD and associated with different retinal diseases in human. Number in the parenthesis denotes the CRX HD position of the corresponding mutated residue. (B) Alignments of HD recognition helix sequences for the indicated HD proteins for which HD missense mutations have been associated with inherited diseases. Accession numbers can be found in Supplementary Table S1. Missense variants in this study (highlighted) are located at highly conserved residues across species and different HD TFs. (C) Spec-seq experimental workflow (STAR Methods). (D) Spec-seq library design of monomeric HD binding sites. (E) EMSA gel images of Spec-seq experiments with different CRX HD species. Bx: Bound. B-: Unbound. (F) Relative binding energy comparison from two different experiments with WT HD. (G) Binding energy model for WT CRX HD. (H-J) Relative binding energy comparison between WT HD and R90W HD (H), E80A HD (I), or K88N HD (J). Consensus sequence is defined to have relative binding energy of 0kT (TAATCC for WT, R90W and E80A, TAATTA for K88N). The diagonal line of identity is represented in grey dash. The orange dashed line shows the best linear fit to the data. (K-M) Binding energy models for R90W HD (K), E80A HD (L), and K88N HD (M). Only sequence variants within two mismatches to the corresponding consensus sequences were used to generate binding models. Negative binding energy is plotted such that bases above the x-axis are preferred bases and bases below the x-axis are unfavorable bases. Constant bases (TAA) carried no information are drawn at arbitrary height in grey.

Intriguingly, several missense mutations within the same HD recognition helix as R90W, including *p.E80A* (E80A) and *p.K88N* (K88N), are linked to severe dominant IRDs^{(20)–(22)} (Figures 1A and 1B). Here, we utilized a multi-omics approach to investigate the functional consequences of the E80A and K88N mutations on CRX regulatory activities and photoreceptor development (Figure S1). Comparison of the *in vitro* HD-DNA binding models of CRX and disease variants generated by Spec-seq revealed unique specificity changes of each mutant protein. Introduction of each mutation into the endogenous *Crx* locus generated knock-in mouse models *Crx*^{E80A} and *Crx*^{K88N} that reproduced dCoRD- and dLCA-like phenotypes. ChIP-seq analysis of CRX binding *in vivo* revealed mutation-specific changes in

CRX targetome, consistent with mutation-specific DNA binding changes *in vitro*. Retinal RNA-seq analysis uncovered two distinct mechanisms by which the two HD missense mutations contribute to altered gene expression programs during photoreceptor differentiation and maturation. Our results highlight the importance of residues E80 and K88 in CRX-mediated transcriptional regulation during photoreceptor development and the diverse mechanisms by which HD missense mutations can affect TF functions and lead to severe dominant neurological diseases.

Results

K88N but not E80A mutation alters CRX HD DNA-binding specificity *in vitro*

CRX belongs to the *paired*-like HD TF family that recognize a 6-bp DNA motif in a stereotypic way^{(23)–(29)}. Extensive studies of the HD have revealed a canonical HD-DNA recognition model where recognition of the 3' region (bases 4–6) of the HD DNA binding site is mediated by specificity determinants within the conserved HD recognition helix, corresponding to CRX residues 80–96^{(23)–(26),(29),(30)} (Figure 1B). In particular, HD residue 50, equivalent to CRX K88 residue (Figure 1A), is the specificity determinant in *paired*-like HD TF-DNA interactions^{(24),(25)}. Since the three disease-associated HD missense mutations, E80A, K88N and R90W, are located within the CRX HD recognition helix, we wondered if these mutations change CRX HD DNA-binding specificity.

We adapted a high-throughput *in vitro* assay, Spec-seq, that determines protein-DNA-binding specificity by sequencing^{(31)–(33)}. Spec-seq was developed based on the traditional electrophoretic mobility shift assay (EMSA) to measure protein-DNA interactions. Spec-seq allows us to measure the relative binding affinities (i.e., specificity) for a library of HD binding motifs in parallel and generate quantitative binding models for different CRX HDs (Figures 1C). Based on the HD-DNA interaction model, we designed and tested a Spec-seq library containing all possible monomeric HD motifs (TAANN) (Figure 1D).

We first obtained the wild-type (WT) CRX HD DNA binding model with Spec-seq using bacterially-expressed and affinity-purified HD peptides (Figures 1E–1G, STAR Methods). Relative binding energies of CRX WT HD from two experiments showed strong correlation (r : 0.984) and noise level (0.114 kT) within the expected range in typical Spec-seq data (Figure 1F). Binding energy model of WT HD was then generated by applying multiple linear regressions on the relative binding energies of all sequences within two base-pair mismatches to the WT CRX consensus (TAATCC)⁽¹³⁾ (Figure 1G, STAR Methods). A clear preference for CC bases at the 3' end of the motif is consistent with known CRX binding preference *in vitro* and *in vivo*^{(34),(35)}.

We next sought to understand how disease mutations affect CRX DNA-binding specificity. We purified all mutant HD peptides following the same protocol as WT HD peptides and verified their DNA binding (Figures S2A–S2D)⁽¹⁷⁾. Comparison of the relative binding energies between each pair of mutant and WT HD revealed distinct effects (Figures 1H–1J). By definition, the consensus DNA binding motif of a testing peptide has a relative binding energy of 0 kT. The relative binding energy difference between nucleotide variants and the consensus motif correlates with the DNA binding specificity of the testing peptide. We found that R90W HD and E80A HD both prefer the same consensus motif as WT (Figures 1K and 1L). R90W HD bound with slightly higher specificity than WT, as demonstrated by most data points falling above the diagonal line (Figures 1H and 1K), suggesting that R90W HD is more sensitive than WT to binding sequence variations. In contrast, when comparing E80A HD with WT HD, many data points fell below the diagonal line and the relative binding energies

regressed towards 0 on the E80A axis (Figures 1I and 1L). This suggests that E80A HD bound with lower specificity than WT HD and thus was more tolerant to base variations in the HD DNA motif. Different from R90W and E80A, K88N mutation dramatically altered CRX HD DNA-binding specificity (r : 0.160) (Figures 1J and 1M). The K88N preferred binding sequence (TAAT/ATT/A) is referred to as N88 HD motif hereafter. K88N HD also had the largest degree of discrimination from its preferred to the weakest binding motif, suggesting that it is most sensitive to variants in the HD DNA motif. As a control, we tested a second library with the TAANN sites on the reverse strand and obtained similar results (Figures S2E-S2L). Together, these results indicate that while E80A mutation does not affect CRX HD DNA binding specificity, the K88N mutation dramatically alters the specificity *in vitro*.

E80A protein binds to WT sites while K88N occupies novel genomic regions with N88 HD motifs *in vivo*

Next, we asked if changes in DNA-binding specificity affected mutant CRX chromatin binding in developing photoreceptors. We first created two human mutation knock-in mouse models, Crx^{E80A} and Crx^{K88N} , each carrying a single base substitution at the endogenous *Crx* locus (Figures S3A and S3B, STAR Methods). We confirmed that *Crx* mRNA was expressed at comparable levels in WT and mutant retinas (Figure S3C), and the full-length CRX proteins were readily detectable in the nuclear extracts from all samples (Figure S3D). We then obtained genome-wide binding profiles for each CRX variant by chromatin immunoprecipitation followed by sequencing (ChIP-seq) on mouse retinas at P14, a time when all retinal cell types are born, photoreceptor specification is completed in WT animals, and prior to any observed cell death in other CRX mutants previously characterized^{(18),(36)}. To focus on changes specific to each mutant CRX protein, only homozygous animals were used for ChIP-seq profiling.

Unsupervised clustering of all CRX binding sites revealed two major clusters (Figure 2A, STAR Methods). Cluster 1 consisted of canonical WT CRX binding sites that are also occupied by CRX E80A protein (Figures 2A and 2B). Similar to WT CRX, CRX E80A binding *in vivo* was mostly enriched in intronic, followed by intergenic and TSS regions (Figure 3D). In contrast, CRX R90W, that also showed similar consensus preference to WT *in vitro*, failed to produce significant DNA binding *in vivo* (Figures 2A-2C, STAR Methods). This suggests that the retinopathy phenotype of $Crx^{R90W/W}$ is likely due to loss of binding at canonical WT CRX binding sites. Intriguingly, while CRX K88N showed loss of binding at canonical CRX binding sites, it gained a small set of binding sites (Cluster 2, Figures 2A-2C). *De novo* motif searching with DREME⁽³⁷⁾ under CRX peaks in each genotype revealed enrichment of monomeric HD motifs (Figure 2E) consistent with those found in Spec-seq (Figures 1G and 1K-1M), highlighted by a change in enriched HD motif from WT CRX HD type to N88 HD type in the $Crx^{K88N/N}$ retinas. Consistency with *in vitro* binding models suggests that *in vivo* changes in CRX chromatin binding were at least in part driven by the intrinsic changes in HD-DNA binding specificity by each individual mutation.

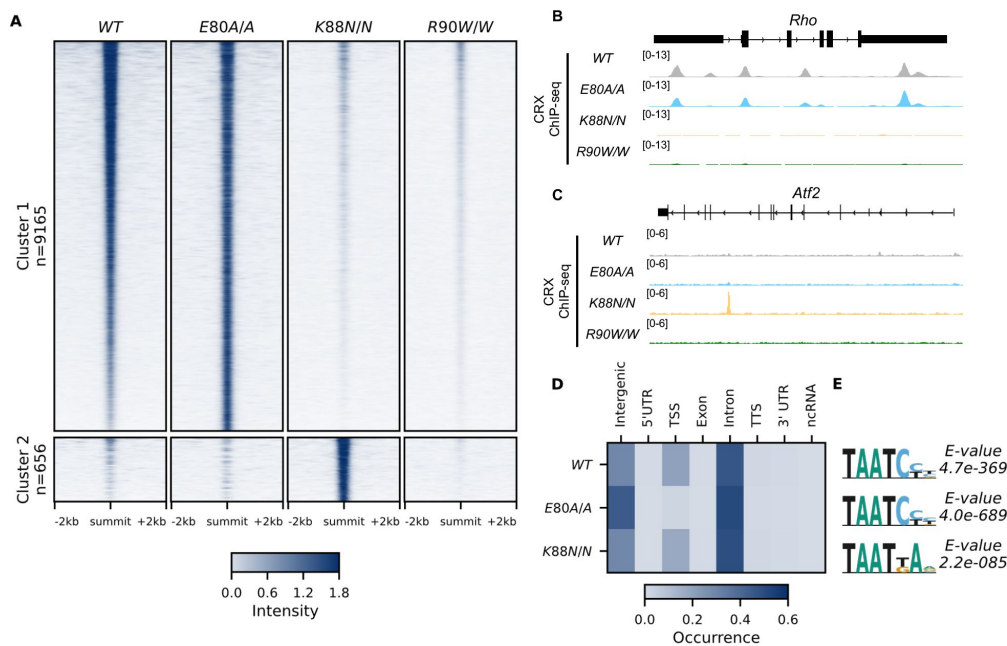


Figure 2.

CRX E80A binds to WT sites while CRX K88N occupies novel genomic regions enriched for N88 HD motif *in vivo*.

(A) Enrichment heat map depicting CRX ChIP-seq normalized reads centered at all possible CRX peaks $\pm$ 2kb, sorted by binding intensity in WT samples. Clusters were defined by hierarchical clustering of CRX binding intensity matrix from all genotypes (STAR Methods).

(B-C) Genome browser representations of ChIP-seq normalized reads for different CRX species in P14 WT and mutant mouse retinas at *Rho* and *Atf2*. (D) Enrichment heatmap showing fraction of CRX ChIP-seq peaks fall in different genomic environments. (E) Logo representations of *de novo* found short HD motifs under CRX ChIP-seq peaks in WT and mutant mouse retinas with DREME *E-value* on the right.

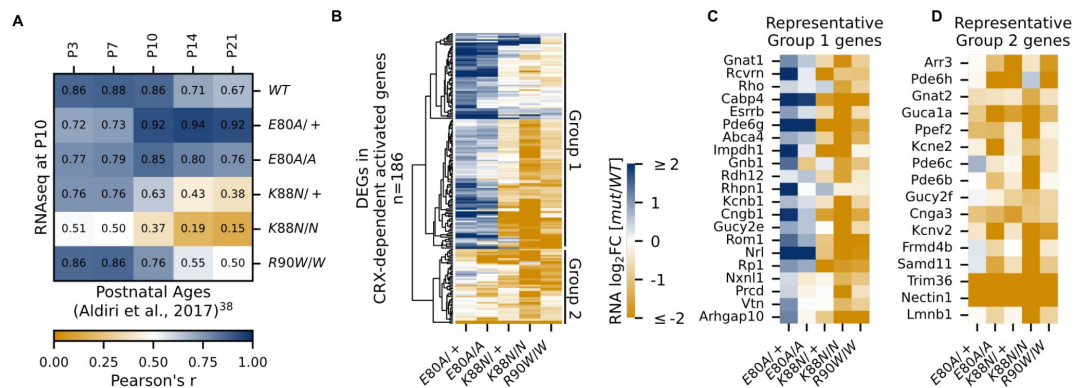


Figure 3.

CRX-dependent activated genes affected in opposite directions in developing *Crx*^{E80A} and *Crx*^{K88N} mutant retinas

(A) Heat map showing sample-wise Pearson correlations of the expression of all CRX-dependent activated genes between P10 WT and HD mutant mouse retinas in this study (rows) with post-natal WT retinas from age P3 to P21 (columns, data from GSE87064). (B) Heat map showing the expression changes of DEGs in CRX-dependent activated gene set in HD mutant mouse retinas at P10. (C-D) Heat maps showing expression changes of selected photoreceptor genes from Group 1 and Group 2. Color scale identical to (B).

E80A and K88N mutations affected the expression of CRX-dependent activated genes in opposite directions in a critical time window of photoreceptor differentiation

To understand how different CRX mutations affected gene expression at canonical and *de novo* binding sites and how these changes impair photoreceptor differentiation, we turned to bulk RNA-seq analysis from the developing retinas at P10. At P10, photoreceptors have started to differentiate, and the expression of many photoreceptor genes undergo exponential increase^{(38),(39)}. To focus on the most relevant expression changes, we first defined a set of genes that most likely depend on CRX activity nearby for expression (Figure S4A, STAR Methods). Briefly, we associated each CRX ChIP-seq peak to the nearest gene, filtered only genes with nearby CRX ChIP-seq peaks, and further narrowed the list of genes to those significantly down-regulated in the loss of function mutant *Crx*^{R90W/W}. Gene ontology (GO) analysis confirmed that this putative CRX-dependent gene set is associated with biological processes related to photoreceptor development and functions (Figure S4B). This set of putative CRX-dependent genes also showed developmental dependent gain in expression, consistent with CRX's primary function as a transcriptional activator (Figure S4C). As a control, CRX-independent genes were constitutively expressed and largely involved in general cellular processes (Figures S4D-S4F). Therefore, the CRX-dependent gene set comprises genes important for photoreceptor differentiation and functional maturation and are dependent on CRX for activation. We denote these genes as “CRX-dependent activated genes” (CRX-DAGs).

Next, we sought to understand how each mutation affected photoreceptor differentiation. One way of measuring the progression of photoreceptor differentiation is to determine the similarity in CRX-DAG expression in a given sample with that of known developmental ages in *WT* control animals. We thus performed sample-wise correlation of CRX-DAG expression obtained in our P10 samples with a previously published RNA-seq dataset of normal mouse retinal development (Figure 3A)⁽³⁸⁾. As expected, our P10 *WT* sample showed strong correlations with all developmental ages in the published *WT* control dataset. A stronger correlation with early ages (P3, P7, P10) and a weaker correlation with later ages (P14, P21) is also an indication of ongoing photoreceptor differentiation at P10. Unlike the *WT* sample, *Crx*^{E80A/+} and *Crx*^{E80A/A} samples both showed a stronger correlation with later developmental ages (P14, P21) but a weaker correlation with earlier postnatal ages (P3, P7). Since the CRX-DAGs are normally developmentally upregulated, this shift toward correlation with later developmental ages suggested that these genes were prematurely upregulated in the P10 *Crx*^{E80A} mutant retinas. In contrast, *Crx*^{K88N/+} and *Crx*^{K88N/N} samples both showed a weaker correlation with all developmental ages when compared with *WT* samples in our dataset. This suggests that early photoreceptor differentiation was compromised in both *Crx*^{K88N} mutants, consistent with their association with early-onset LCA⁽²²⁾. Importantly, *Crx*^{R90W/W}, also associated with LCA-like phenotype^{(18),(21)}, displayed strong correlation with earlier ages (P3, P7) similar to *WT*, but only showed moderate correlation with later ages (P14, P21). This suggests loss of CRX function at canonical binding sites does not affect the initiation of photoreceptor differentiation, but WT CRX activity at these sites is required to sustain differentiation. Since CRX-DAG expression was more severely affected in *Crx*^{K88N} mutants than in *Crx*^{R90W/W}, the photoreceptor differentiation deficits seen in the *Crx*^{K88N} mutants cannot be explained solely by the loss of regulatory activity at canonical CRX binding sites. Overall, our sample-wise correlation analysis with normal retinal development dataset suggests that E80A and K88N mutations affected the expression CRX-dependent activated genes in opposite directions, implicating novel and distinct pathogenic mechanisms from the loss-of-function R90W mutation.

***Crx*^{E80A} retinas show up-regulation of rod genes but down-regulation of cone genes, underlying CoRD-like phenotype**

Upon closer examination, we noted that not all *CRX*-DAGs were up-regulated in *Crx*^{E80A} mutants (Figure 3B). Hierarchical clustering of all *CRX*-DAGs using expression changes revealed two major groups (STAR Methods). In aggregate, when compared to WT, Group 1 genes were up-regulated in *Crx*^{E80A} mutants while Group 2 genes were down-regulated. We noted genes indicative of the two photoreceptor subtypes, rods and cones, could partially define the two groups (Figures 3C and 3D). For example, *Esrrb*⁽⁴⁰⁾ and *Nrl*⁽⁴¹⁾ in Group 1 are important regulators of rod differentiation. Other genes in Group 1 are components of the phototransduction cascade in rods, including *Rcvrn*⁽⁴²⁾, *Rho*⁽⁴³⁾, *Gnat1*^{(44),(45)}, *Pde6g*⁽⁴⁶⁾, *Abca4*^{(47),(48)}, *Gnb1*⁽⁴⁹⁾, *Rdh12*⁽⁵⁰⁾, *Cngb1*⁽⁵¹⁾, *Rp1*^{(52),(53)}. Mis-regulations of many of these genes have been associated with diseases that affect rod development, function, and long-term survival. The increased activation of these genes likely underlies the stronger correlation with later developmental ages in *Crx*^{E80A/+} and *Crx*^{E80A/A} retinas (Figure 3A). In contrast, Group 2 genes, many down-regulated in *Crx*^{E80A} mutants, were implicated in cone development and functions. For example, *Gnat2*^{(54),(55)}, *Pde6c*⁽⁵⁶⁾, and *Pde6h*^{(57),(58)} all act in the cone phototransduction cascade. Mis-regulation of these genes has also been implicated in different retinal dystrophies that primarily affect cone photoreceptors. Comparison of ChIP-seq signal revealed that peaks associated with Group 2 genes showed lower occupancy compared to Group 1 genes in the *Crx*^{E80A/A} retinas (Mann-Whitney U test *p*-value: 9.51e-07) while no difference was observed in WT retinas (Mann-Whitney U test *p*-value: 0.541), suggesting loss of CRX activity likely underlies the down-regulation of Group 2 genes in the *Crx*^{E80A} mutants (Figure S5A). Collectively, the selective down-regulation of cone genes in Group 2 may explain the CoRD-like phenotype in adult *Crx*^{E80A} mutant mice described later.

Additionally, we noticed that a subset of genes not affected in *Crx*^{R90W/W} (CRX-independent genes) were also down-regulated in *Crx*^{E80A} mutants (Figures S5B-S5D). Among these genes were transcription regulators important for early photoreceptor development, such as *Ascl1*⁵⁹, *Rax*⁶⁰, *Sall3*⁽⁶¹⁾, and *Pias3*⁽⁶²⁾. The down-regulation of these factors coincided with the up-regulation of mature rod genes in P10 *Crx*^{E80A} retinas, suggesting that the E80A mutation might hamper the proper timing of photoreceptor differentiation.

***Crx*^{K88N} retinas display greater reduction of rod and cone genes than the loss-of-function mutants**

Crx^{K88N} retinas had the most severe gene expression changes among all mutants with down-regulation of both Group 1 (rod) and Group 2 (cone) genes (Figures 3B-3D). The heterozygous *Crx*^{K88N/+} retina displayed a similar degree of expression reduction as homozygous *Crx*^{R90W/W}, consistent with its association with dLCA⁽²²⁾. Given the normal phenotype of heterozygous loss-of-function mutants -*Crx*^{+/-} and *Crx*^{R90W/+18}, these results suggested that mutant CRX K88N not only failed to activate WT target genes, but also functionally antagonized WT CRX regulatory activity in differentiating photoreceptors. This antagonism might be associated with ectopic CRX K88N activity when bound to regulatory regions with N88 HD DNA motifs (Figures 2A and 2E). In the absence of WT CRX, *Crx*^{K88N/N} retina displayed a more severe expression reduction of *CRX*-DAG than *Crx*^{R90W/W}, raising the possibility that CRX K88N also antagonized the activity of other transcriptional regulators important for photoreceptor differentiation. Supporting this possibility, a set of CRX-independent genes were also mis-regulated in both heterozygous and homozygous *Crx*^{K88N} mutants, we noted that a number of these down-regulated genes are also involved in photoreceptor functional development (Figures S5B, S5E, and S5F). Overall, CRX K88N is

associated with greater gene expression changes than other CRX HD mutants, which is likely attributed to ectopic regulatory activity.

Both *E80A* and *K88N* mutants show compromised rod/cone terminal differentiation in young adults

Since *Crx*^{+/-} and *Crx*^{R90W/+} mutant mouse models showed a late-time recovery in photoreceptor gene expression and function⁽¹⁹⁾, we sought to determine the degree of photoreceptor differentiation in *Crx*^{E80A} and *Crx*^{K88N} mutants at P21. At this age, the normal retina has largely completed terminal differentiation with photoreceptor gene expression reaching a plateau⁽⁶³⁾. However, when P21 CRX HD mutant retinas were examined for the expression of genes under the GO term *detection of light stimulus* (GO:0009583), which comprises genes in both rod and cone phototransduction cascades, many genes failed to reach *WT* levels, despite variable degrees of impact across different HD mutants (Figure 4A). *Crx*^{E80A} mutants, in contrast to the increased rod gene expression at P10, displayed a deficiency in both cone and rod phototransduction genes at P21 (Figure 4B, Supplementary Table 6). This suggests that premature gene mutant CRX E80A transcriptional activity fails to sustain photoreceptor terminal differentiation and ultimately results in non-functional and severely affected photoreceptors. In comparison, *Crx*^{K88N} mutants showed severely reduced expression of rod/cone phototransduction genes at both P10 and P21 (Figures 3B-3D and 4B).

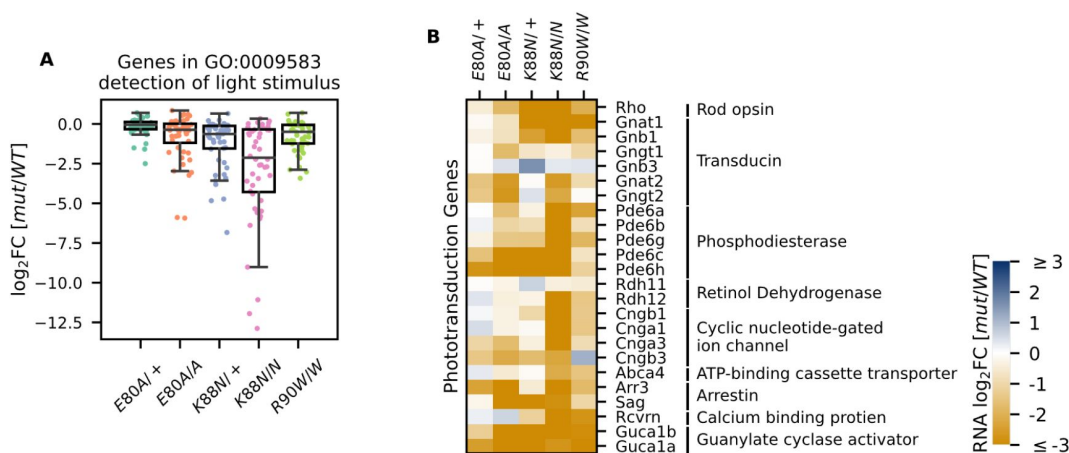


Figure 4.

Photoreceptor genes important for phototransduction are down-regulated in all HD mutants

(A) Box plot showing that genes in the detection of light stimulus GO term were down-regulated and affected to various degrees in different adult (P21) HD mutant mouse retinas. (B) Heat map showing that expression of both cone and rod phototransduction genes were down-regulated in adult (P21) HD mutant mouse retinas. Annotation of rod and cone enrichment of each gene is in Supplementary Table 6. See [Supplementary Figure S6](#) for the developmental expression dynamics of these genes.

To assess retinal morphology and photoreceptor subtype-specific gene expression at the cellular level, we performed immunohistochemistry analysis on P21 retinal sections. In *WT* animals, a hallmark of photoreceptor maturation is the outgrowth of photoreceptor outer segments (OS) filled with proteins necessary for the phototransduction. We thus performed

hematoxylin and eosin (H&E) staining on P21 sagittal retinal sections to visualize changes in retinal layer organization, focusing on photoreceptor layers – outer nuclear layer (ONL) and OS. Compared to the well-organized ONL in *WT* retinas (Figure 5A), all mutants showed variable degrees of ONL disorganization, forming waves, whorls, and rosettes (Figures 5B–5E). The ONL disorganization was more severe in homozygotes than in heterozygotes for both mutations and *Crx*^{K88N} mutants were more severely affected than *Crx*^{E80A} mutants. Photoreceptor OS layer was formed in the *Crx*^{E80A/+} retinas, but absent in *Crx*^{E80A/A}, *Crx*^{K88N/+}, and *Crx*^{K88N/N} mutant retinas. Inner retinal layers, including inner plexiform layer (IPL) and ganglion cell layer (GCL) were not as severely affected as the outer retinal layers, supporting a model that the mutant morphological abnormalities largely originated from the diseased photoreceptors. These morphological abnormalities were distinct from the degenerative phenotypes of other *Crx* mutant models reported previously^{(18),(64),(65)}.

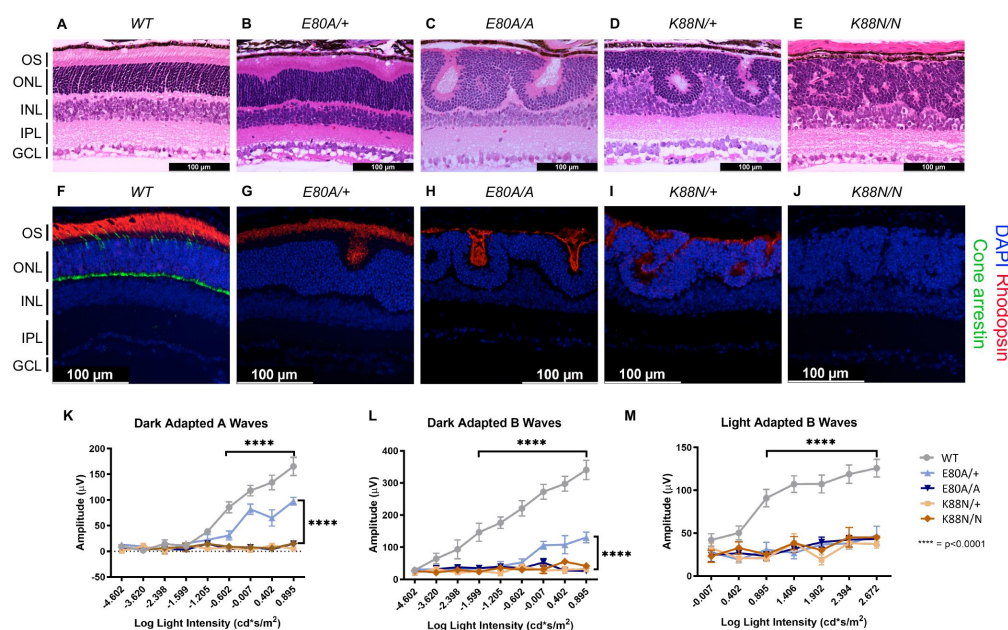


Figure 5.

Only *Crx*^{E80A/+} retinas maintain photoreceptor OS and residual rod ERG response

(A–E) Hematoxylin-eosin (H&E) staining of P21 retina sections show that photoreceptor OS layer is absent in all mutant retinas except *Crx*^{E80A/+}. OS: outer segment; ONL: outer nuclear layer; INL: inner nuclear layer; IPL: inner plexiform layer; GCL: ganglion cell layer. Scale bar, 100µm. (F–J)

Rhodopsin (RHO, red) immunostaining is present in *Crx*^{E80A/+}, *Crx*^{E80A/A}, and *Crx*^{K88N/+} retinas and absent in *Crx*^{K88N/N} retina. Cone arrestin (mCAR, green) immunostaining is absent in all mutant retinas. Nuclei were visualized by DAPI staining (Blue). Scale bar, 100µm. (K–M) The electroretinogram responses (ERG) recorded from 1-month mice. Error bars represent the standard error of the mean (SEM, $n \geq 4$). *p*-value: Two-way ANOVA and Tukey's multiple comparisons. ****: $p \leq 0.0001$. ns: >0.05 .

Next, we sought to determine the expression of the rod-specific visual pigment rhodopsin (RHO) and cone arrestin (mCAR) in the P21 mouse retinas. In *WT* retinas, RHO is trafficked to the rod OS while mCAR is present in the cone OS and IS (inner segment), cell body, and synaptic terminals (Figure 5F). Unlike *WT* retina, all mutants lacked mCAR immunoreactivity (Figures 5G–5J), consistent with the loss of cone gene expression shown by RNAseq (Figure 4B). In *Crx*^{E80A/+} retinas, RHO staining was localized to the OS layer; in *Crx*^{E80A/A} and *Crx*^{K88N/+} retinas, positive RHO staining was observed within the whorls and rosettes; in *Crx*^{K88N/N} mutant retinas, RHO staining was completely absent. Importantly, we did not observe mis-localized RHO staining in the inner retinal layers suggesting that the developmental programs of other retinal cell types were not directly affected by E80A or K88N mutation. Overall, abnormalities in the cone/rod gene expression matched the corresponding human disease diagnosis^{(20),(22)}, and the phenotypic severity correlated with the degree of mis-regulation of CRX target genes in the corresponding RNAseq dataset. Thus,

these results support a model that CRX HD mutation-mediated mis-regulation of gene expression disrupts photoreceptor terminal differentiation and leads to defects in retinal layer organization and OS formation.

***E80A* and *K88N* mouse models show visual function deficits that recapitulate human diseases**

To understand the consequences of disrupted photoreceptor differentiation on visual function, we measured electroretinogram (ERG) responses to light stimuli for *WT* and mutant mice at one month of age (Figures 5K-M). In response to incremental changes of light intensities, *WT* animals showed corresponding amplitude increases in dark adapted A-waves (rod signals) and B-waves (rod-evoked bipolar cell signals), as well as in light-adapted B-waves (cone-evoked bipolar cell signals). The three severe mutants, *Crx*^{E80A/A}, *Crx*^{K88N/+}, and *Crx*^{K88N/N} had no detectable dark-adapted or light-adapted ERG responses, suggesting that these mice have no rod or cone function and are blind at young ages. The null ERG phenotype of the *Crx*^{K88N} animals is consistent with the clinical LCA phenotype in humans⁽²²⁾. In contrast, *Crx*^{E80A/+} animals retained partial rod ERG responses as indicated by the reduced A-wave and B-wave amplitudes (Figures 5K and 5L). Yet, *Crx*^{E80A/+} animals had no detectable cone ERG responses, which is consistent with the CoRD clinical phenotype in humans⁽²⁰⁾ (Figure 5M). Taken together, the visual function impairment in each CRX HD mutant model, coincided with the morphological and molecular changes, suggesting that *Crx*^{E80A} and *Crx*^{K88N} mouse models recapitulate the corresponding human diseases.

CRX E80A has increased transactivation activity and leads to precocious differentiation in *Crx*^{E80A} retinas

Lastly, we asked what might be the molecular mechanism that causes the mis-regulation of photoreceptor genes in the mutant retinas. Previous studies have established that reporter assays with the *rhodopsin* promoter in HEK293T cells can measure changes of CRX transactivation activity and inform the mechanisms by which photoreceptor genes are mis-regulated in CRX mutant retinas⁽¹⁷⁾. We thus tested the transactivation activity of the three CRX HD mutants on the *pRho-Luc* reporter in HEK293T cells (Figure 6A, STAR Methods). Consistent with published studies, R90W mutant had significantly reduced activity compared to WT CRX⁽¹⁷⁾. K88N mutant showed a similarly reduced activity as R90W, consistent with K88N's loss of binding at canonical CRX sites and failure to activate CRX-DAGs *in vivo* (Figures 2A, 2B, 3A-3C). In contrast, E80A mutant, which binds to canonical CRX sites, showed significantly increased transactivation activity on the *Rho* promoter. This hyperactivity of E80A protein at *Rho* promoter correlates with the upregulation of CRX-DAGs in the mutant retinas at P10 (Figures 3A-3C).

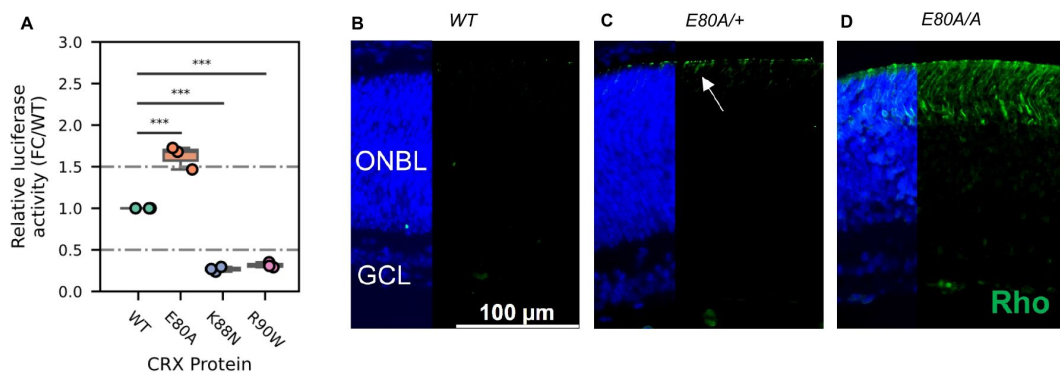


Figure 6.

CRX E80A hyperactivity underlies precocious photoreceptor differentiation in Crx^{E80A} retinas.

(A) Boxplot showing luciferase reporter activities of different CRX variants. *P*-values for one-way ANOVA with Turkey honestly significant difference (HSD) test are indicated. *p*-value: ****: ≤ 0.0001 , ***: ≤ 0.001 , ns: > 0.05 . (B-D) Rhodopsin (RHO, green) immunostaining is absent in P3 WT retina but detected in $Crx^{E80A/+}$ and $Crx^{E80A/A}$ retinas. Nuclei are visualized by DAPI staining (Blue). Arrow indicates the sporadic RHO staining in $Crx^{E80A/+}$ sample. ONBL: outer neuroblast layer; GCL: ganglion cell layer. Scale bar, 100 μ m.

To determine the consequences of E80A hyperactivity on photoreceptor differentiation timing, we compared RHO protein expression during early postnatal retinal development using retinal section immunostaining (Figures 6B-6D). In WT retinas, most rods were born by P3 but had not differentiated (Figure 6B). Previously studies showed that RHO proteins were detected by IHC starting around P7 in WT retinas⁽⁶³⁾. In comparison, both $Crx^{E80A/+}$ and $Crx^{E80A/A}$ retinas showed positive RHO staining at P3 (Figures 6C and 6D). RHO⁺ cells were largely seen in the outer portion of the ONBL layer in $Crx^{E80A/+}$ retinas, and strikingly spread throughout the large presumptive ONL layer in $Crx^{E80A/A}$ retinas. The detection of RHO protein in P3 Crx^{E80A} mutant retinas indicates that photoreceptor differentiation program was precociously activated. Taken together, our results support a model that E80A and K88N mutations each perturbs CRX regulatory activity in a unique way, causes photoreceptor differentiation defects, and ultimately leads to distinct dominant disease phenotypes that recapitulate human diseases.

Discussion

Through molecular characterization of mutant proteins, transcriptome and cellular profiling of developing mutant mouse retinas and ERG testing of adult retinas, we have identified two novel pathogenic mechanisms of CRX HD mutations, E80A and K88N, that are associated with dominant CoRD and dominant LCA in human^{(20),(21)}. Distinct from the previously characterized loss-of-function R90W mutation^{(18),(19)}, E80A and K88N mutations produce altered CRX proteins with gain of regulatory functions -CRX E80A is associated with increased transcriptional activity and CRX K88N has altered DNA binding specificity. Both CRX E80A and CRX K88N proteins impair photoreceptor gene expression, development and produce structural and functional deficits in knock-in mouse models, recapitulating human

diseases⁽⁶⁴⁾ (Figure 7). Thus, precise regulatory activity and target specificity of CRX are both essential for proper photoreceptor development and functional maturation.

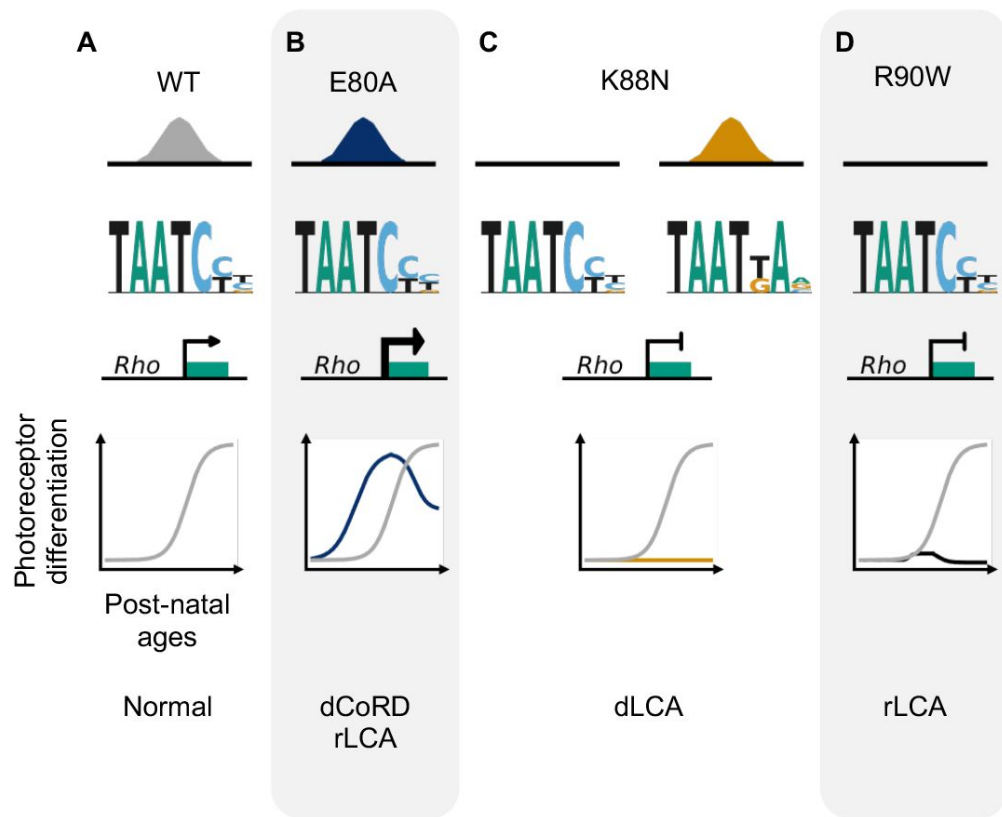


Figure 7.

Missense mutations in CRX HD affect photoreceptor gene expression and leads to distinct retinal disease phenotypes through gain- and loss-of-function mechanisms.

The *Crx*^{E80A} and *Crx*^{K88N} knock-in mouse models recapitulate dominant human CRX-linked retinopathies and present novel morphological deficits. Published *Crx* mutant mouse models have established that decreased expression of CRX or loss of CRX transcriptional activity results in incomplete photoreceptor differentiation and ultimately rapid degeneration of immature photoreceptors^{(15),(18)}. Despite the difference in molecular mechanisms, both *Crx*^{E80A} and *Crx*^{K88N} mouse models show defective photoreceptor terminal differentiation and develop whorls and rosettes in the ONL layers by P21, which are not observed in degenerative CRX mouse models⁽¹⁸⁾. No significant increase in cell death is observed in *Crx*^{E80A} and *Crx*^{K88N} mutant mouse retinas at P21, suggesting distinct pathogenic mechanisms than the photoreceptor degeneration previously reported in other CRX mouse models. P10 RNAseq analysis on cell markers of other neuronal cell types and Muller glia cells did not detect significant changes, suggesting that the developmental programs for these cell types were unaffected. Further studies are needed to systematically characterize changes in photoreceptor differentiation and retinal morphology during early development, and interactions between diseased photoreceptors and other major cell types in *Crx*^{E80A} and *Crx*^{K88N} mutant mouse retinas.

Precisely tuned CRX-mediated transcriptional activity is important to ensure proper photoreceptor development. Previous studies have demonstrated that CRX is activated shortly after cell cycle exit in retinal progenitor cells fated to be photoreceptors^{(66),(67)}. CRX then works with other cell type specific TFs, such as NRL and NR2E3 (rods) or RORB and THRB (cones), to regulate gene expression important for cell-type specific structures and functions and drive the differentiation of photoreceptor precursors^{(68),(69)}. Loss-of-function animal studies have established that CRX is primarily a transcriptional activator since loss of CRX protein or regulatory function is associated with decreased expression of photoreceptor-specific genes regulated by CRX⁽¹⁹⁾. Here, we find that CRX E80A mutant, that

displays increased CRX transcriptional activator activity, is also deleterious to photoreceptor development. CRX E80A maintains WT DNA-binding specificity *in vitro* and binds to canonical CRX sites in mutation knock-in mouse retinas. Consistent with CRX E80A's hyper-activity at canonical CRX target *pRho*, in differentiating *Crx*^{E80A/+} and *Crx*^{E80A/A} retinas, many of the CRX-dependent activated genes implicated in rod functions are prematurely up-regulated at early postnatal ages. The RHO protein is also detected in developing rod photoreceptor in *Crx*^{E80A/+} and *Crx*^{E80A/A} retinas as early as P3. However, at later ages, rod photoreceptor terminal differentiation is impaired, resulting in disrupted retinal morphology and defective visual functions. Photoreceptor differentiation is precisely programmed via sequential and concerted gene expression programs within a defined time window^{(66),(67)}. One explanation for adult *Crx*^{E80A} phenotypes is that CRX E80A associated hyper-activity precociously activates later stage genes in the absence of proper nuclear context and/or subcellular structures, which in turn negatively impacts early events in photoreceptor differentiation. Multiple possibilities could account for CRX E80A associated hyper-activity: (1) E80A mutation increases CRX interactions with its binding sites; (2) E80A mutation increases CRX interactions with the transcription complex or other TFs; (3) certain protein factors that could promote photoreceptor gene expression are up-regulated in CRX E80A expressing cells.

While these possibilities are not mutually exclusive, evidence from the current study favors the first possibility since E80A HD is less discriminative of HD motif variants than WT HD as shown by Spec-seq (Figure 1I). Regulatory sequences of many photoreceptor genes contain both consensus and non-consensus CRX motifs^{(13),(14),(17),(34)}. CRX E80A might interact better with non-consensus CRX motifs within binding sites, resulting in increased transcriptional output of target genes. Large scale reporter assays (e.g. Massively Parallel Reporter Assays) in the appropriate cellular context – neonatal mouse retinas – would be informative to understanding CRX E80A hyper-activity mechanism and the direct impacts of its hyper-activity.

Although a global increase in expression is expected in *Crx*^{E80A} models, a subset of CRX dependent activated genes implicated in cone photoreceptor development and functions is down-regulated in both differentiating and mature mutant retinas. Cones are born prenatally in mice within an earlier time window than rods. Yet, cones undergo terminal differentiation in a similar postnatal window as rods to develop into subtypes of cones – M-cones v.s. S-cones. One model is that CRX E80A improperly activates later stage cone genes at a much earlier time window, disrupting cone terminal differentiation. Supporting this model, similar to the presented *pRho* data, CRX E80A also hyper-activates the *S-cone opsin* promoter (data not shown).

Detecting premature activation of cone genes in *Crx*^{E80A} retinas will require analysis at early developmental ages. An alternative model for lack of cone markers in *Crx*^{E80A} retinas is that cones might be more sensitive to perturbations in CRX transcriptional activity, which impacts cone survival. It is known that cones depend on a different repertoire of TFs than rods for subtype terminal differentiation⁽⁵⁹⁾. It is possible that cone TFs respond differently to mutant CRX E80A hyper-activity, leading to the distinct expression changes in *Crx*^{E80A} mutant retinas. Since cones only make up a very small portion of photoreceptors in mouse retinas⁽⁷⁰⁾, characterization of CRX E80A activity in purified cone population at earlier ages will be necessary to understand its selective effect on the cone differentiation program and help to elucidate WT CRX regulatory principles in early photoreceptor development. Collectively, CRX transcriptional activity at its canonical genomic binding sites need to be maintained within a quantitatively defined range such that CRX dependent target genes are activated in a precise temporal pattern to ensure proper photoreceptor differentiation.

CRX DNA binding specificity constitutes an additional regulatory program required for the fidelity of photoreceptor gene expression during development. Previous work has shown

that subtle differences in HD motif content could have distinct transcriptional and functional outcomes. For instance, studies in *D. melanogaster* demonstrate that the sequential use of HD DNA motif variants defines different classes of enhancers and coordinates the precise timing of gene expression patterns during embryonic development⁽⁷¹⁾. In our current study, the K88N mutation alters CRX DNA-binding specificity that redirects CRX binding to binding sites with N88 HD motifs in the mutant mouse retinas. *Crx*^{K88N} retinas show more severe perturbations in photoreceptor gene expression and development than in loss-of-function mutant *Crx*^{R90W/W}. Concomitantly, *Crx*^{K88N/+} and *Crx*^{K88N/N} mice display the most severe photoreceptor morphological deficits observed in any *Crx* mouse models and show absence of visual functions in young adults. In developing *WT* mouse retinas, the HD motif preferred by HD TFs with a glutamine (Q) at HD50 position encodes quantitatively different activity than the CRX consensus motif suggesting functional difference between HD binding site variants^{(60),(72)}. It is likely that the severe *Crx*^{K88N} phenotypes are attributed to both loss of activity at canonical CRX motifs and ectopic binding and transcriptional activity of CRX K88N proteins at N88 HD motifs. Additionally, since a functional *WT* copy of CRX is retained in *Crx*^{K88N/+} retinas and CRX K88N prefers different motifs than *WT* CRX, the lack of *WT* activity alone cannot explain the severe expression and developmental deficits.

Alternatively, these results suggest involvement of additional regulatory mechanisms: CRX K88N activity at N88 HD motifs might (1) ectopically activate genes whose expression prevents the progression of development or inactivate genes required for development; (2) interfere with other HD TFs that also recognize N88 HD motifs; (3) lead to epigenetic alterations that antagonize normal CRX functions. Many other HD containing TFs are expressed in developing mouse retina, including OTX2, RAX, VSX2, PAX6, SIX3/6, and LHX family⁽⁶⁾. Different from CRX, these HD TFs are essential for gene regulation in retinal progenitor cells and/or in other retinal cell lineages. It is very likely that CRX correctly targets photoreceptor genes at least through its DNA binding specificity and alteration of CRX DNA binding specificity would mis-regulate genes originally targeted by other HD TFs and lead to perturbations in GRNs and development. To date, most studies have focused on CRX activity at cis-regulatory sequences enriched for the *WT* CRX consensus motifs^{(73),(74)}. Systematic comparison of transcriptional activity at N88 HD motifs and *WT* consensus motifs in the context of photoreceptor development in both *WT* and mutant retinas would be needed to substantiate the impact of mutant CRX K88N activity at different HD motifs. These experiments will also help clarify the pathogenic mechanisms in the *Crx*^{K88N} models and extend our knowledge of CRX regulatory grammar and activation of photoreceptor genes during development.

One limitation of this work is that effects of E80A and K88N mutations on CRX HD-DNA interactions have been evaluated at monomeric HD motifs and with homogenous protein species both *in vitro* and *in vivo*. Further evaluation of *WT* and mutant CRX binding at dimeric motifs will be desirable, since selected dimeric HD motifs are known to mediate HD TF cooperativity to ensure gene expression fidelity during development^{(75),(76)}. Relatedly, we also need to address how CRX *WT* and mutant E80A or K88N proteins interact at HD binding motifs – whether they cooperate or compete with each other, whether these interactions are HD motif sequence-dependent, and how gene expression is impacted by CRX cooperativity or competition. While CRX HD mediates both TF-DNA interactions and protein-protein interactions, evaluation of how E80A and K88N mutations impact CRX interaction with other important photoreceptor TFs and how perturbations in these interactions lead to disease phenotypes warrant further study.

Since the spatial structures and HD-DNA contact models of HD proteins are evolutionarily conserved, our study of CRX provides valuable molecular insights for HD mutations implicated in other diseases. For example, *p.E79K* substitution (corresponds to CRX E80) in OTX2 HD is associated with dominant early-onset retinal dystrophy in two Caucasian Canadian families⁽⁷⁷⁾, heterozygous *p.R89G* (corresponds to CRX R90) mutation in OTX2 HD

causes severe ocular malformations⁽⁷⁸⁾, and missense mutations of the CRX K88 and R90 homologous residues in PITX2 HD are associated with Rieger syndrome⁽⁷⁹⁾. It is likely that these mutations affect HD activity in similar ways as observed in CRX, and the exact disease manifestation is determined by cell-type specific mechanisms. The retina is readily accessible and a broad range of molecular tools are available for *ex vivo* and *in vivo* manipulations. We believe that CRX is an ideal model to understand the pathogenic mechanisms of HD mutations and to test therapeutic regimens, which would ultimately benefit the study of HD TFs and their associated diseases in other tissues and organs.

Collectively, our findings support a unifying model in which precise CRX interaction with cis-regulatory sequences is essential for gene expression and functional maturation during photoreceptor development. Mutations that perturb CRX target specificity or transactivation activity impair the integrity of CRX-mediated transcriptional regulation and result in distinct morphological and functional deficits (Figures 7A-7D). Regulatory activity at ectopic N88 HD motifs antagonizes normal CRX functions and leads to aberrant photoreceptor development and early vision loss. This model also emphasizes the importance of tailoring gene therapy regimens to tackle individual pathogenic mechanisms. For instance, while supplementing WT CRX might be sufficient to rescue a hypomorphic/loss-of-function mutant, simultaneous elimination of a gain-of-function CRX product would be necessary to rescue dominant mutants, as exemplified in a recent report of allele-specific gene editing to rescue dominant CRX-associated LCA7 phenotypes in a retinal organoid model⁽⁸⁰⁾. We believe that this principle also applies to other dominant neurological diseases. Additionally, with the refinement of the CRX mechanistic model, when new disease mutations are identified, genetic counsellors can now provide more informed predictions of disease progression and future visual deficits. This information is important for individuals to be psychologically prepared and seek necessary assistance to improve their quality of life.

Author Contributions

S. Chen conceived and supervised the study. S. Chen, P. Ruzycki, C. Sun and Y. Zheng designed the experiments. Y. Zheng performed Spec-seq and luciferase experiments; X. Zhang performed CRX ChIP-seq experiments; C. Sun performed immunochemistry and ERG experiments. Y. Zheng analyzed Spec-seq, CRX ChIP-seq, RNA-seq, and luciferase data; P. Ruzycki assisted in CRX ChIP-seq and RNA-seq data analysis; C. Sun analyzed immunochemistry and ERG data. Y. Zheng wrote the original draft; S. Chen, P. Ruzycki, C. Sun, and Y. Zheng revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Mingyan Yang and Guangyi Ling for technical assistance, Susan Penrose and Mike Casey from the Molecular Genetics Service Core for generating E80A and K88N mutation knock-in mice lines, Inez Oh for RNA-seq sample collection and processing, and J. Hoisington-Lopez and M. Crosby from DNA Sequencing Innovation Lab at the Center for Genome Sciences & Systems Biology for sequencing assistance. This work was supported by NIH grants EY012543 (to S. Chen), EY032136 (to S. Chen), EY002687 (to WU-DOVS), and unrestricted funds from Research to Prevent Blindness (to WU-DOVS). We also thank Mr. Artur Widlak for the generous gift from Widlak Family CRX Research Fund.

Declaration of Interests

The authors declare no competing interests.

Inclusion and diversity

We support inclusive, diverse, and equitable conduct of research.

STAR Methods

Resource Availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Shiming Chen (chenshiming@wustl.edu).

Materials Availability

- All unique/stable reagents generated in this study are available from the lead contact with a completed materials transfer agreement.

Data and Code Availability

- The raw sequencing data and processed data generated in this study have been deposited at NCBI.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Animal study and sample collection

Ethics statement

All procedures involving mice were approved by the Animal Studies Committee of Washington University in St. Louis and performed under Protocol 21-0414 (to SC). Experiments were carried out in strict accordance with recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health (Bethesda, MD), the Washington University Policy on the Use of Animals in Research; and the Guidelines for the Use of Animals in Visual Research of the Association for Research in Ophthalmology and Visual Sciences. Every effort was made to minimize the animals' suffering, anxiety and discomfort.

Mutation knock-in mouse model generation

CRISPR/Cas9 based genome editing was performed to generate the *Crx*^{E80A} and *Crx*^{K88N} mice as previously described⁽⁸¹⁾. The Cas9 guide RNAs (gRNA) were designed based on proximity to the target amino acid and was synthesized using the MEGashortscript™ T7 Transcription Kit (Thermo Fisher Scientific, Waltham, MA). The gRNAs were subsequently tested for cutting efficiency in cell culture by the Washington University Genome Engineering and iPSC Center. The validated gRNA and Cas9 protein were then microinjected into the pronuclei of C57BL/6J-0.5-dpc (days post coitum) zygotes along with the donor DNA, a 190-bp single stranded oligodeoxynucleotide (ssODN) carrying either the *c.239A>G* substitution for *p.E80A* mutation or the *c.264G>T* substitution for *p.K88N* mutation⁽⁸²⁾. Embryos were then transferred into the oviduct of pseudo-pregnant female. Pups were generally delivered ~20 days after microinjection. Tissues from 10-day postnatal (P10) pups were collected by toe biopsy/tail for identification of the targeted allele by restriction digest (HinfI) of PCR amplified DNA first and then confirmed by Sanger sequencing (Genewiz).

Founders carrying the correct alleles were then bred with wild-type C57BL/6J mice (Jackson Laboratories, Bar Harbor, ME, Strain #000664) to confirm transmission. All experimental animals used were backcrossed at least 10 generations. Genotyping of mutation knock-in mice follows cycling conditions: 95°C for 2 min, 94°C for 30 sec, 60°C for 30 sec, 68°C for 60 sec, repeat steps 2-4 for 34 cycles, 68°C for 7 min, and hold at 4°C. After PCR reaction, the amplified DNA fragments were digested with HinfI. Sequences of gRNAs, ssODNs and genotyping primers can be found in Supplementary Table S3. A representative DNA gel of HinfI digested genotyping DNA fragments can be found at [Figure S3B](#).

RNA-seq sample collection and library preparation

For each genotype, three biological replicates, two retinas per replicate from one male and one female mouse were analyzed. All retinas were processed for RNA simultaneously using TRIzol™ Reagent (Invitrogen, Waltham, MA) following the manufacturer's protocol. The quantity and quality of the RNA was assayed using Bioanalyzer (Agilent, Santa Clara, CA). Samples with a minimum RNA integrity number (RIN) score of 8.0 were then selected for library construction as previously described⁽¹⁹⁾.

Chromatin immunoprecipitation and library preparation

CRX chromatin immunoprecipitation was performed as previously published⁽⁸³⁾. Briefly, pooled nuclear extracts from 6 retinæ were cross-linked with formaldehyde prior to immunoprecipitation with anti-CRX antibody A-9 (#sc-377138, Santa Cruz Biotechnology, Dallas, TX). Input controls were included as background. The libraries were prepared following the standard ChIP-seq protocol⁽⁸⁴⁾. The quantity and quality of the ChIP-seq libraries was assayed using Bioanalyzer (Agilent, Santa Clara, CA) prior to sequencing.

ERG and Statistical Analyses

ERGs were performed on 1-month-old mice using UTAS-E3000 Visual Electrophysiology System (LKC Technologies Inc., MD). Mice were dark-adapted overnight prior to the tests. Mouse body temperature was kept at 37 ± 0.5 °C during the tests. Pupils were dilated with 1% atropine sulfate solution (Bausch and Lomb). Platinum 2.0 mm loop electrodes were placed on the cornea of each eye. A reference electrode was inserted under the skin of the mouse's head and a ground electrode was placed under the skin near mouse's tail.

Retinal response to full-field light flashes (10 µs) of increasing intensity were recorded; maximum flash intensity for dark-adapted testing was 0.895 cd*s/m². Following dark

adapted tests, mice were light-adapted under light condition (about 29.2 cd/mm) for 10 minutes and exposed to 10 μ s light flashes of increasing intensity; maximum flash intensity for light-adapted testing was 2.672 cd*s/m². ERG responses of biological replicates were recorded, averaged and analyzed using Graphpad Prism 8 (GraphPad Software, CA). The mean peak amplitudes of dark-adapted A and B waves and light-adapted B waves were plotted against log values of light intensities (cd*s/m²). The statistics were obtained by two-way ANOVA with multiple pairwise comparisons (Tukey's).

Histology and Immunohistology chemistry

Enucleated eyes were fixed at 4°C overnight for paraffin embedded sections. Each retinal cross-section was cut 5 microns thick on a microtome. Hematoxylin and Eosin (H&E) staining was performed to examine retinal morphology. For IHC staining, sections firstly went through antigen retrieval with citrate buffer, and blocked with a blocking buffer of 5% donkey serum, 1% BSA, 0.1% Triton X-100 in 1x PBS (pH-7.4) for 1 h. Sections were then incubated with primary antibodies at 4°C overnight. Sections were washed with 1x PBS containing 0.01% Triton X-100 (PBST) for 30 min, and then incubated with specific secondary antibodies for 1 h. Primary and secondary antibodies were applied with optimal dilution ratios. All slides were mounted with VECTASHIELD® HardSet™ Antifade Mounting Medium with DAPI (Vector Laboratories, Inc., CA). Anti-Rho antibody was obtained from MilliporeSigma, O4886. Anti-mCAR antibody was obtained from MilliporeSigma, AB15282. All images were taken on a Leica DB5500 microscope. All images were acquired at 1000 μ m from ONH for $\geq$ P21 samples and at 500 μ m from ONH for P3, P10 samples.

Biochemistry

Protein expression

Expression plasmids for GST-WT, E80A, and R90W HDs were published previously⁽¹⁷⁾. Plasmid for GST-K88N HD was generated by site-directed mutagenesis from the pGEX4T2-CRX WT HD backbone. *In vivo* protein expression and purification was done as previously described⁽¹⁷⁾. Briefly, 0.05mM IPTG was added to *E. coli* BL-21 (DE3) cell cultures containing different CRX HD constructs at OD600 = 0.6. The cultures were incubated for 2 hours or until OD600 = 2.0 at 34°C and the cells were collected by centrifugation at 6000 rpm and 4°C for 15 minutes. Cell pellets were resuspended in 1x PBS (Corning, Corning, NY) and then lysed by sonication. 5mM DTT (Bio-Rad Laboratories, Inc., Hercules, CA) and 1% Triton X-100 (MilliporeSigma, Burlington, MA) was then added, and the mixtures were incubated with gentle shaking at 4°C for 30 minutes to maximize protein extraction. The separation of proteins from the cellular debris were then performed by centrifugation at 15 000 rpm for 10 minutes and filtered through a 0.45 μ m membrane. Glutathione Sepharose 4B resin (Cytiva, Marlborough, MA) was first equilibrated with PBS before adding to the supernatant. 5x Halt™ Protease Inhibitor Cocktail and PMSF was added to minimize degradation. The mixtures were incubated with gentle shaking at 4°C overnight before loading on GST Spintrap™ columns (Cytiva, Marlborough, MA). The protein was eluted following the manufacturer's protocol and buffer exchanged into CRX binding buffer⁽⁸⁵⁾ using Amicon centrifugal filters (MilliporeSigma, Burlington, MA). The protein stock was supplemented with 10% glycerol before aliquoted and stored at -80°C.

Protein quantification and visualization

The size and integrity of purified GST-CRX HDs were visualized with a native 12% Tris-Glycine SDS-PAGE gel in the absence any reducing agent. Protein concentration was measured by NanoDrop One^c Microvolume UV-Vis Spectrophotometers (ThermoFisher

Scientific, Waltham, MA) and calculated using the equation: $C = (1.55 * A280) - (0.76 * A260)$, where C is the concentration of the protein in mg/ml, A280 and A260 is the absorbance of protein samples at 280nm and 260nm respectively⁽⁸⁶⁾. The protein concentration obtained with this method was comparable with BCA protein quantification assays.

Spec-seq library synthesis and purification

Single-stranded Spec-seq library templates and IRDye 700-labeled reverse complement primers (Supplementary Table S1) were ordered directly from Integrated DNA Technologies (IDT, Coralville, Iowa). The synthesis and purification of the double-stranded libraries followed previously published protocols^{(33),(86)}. Briefly, 100 pmol of template oligos and 125 pmol IRDye 700-labeled reverse complement primer F1 were mixed in Phusion® High-Fidelity PCR Master Mix (NEB, Ipswich, MA). A 15s denaturing at 95°C following a 10-minute extension at 52°C afforded duplex DNAs. Subsequently, the mixture was treated with Exonuclease I (NEB, Ipswich, MA) to remove excess ssDNA. The libraries were purified by MinElute PCR Purification Kit (QIAGEN, Hilden, Germany) and eluted in molecular biology graded water (Corning, Corning, NY).

EMSA and sample preparation for sequencing

The protein-DNA binding reactions was done in 1x CRX binding buffer (60mM KCl, 25mM HEPES, 5% glycerol, 1mM DTT)⁽⁸⁵⁾. A fixed amount (Supplementary Table S1) of IRDye-labelled DNA libraries were incubated on ice for 30 minutes with varying concentrations of wild type or mutant peptides in 20µl reaction volume. The reaction mixtures were run at 4°C in native 12% Tris-Glycine PAGE gel at 160V for 40min. The IRDye-labeled DNA fragments in the bound and unbound fractions were visualized by Odyssey® CLx and Fc Imaging Systems (LI-COR, Inc., Lincoln, NE). The visible bands were excised from the gels and DNAs were extracted with acrylamide extraction buffer (100mM NH₄OAc, 10mM Mg(OAc)₂, 0.1% SDS) then purified with MinElute PCR Purification Kit (QIAGEN, Hilden, Germany). The DNAs were amplified, barcoded by indexed Illumina primers. All indexed libraries were then pooled and sequenced on a single 1×50bp Miseq run at DNA Sequencing Innovation Lab at the Center for Genome Sciences & Systems Biology (CGS&SB, WashU).

qRT-PCR

For each replicate, RNA from 2 retinæ of a mouse was extracted using the NucleoSpin® RNA kits (Takara Bio USA, Inc., San Jose, CA). RNA sample concentration and quality was determined with NanoDrop One^c Microvolume UV-Vis Spectrophotometers (ThermoFisher Scientific, Waltham, MA). 1 µg of RNA was used to synthesize cDNA with iScript cDNA Synthesis Kits (BioRad, Hercules, CA). Primers used in this study were listed in Supplementary Table S2. qRT-PCR reactions were assembled using SsoFast™ EvaGreen® Supermix with Low ROX following manufacturer's protocol (Bio-Rad Laboratories, Inc., Hercules, CA). Data was obtained from Bio-Rad CFX96 Thermal Cycler following a three-step protocol: 1 cycle of 95°C 3 min, 40 cycles of 95°C 10 sec and 60°C 30 sec. Data was exported and further processed with customized python script.

Western blot

Experiments were performed using two biological replicates with two retinas for each replicate. Nuclear extracts were prepared using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher Scientific, Waltham, MA) following manufacturer's instructions. 1x Roche cOmplete™, Mini Protease Inhibitor Cocktail (MilliporeSigma, Burlington, MA) was supplemented in all extraction reagents. 5mM of DTT was added immediately before sample denaturing and protein was separated by running on Invitrogen

NuPAGE™ Novex 4-12% Bis-Tris MiniGels (Invitrogen, Waltham, MA). Membrane transfer was done with the Blot™ mini blot module (Invitrogen, Waltham, MA) following the manufacturer's protocol. Membrane was probed with mouse monoclonal anti-CRX antibody M02 (1:1 000, Abnova Corp., Taipei City, Taiwan) and rabbit polyclonal anti-HDAC1 antibody H51 (1:1 000, Santa Cruz Biotechnology, Dallas, TX), visualized with IRDye® 680RD goat anti-rabbit IgG and IRDye® 800CW goat anti-mouse IgG secondary antibodies (1:10 000, LI-COR, Inc., Lincoln, NE). The membrane was then imaged using Odyssey® CLx and Fc Imaging Systems (LI-COR, Inc., Lincoln, NE).

Transient transfection luciferase reporter assays

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum (FBS) and Penicillin-Streptomycin following the manufacturer's protocol. Cells were transfected with calcium phosphate transfection protocol in 6-well plates as previously described^{(17),(18)}. Experimental plasmids and usage amount are described in Supplementary Table S1. Typically, 48 hours after transfection, cells were harvested, digested and assayed for luciferase activity using Dual-Luciferase Reporter Assay System (Promega, Madison, WI) following the manufacturer's protocol. Data were collected using TD-20/20 Luminometer (Turner Designs, East Lyme, CT) and further processed with customized python scripts.

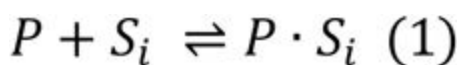
Data Analysis

Homeodomain sequence alignment

The full-length protein sequences for the selected TFs were first aligned with Clustal Omega (EMBL-EBI, UK). Aligned sequences of the homeodomain helix three were then extracted to generate [Figure 1B](#) using Jalview (v2.11.1.7). A list of the accession numbers for the selected TFs can be found in Supplementary Table S1.

Determination of relative binding affinity with Spec-seq

For a biomolecular interaction between a protein P and a particular DNA sequence, S_i , the interaction can be diagrammed as:



where $P \cdot S_i$ refers to the protein-DNA complex. The affinity of the protein P to sequence S_i is defined as the association constant K_A , or its reciprocal, the dissociation constant K_D . The K_A of the protein-DNA interaction is determined by measuring the equilibrium concentrations of each reactant and the complex:

$$K_A(S_i) = \frac{[P \cdot S_i]}{[P] \cdot [S_i]} \quad (2)$$

where [...] refers to concentrations. As in a typical Spec-seq experiment, thousands of different DNA sequences compete for the same pool of proteins, their relative binding affinities (the ratio of their K_A) can be determined by measuring the concentrations of each

sequence in the bound and unbound fractions without measuring the free protein concentrations, which is often the most difficult to measure accurately:

$$K_A(S_1) : K_A(S_2) : \dots : K_A(S_n) = \frac{[P \cdot S_1]}{[S_1]} : \frac{[P \cdot S_2]}{[S_2]} : \dots : \frac{[P \cdot S_n]}{[S_n]} \quad (3)$$

In a binding reaction involving TF and a library of DNAs, the concentration of bound and unbound species are directly proportional to the number of individual DNA molecules in each fraction which can be obtained directly from sequencing data. With enough counts in each fraction, we can accurately estimate the ratios of concentrations from counts with the relationship:

$$\frac{[P \cdot S_i]}{[P \cdot S_j]} \approx \frac{N_B(S_i)}{N_B(S_j)} \quad \text{and} \quad \frac{[S_i]}{[S_j]} \approx \frac{N_U(S_i)}{N_U(S_j)} \quad (4)$$

where N_U denotes counts in the unbound fraction and N_B denotes counts in the bound fraction. Therefore, the binding affinity of a sequence variant S_x relative to the reference sequence S_{ref} can be calculated by:

$$\frac{K_A(S_x)}{K_A(S_{ref})} \approx \frac{[P \cdot S_x]}{[P \cdot S_{ref}]} \frac{[S_{ref}]}{[S_x]} \approx \frac{N_B(S_x)}{N_B(S_{ref})} \frac{N_U(S_{ref})}{N_U(S_x)} \quad (4)$$

The natural logarithms of these ratios are the relative binding free energies in the units of *kcal/mol*. The relative free energy of the reference site for each CRX HD was set to 0.

Spec-seq data analysis and energy logo visualization

The sequencing results were first filtered and sorted based on conserved regions and barcodes. Reads with any mismatch in the conserved regions were discarded prior to further analysis as described previously^{(32),(86),(87)}. The ratio of individual sequence in bound and unbound reads was calculated as a measurement of relative binding affinity (Equation. 4) compared to the consensus sequence. The relative binding energy was then derived from the natural logarithm of the relative binding affinity and represented in *kcal/mol* units.

For wild type CRX HD and all the mutants, the energy weight matrices (ePWMs) were generated based on the regression of the TF's binding energy to its reference sequence. Only sequences within two mismatches to the reference were used to generate the ePWMs. Energy logos were generated from ePWMs after normalizing the sum of energy on each position to 0 and the negative energy values were plotted such that preferred bases are on top. The sequence logos were generated from ePWMs with python package logomaker (v0.8). The ePWMs for all CRX HDs are listed in Supplementary Table S3.

ChIP-seq data analysis

2×150bp reads from Illumina NovaSeq were obtained for all samples with a minimum depth of 22M reads at Novogene (Beijing, China). For each sample, reads from two sequencing lanes were first concatenated and run through Trim Galore (v0.6.1)⁽⁸⁸⁾ to remove adapter sequences and then QC by FastQC (v0.11.5)⁽⁸⁹⁾. The trimmed reads were then mapped to the

mm10 genome using Bowtie2 (v 2.3.4.1)⁽⁹⁰⁾ with parameters -X 2000 --very-sensitive. Only uniquely mapped and properly paired reads were retained with samtools (v1.9)⁽⁹¹⁾ with parameters -f 0×2 -q 30. Mitochondria reads were removed with samtools (v1.9)⁽⁹¹⁾. Duplicated reads were marked and removed with Picard (v2.21.4)⁽⁹²⁾. Last, reads mapped to the mm10 blacklist regions were removed by bedtools (v2.27.1)⁽⁹³⁾ by intersect -v. bigWig files were generated with deeptools (v3.0.0)⁽⁹⁴⁾ and visualized on IGV Web App⁽⁹⁵⁾. For each genotype, an average binding intensity bigWig file from two replicates was generated with deeptools (v3.0.0)⁽⁹⁴⁾ command bamCompare --operation mean with default parameters.

Peak-calling was done with MACS2 (v2.1.1.20160309)⁽⁹⁶⁾ on individual replicate with the default parameters. For each genotype, we then generated a genotype specific high confidence peakset by intersection of peaks called in two replicates. IDR framework (v2.0.4)⁽⁹⁷⁾ were used to generate quality metrics for the processed ChIPseq data. R package DiffBind (v3.0.15)^{(98),(99)} and DEseq2 (v1.30.1)⁽¹⁰⁰⁾ were then used to re-center peaks to ±200bp regions surrounding summit, generate normalized binding intensity matrix, and differential binding matrix. We defined differentially bound peaks between each mutant and wild type sample if the absolute log2FC is more than 1.0, corresponding to two-fold, and the FDR is smaller than 5e-2.

To associate peaks to genes, we used Genomic Regions Enrichment of Annotations Tool (GREAT v4.0.4)⁽¹⁰¹⁾ through the R package rGREAT (v1.19.2)⁽¹⁰²⁾. Each peak was assigned to the closest TSS within 100kb.

Binding intensity heatmap and clustering

To generate the binding intensity heatmap in [Figure 2A](#), we first compiled the genotype specific high confidence peakset for all genotypes into a single consensus peakset and only peaks with at least 5 cpm in all genotypes were retained. Python package fastcluster (v1.1.26)⁽¹⁰³⁾ was used to perform hierarchical clustering of the consensus peakset intensity matrix with parameters method='single', metric='euclidean'. The genomic regions corresponding to the two major clusters were exported and used to generate binding intensity heatmaps with deeptools (v3.0.0)⁽⁹⁴⁾.

Genomic region enrichment of CRX peaks

Peak annotation in [Figure 2D](#) were obtained using annotatePeaks.pl from HOMER (v4.8).

De novo motif searching

The mm10 fasta sequences for each genotype specific peaks were obtained using R package BSgenome (v 1.58.0)⁽¹⁰⁴⁾. De novo motif enrichment analysis for each set of sequences were then performed with MEME-ChIP in MEME Suite (v5.0.4)⁽¹⁰⁵⁾ using order 1 Markov background model and default parameters. Since homeodomain motifs are relatively short and can be repetitive (e.g. K88N motif), we reported DREME⁽³⁷⁾ found motifs for [Figure 2E](#), which is more sensitive than MEME to find short, repetitive motifs.

RNA-seq data analysis

2×150bp reads from Illumina NovaSeq were obtained for all samples with a minimum depth of 17 M reads at Novogene (Beijing, China). Sequencing reads were first run through Trim Galore (v0.6.1)⁽⁸⁸⁾ to remove adapter sequences and then QC by FastQC (v0.11.5)⁽⁸⁹⁾. Trimmed reads were then mapped to the mm10 genome and quantified with kallisto (v0.46.2)⁽¹⁰⁶⁾. Kallisto output transcript-level abundance matrices were then imported and summarized into gene-level matrices with R package tximport (v1.18.0)⁽¹⁰⁷⁾. DEseq2 (v1.30.1)

(100) was then used for normalization and differential expression analysis. The normalized count and differential expression matrices were then exported and further processed with customized python scripts.

We defined differentially expressed genes between each mutant and wild type sample if the absolute log2FC is more than 1.0, corresponding to two-fold, and the FDR is smaller than $1e-2$. For comparison between heterozygous and homozygous mutants, we first filtered genes with at least 5 cpm and then those that were called differentially expressed compared with wild type in at least one mutant genotype. We retrieved gene names in supplementary tables 3-5 from the Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.8) (108).

Definition of CRX-dependent and -independent gene set

We first identified CRX peaks that were bound in the *WT* sample but lost in *R90W/W* sample ($\log_2FC < -1$ and $FDR < 5e-2$). This yielded a total of 7677 peaks. We then found the genes associated with these peaks. We defined a gene to be CRX-dependent activated if its expression was down in adult (P21) *R90W/W* RNA-seq sample ($\log_2FC < -0.6$ and $FDR < 1e-5$). Similarly, a gene is defined as CRX-dependent suppressed if its expression was up in adult *R90W/W* RNA-seq sample ($\log_2FC > 0.6$ and $FDR < 1e-5$). A gene is defined as CRX-independent if its expression was not significantly affected in adult *R90W/W* RNA-seq samples. There were 617 CRX-dependent activated, 135 CRX-dependent suppressed, and 5565 CRX-independent genes. Manual inspection of the CRX-dependent suppressed genes revealed no clear association with photoreceptor development. Therefore, we did not further pursue this gene set. The complete list of CRX-dependent activated genes that showed differential expression in at least one of the HD mutant retinas can be found in Supplementary Table S5. The lists for CRX-independent genes that showed differential expression in *Cr^x^{E80A}* or *Cr^x^{K88N}* mutant retinas can be found in Supplementary Tables S6 and S7 respectively.

Gene ontology analysis

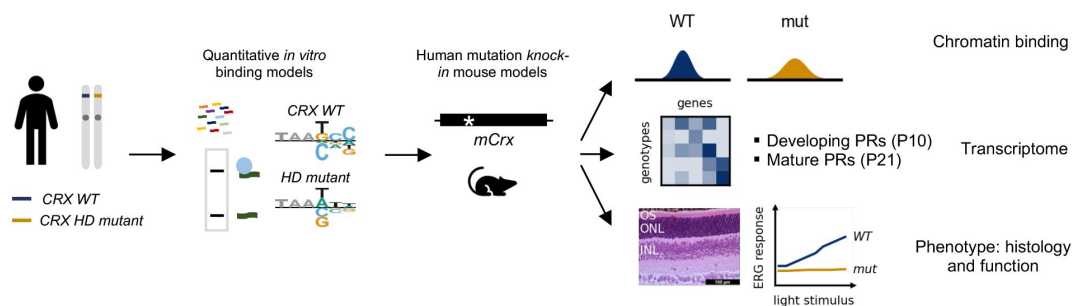
Gene ontology analysis in Figures S4B, S4E, and S5F were performed using R package clusterProfiler (v4.0.5) (109),(110) with the genome wide annotation package org.Mm.eg.db (v3.12.0) (111). Redundant enriched GO terms were removed using simplify() function with parameters cutoff=0.7, by="p.adjust". The enrichment analysis results were then exported in table format and further processed for plotting with python.

Aldiri et al. RNA-seq data re-analysis

The RNA-seq data from Aldiri et al. (38) were obtained from GEO under accession numbers GSE87064. The reads were processed similarly as all other RNA-seq data generated in this study. For Figures S4C, S4F, S5C, S5E, S6A, S6B, expression row z-scores were calculated using average cpm from replicates at each age.

Statistical analysis

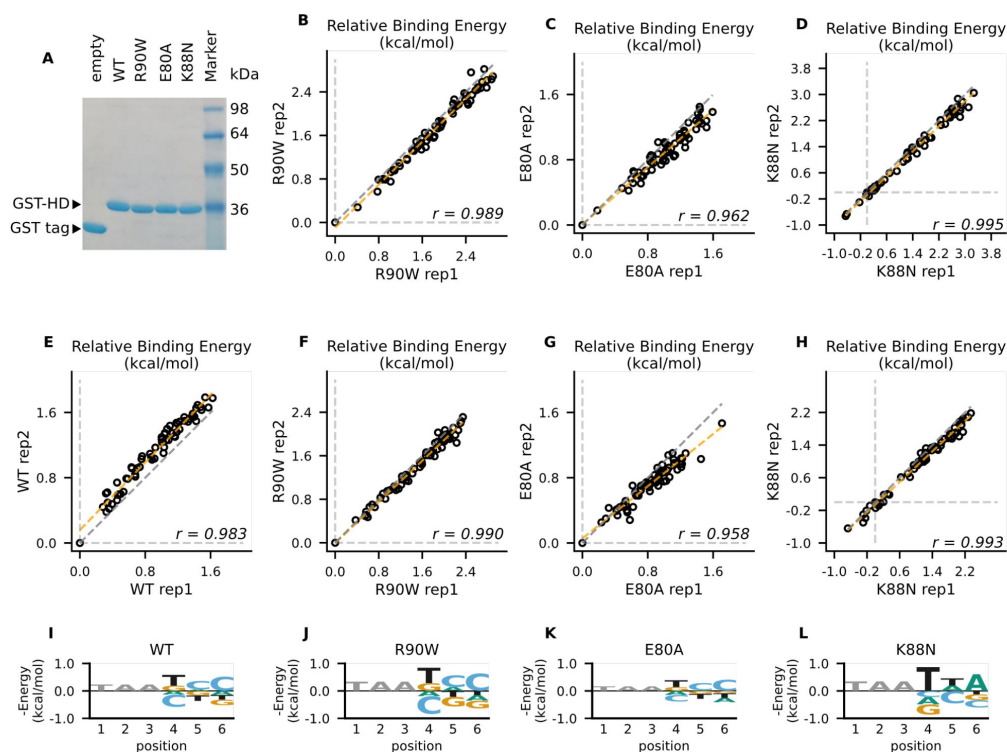
One-way ANOVA with Turkey honestly significant difference (HSD) test in Figures S3C and 6A were performed with python packages scipy (v1.8.1) (112) and scikit_posthocs (v0.7.0) (113). Two-sided Mann-Whitney U test in Figure S5A was performed with python package scipy (v1.8.1) (112).



Supplementary Figure S1.

Multi-omics approach to investigate the functional consequences of dominant disease mutations on CRX regulatory activities and photoreceptor development.

Human retinopathy associated CRX HD mutant is first tested *in vitro* for HD-DNA interactions by Spec-seq. Quantitative binding models are generated for WT and mutant HDs. Each mutation is then introduced into endogenous *mCrx* locus to generate human mutation knock-in mouse models. ChIP-seq is employed to characterize CRX chromatin binding in *WT* and mutant mouse retinas. Bulk RNA-seq is then applied to determine transcriptomic changes in developing (P10) and mature (P21) photoreceptors in both *WT* and mutant mouse retinas. Last, phenotypic characterization on retinal morphology and visual functions is carried out to understand the consequences of mutant CRX chromatin binding and associated transcriptomic alterations.



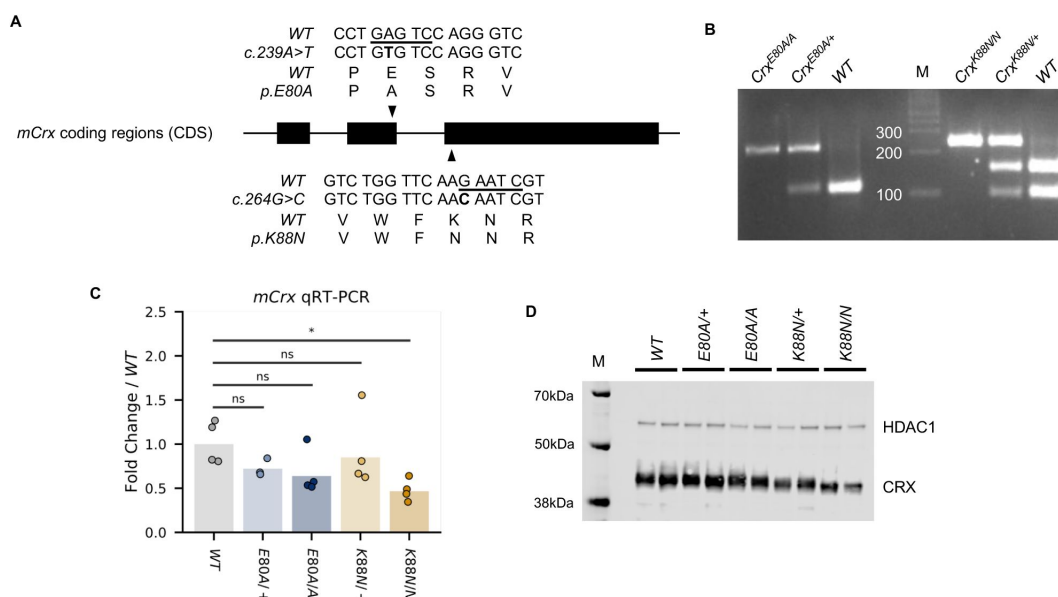
Supplementary Figure S2.

Reversed-strand Spec-seq library showed similar changes in mutant CRX HD DNA-binding specificity.

(A) Native SDS-PAGE gel image of affinity purified empty GST tag and GST-CRX HDs.

(B-D) Relative binding energy comparison from two different experiments for R90W HD (B), E80A HD (C), and K88N HD (D) on the same Spec-seq library as in Figure 2.

(E-L) Spec-seq experiments of a second library with the TAANN sites on the reverse strand show similar results. (E-H) Relative binding energy comparison from two different experiments for WT HD (E), R90W HD (F), E80A HD (G), and K88N HD (H) on the reversed monomeric library. The diagonal line of identity is represented in grey dash. The orange dashed line shows the best linear fit to the data. (I-L) Binding energy models for WT HD (I), R90W HD (J), E80A HD (K), and K88N HD (L) obtained from the reversed monomeric library. Quantitative difference with models obtained from forward-oriented library likely comes from difference in sequences immediately flanking the TAANN variable region.



Supplementary Figure S3.

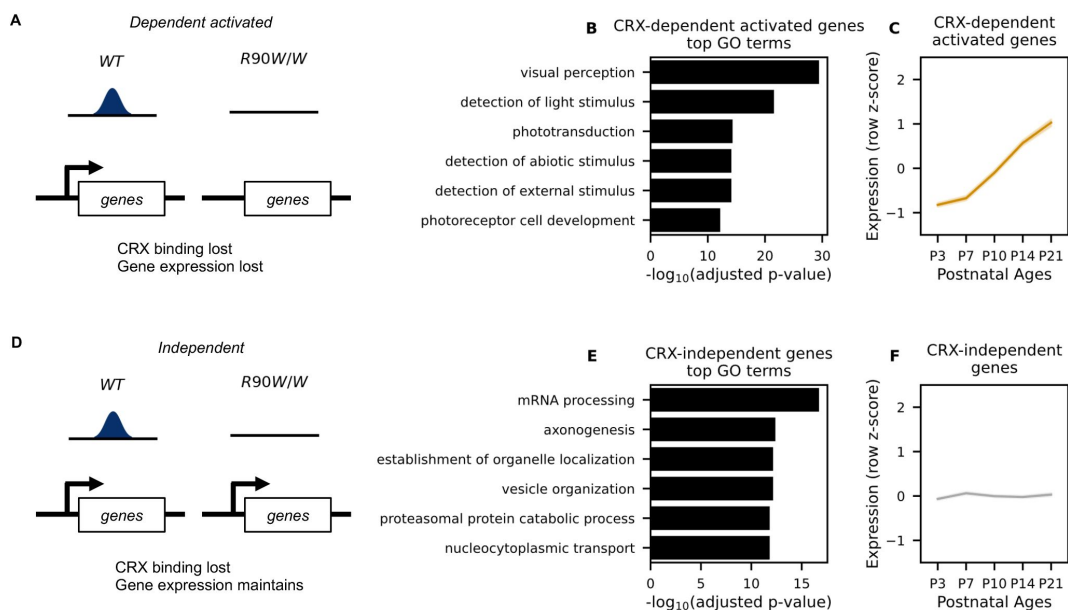
WT and mutation knock-in mouse CRX sequences and genotyping identifications

(A) Alignment of *mCrX* cDNA and protein sequences showing the nucleotide substitutions and amino acid changes of *CrX*^{E80A} (top) and *CrX*^{K88N} (bottom) alleles. Only coding regions of *mCrX* exons are shown and the diagram is not to scale. Underlined bases in WT sequences indicate the restriction enzyme *HinfI* cut sites used in the genotyping PCR.

(B) Representative mutation knock-in mouse genotyping gel image.

(C) Bar chart and stripplot showing *mCrX* mRNA expression levels in P14 WT and mutant mouse retinas. *P*-values for one-way ANOVA with Turkey honestly significant difference (HSD) test are indicated. *p*-value: ****: ≤0.0001, ***: ≤0.001, **: ≤0.01, *: ≤0.05, ns: >0.05.

(D) Immunoblots of nuclear extracts generated from P14 WT and mutant mouse retinas showing that full-length CRX protein are produced and localized to the nucleus fraction in all mutant mouse retinas. HDAC1 was used as a loading control.



Supplementary Figure S4.

Definition of CRX-dependent activated and CRX-independent gene sets.

(A) Schematic representation of CRX-dependent activated genes where CRX binding nearby is required for the expression of these genes in mature *WT* retinas.

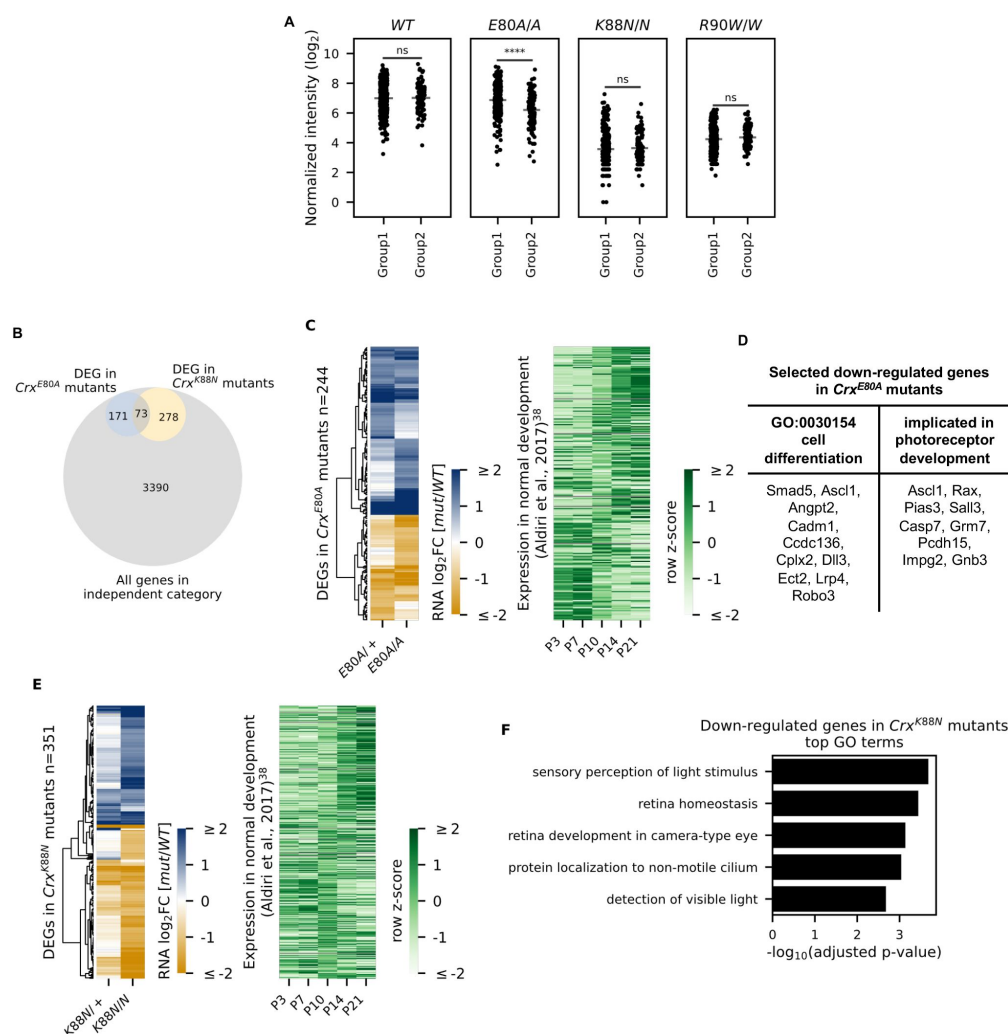
(B) Top GO terms associated with CRX-dependent activated genes. Benjamini-Hochberg adjusted p-values are shown.

(C) Line plot showing average expression pattern of CRX-dependent activated genes during normal post-natal retina.

(D) Definition of CRX-independent genes where CRX binding nearby is dispensable for the expression of these genes in mature *WT* retinas.

(E) Top GO terms associated with CRX-independent genes.

(F) Line plot showing average expression pattern of CRX-independent genes during normal retina development. RNA-seq data in (C) and (F) were retrieved from GSE87064.



Supplementary Figure S5.

E80A and K88N mutation each causes novel gene expression changes in the CRX-independent category

(A) Strip plots showing normalized CRX ChIP-seq intensity at peaks associated with Group1 or Group2 genes in WT and CRX mutant mouse retinas. *P*-values for two-sided Mann-Whitney U test are indicated. *p*-value: ****: ≤ 0.0001 , ***: ≤ 0.001 , **: ≤ 0.01 , *: ≤ 0.05 , ns: > 0.05 .

(B) Venn diagram showing the overlap of genes differentially expressed (DEGs) in *Crx*^{E80A} (pale blue) and *Crx*^{K88N} (pale yellow) but not in *Crx*^{R90W/W} (grey) mutant retinas. For DEGs in *Crx*^{E80A} and *Crx*^{K88N} mutants, genes that were differentially ex-

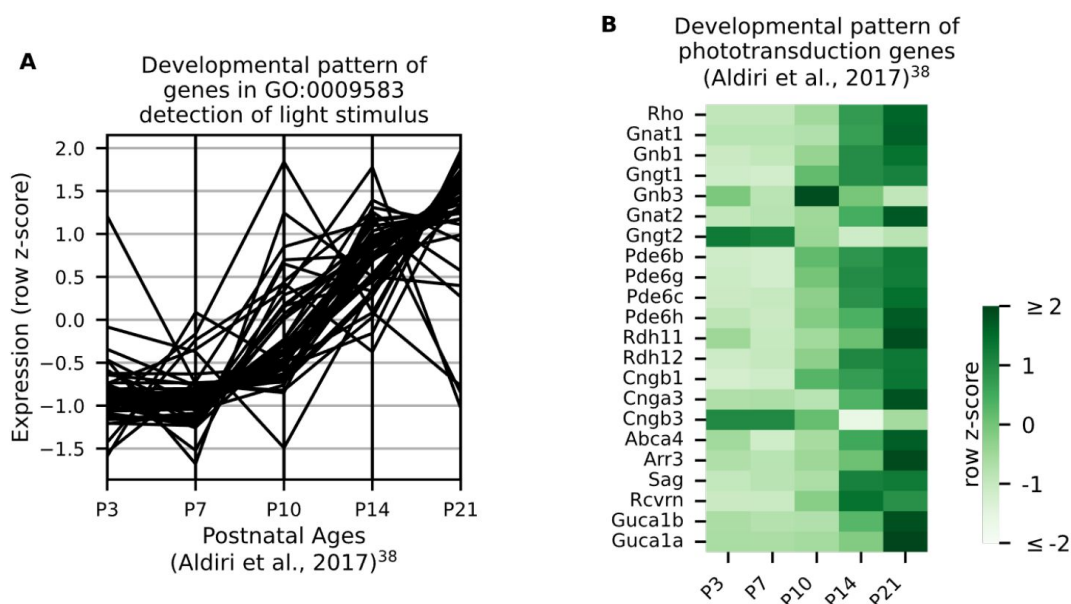
pressed in either heterozygotes or homozygotes, or both were counted.

(C) Heat map showing the expression changes of CRX-independent genes that are DEGs in at least one of the *Crx*^{E80A} mutants ($n = 244$, left). Heat map on the right shows the expression pattern of these genes during normal post-natal development (data from GSE87064).

(D) Table showing selected genes down-regulated in *Crx*^{E80A} mutants that have been implicated in cell differentiation or photoreceptor development.

(E) Heat map showing the expression changes of CRX-independent genes that are DEGs in at least one of the *Crx*^{K88N} mutants ($n = 351$, left). Heat map on the right shows the expression pattern of these genes during normal post-natal development (data from GSE87064).

(F) Bar chart showing GO term enrichment of DEGs in *Crx*^{K88N} mutants.



Supplementary Figure S6.

Developmental expression pattern of phototransduction genes in WT animals

(A) Parallel coordinates plot showing expression pattern of genes in GO:0009583 during normal post-natal retina development (data from GSE87064). Normalized row z-score is shown

(B) Heatmap showing expression patterns of phototransduction genes during normal post-natal retina development (data from GSE87064). Normalized row z-score is shown.

References

1. Mark M. , Rijli F.M. , Chambon P. (1997) **Homeobox Genes in Embryogenesis and Pathogenesis** *Pediatric Research* **42**:421–429
2. Lewis E.B. (1978) **A gene complex controlling segmentation in Drosophila** *Nature* **276**:565–570
3. Leung R.F. (2022) **Genetic Regulation of Vertebrate Forebrain Development by Homeobox Genes** *Frontiers in Neuroscience* **16**:
4. Chi Y.-I. (2005) **Homeodomain revisited: a lesson from disease-causing mutations** *Human Genetics* **116**:433–444
5. Zagozewski J.L. , Zhang Q. , Pinto V.I. , Wigle J.T. , Eisenstat D.D. (2014) **The role of homeobox genes in retinal development and disease** *Developmental Biology* **393**:195–208

6. Diacou R. (2022) **Cell fate decisions, transcription factors and signaling during early retinal development** *Progress in Retinal and Eye Research* **91**:101093
7. Beby F. , Lamonerie T. (2013) **The homeobox gene Otx2 in development and disease** *Experimental Eye Research* **111**:9–16
8. Henderson R.H. (2009) **A rare de novo nonsense mutation in OTX2 causes early onset retinal dystrophy and pituitary dysfunction** *Mol Vis* **15**:2442–7
9. Abouzeid H. (2009) **PAX6 aniridia and interhemispheric brain anomalies** *Mol Vis* **15**:2074–83
10. Lima Cunha D. , Arno G. , Corton M. , Moosajee M. (2019) **The Spectrum of PAX6 Mutations and Genotype-Phenotype Correlations in the Eye** *Genes (Basel)* **10**:
11. Voronina V.A. (2003) **Mutations in the human RAX homeobox gene in a patient with anophthalmia and sclerocornea** *Human Molecular Genetics* **13**:315–322
12. Abouzeid H. (2012) **RAX and anophthalmia in humans: evidence of brain anomalies** *Mol Vis* **18**:1449–56
13. Chen S. (1997) **Crx, a Novel Otx-like Paired-Homeodomain Protein, Binds to and Transactivates Photoreceptor Cell-Specific Genes** *Neuron* **19**:1017–1030
14. Furukawa T. , Morrow E.M. , Cepko C.L. (1997) **Crx, a Novel otx-like Homeobox Gene, Shows Photoreceptor-Specific Expression and Regulates Photoreceptor Differentiation** *Cell* **91**:531–541
15. Furukawa T. , Morrow E.M. , Li T. , Davis F.C. , Cepko C.L. (1999) **Retinopathy and attenuated circadian entrainment in Crx-deficient mice** *Nature Genetics* **23**:466
16. Chau K.Y. , Chen S. , Zack D.J. , Ono S.J. (2000) **Functional domains of the cone-rod homeobox (CRX) transcription factor** *J Biol Chem* **275**:37264–70
17. Chen S. (2002) **Functional analysis of cone-rod homeobox (CRX) mutations associated with retinal dystrophy** *Human Molecular Genetics* **11**:873–884
18. Tran N.M. (2014) **Mechanistically Distinct Mouse Models for CRX-Associated Retinopathy** *PLOS Genetics* **10**:
19. Ruzycki P.A. , Tran N.M. , Kolesnikov A.V. , Kefalov V.J. , Chen S. (2015) **Graded gene expression changes determine phenotype severity in mouse models of CRX-associated retinopathies** *Genome Biology* **16**:171
20. Freund C.L. (1997) **Cone-Rod Dystrophy Due to Mutations in a Novel Photoreceptor-Specific Homeobox Gene (CRX) Essential for Maintenance of the Photoreceptor** *Cell* **91**:543–553

21. Swaroop A. (1999) **Leber Congenital Amaurosis Caused by a Homozygous Mutation (R90W) in the Homeodomain of the Retinal Transcription Factor CRX: Direct Evidence for the Involvement of CRX in the Development of Photoreceptor Function** *Human Molecular Genetics* **8**:299–305
22. Nichols L.L. (2010) **Two novel CRX mutant proteins causing autosomal dominant Leber congenital amaurosis interact differently with NRL** *Hum Mutat* **31**:–83
23. Desplan C. , Theis J. , O’Farrell P.H. (1988) **The sequence specificity of homeodomain-DNA interaction** *Cell* **54**:1081–1090
24. Trelisman J. , Gönczy P. , Vashishtha M. , Harris E. , Desplan C. (1989) **A single amino acid can determine the DNA binding specificity of homeodomain proteins** *Cell* **59**:553–562
25. Hanes S.D. , Brent R. (1989) **DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9** *Cell* **57**:1275–1283
26. Hanes S.D. , Brent R. (1991) **A genetic model for interaction of the homeodomain recognition helix with DNA** *Science* **251**:426
27. Ades S.E. , Sauer R.T. (1995) **Specificity of Minor-Groove and Major-Groove Interactions in a Homeodomain-DNA Complex** *Biochemistry* **34**:14601–14608
28. Wilson D.S. , Guenther B. , Desplan C. , Kuriyan J. (1995) **High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA** *Cell* **82**:709–719
29. Gehring W.J. (1994) **Homeodomain-DNA recognition** *Cell* **78**:211–223
30. Noyes M.B. (2008) **Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites** *Cell* **133**:1277–89
31. Zuo Z. , Stormo G.D. (2014) **High-Resolution Specificity from DNA Sequencing Highlights Alternative Modes of Lac Repressor Binding** *Genetics* **198**:1329
32. Stormo G.D. , Zuo Z. , Chang Y.K. (2015) **Spec-seq: determining protein–DNA-binding specificity by sequencing** *Briefings in Functional Genomics* **14**:30–38
33. Zuo Z. , Roy B. , Chang Y.K. , Granas D. , Stormo G.D. (2017) **Measuring quantitative effects of methylation on transcription factor-DNA binding affinity** *Science advances* **3**:
34. Corbo J.C. (2010) **CRX ChIP-seq reveals the cis-regulatory architecture of mouse photoreceptors** *Genome Res* **20**:1512–25
35. Kwasnieski J.C. , Mogno I. , Myers C.A. , Corbo J.C. , Cohen B.A. (2012) **Complex effects of nucleotide variants in a mammalian cis;-regulatory element** *Proceedings of the National Academy of Sciences* **109**:19498

36. Bassett E.A. , Wallace V.A. (2012) **Cell fate determination in the vertebrate retina** *Trends Neurosci* **35**:565–73
37. Bailey T.L. (2011) **DREME: motif discovery in transcription factor ChIP-seq data** *Bioinformatics* **27**:1653–1659
38. Aldiri I. (2017) **The Dynamic Epigenetic Landscape of the Retina During Development, Reprogramming, and Tumorigenesis** *Neuron* **94**:550–568
39. Kim J.W. (2016) **NRL-Regulated Transcriptome Dynamics of Developing Rod Photoreceptors** *Cell Rep* **17**:2460–2473
40. Onishi A. (2010) **The orphan nuclear hormone receptor ERR β controls rod photoreceptor survival** *Proceedings of the National Academy of Sciences* **107**:11579
41. Yoshida S. (2004) **Expression profiling of the developing and mature Nrl $^{-/-}$ mouse retina: identification of retinal disease candidates and transcriptional regulatory targets of Nrl** *Hum Mol Genet* **13**:1487–503
42. Zang J. , Neuhauss S.C.F. (2018) **The Binding Properties and Physiological Functions of Recoverin** *Frontiers in Molecular Neuroscience* **11**:
43. Palczewski K. (2006) **G protein-coupled receptor rhodopsin** *Annu Rev Biochem* **75**:743–67
44. Dryja T.P. , Hahn L.B. , Reboul T. , Arnaud B. (1996) **Missense mutation in the gene encoding the α subunit of rod transducin in the Nougaret form of congenital stationary night blindness** *Nature Genetics* **13**:358–360
45. Carrigan M. (2016) **A novel homozygous truncating GNAT1 mutation implicated in retinal degeneration** *British Journal of Ophthalmology* **100**:495
46. Dvir L. (2010) **Autosomal-Recessive Early-Onset Retinitis Pigmentosa Caused by a Mutation in PDE6G, the Gene Encoding the Gamma Subunit of Rod cGMP Phosphodiesterase** *The American Journal of Human Genetics* **87**:258–264
47. Allikmets R. (1997) **Mutation of the Stargardt Disease Gene (ABCR) in Age-Related Macular Degeneration** *Science* **277**:1805–1807
48. Nasonkin I. (1998) **Mapping of the rod photoreceptor ABC transporter (ABCR) to 1p21-p22.1 and identification of novel mutations in Stargardt's disease** *Human Genetics* **102**:21–26
49. Kitamura E. (2006) **Disruption of the Gene Encoding the β 1-Subunit of Transducin in the Rd4/+ Mouse** *Investigative Ophthalmology & Visual Science* **47**:1293–1301
50. Janecke A.R. (2004) **Mutations in RDH12 encoding a photoreceptor cell retinol dehydrogenase cause childhood-onset severe retinal dystrophy** *Nature Genetics* **36**:850–854

51. Bareil C. (2001) **Segregation of a mutation in CNGB1 encoding the β -subunit of the rod cGMP-gated channel in a family with autosomal recessive retinitis pigmentosa** *Human Genetics* **108**:328–334
52. Pierce E.A. (1999) **Mutations in a gene encoding a new oxygen-regulated photoreceptor protein cause dominant retinitis pigmentosa** *Nature Genetics* **22**:248–254
53. Sullivan L.S. (1999) **Mutations in a novel retina-specific gene cause autosomal dominant retinitis pigmentosa** *Nature Genetics* **22**:255–259
54. Kohl S. (2002) **Mutations in the Cone Photoreceptor G-Protein Subunit Gene GNAT2 in Patients with Achromatopsia** *The American Journal of Human Genetics* **71**:422–425
55. Rosenberg T. (2004) **Variant Phenotypes of Incomplete Achromatopsia in Two Cousins with GNAT2 Gene Mutations** *Investigative Ophthalmology & Visual Science* **45**:4256–4262
56. Thiadens A.A.H.J. (2009) **Homozygosity Mapping Reveals PDE6C Mutations in Patients with Early-Onset Cone Photoreceptor Disorders** *The American Journal of Human Genetics* **85**:240–247
57. Kohl S. (2012) **A Nonsense Mutation in PDE6H Causes Autosomal-Recessive Incomplete Achromatopsia** *The American Journal of Human Genetics* **91**:527–532
58. Piri N. (2005) **A substitution of G to C in the cone cGMP-phosphodiesterase γ subunit gene found in a distinctive form of cone dystrophy** *Ophthalmology* **112**:159–166
59. Kaufman M.L. (2019) **Transcriptional profiling of murine retinas undergoing semi-synchronous cone photoreceptor differentiation** *Developmental Biology* **453**:155–167
60. Irie S. (2015) **Rax Homeoprotein Regulates Photoreceptor Cell Maturation and Survival in Association with Crx in the Postnatal Mouse Retina** *Molecular and Cellular Biology* **35**:2583–2596
61. de Melo J. , Peng G.-H. , Chen S. , Blackshaw S. (2011) **The Spalt family transcription factor Sall3 regulates the development of cone photoreceptors and retinal horizontal interneurons** *Development* **138**:2325–2336
62. Onishi A. (2009) **Pias3-Dependent SUMOylation Directs Rod Photoreceptor Development** *Neuron* **61**:234–246
63. Aavani T. , Tachibana N. , Wallace V. , Biernaskie J. , Schuurmans C. (2017) **Temporal profiling of photoreceptor lineage gene expression during murine retinal development** *Gene Expression Patterns* **23**:32–44
64. Tran N.M. , Chen S. (2014) **Mechanisms of blindness: Animal models provide insight into distinct CRX-associated retinopathies** *Developmental Dynamics* **243**:1153–1166
65. Roger J.E. (2014) **OTX2 loss causes rod differentiation defect in CRX-associated congenital blindness** *The Journal of Clinical Investigation* **124**:631–643

66. Brzezinski J.A. , Reh T.A. (2015) **Photoreceptor cell fate specification in vertebrates** *Development* **142**:3263–73
67. Wang S. , Cepko C.L. (2016) **Photoreceptor Fate Determination in the Vertebrate Retina** *Investigative ophthalmology & visual science* **57**:
68. Swaroop A. , Kim D. , Forrest D. (2010) **Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina** *Nature Reviews Neuroscience* **11**:563–576
69. Hennig A.K. , Peng G.H. , Chen S. (2008) **Regulation of photoreceptor gene expression by Crx-associated transcription factor network** *Brain Res* **1192**:114–33
70. Carter-Dawson L.D. , LaVail M.M. (1979) **Rods and cones in the mouse retina. I. Structural analysis using light and electron microscopy** *J Comp Neurol* **188**:245–62
71. Datta R.R. (2018) **A feed-forward relay integrates the regulatory activities of Bicoid and Orthodenticle via sequential binding to suboptimal sites** *Genes & Development* **32**:723–736
72. Kimura A. (2000) **Both PCE-1/RX and OTX/CRX Interactions Are Necessary for Photoreceptor-specific Gene Expression** * *Journal of Biological Chemistry* **275**:1152–1160
73. Hughes A.E.O. , Myers C.A. , Corbo J.C. (2018) **A massively parallel reporter assay reveals context-dependent activity of homeodomain binding sites in vivo** *Genome Res* **28**:1520–1531
74. White Michael A. (2016) **A Simple Grammar Defines Activating and Repressing cis-Regulatory Elements in Photoreceptors** *Cell Reports* **17**:1247–1254
75. Rister J. (2015) **Single-base pair differences in a shared motif determine differential Rhodopsin expression** *Science* **350**:1258–61
76. Tucker S.C. , Wisdom R. (1999) **Site-specific Heterodimerization by Paired Class Homeodomain Proteins Mediates Selective Transcriptional Responses*** *Journal of Biological Chemistry* **274**:32325–32332
77. Vincent A. (2014) **OTX2 mutations cause autosomal dominant pattern dystrophy of the retinal pigment epithelium** *Journal of Medical Genetics* **51**:797
78. Ragge N.K. (2005) **Heterozygous Mutations of OTX2 Cause Severe Ocular Malformations** *The American Journal of Human Genetics* **76**:1008–1022
79. Perveen R. (2000) **Phenotypic Variability and Asymmetry of Rieger Syndrome Associated with PITX2 Mutations** *Investigative Ophthalmology & Visual Science* **41**:2456–2460
80. Chirco K.R. , Chew S. , Moore A.T. , Duncan J.L. , Lamba D.A. (2021) **Allele-specific gene editing to rescue dominant CRX-associated LCA7 phenotypes in a retinal organoid model** *Stem Cell Reports* **16**:2690–2702

81. Yang H. , Wang H. , Jaenisch R. (2014) **Generating genetically modified mice using CRISPR/Cas-mediated genome engineering** *Nature Protocols* **9**:1956–1968
82. Cho A. , Haruyama N. , Kulkarni A.B. (2009) **Generation of transgenic mice** *Curr Protoc Cell Biol* **19**:
83. Chen S. (2004) **Interference of Crx-dependent transcription by ataxin-7 involves interaction between the glutamine regions and requires the ataxin-7 carboxy-terminal region for nuclear localization** *Human Molecular Genetics* **13**:53–67
84. Schmidt D. (2009) **ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions** *Methods* **48**:240–8
85. Lee J. , Myers C.A. , Williams N. , Abdelaziz M. , Corbo J.C. (2010) **Quantitative fine-tuning of photoreceptor cis-regulatory elements through affinity modulation of transcription factor binding sites** *Gene Therapy* **17**:1390–1399
86. Roy B. , Zuo Z. , Stormo G.D. (2017) **Quantitative specificity of STAT1 and several variants** *Nucleic acids research* **45**:8199–8207
87. Zuo Z. , Stormo G.D. (2014) **High-Resolution Specificity from DNA Sequencing Highlights Alternative Modes of Lac Repressor Binding** *Genetics* **198**:1329–1343
88. Felix Krueger F.J. , Ewels Phil , Afyounian Ebrahim , Weinstein Michael , Schuster-Boeckler Benjamin , Hulselmans Gert , sclamons (2023) **FelixKrueger/TrimGalore: v0.6.9 - fix declaration bug (0.6.9)**
89. (2015) **FastQC. (2015).**
90. Langmead B. , Salzberg S.L. (2012) **Fast gapped-read alignment with Bowtie 2** *Nature Methods* **9**:357–359
91. Li H. (2009) **The Sequence Alignment/Map format and SAMtools** *Bioinformatics* **25**:2078–2079
92. Picard toolkit (2019) **Broad Institute**
93. Quinlan A.R. , Hall I.M. (2010) **BEDTools: a flexible suite of utilities for comparing genomic features** *Bioinformatics* **26**:841–842
94. Ramírez F. (2016) **deepTools2: a next generation web server for deep-sequencing data analysis** *Nucleic Acids Research* **44**:
95. Robinson J.T. (2011) **Integrative genomics viewer** *Nature Biotechnology* **29**:24–26
96. Zhang Y. (2008) **Model-based Analysis of ChIP-Seq (MACS)** *Genome Biology* **9**:

97. Li Q. , Brown J.B. , Huang H. , Bickel P.J. (2011) **Measuring reproducibility of high-throughput experiments** *The Annals of Applied Statistics* **5**:1752–1779
98. Stark R. , Brown G. (2012) **Stark, R. & Brown, G. DiffBind: Differential binding analysis of ChIP-Seq peak data. (2012).**
99. Ross-Innes C.S. (2012) **Differential oestrogen receptor binding is associated with clinical outcome in breast cancer** *Nature* **481**:389–393
100. Love M.I. , Huber W. , Anders S. (2014) **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2** *Genome Biology* **15**:550
101. McLean C.Y. (2010) **GREAT improves functional interpretation of cis-regulatory regions** *Nat Biotechnol* **28**:495–501
102. Gu Z. , Hübschmann D. (2022) **rGREAT: an R/bioconductor package for functional enrichment on genomic regions** *Bioinformatics* **39**:
103. Müllner D. (2013) **fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python** *Journal of Statistical Software* **53**:1–18
104. Pagès H. (2020) **Pagès, H. BSgenome: Software infrastructure for efficient representation of full genomes and their SNPs. (2020).**
105. Bailey T.L. , Johnson J. , Grant C.E. , Noble W.S. (2015) **The MEME Suite** *Nucleic Acids Research* **43**:
106. Bray N.L. , Pimentel H. , Melsted P. , Pachter L. (2016) **Near-optimal probabilistic RNA-seq quantification** *Nature Biotechnology* **34**:525–527
107. Soneson C. , Love M.I. , Robinson M.D. (2015) **Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences** *F1000Res* **4**:1521
108. Sherman B.T. (2022) **DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)** *Nucleic Acids Res* **50**:–21
109. Yu G. , Wang L.G. , Han Y. , He Q.Y. (2012) **clusterProfiler: an R package for comparing biological themes among gene clusters** *Omics* **16**:284–7
110. Wu T. (2021) **clusterProfiler 4.0: A universal enrichment tool for interpreting omics data** *Innovation (Camb)* **2**:100141
111. Carlson M. (2019) **Carlson, M.org.Mm.eg.db: Genome wide annotation for Mouse. (2019).**
112. Virtanen P. (2020) **SciPy 1.0: fundamental algorithms for scientific computing in Python** *Nature Methods* **17**:261–272

113. Terpilowski M.A. (2019) **scikit-posthocs: Pairwise multiple comparison tests in Python** *Journal of Open Source Software* **4**:1169
114. Goujon M. (2010) **A new bioinformatics analysis tools framework at EMBL-EBI** *Nucleic Acids Research* **38**:
115. Sievers F. (2011) **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega** *Molecular Systems Biology* **7**:539
116. Waterhouse A.M. , Procter J.B. , Martin D.M.A. , Clamp M. , Barton G.J. (2009) **Jalview Version 2—a multiple sequence alignment editor and analysis workbench** *Bioinformatics* **25**:1189–1191
117. Tareen A. , Kinney J.B. (2019) **Logomaker: beautiful sequence logos in Python** *Bioinformatics* **36**:2272–2274
118. Hunter J.D. (2007) **Matplotlib: A 2D Graphics Environment** *Computing in Science & Engineering* **9**:90–95
119. Jeff Reback j. , McKinney Wes , Van den Bossche Joris , Augspurger Tom , Roeschke Matthew , Hawkins Simon , Cloud Phillip , gfyong Sinhrks , Hoefler Patrick , Klein Adam , Petersen Terji , Tratner Jeff , She Chang , Ayd William , Naveh Shahar , Darbyshire JHM , Garcia Marc , Shadrach Richard , Schendel Jeremy , Hayden Andy , Saxton Daniel , Gorelli Marco Edward , Li Fangchen , Zeitlin Matthew , Jancauskas Vytautas , McMaster Ali , Wörtwein Torsten , Battiston Pietro (2022) **Jeff Reback, j., WesMcKinney, JorisVan den Bossche, TomAugspurger, MatthewRoeschke, SimonHawkins, PhillipCloud, gfyong, Sinhrks, PatrickHoefler, AdamKlein, TerjiPetersen, JeffTratner, ChangShe, WilliamAyd, ShaharNaveh, JHMDarbyshire, MarcGarcia, RichardShadrach, JeremySchendel, AndyHayden, DanielSaxton, Marco EdwardGorelli, FangchenLi, MatthewZeitlin, VytautasJancauskas, AliMcMaster, TorstenWörtwein, PietroBattiston. pandas-dev/pandas: Pandas 1.4.2. (2022).**
120. Van Rossum G.a.D.J. , Fred L. (1995) **Python reference manual**
121. Waskom M.L. (2021) **seaborn: statistical data visualization** *Journal of Open Source Software* **6**:3021

Author information

1. Yiqiao Zheng

Molecular Genetic and Genomics Graduate program, Division of Biological and Biomedical Sciences, Washington University in St Louis, Saint Louis, Missouri, USA, Department of Ophthalmology and Visual Sciences, Washington University in St Louis, Saint Louis, Missouri, USA
ORCID iD: [0000-0003-4133-0439](https://orcid.org/0000-0003-4133-0439)

2. Chi Sun

Department of Ophthalmology and Visual Sciences, Washington University in St Louis, Saint Louis, Missouri, USA

3. Xiaodong Zhang

Department of Ophthalmology and Visual Sciences, Washington University in St Louis, Saint Louis, Missouri, USA

4. Philip A. Ruzycki

Department of Ophthalmology and Visual Sciences, Washington University in St Louis, Saint Louis, Missouri, USA, Department of Genetics, Washington University in St Louis, Saint Louis, Missouri, USA

5. Shiming Chen

Molecular Genetic and Genomics Graduate program, Division of Biological and Biomedical Sciences, Washington University in St Louis, Saint Louis, Missouri, USA, Department of Ophthalmology and Visual Sciences, Washington University in St Louis, Saint Louis, Missouri, USA, Department of Developmental Biology, Washington University in St Louis, Saint Louis, Missouri, USA

For correspondence:

chenshiming@wustl.edu

Editors

Reviewing Editor

Paschalis Kratsios

University of Chicago, United States of America

Senior Editor

Claude Desplan

New York University, United States of America

Reviewer #1 (Public Review):

The manuscript by Zheng et al. examined the disease-causing mechanisms of two missense mutations within the homeodomain (HD) of CRX protein. Both mutations were found in humans and can produce severe dominant retinopathy. The authors investigated the two CRX HD mutants via in vitro DNA-binding assay (Spec-seq), in vivo chromatin-binding assay (ChIP-seq), in vivo expression assay of downstream target genes (RNA-seq), and retinal histological and functional assays. They concluded that p.E80A increased the transactivation activity of CRX and resulted in precocious photoreceptor differentiation, whereas p.K88N significantly changed the binding specificity of CRX and led to defects in photoreceptor differentiation and maintenance. The authors performed a significant amount of analyses. The claims are sufficiently supported by the data. The results not only uncovered the underlying disease-causing mechanisms, but also can significantly improve our understanding of the interaction between HD-TF and DNA during development.

Minor concerns:

1. The E80A, K88N and R90W (previously reported by the same group) mutations are located very close to each other in the homeodomain (Figure 1A), but had distinct effects on the activity of CRX. Has the structure of the homeodomain (of CRX) been resolved? If so, could the authors discuss this phenomenon (mutations close to each other but have distinct effects) based on the HD-DNA structure? In addition, has this phenomenon been observed in other homeodomain TFs?
2. The authors should briefly summarize the effects/disease-causing-mechanisms of all the reported CRX mutations in the discussion part. The readers can then have a better overview of the topic.
3. CRX can also function as a pioneer factor (reported by the same group). Would these HD mutations distinctively affect chromatin accessibility (which then leads to ectopic binding on the genome)?
4. The discussion part can be shortened and simplified.

Reviewer #2 (Public Review):

Zheng et al., investigated the molecular and functional mechanisms of two homeodomain missense mutations causing human retinal photoreceptor degeneration diseases in photoreceptor development regulated by the CRX transcription factor. They analyzed the E80A mutation associated with dominant cone-rod dystrophy (CRD) and the K88N mutation associated with dominant Leber Congenital Amaurosis (LCA). The authors found that E80A CRX binds to the same target DNA sites as WT CRX, but the binding specificity of K88N CRX is altered from that of WT in an in vitro assay. They generated Crx(E80A) and Crx(K88N) KI mice and performed ChIP assay and observed that K88N CRX binds to novel genomic regions from the WT-binding sites, while E80A binds to the WT sites. In addition, using the KI mice, they found that E80A and K88N differently affect the expression of Crx target genes. This study is well executed with proper and solid methodologies, and the manuscript is clearly written. This study gives us the insights how single missense CRX mutations lead to different types of human retinal photoreceptor degeneration diseases.

While the study has strengths in principle, it has a couple of weaknesses. One is how well E80A KI mice function as a pathological model of dominant CRD, in which cones are mainly first affected, is not clearly shown in this study. More data investigating how cones are affected by performing histological, molecular, and physiological analyses will be helpful and useful. For example, in the Discussion, the authors describe that E80A associates with S-cone opsin promoter results is "data now shown". This data must be presented for the readers. In addition, more molecular insights as to how E80A affects cones will strengthen this study. Another point is that it will be very valuable if the authors could show how E80A and K88N differently affect the 3D structure of the CRX homeodomain. Even a simulation model would be valuable.

Author Response

Reviewer #1 (Public Review):

The manuscript by Zheng et al. examined the disease-causing mechanisms of two missense mutations within the homeodomain (HD) of CRX protein. Both mutations were found in humans and can produce severe dominant retinopathy. The authors investigated the two CRX HD mutants via in vitro DNA-binding assay (Spec-seq), in vivo chromatin-binding assay (ChIP-seq), in vivo expression assay of downstream target genes (RNA-seq), and retinal histological and functional assays. They concluded that p.E80A increased the transactivation activity of CRX

and resulted in precocious photoreceptor differentiation, whereas p.K88N significantly changed the binding specificity of CRX and led to defects in photoreceptor differentiation and maintenance. The authors performed a significant amount of analyses. The claims are sufficiently supported by the data. The results not only uncovered the underlying disease-causing mechanisms, but also can significantly improve our understanding of the interaction between HD-TF and DNA during development.

Thank you for summarizing the key findings and strengths of our manuscript.

Minor concerns:

1. *The E80A, K88N and R90W (previously reported by the same group) mutations are located very close to each other in the homeodomain (Figure 1A), but had distinct effects on the activity of CRX. Has the structure of the homeodomain (of CRX) been resolved? If so, could the authors discuss this phenomenon (mutations close to each other but have distinct effects) based on the HD-DNA structure?*

In paragraphs 2, 4, 5 of the discussion section, we have added explanations on how each mutation could affect CRX HD-DNA interactions differently based on published structural studies. And we further explain how these biochemical changes relate to the molecular perturbations and cellular phenotypes seen in vivo.

In addition, has this phenomenon been observed in other homeodomain TFs?

Disease associated missense mutations at residues HD50 (K88) and HD52 (R90) have also been reported in other HD TFs implicated in CNS development (see discussion paragraph 7). Distinctively, different substitutions at CRX E80 residue have been reported in multiple CoRD cases, suggesting its essential role in HD-DNA-mediated regulation during retinal development. These new points are now included in the discussion section.

2. *The authors should briefly summarize the effects/disease-causing-mechanisms of all the reported CRX mutations in the discussion part. The readers can then have a better overview of the topic.*

We have added a concise summary of previously proposed CRX mutation classification scheme, all characterized Crx mutant mouse models and their pathogenic mechanisms. Please see paragraph 9 in the discussion section.

3. *CRX can also function as a pioneer factor (reported by the same group). Would these HD mutations distinctively affect chromatin accessibility (which then leads to ectopic binding on the genome)?*

Prior evidence has demonstrated that regulatory regions for many photoreceptor genes failed to stay accessible upon loss of CRX in the Crx^{-/-} model (PMID: 30068366). It is unclear with the existing data whether CRX could initiate the chromatin remodeling (true pioneering function) of these regions, or it simply maintains the accessibility once these regions became accessible. Future studies comparing epigenomic landscape changes in mutant Crx KI models at various ages can be informative, particularly for the CRX K88N ectopic binding events. Determining how the CRX K88N mutant protein alters chromatin landscape important for photoreceptor fate and/or differentiation during development would shed light on the nature of these ectopic binding events.

4. *The discussion part can be shortened and simplified.*

We have re-written the discussion section to make it concise and to incorporate discussions on mutant CRX HD structures. Please see the revised manuscript.

Reviewer #2 (Public Review):

Zheng et al., investigated the molecular and functional mechanisms of two homeodomain missense mutations causing human retinal photoreceptor degeneration diseases in photoreceptor development regulated by the CRX transcription factor. They analyzed the E80A mutation associated with dominant cone-rod dystrophy (CRD) and the K88N mutation associated with dominant Leber Congenital Amaurosis (LCA). The authors found that E80A CRX binds to the same target DNA sites as WT CRX, but the binding specificity of K88N CRX is altered from that of WT in an in vitro assay. They generated Crx(E80A) and Crx(K88N) KI mice and performed ChIP assay and observed that K88N CRX binds to novel genomic regions from the WT-binding sites, while E80A binds to the WT sites. In addition, using the KI mice, they found that E80A and K88N differently affect the expression of Crx target genes. This study is well executed with proper and solid methodologies, and the manuscript is clearly written. This study gives us the insights how single missense CRX mutations lead to different types of human retinal photoreceptor degeneration diseases.

We greatly appreciate the reviewer's summary and positive comments.

While the study has strengths in principle, it has a couple of weaknesses. One is how well E80A KI mice function as a pathological model of dominant CRD, in which cones are mainly first affected, is not clearly shown in this study. More data investigating how cones are affected by performing histological, molecular, and physiological analyses will be helpful and useful. For example, in the Discussion, the authors describe that E80A associates with S-cone opsin promoter results is "data now shown". This data must be presented for the readers. In addition, more molecular insights as to how E80A affects cones will strengthen this study.

The mouse retina is rod dominant and contains only a small number of cones (3% of all photoreceptors) that are born prenatally. This poses technical challenges to appropriately assess cone-specific changes during disease initiation/progression. We are in the process of developing cellular/molecular tools to investigate how cones are being affected in Crx E80A KI model, but this is beyond the scope of the current study.

At the same time, we have added a supplemental panel showing that, based on P0 retinal immunostaining of the early cone marker RXRy, cones were initially born, and fate specified in CrxE80A retinas (see Figure S7A). Since the E80A protein also hyper-activated S-cone opsin promoter-luciferase (Sop-luc) reporter in HEK293 cells (see Figure S7B), we predict that CRX E80A affects cone photoreceptor differentiation in a similar manner as rod photoreceptors. Furthermore, the cone transcriptional program might be more prone to perturbations by abnormal CRX activities. These possibilities require future investigations. For this manuscript, we have included all these points in the discussion section.

Another point is that it will be very valuable if the authors could show how E80A and K88N differently affect the 3D structure of the CRX homeodomain. Even a simulation model would be valuable.

Please see our answer to Point 1 of Reviewer #1. In short, we have added in the discussion section our explanations on how each mutation could affect CRX HD-DNA interactions differently based on structural studies. We further explain how these biochemical changes relate to the molecular perturbations and cellular phenotypes seen in vivo. Additionally, since TF-DNA interactions are diverse and dynamic across binding sites with different sequence features and genomic environments, future studies that systematically and quantitatively evaluate CRX transcriptional activity at different regulatory sequences would be important.