

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Aberrant homeodomain-DNA cooperative dimerization underlies distinct developmental defects in two dominant *CRX* retinopathy models

Yiqiao Zheng^{1,2,§}, Gary D. Stormo^{4*}, Shiming Chen^{2,3*}

¹ Molecular Genetics and Genomics Graduate Program, Division of Biological and Biomedical Sciences

² Department of Ophthalmology and Visual Sciences ³ Department of Developmental Biology ⁴ Department of Genetics, Washington University in St Louis, Saint Louis, Missouri, 63110, USA

* To whom correspondence should be addressed.

Dr. Shiming Chen

Tel. 314 747 4350

Fax. 314 747 4211

chenshiming@wustl.edu

Dr. Gary D. Stormo

Tel. 314 747 5534

stormo@wustl.edu

[§] Current address: Department of Biology, Massachusetts Institute of Technology

Keywords: *CRX*, homeodomain, transcription factor, missense mutations, DNA binding specificity, DNA binding cooperativity, inherited retinal disease, photoreceptor development, chromatin remodeling

Running title:
Aberrant *CRX* homeodomain cooperative dimerization

25 **ABSTRACT**

26 Paired-class homeodomain transcription factors (HD TFs) play essential roles in vertebrate
27 development, and their mutations are linked to human diseases. One unique feature of paired-class HD is
28 cooperative dimerization on specific palindrome DNA sequences. Yet, the functional significance of HD
29 cooperative dimerization in animal development and its dysregulation in diseases remain elusive. Using the
30 retinal TF Cone-rod Homeobox (CRX) as a model, we have studied how blindness-causing mutations in the
31 paired HD, p.E80A and p.K88N, alter CRX's cooperative dimerization, lead to gene misexpression and
32 photoreceptor developmental deficits in dominant manners. CRX^{E80A} maintains binding at monomeric WT
33 CRX motifs but is deficient in cooperative binding at dimeric motifs. CRX^{E80A}'s cooperativity defect impacts
34 the exponential increase of photoreceptor gene expression in terminal differentiation and produces immature,
35 non-functional photoreceptors in the *Crx*^{E80A} retinas. CRX^{K88N} is highly cooperative and localizes to ectopic
36 genomic sites with strong enrichment of dimeric HD motifs. CRX^{K88N}'s altered biochemical properties disrupt
37 CRX's ability to direct dynamic chromatin remodeling during development to activate photoreceptor
38 differentiation programs and silence progenitor programs. Our study here provides *in vitro* and *in vivo*
39 molecular evidence that paired-class HD cooperative dimerization regulates neuronal development and
40 dysregulation of cooperative binding contributes to severe dominant blinding retinopathies.

41 INTRODUCTION

42 Homeodomain transcription factors (HD TFs) are essential for diverse biological processes in
43 vertebrate development, including body plan specification, pattern formation, and cell fate specification (Mark
44 et al. 1997; Hobert 2021; Leung et al. 2022). Paradoxically, for a protein domain that has evolved numerous
45 functional specificities, it binds with high affinity to closely related DNA motifs that are typically only 5-6
46 base pairs long (Treisman et al. 1992; Wilson et al. 1996; Noyes et al. 2008; Bürglin and Affolter 2016). Thus,
47 additional mechanisms are required to achieve the individual functions of homeoproteins (Wilson et al. 1993).

48 The paired-class HD family possesses a unique feature in that members of this class confer
49 cooperative dimerization on specific dimeric DNA sequences. The “cooperative” interaction here is where the
50 first HD-DNA half-site binding greatly enhances the binding of a second molecule to the other half-site (Fig.
51 1A). The paired-class HD cooperative dimerization solely relies on the 60-amino-acid HD, distinguishing it
52 from the other HD families that require additional domains to form higher-order DNA binding complexes
53 (Hayashi and Scott 1990; Wilson et al. 1993). Yet, the functional importance of paired-class HD cooperative
54 dimerization in development and its dysregulation in human diseases remains elusive.

55 We have studied CRX, a paired-class HD TF essential for photoreceptor cells in the retina, as a model
56 to understand HD-DNA interactions in normal development and dominant blinding diseases (Tran and Chen
57 2014; Tran et al. 2014; Zheng et al. 2023; Zheng and Chen 2024). Photoreceptors are specialized neurons in
58 the retina that sense light and initiate vision through the phototransduction process. In vertebrates,
59 photoreceptors come in two major classes, rods and cones, that mediate vision in dim and bright light,
60 respectively. Animal studies have demonstrated that *Crx* expression is activated in post-mitotic photoreceptor
61 precursors and maintained throughout adult life (Chen et al. 1997; Furukawa et al. 1997; Muranishi et al.
62 2011). In the cellular context, disruption of CRX functions leads to significantly reduced enhancer activity of
63 photoreceptor gene regulatory elements and the loss of gene expression which ultimately results in
64 degeneration of immature, non-functional photoreceptors (Furukawa et al. 1999; Roger et al. 2014; Tran et al.
65 2014; Ruzycski et al. 2015; Ruzycski et al. 2018). Thus, CRX mainly functions as a positive regulator of
66 photoreceptor development and functions *in vivo*. Coding sequence mutations in human *CRX* have been
67 associated with at least three inherited retinal diseases (IRDs) that primarily affect photoreceptors: Leber

68 congenital amaurosis 7 (LCA7, OMIM: 613829), Cone-rod dystrophy 2 (CoRD2, OMIM: 120970), retinitis
69 pigmentosa (RP, OMIM: 268000). *CRX*-associated retinopathies vary significantly in the ages of onset,
70 severity, and disease progression (Tran and Chen 2014; Zheng and Chen 2024). The phenotype heterogeneity
71 suggests that individual mutation may cause disease via distinct pathogenic mechanisms. Deciphering these
72 mechanisms is, therefore, critical for informing the future development of therapeutic approaches.

73 *CRX* has two functional domains – the N-terminal DNA-binding domain (homeodomain, HD) and the
74 C-terminal transcription effector domain. Disease-associated mutations are distributed across both domains,
75 with amino acid substitutions primarily observed in the *CRX* HD (Tran and Chen 2014; Zheng and Chen
76 2024). To understand how HD mutations alter *CRX*-DNA interactions and lead to photoreceptor diseases, we
77 have previously reported two human mutation *knock-in* mouse models (Zheng et al. 2023), each carrying a
78 gain-of-function mutation, p.E80A (E80A) and p.K88N (K88N), that are associated with dominant LCA and
79 dominant CoRD, respectively. Using an integrated approach that combines quantitative *in vitro* biochemical
80 models, functional genomics, cellular profiling, and functional testing in mouse models, we found that E80A
81 and K88N alter *CRX* DNA binding specificity and produce distinct photoreceptor deficits in mutant mouse
82 retinas. Yet, the proposed mechanisms were primarily based on analyzing *CRX*-DNA interactions at
83 monomeric HD motifs both *in vitro* and *in vivo*.

84 Given that *CRX* is a paired-class HD TF, it is unclear whether E80A and K88N mutations affect *CRX*
85 HD's cooperative dimerization and how mutant *CRX* activity interferes with WT *CRX* functions when both
86 alleles are present, leading to the severe dominant photoreceptor deficits in developing mouse retinas. Here,
87 we extend our multi-omics approach to further elucidate the consequences of E80A and K88N mutations on
88 *CRX* regulatory activities in developing photoreceptors. We will discuss how a single transcription factor,
89 *CRX*, through differential interactions with monomeric and dimeric DNA motifs, regulates different
90 biological functions at different developmental stages and possibly in different cell types.

91

92 **RESULTS**

93 ***CRX* K88N but not WT or E80A HD confers strong cooperative dimerization on *pRho BAT-1* probe**

94 Paired-class HDs bind both monomeric and dimeric HD DNA motifs (Fig. 1A). Uniquely, paired-
95 class HDs can cooperatively dimerize on specific dimeric motifs, historically known as P3 sequences (Wilson
96 et al. 1993; Tucker and Wisdom 1999) (Fig. 1A bottom). In a P3 sequence, the two half-site core motifs 5'-
97 TAAT-3' are separated by a 3bp (base pair) spacer and form an approximate palindrome that places the two
98 homeodomains in a head-to-head arrangement (Wilson et al. 1995). Recognition helix residues that determine
99 paired-class HD's DNA binding specificity at monomeric motifs also confer distinct cooperative dimerization
100 properties, including the preferred spacer length and identity between the two 5'-TAAT-3' core motifs and the
101 magnitude of cooperativity. In the previous report, we found that disease-causing mutations E80A, K88N, and
102 R90W, all located within the CRX HD recognition helix, differentially affect CRX HD DNA-binding
103 specificity at monomeric sequences (Zheng et al. 2023). Here, we sought to understand whether any of the
104 three mutations affect CRX HD's cooperative dimeric binding using electrophoretic mobility shift assays
105 (EMSAs).

106 We chose the *BAT-1* EMSA probe, an established model template to assay CRX HD's dimeric DNA
107 binding (Chen et al. 1997; Chen et al. 2002) (Fig. 1B). The *BAT-1* sequence is a fragment in the promoter of
108 rhodopsin, a gene that encodes the rod-specific photopigment and is a direct target of CRX *in vivo*. The *BAT-1*
109 fragment contains a central dimeric P3 sequence TAATCATATTA and additional overlapping monomeric
110 HD motifs (Fig. 1B). We generated a series of *BAT-1* variant fragments to explicitly interrogate the
111 interactions between the two half-sites constituting the dimeric P3 motif (Fig. 1B). Specifically, in the *BAT-1*
112 P5 GA variant, the two half-sites are separated by a guanine (G) nucleotide, with each 6mer sequence
113 preserved. The P5 configuration (5bp spacer) has been shown to abolish cooperative dimerization of paired-
114 class homeoproteins (Wilson et al. 1993) and thus was used as a control to visualize non-cooperative dimeric
115 binding events.

116 WT HD bound strongly to *BAT-1* (WT) and P5 probes as monomeric and dimeric complexes but
117 demonstrated no obvious cooperativity (Fig. 1C,D), as exemplified by the saturation of the monomeric band
118 (M) before the gradual formation of the dimeric band (D). The difference in WT HD bound monomeric *versus*
119 dimeric band intensities between P3 and P5 probes may be attributed to DNA shape features of the two half-
120 sites that intrinsically affect HD-DNA interactions without explicitly changing the underlying sequences

121 (Mathelier et al. 2016; Li et al. 2024). K88N HD showed strong dimeric binding with weak monomeric
 122 binding at the *BAT-1* (WT) probe (Fig. 1C,D; Supplemental Fig. S1C), characteristic of cooperative
 123 dimerization where the binding of one molecule stimulates the binding of a second molecule (Wilson et al.
 124 1993; Tucker and Wisdom 1999). Increasing the spacer length (*BAT-1* P5 GA, Fig. 1D) or abolishing either
 125 half-site of the P3 sequence (*BAT-1* mA and mB, Supplemental Fig. S1C) resulted in diminished K88N HD
 126 dimeric binding, corroborating the essential P3 configuration and intact palindromic half-sites in mediating
 127 K88N HD's cooperative dimerization. The diminished K88N HD dimeric binding at *BAT-1* P5, mA, and mB
 128 probes unequivocally argues that K88N mutation does not render CRX^{K88N} proteins obligatory dimers, and
 129 cooperative dimerization is mediated through specific HD-DNA interactions. In comparison, E80A HD bound
 130 stronger as monomeric complexes and weaker as dimeric complexes at both *BAT-1* (WT) and P5 probes when
 131 compared to WT HD (Supplemental Fig. S1A,B), which may inherently relate to E80A HD's reduced binding
 132 specificity at monomeric sequences (Zheng et al. 2023); R90W HD bound poorly to either *BAT-1* (WT) or P5
 133 probe, consistent with R90W being a loss-of-function mutation that reduces CRX HD's overall DNA binding
 134 affinity (Chen et al. 2002). In summary, only CRX K88N but not WT or E80A HD confers strong cooperative
 135 dimerization on the *BAT-1* probe containing a P3 dimeric HD motif.

136 **CRX^{K88N} cooperative dimerization mediates strong reporter gene activation in HEK293T cells**

137 To determine the consequences of K88N HD's enhanced cooperative dimerization on CRX's
 138 transactivation activity, we performed luciferase reporter assays in the HEK293T cells with enhancers
 139 harboring three tandem repeats of *BAT-1* or variant sequences identical to that used in EMSAs. Consistent
 140 with EMSA results, CRX^{K88N} demonstrated strong activator activity at 3×*BAT-1* WT enhancer harboring the
 141 intact P3 dimeric sequence but much weaker activity at all *BAT-1* variants (Fig. 1E; Supplemental Fig. S1D).
 142 Thus, CRX^{K88N} is a competent transcription activator, and CRX^{K88N}'s cooperative dimerization on *BAT-1* P3
 143 sequence mediates strong gene activation likely through stabilizing the dimeric binding complexes.

144 CRX^{WT} only significantly activated the 3×*BAT-1* P5 and mA enhancers but not the 3×*BAT-1* WT and
 145 mB enhancers (Fig. 1E; Supplemental Fig. S1D). In the *BAT-1* mA sequence, the sub-optimal 5'-TAATCA-
 146 3'(A-f) is destroyed, and a single WT CRX consensus monomeric motif 5'-TAATCC-3' (B-f) remains intact.
 147 Since typical B-form DNA is arranged in helical turns of 10.5bp and paired-class HD's cooperative

148 dimerization involves DNA conformational changes (Wilson et al. 1995), the orientation of CRX protein
149 binding relative to the transcription start site and the relative orientation of two CRX proteins in the dimeric
150 binding complex are different in different *BAT-1* variants. These orientation differences may alter the
151 interaction surfaces for additional factors that bind to CRX and ultimately lead to the creation of the
152 transcription initiation complex.

153 **K88N enhances but E80A impairs CRX HD's cooperative dimerization at various P3 sequences *in vitro***

154 Although WT CRX HD did not show apparent cooperative dimerization at the *BAT-1* (WT) probe, it
155 is unclear whether it can form cooperative dimers on other P3 sequences. To characterize CRX HD's DNA
156 binding cooperativity unbiasedly, we adapted a high-throughput *in vitro* assay, Coop-seq, that determines
157 protein-DNA-binding cooperativity by sequencing (Chang et al. 2016; Hu et al. 2017) (Fig. 2A,B). Coop-seq
158 was developed based on traditional EMSAs, which allows the physical separation of distinct binding
159 complexes. Coop-seq enables us to accurately measure the cooperativity parameters – interactions between
160 two HD half-sites – for a library of dimeric HD DNA motifs in parallel and quantitatively compare the
161 cooperativity of different CRX HDs. Based on HD-DNA interaction models and previous Spec-seq-generated
162 monomeric CRX HD-DNA binding models (Zheng et al. 2023), we designed a Coop-seq library containing all
163 possible dimeric P3 spacer sequences TAATNNNATTA (Fig. 2C).

164 As a control, we first tested a P5 library (TAATNNGNNATTA) (Fig. 2C), which is expected to limit
165 cooperativity between half-sites (Wilson et al. 1993; Tucker and Wisdom 1999) (Fig. 1D; Supplemental Fig.
166 S1B). We obtained WT and mutant HDs' DNA binding cooperativity index ω at the P5 library using
167 bacterially-expressed and affinity-purified HD peptides as previously described (Methods). As expected, WT
168 and mutant CRX HDs showed weak cooperativity at P5 sequences (Supplemental Fig. S2A-D; Supplemental
169 Table S2). It supports the idea that CRX HDs primarily bind non-overlapping half-sites independently,
170 consistent with homeoproteins' high affinity binding at monomeric HD motifs.

171 Next, we compared CRX HDs' cooperativity profiles on the preferred P3 configuration sequences. To
172 visualize the specific cooperative interactions elicited by the P3 configuration, we normalized the
173 cooperativity index ω at P3 sequences (ω_{p3}) by their P5 counterparts (ω_{p5}) (Fig. 2C; Methods). WT HD
174 showed moderate cooperativity at a small subset of P3 sequences (Fig. 2D). Close examination of the P3

175 sequence spacers of this subset revealed an enrichment of guanine (G) and cytosine (C) bases (Supplemental
176 Table S3), which are known to be preferred by Lys50 (K₅₀) HD subfamily of paired-class homeoproteins,
177 including CRX. R90W HD showed a similar cooperativity profile as WT HD with slightly increased
178 cooperativity at a few P3 sequences. This is consistent with R90W being a loss-of-function mutation that
179 reduces the overall HD-DNA binding affinity without selectively altering the specific HD-DNA contacts
180 (Chen et al. 2002; Zheng et al. 2023). E80A HD exhibited reduced cooperativity compared to WT HD, while
181 K88N HD showed enhanced cooperativity at all P3 sequences. A previous study found that a Gln50 (Q₅₀)
182 paired-class HD shows more than 10-fold stronger cooperativity than a K₅₀ HD (Wilson et al. 1993). Since
183 K88N mutation alters CRX HD's DNA binding specificity at monomeric motifs to one that mimics a natural
184 Q₅₀ HD (Zheng et al. 2023), K88N HD's enhanced cooperativity at dimeric motifs can be attributed to its
185 similarity to a Q₅₀ HD.

186 Collectively, *in vitro* protein-DNA binding results indicate that E80A mutation impairs CRX HD's
187 DNA binding cooperativity at specific dimeric HD motifs while K88N mutation drastically alters both
188 specificity and cooperativity. The luciferase reporter assays highlight the functional distinctions between
189 cooperative dimeric binding at P3 sequences in contrast to non-cooperative dimeric binding and individual
190 binding to monomeric sites.

191 ***Crx*^{K88N/+} and *Crx*^{K88N/N} retinas show severely decreased accessibility at CREs enriched for K₅₀ HD motifs**

192 Next, we sought to understand the roles of CRX's monomeric and cooperative dimeric binding on the
193 regulation of photoreceptor development using *Crx*^{E80A} and *Crx*^{K88N} mouse models established in our previous
194 study (Zheng et al. 2023). We asked how E80A and K88N mutations affect CRX's ability to facilitate
195 chromatin remodeling by performing retinal ATAC-seq on post-natal day 14 (P14) WT, heterozygous, and
196 homozygous mutant retinas. For concision, we use *Crx*^{E80A} and *Crx*^{K88N} when both heterozygous and
197 homozygous mutants are discussed. The ATAC-seq results were analyzed in conjunction with our published
198 P14 CRX ChIP-seq data. We asked how E80A and K88N mutation-specific changes in HD-DNA interactions
199 affect the chromatin landscape in individual mutant models and how perturbed epigenome relates to
200 photoreceptor gene mis-expression.

201 Since CRX^{WT} and CRX^{K88N} prefer different monomeric HD motifs *in vitro* and *in vivo* (Zheng et al.
 202 2023), we predicted that the *Crx*^{K88N/N} retinas show similar chromatin remodeling defects as the loss-of-
 203 function *Crx*^{R90W/W} model and CRX^{WT} proteins in the *Crx*^{K88N/+} retinas can bind canonical CRX binding sites
 204 and facilitate chromatin remodeling. We found that the *Crx*^{K88N/N} retinas showed more significant chromatin
 205 accessibility loss at canonical CRX binding sites than the loss-of-function *Crx*^{R90W/W} retinas and the *Crx*^{K88N/+}
 206 retinas showed reduced accessibility similar to the *Crx*^{R90W/W} retinas (Fig. 3A). Genomic regions that showed
 207 defective chromatin remodeling in the *Crx*^{K88N/+} and *Crx*^{K88N/N} retinas were enriched for the K₅₀ HD motifs
 208 (Fig. 3B), gain accessibility in normal post-natal retinal development (Fig. 3C), and regulate genes important
 209 for photoreceptor structures, functions, and maintenance (Fig. 3D). The impaired chromatin remodeling at
 210 canonical CRX binding sites led to significant photoreceptor gene downregulation in the developing *Crx*^{K88N}
 211 retinas (Fig. 3E). Thus, defects in chromatin remodeling at regions with K₅₀ HD motifs underlie defective
 212 photoreceptor differentiation in young adults in the *Crx*^{K88N/+} and *Crx*^{K88N/N} mice (Fig. 3F) (Zheng et al. 2023).

213 ***Crx*^{K88N/+} and *Crx*^{K88N/N} retinas show increased accessibility at CREs enriched for Q₅₀ HD motifs**

214 Since the *Crx*^{R90W/+} retinas, expressing a single dose of CRX^{WT} proteins, show largely WT
 215 phenotypes (Tran et al. 2014), the severe chromatin remodeling defects and photoreceptor gene misexpression
 216 in the *Crx*^{K88N} retinas likely attributed to CRX^{K88N} ectopic activities. The *Crx*^{K88N}-increased accessibility
 217 ATAC-seq peaks coincided with ectopic CRX^{K88N} binding but showed comparably low accessibility in the
 218 *Crx*^{R90W/W} and WT retinas (Fig. 4A). *de novo* motif searching on sequences under *Crx*^{K88N}-increased ATAC-
 219 seq peaks revealed both dimeric and monomeric Q₅₀ HD motifs (Fig. 4B), consistent with *in vitro* found
 220 alterations in K88N HD-DNA binding specificity and cooperativity. The *Crx*^{K88N} retina ectopically enriched
 221 dimeric HD motifs exhibit no base preferences within the 3bp spacer (Fig. 4B, positions 5-7), in line with
 222 K88N HD's strong cooperativity at nearly all P3 Coop-seq library sequences (Fig. 2D). Collectively, these
 223 observations suggest that ectopic CRX^{K88N} activity at Q₅₀ HD motifs rather than loss of CRX^{WT} activity
 224 mediates chromatin accessibility increase at ectopic sites in the *Crx*^{K88N/+} and *Crx*^{K88N/N} retinas.

225 ***Crx*^{K88N}-increased accessibility CREs show progenitor cell regulatory signatures and are**
 226 **developmentally silenced during photoreceptor differentiation**

227 Last, we sought to understand the functional significance of CRX^{K88N}-associated chromatin
228 accessibility increase. Different from previously characterized mutant CRX proteins that recognize K₅₀ HD
229 motifs, CRX^{K88N} prefers Q₅₀ HD motifs that are recognized by Q₅₀ HD TFs highly expressed in retinal
230 progenitor cells (Bassett and Wallace 2012; Zagozewski et al. 2014). In the developing mouse retinas, these
231 Q₅₀ HD TFs coordinate the dynamic chromatin landscape changes during retinal neurogenesis and are
232 downregulated in differentiating photoreceptors (Zibetti et al. 2019; Bian et al. 2022). We predicted that
233 CRX^{K88N} ectopic activities affect the chromatin landscape changes at progenitor CREs that are usually bound
234 by progenitor Q₅₀ homeoproteins.

235 Using a previously published ATAC-seq dataset of normal mouse retinal development (Aldiri et al.
236 2017), we found that *Crx*^{K88N}-increased ATAC-seq peaks showed the strongest accessibility at neonatal ages
237 P0 and P3, followed by a gradual decrease in accessibility as photoreceptors undergo differentiation (Fig. 4C).
238 Re-analysis of a published VSX2 (Q₅₀ HD TF) retinal ChIP-seq data (Bian et al. 2022) showed that embryonic
239 day 14.5 (E14.5) but not adult VSX2 binding is enriched at the *Crx*^{K88N}-increased ATAC-seq peaks (Fig. 4D).
240 In the embryonic retina, VSX2 is expressed in retinal progenitor cells; in the adult retina, VSX2 expression is
241 maintained in bipolar cells and Müller glia but lost in mature photoreceptors (Liu et al. 1994; Burmeister et al.
242 1996; Rowan and Cepko 2004). Thus, the enrichment of embryonic VSX2 binding and the depletion of adult
243 VSX2 binding at *Crx*^{K88N}-increased ATAC-seq peaks suggests that CRX^{K88N} ectopic activity is likely
244 impeding the silencing of progenitor chromatin states instead of directing photoreceptor precursors into
245 alternative cell lineage programs. Additionally, *de novo* motif enrichment analysis of sequences under
246 *Crx*^{K88N}-increased ATAC-seq peaks found patterns characteristic of basic helix-loop-helix (bHLH) neurogenic
247 transcription factor consensus (Fig. 4E), corroborating the potential functionality of *Crx*^{K88N}-increased
248 ATAC-seq peaks in regulating neurogenic programs during normal development. Gene Ontology (GO)
249 analysis revealed that genes adjacent to *Crx*^{K88N}-increased ATAC-seq peaks were implicated in general
250 neuronal development (Fig. 4F). Collectively, these pieces of evidence suggest that instead of initiating
251 accessibility *de novo*, CRX^{K88N}'s ectopic activities at Q₅₀ HD motifs maintain the accessibility of chromatin
252 regions that are developmentally closed which may create a chromatin environment that is inhibitive to TFs
253 that regulate photoreceptor differentiation.

254 ***Crx*^{E80A} retinas show opposite chromatin accessibility changes at CREs enriched for monomeric and**
 255 **dimeric K₅₀ HD motifs**

256 Different from the K88N mutation that drastically alters CRX HD's DNA binding specificity, the
 257 E80A mutation does not affect CRX HD's DNA sequence preference *per se* but reduces binding specificity at
 258 monomeric motifs (Zheng et al. 2023) and impairs binding cooperativity at specific dimeric motifs (Fig. 2C).
 259 We asked if these changes differentially affected CRX^{E80A}'s ability to modulate chromatin remodeling.
 260 Among the consensus ATAC-seq peaks identified in WT, *Crx*^{E80A/+}, and *Crx*^{E80A/A} retinas, a set of peaks
 261 (n=1681) showed significantly increased ATAC-seq signal intensity in the *Crx*^{E80A/A} retinas compared to WT,
 262 with the *Crx*^{E80A/+} retinas showing intermediate intensity at these loci (Fig. 5A). The *Crx*^{E80A}-increased ATAC-
 263 seq peaks showed increased CRX^{E80A} binding in the *Crx*^{E80A/A} retinas and were enriched for the WT CRX
 264 consensus monomeric K₅₀ HD motifs (Fig. 5B). No additional TF motif was found significantly enriched in
 265 these peaks suggesting a major contribution from CRX^{E80A} binding and activity. A smaller subset of peaks
 266 (n=485) showed significantly decreased ATAC-seq signals in the *Crx*^{E80A/A} retinas compared to WT, again
 267 with the *Crx*^{E80A/+} retinas showing intermediate signal reduction (Fig. 5A). The *Crx*^{E80A}-reduced ATAC-seq
 268 peaks showed diminished CRX^{E80A} binding in the *Crx*^{E80A/A} retinas correlating with defective chromatin
 269 remodeling. *de novo* motif searching of sequences under the *Crx*^{E80A}-reduced ATAC-seq peaks identified a
 270 dimeric P3 sequence pattern resembling dimeric K₅₀ HD motifs, highlighted by a preference for cytosine (C)
 271 at the spacer positions 5 and 6 (Fig. 5C).

272 Since CRX proteins can, theoretically, bind non-cooperatively to the two half-sites constituting a P3
 273 dimeric motif as observed in EMSAs with *BAT-1* probes (Fig. 1B), we sought to determine what sequence
 274 features of the dimeric motifs under *Crx*^{E80A}-reduced ATAC-seq peaks rendered them susceptible to defects in
 275 CRX's cooperative dimerization. We first identified instances of dimeric K₅₀ HD motifs under *Crx*^{E80A}-
 276 reduced ATAC-seq peaks by scanning the DNA sequences with FIMO (Grant et al. 2011) using the dimeric
 277 K₅₀ HD motif PWM (position weight matrix) in Fig. 5C at a *p-value* threshold of 1×10^{-3} (Methods). We then
 278 estimated how well can WT CRX proteins bind individually to the two half-sites by calculating their relative
 279 binding affinities using a CRX monomeric binding PWM model (Lee et al. 2010). We found that these
 280 dimeric motifs are often consist with two low-affinity half-sites, suggesting cooperative dimerization may be

281 crucial to facilitate CRX's stable binding (Supplemental Fig. S3A). This is distinct from previous analyses
282 that have focused mostly on the highest affinity (not the strongest cooperativity) dimeric K₅₀ HD motifs on
283 non-chromatin templates (Wilson et al. 1993; Hughes et al. 2018). Our observations suggest that cooperative
284 dimerization is critical for CRX regulatory activities at specific dimeric K₅₀ motifs *in vivo* and in the
285 chromatin context. Impaired cooperative dimerization likely reduces CRX^{E80A}'s stable binding at these
286 dimeric K₅₀ HD motifs and in turn, affects the chromatin remodeling activity.

287 ***Crx*^{E80A}-differentially accessible CREs exhibit distinct chromatin remodeling kinetics in normal**
288 **development**

289 Previously, we observed that photoreceptor differentiation is perturbed in a cell-type and
290 developmental stage-specific manner in *Crx*^{E80A} retinas (Zheng et al. 2023). Rod photoreceptors show
291 precocious early differentiation but defective terminal differentiation, while cone photoreceptors show a lack
292 of differentiation at both stages. We asked whether these patterns were associated with CRX^{E80A}'s differential
293 impacts on genomic regions of different regulatory functions. It is important to note that the mouse retina is
294 rod-dominant and thus, global patterns in bulk ATAC-seq signals mainly reflect chromatin accessibility
295 landscape in rods, and regulatory elements for cones need to be evaluated in a gene-specific manner. Using a
296 previously published ATAC-seq dataset of normal mouse retinal development (Aldiri et al. 2017), we found
297 that both *Crx*^{E80A}-increased and *Crx*^{E80A}-reduced ATAC-seq peaks gain accessibility during post-natal retinal
298 development (Fig. 5D,E). Yet, the two sets of peaks differ in their kinetics of accessibility gain. In WT retinas,
299 the *Crx*^{E80A}-increased ATAC-seq peaks exhibit a strong increase in accessibility during early photoreceptor
300 development, from post-natal day 0 (P0) to P7, followed by a moderate change from P7 to P21. The *Crx*^{E80A}-
301 reduced ATAC-seq peaks are characterized by an exponential gain in accessibility between P10 and P14 with
302 smaller changes prior to P10 and after P14. The P10-P14 time points represent a critical window of
303 photoreceptor differentiation characterized by a significant change in the photoreceptor transcriptome (Kim et
304 al. 2016) and the elaboration of Outer Segments (OSs), the subcellular structures where phototransduction
305 occurs (Swaroop et al. 2010). The distinct developmental accessibility kinetics suggest that CRX^{E80A} activity
306 at monomeric and dimeric K₅₀ HD motifs might affect gene expression at different stages of photoreceptor
307 development in the mutant mouse retinas.

308 **Monomeric and dimeric K₅₀ HD motifs associated with stage-specific photoreceptor gene mis-**
309 **expression in the *Crx*^{E80A} retinas**

310 To identify the likely direct impacts of CRX^{E80A} mutant activity on gene expression, we focused on
311 the “CRX-dependent activated genes” (CRX-DAGs) defined previously (Zheng et al. 2023). CRX-DAGs are
312 adjacent to at least one CRX binding site, and their expressions are significantly reduced in the loss-of-
313 function model *Crx*^{R90W/W}. CRE-gene association analysis using ATAC-seq identified CREs that also overlap
314 with CRX binding revealed that many CRX-DAGs are potentially under combinational regulations of
315 monomeric and dimeric K₅₀ HD motifs (Supplemental Table S6; Methods). Specifically, 28.87% of CRX-
316 DAGs are associated with CREs that only contain monomeric K₅₀ HD motifs while 61.27% of CRX-DAGs
317 are associated with one or more CREs that contain both monomeric and dimeric K₅₀ HD motifs (Fig. 6A). By
318 analyzing a previously published RNA-seq dataset of normal mouse retinal development (Aldiri et al. 2017),
319 we found that genes regulated by both monomeric and dimeric K₅₀ HD motifs display a broader range of
320 expression changes with genes critical for rod photoreceptor terminal differentiation displaying an exponential
321 increase, such as *Esrrb*, *Gnat1*, and *Rho* (Supplemental Fig. S4A,B). This suggests that CRX’s action on
322 dimeric motifs may be associated with unique gene expression changes for a specific subset of target genes.

323 To understand how altered CRX^{E80A}-DNA interactions at monomeric and dimeric K₅₀ HD motifs
324 relate to photoreceptor gene mis-expression in the developing (P10) and mature (P21) *Crx*^{E80A} mouse retinas,
325 we reanalyzed the RNA-seq data generated in our previous study (Zheng et al. 2023). In the P10 *Crx*^{E80A/A}
326 retinas, CRX-DAGs associated either solely with monomeric K₅₀ HD motifs or also with dimeric K₅₀ HD
327 motifs showed over-expression compared to WT (Fig. 6B, P10 columns). Since regulatory elements enriched
328 for the monomeric K₅₀ HD motifs display an early chromatin remodeling profile in normal development (Fig.
329 5D), it is likely that CRX^{E80A} promiscuous binding at monomeric K₅₀ HD motifs (Zheng et al. 2023)
330 accelerated the chromatin remodeling and/or directly enhanced the expression of CRX-DAGs in the
331 developing *Crx*^{E80A/A} mutant retinas. In the P21 adult *Crx*^{E80A/A} retinas, expression of the P10 *Crx*^{E80A/A}-
332 overexpressed genes showed two patterns: genes associated solely with monomeric K₅₀ HD motifs became
333 comparable to WT or remained over-expressed but of a much lower magnitude; genes associated with both
334 monomeric and dimeric K₅₀ HD motifs became comparable to WT and even significantly downregulated (Fig.

335 6B, P21 columns). The $Crx^{E80A/+}$ retinas showed similar patterns of gene mis-expression but at a less severe
336 degree (Supplemental Fig. S5A). The selective downregulation of dimeric K₅₀ HD motif-associated genes
337 may be explained by CRX^{E80A}'s impaired cooperative dimerization and subsequently defective chromatin
338 remodeling at regulatory elements enriched for the dimeric HD motifs. Genes encoding structural proteins of
339 photoreceptor OSs and molecules involved in the second messenger cascade of the visual cycle are enriched
340 in this specific set of CRX target genes (Supplemental Fig. S5B; Supplemental Table S6). The precise
341 regulation of these genes is fundamental to the integrity of photoreceptor OSs and functions (Purves and
342 Williams 2001). Perturbations of these genes have been associated with inherited retinal dystrophies that
343 affect rods, cones, or both, and cause blindness (García Bohórquez et al. 2021). Thus, the under-expression of
344 these genes likely underlies the defective photoreceptor terminal differentiation and functions in the Crx^{E80A}
345 retinas.

346 There is a subset of dimeric motif-associated CRX-DAGs that were downregulated at P10 and
347 become more severely down at P21 (Fig. 6B; Supplemental Fig. S5A). This gene set is implicated in cone
348 photoreceptor structures and functions. In the mouse retinas, cones are born embryonically while most rods
349 are born postnatally, and their differentiation progresses differently. At the postnatal ages examined, it is
350 likely that we captured different stages of cone and rod photoreceptor differentiation.

351 **Monomeric and dimeric K₅₀ HD motifs demonstrate regulatory activity changes in *ex plant* retinas**

352 The analyses above clearly demonstrate a close relationship between dysregulated CRX^{E80A} activity at
353 SUPPdifferent K₅₀ HD motifs and gene mis-expression in the Crx^{E80A} retina. Yet, it remains unclear whether
354 CRX^{E80A} acts directly or through modulating chromatin accessibility to affect gene expression. To evaluate the
355 relative contribution of CRX^{E80A}-DNA interactions to transcription regulation of non-chromatin templates, we
356 selected and tested a representative set of CRX-bound *cis*-regulatory elements (CREs) in episomal plasmids
357 with Massively Parallel Reporter Assays (MPRAs) in *ex plant* WT and mutant mouse retinas (Fig. 7A;
358 Supplemental Fig. S7A,B; Methods). We measured the regulatory activities of genomic CREs with intact HD
359 motifs and with activity abolishing mutations in the HD motifs. The regulatory activity difference between
360 each pair of CREs with intact and mutated HD motifs yields an HD motif activity measurement that reflects
361 the direct functional consequences of altered CRX-HD motif interactions (Methods). Comparison between

362 genotypes found a significant increase in monomeric K₅₀ HD motif activity in the *Crx*^{E80A/A} retinas (Fig. 7B),
363 suggesting that CRX^{E80A} binding at monomeric motifs can directly drive increased target gene expression in
364 the developing retinas. Different from the monomeric motifs, dimeric K₅₀ HD motifs showed similar degree of
365 activity reduction in both *Crx*^{E80A/+} and *Crx*^{E80A/A} retinas (Fig. 7C). Since K₅₀ HD's dimeric binding is of much
366 lower affinity compared to its monomeric binding (Wilson et al. 1993), it is possible that proteins generated
367 from the single allele of *Crx*^{WT} is insufficient to produce stable binding at dimeric K₅₀ HD motifs in the
368 *Crx*^{E80A/+} retinas. The MPRA results, combined with alterations of photoreceptor epigenome *in vivo*, suggest
369 that CRX^{E80A} mis-regulates gene expression by acting through both chromatin remodeling and direct
370 transactivation pathways.

371

372 DISCUSSION

373 Homeodomain transcription factors (HD TFs) control the development and functions of many tissues.
374 The retina is an excellent system for understanding how a single HD TF achieves functional specification in
375 different cell types and at different times in development. The development and homeostasis of retinal
376 photoreceptor cells are controlled by a master HD TF, CRX. Human mutations in *CRX* cause a spectrum of
377 inherited retinal diseases (IRDs) that show significant heterogeneity in clinical phenotypes.

378 Since *CRX* is expressed in both cone and rod photoreceptors, both during development and in adults,
379 how does one explain the differential consequences of *CRX* mutations? In this study, we have extended our
380 previous investigations on two *CRX* HD missense mutations, E80A and K88N, that are associated with
381 dominant CoRD and dominant LCA in humans. Besides the previously identified differential impacts on DNA
382 binding specificity at monomeric motifs (Zheng et al. 2023), E80A and K88N mutations also differently alter
383 *CRX* HD's cooperative binding at dimeric HD sequences. The mutation-specific effects on DNA binding
384 specificity and cooperativity underlie the chromatin landscape changes that explain the distinct dominant
385 photoreceptor gene misexpression patterns in the mutation *knock-in* mouse retinas. Unlike other *Crx* mouse
386 models, no obvious photoreceptor degeneration was observed in either *Crx*^{E80A} or *Crx*^{K88N} retinas at the ages
387 examined (Zheng et al. 2023), suggesting that the defects in chromatin remodeling and gene expression
388 largely attributed to changes in *CRX*'s intrinsic regulatory functions instead of changes in cell number.

389 Analysis of the epigenome and transcriptome dynamics in normal development and in the Crx^{E80A} and Crx^{K88N}
390 mutant retinas highlight an underappreciated role of DNA-mediated CRX cooperative dimerization in
391 ensuring proper temporal chromatin and gene expression changes in photoreceptor development.

392 HD residue 50 (corresponding to CRX K88) determines paired-class HDs' DNA binding specificity at
393 monomeric and dimeric HD motifs as well as HD's cooperative binding at palindrome dimeric motifs (Hanes
394 and Brent 1989; Trelisman et al. 1989; Hanes and Brent 1991; Wilson et al. 1993). In addition to the difference
395 in sequence preference at their respective monomeric consensus motifs, a K_{50} HD bound with an order of
396 magnitude higher affinity and more than 100-fold in the half-life of the binding complex than a Q_{50} HD (Ades
397 and Sauer 1994); at dimeric consensus motifs, a Q_{50} HD bound with 10-fold stronger cooperativity than a K_{50}
398 HD (Wilson et al. 1993). Asparagine (N) is structurally similar to glutamine (Q) in that both of them contain
399 amide (NH₂) groups in their respective side chains, and they differ only by one methylene group. Expectedly,
400 the K88N mutation alters CRX HD's DNA binding specificity from K_{50} to that resembling a Q_{50} HD (Zheng
401 et al. 2023) and enhances cooperative dimerization both at the *BAT-1* P3 probe (Fig. 1C) and at Coop-seq P3
402 oligos (Fig. 2D). The drastic difference between WT and K88N HDs in the scale and spectrum of
403 cooperativity agrees with the random-site selection assays (Wilson et al. 1993) where a Q_{50} HD bound
404 strongly to many P3 sequences yielding a 5'-TAATPyNPuATTA-3' consensus while a K_{50} HD only selected
405 the monomeric consensus 5'-TAATCC-3' under the same conditions as the Q_{50} HD and three additional
406 rounds of selection were required to recover the dimeric consensus 5'-TAATCCGATTA-3'. In sum, CRX
407 K88N HD resembles a natural Q_{50} HD in both monomeric and cooperative dimeric binding.

408 In the developing WT mouse retinas, CRX binding at K_{50} HD motifs is essential for photoreceptor
409 differentiation by facilitating chromatin remodeling and regulating target gene expression (Ruzycki et al.
410 2018). CRX^{K88N}'s diminished activity at canonical CRX motifs leads to defective chromatin remodeling at
411 photoreceptor CREs and loss of target gene expression in the $Crx^{K88N/N}$ retinas (Fig. 3) (Zheng et al. 2023). Yet,
412 K88N mutation's dominant inheritance pattern and the more severe gene mis-expression in the $Crx^{K88N/N}$
413 retinas than the loss-of-function $Crx^{R90W/W}$ retinas (Fig. 3E) suggests that CRX^{K88N}'s ectopic activities at Q_{50}
414 HD motifs have a significant impact on photoreceptor development. Q_{50} HD motifs are recognized by many
415 HD containing TFs expressed in the retinal progenitor cells (Bassett and Wallace 2012; Heavner and Pevny

416 2012; Zagozewski et al. 2014). These progenitor Q₅₀ HD TFs promote retinal progenitor cell proliferation and
417 thus are inhibitory to neurogenesis and subsequently post-mitotic differentiation (Burmeister et al. 1996;
418 Green et al. 2003; Gordon et al. 2013). In *Crx*^{K88N} retinas, CRX^{K88N} may function as an ectopic Q₅₀ HD TF in
419 post-mitotic photoreceptors to maintain the cells at an undifferentiated state. Mechanistically, it remains
420 unclear what are the molecular pathways that respond to CRX^{K88N}'s ectopic activities and in turn modulate the
421 epigenetic status of *cis*-regulatory elements crucial for photoreceptor development. Differentially expressed
422 genes near CRX^{K88N}'s ectopic accessible regions do not show significant enrichment in a single Gene
423 Ontology or pathway. This suggests the epigenomic and transcriptomic changes in *Crx*^{K88N} retinas may be a
424 collective consequence of perturbations on many biological processes. In addition, data in our study were
425 generated from whole retina samples which can mask small cell-type specific changes, and the timepoints
426 investigated in this study do not capture photoreceptor precursors which may be the critical stage of
427 CRX^{K88N}'s actions. Single-cell profiling or molecular characterization of purified photoreceptors from WT
428 and *Crx*^{K88N} mutant retinas at different ages will shed light on the antagonistic interplays between CRX^{K88N}
429 and CRX^{WT} functions. These targeted analyses of pure cell populations will be crucial to understanding
430 whether the developmental accessibility pattern of *Crx*^{K88N}-ectopic regions in Figure 4C was due to changes in
431 cell type proportion or specific changes in post-mitotic photoreceptors. In summary, CRX target specificity is
432 not only critical for activating photoreceptor developmental programs but also crucial for the proper silencing
433 of early programs that could be inhibitory to photoreceptor differentiation at later stages.

434 Different from K88N's global impacts on CRX HD-DNA interactions, E80A mutation increases HD's
435 promiscuous binding at monomeric K₅₀ HD motifs (Zheng et al. 2023) but reduces cooperative binding at
436 dimeric K₅₀ HD motifs (Fig. 2D). CRX^{E80A}'s differential interactions with K₅₀ HD motif sub-types parallel the
437 hyper-activation of rod photoreceptor genes in developing *Crx*^{E80A} retinas and the hypo-activation of the same
438 set of genes in adult *Crx*^{E80A} retinas (Zheng et al. 2023). These observations suggest a K₅₀ HD motif division-
439 of-labor model where early-stage photoreceptor development is mediated by CRX interactions with
440 monomeric K₅₀ HD motifs, while terminal differentiation additionally relies on interactions with dimeric K₅₀
441 HD motifs that specifically requires cooperative dimerization (Fig. 8). This dichotomy of K₅₀ HD motif usage
442 likely relates to the different CRX-chromatin interaction kinetics collectively determined by the underlying

443 DNA sequences and CRX concentrations (Fig. 5D,E; Supplemental Fig. S4C,D). Specifically, CRX binds
444 monomeric K_{50} HD consensus motifs with high affinity which can activate gene expression at relatively low
445 CRX concentration, such as in photoreceptor precursors. However, CRX binding easily saturates at high
446 affinity monomeric motifs resulting in a small dynamic range of regulatory activity. Compared to binding at
447 monomeric motifs, K_{50} HD's dimeric binding to P3 sequences is of much lower affinity (Wilson et al. 1993)
448 and dimeric K_{50} HD motifs associated with terminal stage photoreceptor gene expression are made up of
449 individually low-affinity half-sites. Thus, these dimeric K_{50} HD motifs are less active at low CRX
450 concentration and remain responsive to higher and a wider range of CRX concentrations. The requirement of
451 CRX's cooperative dimerization for activity suggests that these low-affinity half-sites are poorly bound
452 individually under the physiological CRX concentration and cooperativity is crucial for stabilizing the binding
453 complex. Many CRX-DAGs that are regulated by monomeric and dimeric K_{50} HD motifs encode proteins in
454 the phototransduction pathway. Proteins in this pathway need to express robustly and remain dynamic in
455 response to changes in ambient illumination. Thus, it is likely that combinational regulation of monomeric and
456 dimeric K_{50} HD motifs imparts functional specification of CRX to regulate a subset of target genes that play
457 specific physiological roles in photoreceptor biology. Collectively, precision in CRX-DNA interactions is
458 important for not only the quantitative regulation but also the temporal control of photoreceptor gene
459 expression.

460 Although the K_{50} HD motif division-of-labor model explains rod gene expression alternations in the
461 Crx^{E80A} retinas, it remains unclear why the subset of cone genes is significantly downregulated at both ages
462 examined (Fig. 6B; Supplemental Fig. S6A). Since cones were born in Crx^{E80A} mutant retinas, the loss of cell-
463 type specific gene expression in early post-natal development suggests defective differentiation (Zheng et al.
464 2023). Natively, cones and rods are generated from distinct pools of retinal progenitor cells that are inherently
465 different in their competence (Hafler et al. 2012; Cepko 2015; Wang and Cepko 2016). It is conceivable that
466 photoreceptor precursors generated from these distinct progenitor pools are different in their baseline
467 epigenome architectures, which have a fundamental influence on the genetic programs required to reprogram
468 the epigenome during cone and rod differentiation. In support of this prediction, 81.25% of the cone-enriched
469 CRX-DAGs are associated with dimeric K_{50} HD motifs as compared to only 55.45% of rod-enriched genes

470 (Supplemental Fig. S5A) suggesting cone genes may be more dependent on CRX activity at dimeric K₅₀ HD
471 motifs. An alternative but not mutually exclusive model is that cone gene expressions are more sensitive to
472 perturbations in CRX activity. In the developing and mature mouse retinas, cones are dependent on a different
473 repertoire of TFs than rods for differentiation and functions (Swaroop et al. 2010; Forrest and Swaroop 2012;
474 Emerson et al. 2013; Sapkota et al. 2014; Jean-Charles et al. 2018). Many rod-specific TFs collaborate with
475 CRX in strongly activating rod gene expressions. Small perturbations in CRX activity may be dynamically
476 compensated by CRX collaborating factors. In contrast, although many nuclear receptor family TFs have been
477 identified to mediate M- vs S-cone subtype differentiation, these factors are dispensable for cone cell genesis,
478 development, or survival in early post-natal ages (Forrest and Swaroop 2012). It is possible that CRX plays a
479 major role in regulating general cone cell development and functions. The mouse retina is rod-dominant and
480 thus is limited in the resolution of cone-related mechanistic understandings. Quantitative characterization of
481 CRX molecular functions in a pure cone population and comparison with the *Crx*^{E80A} model warrants further
482 study to elucidate regulatory principles in early photoreceptor development and in *CRX*-linked dominant
483 CoRD.

484 Besides the identification of two novel pathogenic mechanisms, our comparative analysis of the
485 epigenome and transcriptome in WT and *Crx* mutant mouse retinas suggests two regulatory principles at
486 important photoreceptor developmental transitions (Fig. 8). First, the transition from proliferating retinal
487 progenitor cells to committed photoreceptors is accompanied by a shift from highly expressed Q₅₀ to K₅₀
488 paired-class HD TFs and from regulatory elements enriched with Q₅₀ to those with high-affinity monomeric
489 K₅₀ HD motifs. Second, the transition from early to late-stage photoreceptor development requires the
490 utilization of dimeric K₅₀ HD motifs in addition to monomeric K₅₀ HD motifs. The first transition involves a
491 sharp change in DNA binding specificity which likely confers sensitivity in newly post-mitotic photoreceptor
492 precursors to respond to low concentrations of CRX and quickly fix to a committed photoreceptor precursor
493 status. This strategy may also ensure the rapid elimination of progenitor epigenetic features due to a lack of
494 interacting TFs. In contrast, the second principle implies a functional specialization of CRX at a specific set of
495 target genes whose expressions need to remain dynamic and robust in mature photoreceptors. Related to the
496 differential regulatory activities of different HD motifs, episomal reporter gene assays in the retinal *ex plant*

497 system have shown that CRX-bound CREs containing multiple copies of high-affinity monomeric consensus
498 K₅₀ HD motifs are likely to act as silencers of gene expression (White et al. 2016; Friedman et al. 2021). Since
499 these elements were tested outside of the native genomic context in an *ex vivo* system, whether they play a
500 similar role in CRX's regulation of photoreceptor chromatin and gene expression *in vivo* warrants further
501 study. Collectively, our findings support a unifying model in which differential CRX interactions with
502 different HD motifs underlie cell-stage specific chromatin remodeling and temporal gene regulation during
503 photoreceptor development. A similar mechanism has been described for the SOX9 transcription factor where
504 its DNA-dependent cooperative dimerization is crucial for regulating genes for chondrogenesis but not for sex
505 determination (Bernard et al. 2003). The examples of CRX and SOX9 suggest that some transcription factors
506 have evolved distinct modes of DNA interactions which allow them to regulate diverse biological processes in
507 different cellular contexts.

508 Our study here refines the CRX mechanistic model in photoreceptor development, expands our
509 knowledge of the diverse mechanisms that CRX mutations lead to severe, dominant retinopathies, and lays the
510 foundation for the future development of therapeutic strategies targeting different pathogenic mechanisms.
511 Our CRX mechanistic model emphasizes the importance of considering interactions between coding CRX
512 variants and non-coding variants in CRX binding sites in modifying clinical phenotypes. Our study also
513 demonstrates, in addition to its unique biochemical properties (Wilson et al. 1993), that paired-class HD
514 cooperative dimerization plays a crucial role in development and its dysregulation can lead to distinct human
515 diseases. The adaption of Coop-seq enables the unbiased identification of CRX HD cooperative dimerization
516 as opposed to non-cooperative co-binding, which is not easily separable in selection-based TF-DNA binding
517 assays. Our study provides hints towards understanding the structural basis of paired-class HD cooperative
518 dimerization on palindrome DNA sequences (Wilson et al. 1995). Multiple missense mutations at the CRX
519 E80 residue have been reported in dominant CoRD cases, including p.E80K (ClinVar VCV000099599),
520 p.E80G (VCV000865803), and p.E80A (VCV000007416). Systematic investigation on how disease-
521 associated variants affect CRX HD-DNA contacts and/or intramolecular contacts with other HD residues
522 would guarantee new structural insights. Lastly, given that HD TF molecular mechanisms of action are

523 conserved across evolution and in different tissues and organs, we envision our CRX study will also shed light
524 on the study of other homeoproteins that hopefully lead to advances in medicine for the associated diseases.

525

526 **METHODS**

527 *Ethics statement*

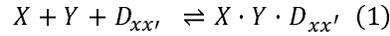
528 All procedures involving mice were approved by the Animal Studies Committee of Washington
529 University in St. Louis and performed under Protocol 21-0414 (to SC). Experiments were carried out in strict
530 accordance with recommendations in the Guide for the Care and Use of Laboratory Animals of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

535 *Coop-seq library preparation and EMSA*

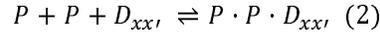
536 *E. coli* expression and purification of GST-CRX HD peptides and preparation of Coop-seq libraries
537 were performed as described previously (Chen et al. 2002; Zheng et al. 2023). The HD-DNA binding
538 reactions were performed in 1×CRX binding buffer (60mM KCl, 25mM HEPES, 5% glycerol, 1mM DTT).
539 The reaction mixtures were run at 4°C in native 12% Tris-Glycine PAGE gel (Invitrogen™) at 160V for
540 40min. The visible bands were excised from the gels. The extracted Coop-seq DNAs were extracted, purified,
541 and PCR amplified to tail on indexing barcodes and sequencing adapters. All Coop-seq samples were pooled
542 and sequenced on a single 1×50bp Miseq run at the DNA Sequencing Innovation Lab at the Center for
543 Genome Sciences & Systems Biology (CGS&SB, WashU).

544 *Determination of relative cooperativity with Coop-seq*

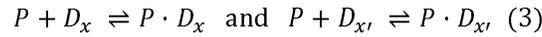
545 For a combinatorial interaction of two (or more) protein species X and Y , and a particular DNA
546 sequence, $D_{xx'}$, which is a combination of two half-sites, x and x' with a spacer, z , between them, the dimeric
547 binding interaction can be diagrammed as:



548 where $X \cdot Y \cdot D_{xx'}$ refers to the dimeric protein-DNA complex. Specifically, in our study, we are interested in
549 single protein species P binding to a dimeric DNA sequence $D_{xx'}$. We can rewrite the equation above as:



550 In addition, in our system, the HD peptides (protein species P) is also competent to bind each half-site as a
551 monomeric protein-DNA complex. There are two additional states of the DNA sequence being bound:



552 At equilibrium, the probability of a DNA molecule being in the unbound (U), monomerically bound (B_m), or
553 dimerically bound (B_d) states (Fig. 2B) are:

$$P(U|D_{xx'}) = \frac{[D_{xx'}]}{[D_{xx'}]_T} \quad (4)$$

$$P(B_m|D_{xx'}) = \frac{[P \cdot D_x] + [P \cdot D_{x'}]}{[D_{xx'}]_T} \quad (5)$$

$$P(B_d|D_{xx'}) = \frac{[P \cdot P \cdot D_{xx'}]}{[D_{xx'}]_T} \quad (6)$$

554 where [...] refers to concentrations and $[D_{xx'}]_T$ as the sum of all the states. The affinity of the protein P to a
555 sequence D_x is defined as the association constant K_A . The ratios of the probabilities above are related to the
556 association constants as:

$$\frac{P(B_m|D_{xx'})}{P(U|D_{xx'})} = \frac{[P \cdot D_x] + [P \cdot D_{x'}]}{[D_{xx'}]} = [P](K_x + K_{x'}) \quad (7)$$

$$\frac{P(B_d|D_{xx'})}{P(U|D_{xx'})} = \frac{[P \cdot P \cdot D_{xx'}]}{[D_{xx'}]} = [P]^2 K_{xx'} = [P]^2 (\omega_i K_x K_{x'}) \quad (8)$$

557 where ω_i is the cooperativity and is unique to each dimeric sequence $D_{xx'}$ with different half-site combination
558 and spacer. The cooperativity for each unique $D_{xx'}$ can be written as:

$$\omega_i = \frac{K_{xx'}}{K_x K_{x'}} = \frac{[P \cdot P \cdot D_{xx'}]}{[P]^2 [D_{xx'}] K_x K_{x'}} \quad (9)$$

559 To obtain relative cooperativities relative to some reference sequence $D_{rr'}$:

$$\frac{\omega_i}{\omega_r} = \frac{K_{xx'} K_r K_{r'}}{K_{rr'} K_x K_{x'}} = \frac{[P \cdot P \cdot D_{xx'}] [D_{rr'}] K_r K_{r'}}{[P \cdot P \cdot D_{rr'}] [D_{xx'}] K_x K_{x'}} \quad (10)$$

560 Since we are interested in comparing the cooperativity index ω_i for half-site matched dimeric sequences with
 561 a 3bp (P3) *versus* 5bp (P5) spacer (Fig. 2C,D), the $\frac{K_r K_{r'}}{K_x K_{x'}}$ term can be neglected assuming the monomeric
 562 binding to half-sites are the same for the two sequence configurations. In a binding reaction involving TF and
 563 a library of DNAs, the concentration of the bound and the unbound species are directly proportional to the
 564 number of individual DNA molecules in each fraction obtained directly from the sequencing data. With
 565 sequencing counts in each fraction, we can accurately estimate the ratios of concentrations from counts with
 566 the relationship:

$$\frac{[P \cdot P \cdot D_{xx'}]}{[P \cdot P \cdot D_{rr'}]} \approx \frac{N_{B_d}(D_{xx'})}{N_{B_d}(D_{rr'})} \quad \text{and} \quad \frac{[D_{xx'}]}{[D_{rr'}]} \approx \frac{N_U(D_{xx'})}{N_U(D_{rr'})} \quad (11)$$

567 where N_U denotes counts in the unbound fraction and N_{B_d} denotes counts in the dimerically bound fraction.
 568 We can then rewrite equation 10 as:

$$\frac{\omega_i}{\omega_r} = \frac{N_{B_d}(D_{xx'}) N_U(D_{rr'})}{N_{B_d}(D_{rr'}) N_U(D_{xx'})} \quad (12)$$

569 *Coop-seq data analysis and cooperativity index calculation*

570 The raw sequencing data for monomer and unbound (free) bands were included in our previous
 571 publication under GEO accession number GSE223658. The sequencing results were filtered and converted to
 572 count ratios following the Spec-seq analysis pipeline previously described (Zheng et al. 2023). Briefly, reads
 573 with any mismatch in the conserved regions were discarded and sequences with less than 50 raw read counts
 574 were discarded. The relative cooperativity index ω_i for each dimeric sequence was calculated following
 575 Equation 12. In Fig. 2D, pair-wise relative cooperativity of half-site matched P3-P5 library members is
 576 presented. The relative cooperativity was ordered by hierarchical clustering using the Python package SciPy
 577 (v1.11.2) (Virtanen et al. 2020) with linkage method “complete” and distance metric “euclidean”. In
 578 Supplemental Fig. S2A-D, a single reference sequence 5'-TAATGCGCTATTA-3' was used to calculate
 579 relative cooperativity. We chose this sequence based on previous Spec-seq results that WT, E80A and K88N

580 HDs all bind with reasonable affinity to either half-site. The full relative cooperativity index table for all
581 library members can be found in Supplemental Table S2. The P3-P5 matched relative cooperativity index
582 table sorted as in Fig. 2D can be found in Supplemental Table S3.

583 *ATAC-seq*

584 For each genotype, three biological replicates and two retinas per replicate from one male and one
585 female were pooled. The assay for transposase-accessible chromatin with sequencing was performed as
586 previously published (Buenrostro et al. 2015). The quantity and quality of the ATAC-seq libraries were
587 assayed using the Qubit™ 3 Fluorometer (Invitrogen™) and the Bioanalyzer (Agilent, Santa Clara, CA) prior
588 to sequencing. All ATAC-seq sequencing libraries were pooled and sequenced on the Illumina NovaSeq2000
589 platform (2×150bp reads) with an average depth of 54M reads at the Genome Technology Access Center at
590 the McDonnell Genome Institute (GTAC@MGI, WashU). The mm10 FASTA sequences for ATAC-seq
591 peaks were obtained using R package BSgenome (v1.66.3) (Pagès 2020). *De novo* motif enrichment analysis
592 was performed with MEME-CHIP in MEME Suite (v5.5.2) (Bailey et al. 2015) using order 1 Markov
593 background model and default parameters. Instances of K₅₀ and Q₅₀ HD monomeric and dimeric motifs were
594 identified with FIMO in MEME Suite (v5.5.2) using order 1 Markov background model and --thresh 1.0E-3.

595 *MPRA library construction, electroporation, and sequencing*

596 The library of 200mer oligonucleotides, each containing a 134bp testing cis-regulatory element (CRE)
597 sequence and a unique 10bp barcode, was ordered directly from Twist Bioscience (South San Francisco, CA).
598 The MPRA plasmid library was constructed using the 200mer oligonucleotides following published protocols
599 (Hughes et al. 2018). For each retinal electroporation, 30μg of MPRA plasmid library DNA per retina (3
600 retinas per replicate, a total of 3-4 replicates per genotype) was used. The electroporated retinas were cultured
601 for 8 days in the incubator (37°C, 5% CO₂) before being harvested for RNA and DNA extraction using
602 TRIzol™ Reagent (Invitrogen™). The MPRA sequencing libraries were prepared as previously described
603 (Friedman et al. 2021). All MPRA sequencing libraries were pooled and sequenced on the Illumina
604 NovaSeq2000 platform (2×150bp reads) at the Genome Technology Access Center at the McDonnell Genome
605 Institute (GTAC@MGI, WashU). Libraries prepared from un-electroporated plasmid DNAs were sequenced

606 to 50M reads in two technical replicates (approx. 2500×). Libraries prepared from retinal *ex plant* extracted
607 RNA (n=3/4 per genotype) and DNA (n=1 per genotype) were sequenced to an average depth of 30M reads
608 (approx. 1600× depth).

609 *MPRA data analysis*

610 The pre-processing of MPRA reads followed the pipeline previously described (Friedman et al. 2021).
611 After dropping low-quality CREs, counts of individual RNA libraries were normalized by the average counts
612 of the plasmid libraries to obtain “raw activity score” for individual barcodes. Genotype average raw activity
613 scores by unique testing CREs were calculated by averaging the barcode raw activity scores for each CRE. A
614 coefficient of variation threshold of 1.0 was used to filter out CREs whose barcode activity varies greatly
615 among genotype replicates. The “regulatory activity score” was calculated by normalizing the raw activity
616 score of each CRE to the average raw activity score of all scrambled control CREs. In the MPRA library
617 design, for each testing genomic CRE sequence (WT) that contains HD motif(s), one or more mutated CRE
618 versions were generated by mutating the HD motif(s) (Supplemental Methods). The “HD motif activity score”
619 was calculated as the “regulatory activity score” difference between a mutant CRE version and its matched
620 genomic CRE sequence.

621

622 **DATA ACCESS**

623 All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene
624 Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE256215.
625 Customized scripts and any additional information required to reproduce the analysis in this paper are
626 available as Supplemental Code and at GitHub at https://github.com/YiqiaoZHENG/CRXHD_epigenome.git
627 [data visualization] and https://github.com/YiqiaoZHENG/CRXHD_mpra.git [dedicated MPRA design,
628 sequencing library preparation, and data analysis].

629

630 **COMPETING INTEREST STATEMENT**

631 The authors declare no competing interests.

632

633 **ACKNOWLEDGMENTS**

634 This work was supported by National Institutes of Health grants EY012543 to S.C., EY032136 to S.C.,
635 EY027784 to S.C. & B.A.C., EY002687 to WU-DOVS; the Stein Innovation Award from Research to Prevent
636 Blindness to S.C.; and unrestricted funds from Research to Prevent Blindness to WU-DOVS. We thank
637 Mingyan Yang for technical assistance; Mike Casey from the Molecular Genetics Service Core for generating
638 luciferase reporter assay plasmids and MPRA oligo library; J. Hoisington-Lopez and M. Crosby from DNA
639 Sequencing Innovation Lab at the Center for Genome Sciences & Systems Biology (CGS&SB) as well as
640 Genome Technology Access Center at the McDonnell Genome Institute (GTAC@MGI) for sequencing
641 assistance. We also thank Mr. Artur Widlak for the generous gift from Widlak Family CRX Research Fund.

642 *Author contributions:* S. Chen and Y. Zheng conceived the study. S. Chen and G.D. Stormo
643 supervised the study. S. Chen and Y. Zheng designed the experiments. Y. Zheng performed all the
644 experiments, data analysis and visualization. G.D. Stormo assisted in Coop-seq data analysis. Y. Zheng wrote
645 the original draft. S. Chen, G.D. Stormo, and Y. Zheng revised the manuscript. All authors read and approved
646 the final manuscript.

647

648 **REFERENCES**

- 649 Ades SE, Sauer RT. 1994. Differential DNA-binding specificity of the engrailed homeodomain: the role of
650 residue 50. *Biochemistry* **33**: 9187-9194.
- 651 Aldiri I, Xu B, Wang L, Chen X, Hiler D, Griffiths L, Valentine M, Shirinifard A, Thiagarajan S, Sablauer A
652 et al. 2017. The Dynamic Epigenetic Landscape of the Retina During Development, Reprogramming,
653 and Tumorigenesis. *Neuron* **94**: 550-568 e510.
- 654 Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Research* **43**: W39-W49.
- 655 Bassett EA, Wallace VA. 2012. Cell fate determination in the vertebrate retina. *Trends Neurosci* **35**: 565-573.
- 656 Bernard P, Tang P, Liu S, Dewing P, Harley VR, Vilain E. 2003. Dimerization of SOX9 is required for
657 chondrogenesis, but not for sex determination. *Human Molecular Genetics* **12**: 1755-1765.

- 658 Bian F, Daghnsni M, Lu F, Liu S, Gross JM, Aldiri I. 2022. Functional analysis of the Vsx2 super-enhancer
659 uncovers distinct cis-regulatory circuits controlling Vsx2 expression during retinogenesis.
660 *Development* **149**.
- 661 Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: A Method for Assaying Chromatin
662 Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**: 21.29.21-21.29.29.
- 663 Bürglin TR, Affolter M. 2016. Homeodomain proteins: an update. *Chromosoma* **125**: 497-521.
- 664 Burmeister M, Novak J, Liang M-Y, Basu S, Ploder L, Hawes NL, Vidgen D, Hoover F, Goldman D, Kalnins
665 VI et al. 1996. Ocular retardation mouse caused by Chx10 homeobox null allele: impaired retinal
666 progenitor proliferation and bipolar cell differentiation. *Nature Genetics* **12**: 376.
- 667 Cepko CL. 2015. The Determination of Rod and Cone Photoreceptor Fate. *Annual Review of Vision Science* **1**:
668 211-234.
- 669 Chang YK, Srivastava Y, Hu C, Joyce A, Yang X, Zuo Z, Havranek JJ, Stormo GD, Jauch R. 2016.
670 Quantitative profiling of selective Sox/POU pairing on hundreds of sequences in parallel by Coop-seq.
671 *Nucleic Acids Research* **45**: 832-845.
- 672 Chen S, Wang Q-L, Nie Z, Sun H, Lennon G, Copeland NG, Gilbert DJ, Jenkins NA, Zack DJ. 1997. Crx, a
673 Novel Otx-like Paired-Homeodomain Protein, Binds to and Transactivates Photoreceptor Cell-
674 Specific Genes. *Neuron* **19**: 1017-1030.
- 675 Chen S, Wang Q-L, Xu S, Liu I, Li LY, Wang Y, Zack DJ. 2002. Functional analysis of cone-rod homeobox
676 (CRX) mutations associated with retinal dystrophy. *Human Molecular Genetics* **11**: 873-884.
- 677 Emerson MM, Surzenko N, Goetz JJ, Trimarchi J, Cepko CL. 2013. Otx2 and Onecut1 promote the fates of
678 cone photoreceptors and horizontal cells and repress rod photoreceptors. *Dev Cell* **26**: 59-72.
- 679 Forrest D, Swaroop A. 2012. Minireview: the role of nuclear receptors in photoreceptor differentiation and
680 disease. *Mol Endocrinol* **26**: 905-915.
- 681 Friedman RZ, Granas DM, Myers CA, Corbo JC, Cohen BA, White MA. 2021. Information content
682 differentiates enhancers from silencers in mouse photoreceptors. *eLife* **10**: e67403.
- 683 Furukawa T, Morrow EM, Cepko CL. 1997. Crx, a Novel otx-like Homeobox Gene, Shows Photoreceptor-
684 Specific Expression and Regulates Photoreceptor Differentiation. *Cell* **91**: 531-541.

- 685 Furukawa T, Morrow EM, Li T, Davis FC, Cepko CL. 1999. Retinopathy and attenuated circadian
686 entrainment in Crx-deficient mice. *Nature Genetics* **23**: 466.
- 687 García Bohórquez B, Aller E, Rodríguez Muñoz A, Jaijo T, García García G, Millán JM. 2021. Updating the
688 Genetic Landscape of Inherited Retinal Dystrophies. *Frontiers in Cell and Developmental Biology* **9**.
- 689 Gordon PJ, Yun S, Clark AM, Monuki ES, Murtaugh LC, Levine EM. 2013. Lhx2 balances progenitor
690 maintenance with neurogenic output and promotes competence state progression in the developing
691 retina. *J Neurosci* **33**: 12197-12207.
- 692 Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**:
693 1017-1018.
- 694 Green ES, Stubbs JL, Levine EM. 2003. Genetic rescue of cell number in a mouse model of microphthalmia:
695 interactions between Chx10 and G1-phase cell cycle regulators. *Development* **130**: 539-552.
- 696 Hafler BP, Surzenko N, Beier KT, Punzo C, Trimarchi JM, Kong JH, Cepko CL. 2012. Transcription factor
697 Olig2 defines subpopulations of retinal progenitor cells biased toward specific cell fates. *Proc Natl*
698 *Acad Sci U S A* **109**: 7882-7887.
- 699 Hanes SD, Brent R. 1989. DNA specificity of the bicoid activator protein is determined by homeodomain
700 recognition helix residue 9. *Cell* **57**: 1275-1283.
- 701 Hanes SD, Brent R. 1991. A genetic model for interaction of the homeodomain recognition helix with DNA.
702 *Science* **251**: 426.
- 703 Hayashi S, Scott MP. 1990. What determines the specificity of action of Drosophila homeodomain proteins?
704 *Cell* **63**: 883-894.
- 705 Heavner W, Pevny L. 2012. Eye development and retinogenesis. *Cold Spring Harb Perspect Biol* **4**.
- 706 Hobert O. 2021. Homeobox genes and the specification of neuronal identity. *Nature Reviews Neuroscience* **22**:
707 627-636.
- 708 Hu C, Malik V, Chang YK, Veerapandian V, Srivastava Y, Huang Y-H, Hou L, Cojocaru V, Stormo GD,
709 Jauch R. 2017. Coop-Seq Analysis Demonstrates that Sox2 Evokes Latent Specificities in the DNA
710 Recognition by Pax6. *Journal of Molecular Biology* **429**: 3626-3634.

- 711 Hughes AEO, Myers CA, Corbo JC. 2018. A massively parallel reporter assay reveals context-dependent
712 activity of homeodomain binding sites in vivo. *Genome Res* **28**: 1520-1531.
- 713 Jean-Charles N, Buenaventura DF, Emerson MM. 2018. Identification and characterization of early
714 photoreceptor cis-regulatory elements and their relation to Onecut1. *Neural Dev* **13**: 26.
- 715 Kim JW, Yang HJ, Brooks MJ, Zelinger L, Karakulah G, Gotoh N, Boleda A, Gieser L, Giuste F, Whitaker
716 DT et al. 2016. NRL-Regulated Transcriptome Dynamics of Developing Rod Photoreceptors. *Cell*
717 *Rep* **17**: 2460-2473.
- 718 Lee J, Myers CA, Williams N, Abdelaziz M, Corbo JC. 2010. Quantitative fine-tuning of photoreceptor cis-
719 regulatory elements through affinity modulation of transcription factor binding sites. *Gene Therapy* **17**:
720 1390-1399.
- 721 Leung RF, George AM, Roussel EM, Faux MC, Wigle JT, Eisenstat DD. 2022. Genetic Regulation of
722 Vertebrate Forebrain Development by Homeobox Genes. *Frontiers in Neuroscience* **16**.
- 723 Li J, Chiu T-P, Rohs R. 2024. Predicting DNA structure using a deep learning method. *Nature*
724 *Communications* **15**: 1243.
- 725 Liu IS, Chen JD, Ploder L, Vidgen D, van der Kooy D, Kalnins VI, McInnes RR. 1994. Developmental
726 expression of a novel murine homeobox gene (Chx10): evidence for roles in determination of the
727 neuroretina and inner nuclear layer. *Neuron* **13**: 377-393.
- 728 Mark M, Rijli FM, Chambon P. 1997. Homeobox Genes in Embryogenesis and Pathogenesis. *Pediatric*
729 *Research* **42**: 421-429.
- 730 Mathelier A, Xin B, Chiu T-P, Yang L, Rohs R, Wasserman Wyeth W. 2016. DNA Shape Features Improve
731 Transcription Factor Binding Site Predictions In Vivo. *Cell Systems* **3**: 278-286.e274.
- 732 Muranishi Y, Terada K, Inoue T, Katoh K, Tsujii T, Sanuki R, Kurokawa D, Aizawa S, Tamaki Y, Furukawa
733 T. 2011. An essential role for RAX homeoprotein and NOTCH-HES signaling in Otx2 expression in
734 embryonic retinal photoreceptor cell fate determination. *J Neurosci* **31**: 16792-16807.
- 735 Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. 2008. Analysis of
736 homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* **133**:
737 1277-1289.

- 738 Pagès H. 2020. BSgenome: Software infrastructure for efficient representation of full genomes and their SNPs.
- 739 Purves D, Williams SM. 2001. *Neuroscience. 2nd edition*. Sinauer Associates 2001.
- 740 Roger JE, Hiriyanna A, Gotoh N, Hao H, Cheng DF, Ratnapriya R, Kautzmann M-AI, Chang B, Swaroop A.
- 741 2014. OTX2 loss causes rod differentiation defect in CRX-associated congenital blindness. *The*
- 742 *Journal of Clinical Investigation* **124**: 631-643.
- 743 Rowan S, Cepko CL. 2004. Genetic analysis of the homeodomain transcription factor Chx10 in the retina
- 744 using a novel multifunctional BAC transgenic mouse reporter. *Dev Biol* **271**: 388-402.
- 745 Ruzycki PA, Tran NM, Kolesnikov AV, Kefalov VJ, Chen S. 2015. Graded gene expression changes
- 746 determine phenotype severity in mouse models of CRX-associated retinopathies. *Genome Biology* **16**:
- 747 171.
- 748 Ruzycki PA, Zhang X, Chen S. 2018. CRX directs photoreceptor differentiation by accelerating chromatin
- 749 remodeling at specific target sites. *Epigenetics Chromatin* **11**: 42.
- 750 Sapkota D, Chintala H, Wu F, Fliesler SJ, Hu Z, Mu X. 2014. Onecut1 and Onecut2 redundantly regulate
- 751 early retinal cell fates during development. *Proc Natl Acad Sci U S A* **111**: E4086-4095.
- 752 Swaroop A, Kim D, Forrest D. 2010. Transcriptional regulation of photoreceptor development and
- 753 homeostasis in the mammalian retina. *Nature Reviews Neuroscience* **11**: 563-576.
- 754 Tran NM, Chen S. 2014. Mechanisms of blindness: Animal models provide insight into distinct CRX-
- 755 associated retinopathies. *Developmental Dynamics* **243**: 1153-1166.
- 756 Tran NM, Zhang A, Zhang X, Huecker JB, Hennig AK, Chen S. 2014. Mechanistically Distinct Mouse
- 757 Models for CRX-Associated Retinopathy. *PLOS Genetics* **10**: e1004111.
- 758 Treisman J, Harris E, Wilson D, Desplan C. 1992. The homeodomain: a new face for the helix-turn-helix?
- 759 *Bioessays* **14**: 145-150.
- 760 Treisman J, Gönczy P, Vashishtha M, Harris E, Desplan C. 1989. A single amino acid can determine the DNA
- 761 binding specificity of homeodomain proteins. *Cell* **59**: 553-562.
- 762 Tucker SC, Wisdom R. 1999. Site-specific Heterodimerization by Paired Class Homeodomain Proteins
- 763 Mediates Selective Transcriptional Responses*. *Journal of Biological Chemistry* **274**: 32325-32332.

- 764 Virtanen P Gommers R Oliphant TE Haberland M Reddy T Cournapeau D Burovski E Peterson P Weckesser
765 W Bright J et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature*
766 *Methods* **17**: 261-272.
- 767 Wang S, Cepko CL. 2016. Photoreceptor Fate Determination in the Vertebrate Retina. *Investigative*
768 *ophthalmology & visual science* **57**: ORSFe1-ORSFe6.
- 769 White Michael A, Kwasnieski Jamie C, Myers Connie A, Shen Susan Q, Corbo Joseph C, Cohen Barak A.
770 2016. A Simple Grammar Defines Activating and Repressing *cis*-Regulatory Elements in
771 Photoreceptors. *Cell Reports* **17**: 1247-1254.
- 772 Wilson D, Sheng G, Lecuit T, Dostatni N, Desplan C. 1993. Cooperative dimerization of paired class homeo
773 domains on DNA. *Genes & Development* **7**: 2120-2134.
- 774 Wilson DS, Guenther B, Desplan C, Kuriyan J. 1995. High resolution crystal structure of a paired (Pax) class
775 cooperative homeodomain dimer on DNA. *Cell* **82**: 709-719.
- 776 Wilson DS, Sheng G, Jun S, Desplan C. 1996. Conservation and diversification in homeodomain-DNA
777 interactions: a comparative genetic analysis. *Proc Natl Acad Sci U S A* **93**: 6886-6891.
- 778 Zagozewski JL, Zhang Q, Pinto VI, Wigle JT, Eisenstat DD. 2014. The role of homeobox genes in retinal
779 development and disease. *Developmental Biology* **393**: 195-208.
- 780 Zheng Y, Chen S. 2024. Transcriptional precision in photoreceptor development and diseases – Lessons from
781 25 years of CRX research. *Frontiers in Cellular Neuroscience* **18**.
- 782 Zheng Y, Sun C, Zhang X, Ruzycski PA, Chen S. 2023. Missense mutations in CRX homeodomain cause
783 dominant retinopathies through two distinct mechanisms. *eLife* **12**: RP87147.
- 784 Zibetti C, Liu S, Wan J, Qian J, Blackshaw S. 2019. Epigenomic profiling of retinal progenitors reveals LHX2
785 is required for developmental regulation of open chromatin. *Communications Biology* **2**: 142.
- 786

787 **Figure 1.** K88N mutation significantly increased CRX HD's cooperative binding and transactivation activity
 788 on BAT-1 sequence containing a P3 dimeric HD motif. (A) Diagrams depicting K_{50} HD preferred monomeric
 789 and dimeric P3 motifs. (B) Alignments showing WT BAT-1 probe sequences and variants. The P3 dimeric
 790 HD motif and four monomeric HD core motifs 5'-TAAT-3' are labelled. f and r indicates whether the core
 791 motif is on the forward or reverse strand. In BAT-1 variants, the mutated nucleotides are italicized and
 792 underlined. (C,D) EMSA gel images showing increasing amount of WT or K88N HD peptides bound to a
 793 fixed amount of BAT-1 (WT) or P5 GA control probes. The cartoon underneath each gel image shows the
 794 dimeric HD motif configuration and is labelled with the spacer length. (E) Schematics and barcharts of
 795 luciferase reporter assays comparing CRX WT and K88N transactivation activity at BAT-1 and variant
 796 enhancer sequences. Bars represent the mean. *p-values* of one-way ANOVA are annotated: ns: >5E-02; ***:
 797 <=1E-03.

798

799 **Figure 2.** E80A and K88N mutations differently affect CRX HD DNA binding cooperativity at P3 sequences.
 800 (A,B) Schematics showing the Coop-seq experimental pipelines. dsDNA oligo pools of P3 and/or P5 Coop-
 801 seq library are incubated with different HD peptides. The dimeric and monomeric binding complexes are
 802 separated from unbound DNAs by EMSA. DNAs are extracted from all three DNA bands and subjected to
 803 quantification by Illumina sequencing. B_d: dimeric band; B_m: monomeric band; U: unbound band. (C)
 804 Diagram depicting the Coop-seq library design and strategy to match a P3 sequence with a P5 counterpart.
 805 Exact oligo sequences can be found in Supplemental Table S1. (D) Heatmap comparing the relative
 806 cooperativity of WT and variant CRX HDs on P3 and P5 libraries (ω_{p3}/ω_{p5}). Note the relative cooperativity is
 807 presented in the Logarithmic scale and ordered by unsupervised hierarchical clustering. The ordered relative
 808 cooperativity matrix can be found in Supplemental Table S3.

809

810 **Figure 3.** $Crx^{K88N/+}$ retinas show defective chromatin remodeling at photoreceptor CREs enriched with K_{50}
 811 HD motifs. (A) Heatmaps depicting the normalized ATAC-seq or CRX ChIP-seq signal intensities at Crx^{K88N} -
 812 reduced accessible ATAC-seq peaks. (B) PWM logo and enrichment significance *E-value* of the STREME *de*
 813 *novo* discovered HD motif. (C) Line plot showing the average developmental accessibility kinetics of Crx^{K88N} -
 814 reduced ATAC-seq peaks. The developmental ATAC-seq data is from Aldiri et al. 2017. (D) Barchart
 815 showing Biological Process (BP) Gene Ontology (GO) term enrichment of differentially expressed genes
 816 adjacent to Crx^{K88N} -reduced ATAC-seq peaks. (E) Heatmap comparing the P10 expression changes of
 817 Crx^{K88N} -reduced ATAC-seq peaks associated genes in different *Crx* mutant retinas. The gene set is identical to
 818 that in (D). (F) Schematics depicting chromatin remodelling defects at photoreceptor regulatory regions in the
 819 $Crx^{K88N/+}$, $Crx^{K88N/N}$, and $Crx^{R90W/W}$ retinas.

820

821 **Figure 4.** CRX K88N ectopic activity at Q_{50} HD motifs impedes the silencing of progenitor regulatory
 822 programs in developing photoreceptors. (A) Heatmaps depicting the normalized ATAC-seq or CRX ChIP-seq
 823 signal intensities at Crx^{K88N} -increased accessible ATAC-seq peaks. (B) PWM logo and enrichment
 824 significance *E-value* of STREME *de novo* discovered HD motifs. (C) Line plot showing the average
 825 developmental accessibility kinetics of Crx^{K88N} -increased ATAC-seq peaks. The developmental ATAC-seq
 826 data is from Aldiri et al. 2017. (D) Heatmap depicting the log odds ratio enrichment of embryonic day 14.5
 827 (e14.5) or adult retinal VSX2 binding sites under Crx^{K88N} -increased ATAC-seq peaks. *p-values* of Fisher's
 828 exact tests are indicated. The VSX2 ChIP-seq data is from Bian et al. 2022. (E) PWM logo and significance
 829 *E-value* of STREME *de novo* discovered basic helix-loop-helix (bHLH) motif under Crx^{K88N} -increased
 830 ATAC-seq peaks. PWM logos of selected retinal progenitor/neurogenic bHLH TFs are given for comparison.
 831 JASPAR IDs of the selected TFs can be found in Methods. (F) Barchart showing Biological Process (BP)
 832 Gene Ontology (GO) term enrichment of genes adjacent to Crx^{K88N} -reduced ATAC-seq peaks.

833

834 **Figure 5.** Crx^{E80A} retinas show defective chromatin remodelling at CREs enriched for dimeric K_{50} HD motifs.
 835 (A) Heatmaps depicting the normalized ATAC-seq or CRX ChIP-seq signal intensities at Crx^{E80A} differentially
 836 accessible ATAC-seq peaks. (B,C) PWM logos and enrichment significance E -values of STREME *de novo*
 837 discovered HD motifs under Crx^{E80A} differentially accessible ATAC-seq peaks. (D,E) Line plots showing the
 838 average developmental accessibility kinetics of Crx^{E80A} differentially accessible ATAC-seq peaks. The
 839 developmental ATAC-seq data is from Aldiri et al. 2017.

840

841 **Figure 6.** CRX E80A differential activity at monomeric and dimeric K_{50} HD motifs contributes to gene mis-
 842 expression at different stages of photoreceptor development. (A) Diagrams of photoreceptor genes regulated
 843 solely by monomeric K_{50} HD motifs (top) or combinationally by monomeric and dimeric K_{50} HD motifs
 844 (bottom). For simplicity, representative motif logos are shown. The relative position of the motifs is arbitrary.
 845 (B) Heatmap comparing the CRX-DAG expression changes in the $Crx^{E80A/A}$ retinas at ages of post-natal day
 846 10 (P10) and day 21 (P21). The gene sets in the heatmaps are as defined in (A). (C,D) Schematics
 847 demonstrating the K_{50} HD division-of-labor model in regulating photoreceptor epigenome and transcriptome
 848 at different stages of development.

849

850 **Figure 7.** Monomeric and dimeric K_{50} HD motifs show different regulatory activity changes in retinal *ex plant*
 851 MPRA. (A) Schematics showing *ex plant* retinal MPRA experimental pipeline. (B,C) Box and strip plots
 852 comparing monomeric (B) or dimeric (C) K_{50} HD motif activities in *ex plant* cultured WT, $Crx^{E80A/+}$ and
 853 $Crx^{E80A/A}$ retinas. In panel B, CREs overlapped with ATAC-seq peaks that were not significantly reduced in
 854 the Crx^{E80A} retinas are plotted. In panel C, CREs overlapped with ATAC-seq peaks that were significantly
 855 reduced in the Crx^{E80A} retinas are plotted. p -values of Mann-Whitney-Wilcoxon tests are annotated.

856

857 **Figure 8.** Differential HD motif usage in photoreceptor development and diseases. (A) Q_{50} and K_{50} paired-
 858 class HDs prefer different monomeric and dimeric DNA motifs. (B&C) In post-natal photoreceptor precursors,
 859 regulatory elements with Q_{50} HD motifs are gradually closed (silenced) due a lack of interacting TFs. High-
 860 affinity monomeric K_{50} HD motifs respond to low concentration of CRX, drive chromatin remodeling in
 861 early-stage development, and plateau in their regulatory output at a later stage. Dimeric K_{50} HD motifs of
 862 individually low-affinity half-sites respond at high CRX concentration and are associated with chromatin
 863 remodeling and gene expression regulation at photoreceptor terminal differentiation. (D) In Crx^{K88N} mutant
 864 retinas, CRX^{K88N}'s ectopic binding impedes the silencing of regulatory elements with Q_{50} HD motifs. In
 865 Crx^{E80A} mutant retinas, CRX^{E80A}'s enhanced interactions with monomeric K_{50} HD motifs and defective
 866 cooperative binding at dimeric K_{50} HD motifs leads to accelerated chromatin remodeling in early-stage
 867 development but defective remodeling at terminal differentiation stage. dLCA: dominant Leber congenital
 868 amaurosis. dCoRD: dominant Cone-rod dystrophy.

869

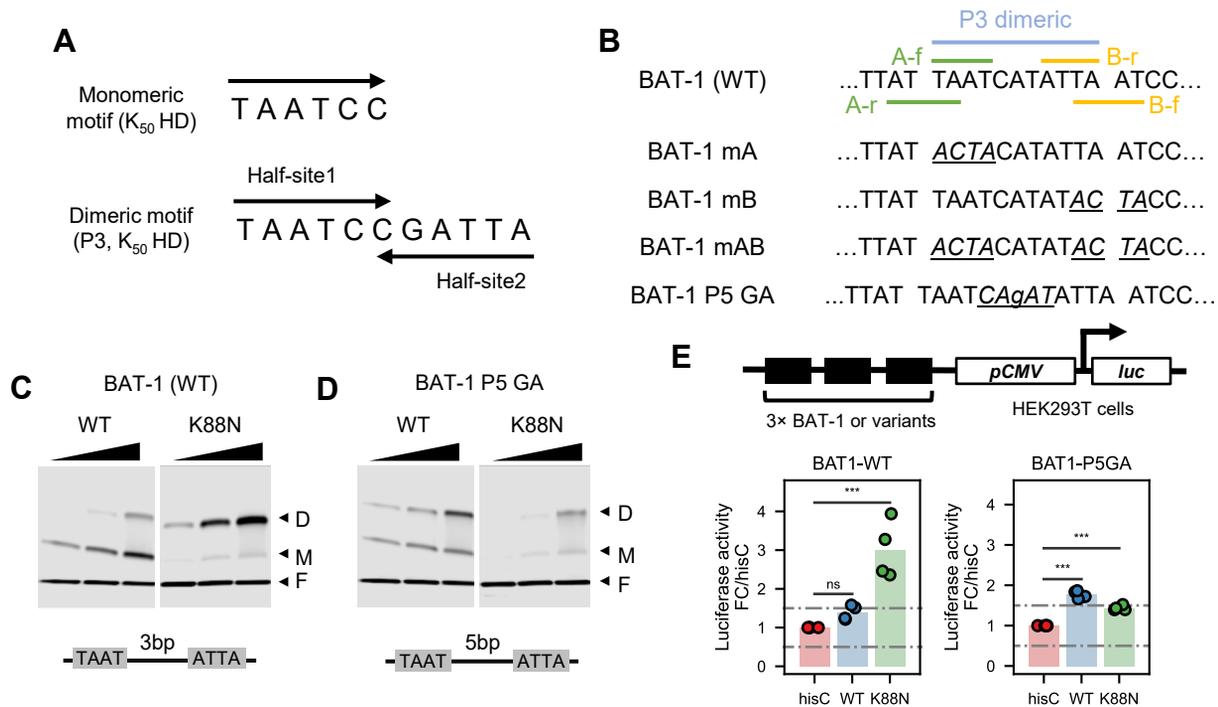


Figure 1

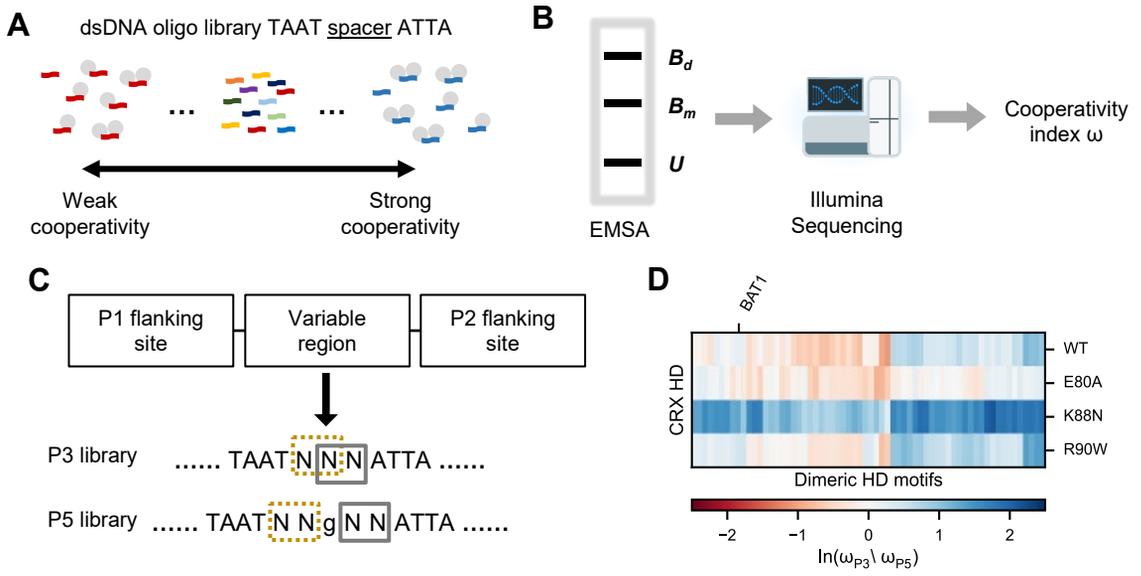


Figure 2

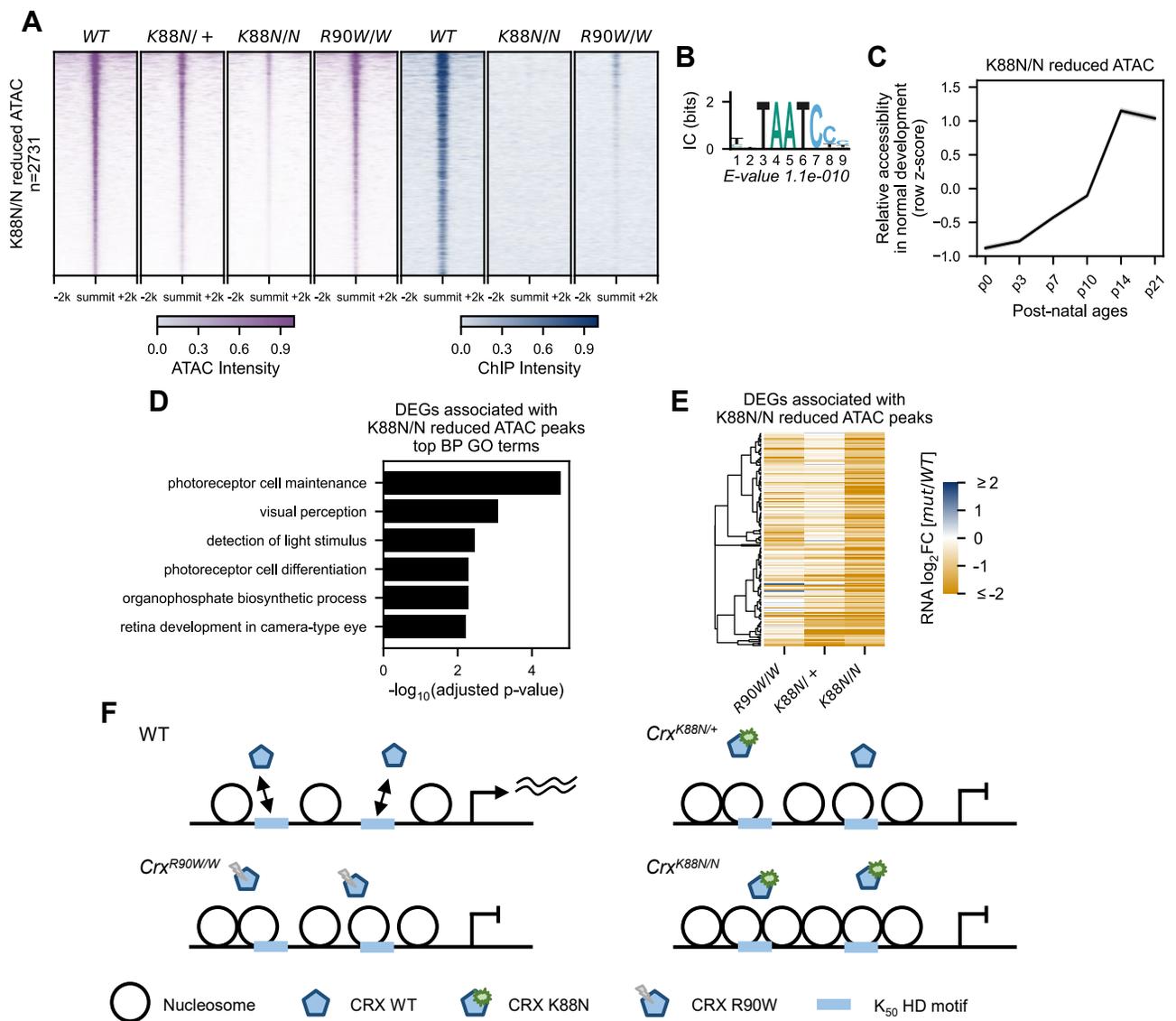


Figure 3

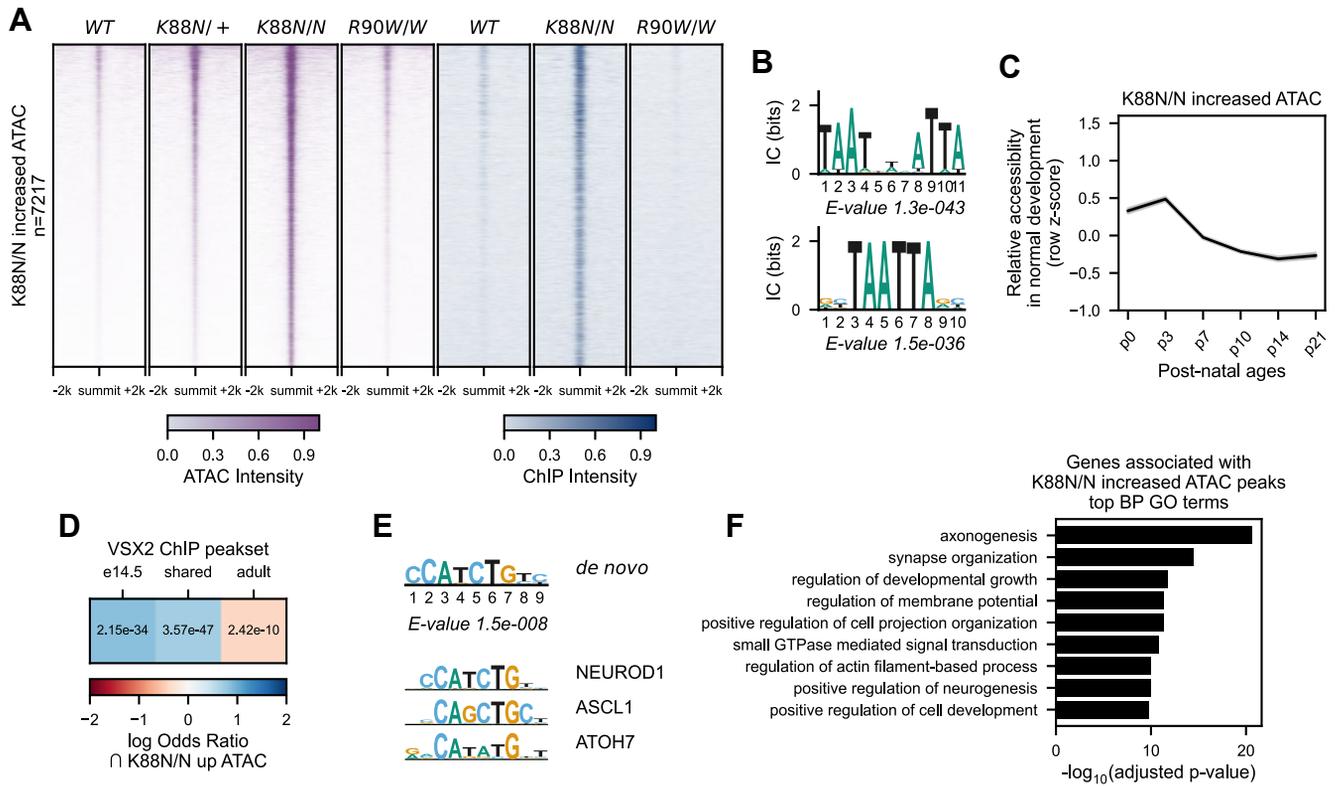


Figure 4

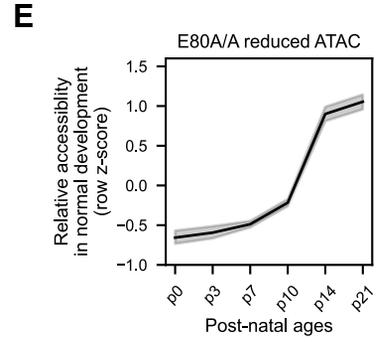
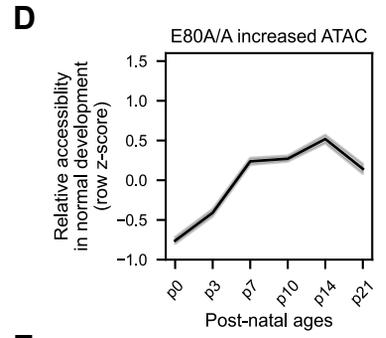
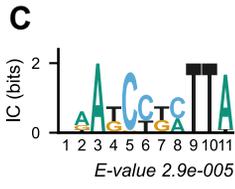
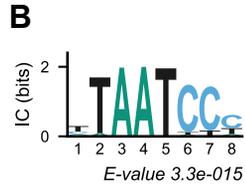
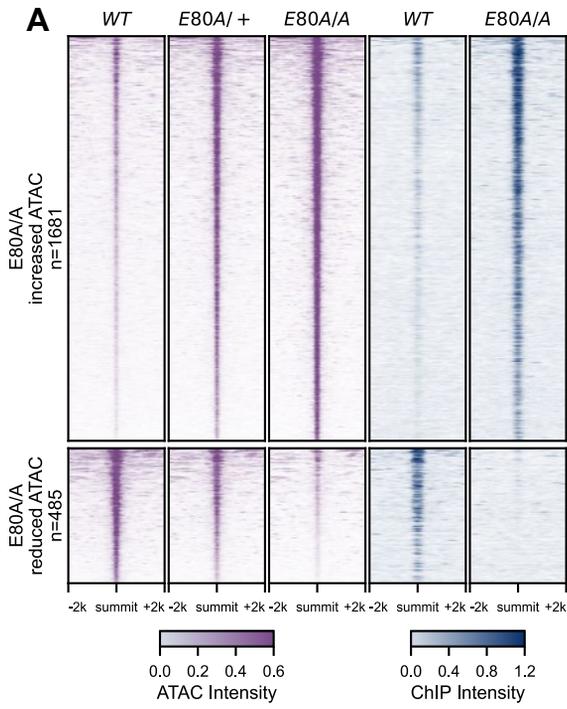


Figure 5

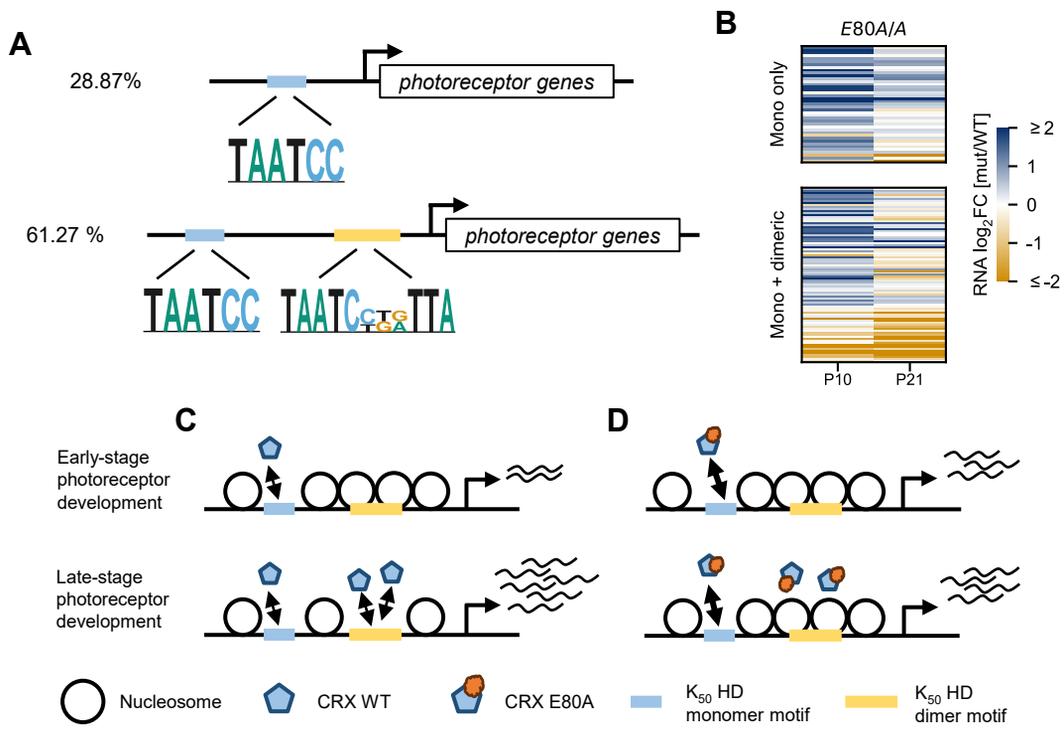


Figure 6

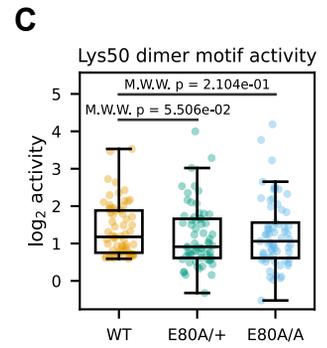
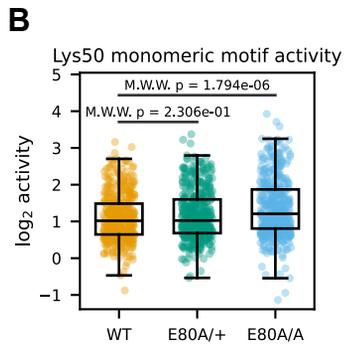
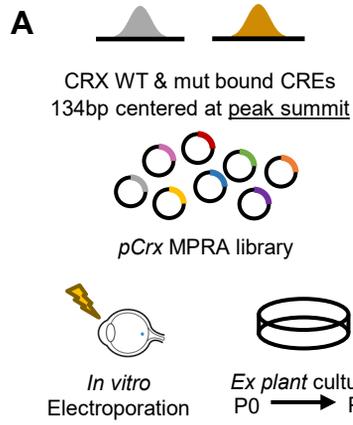


Figure 7

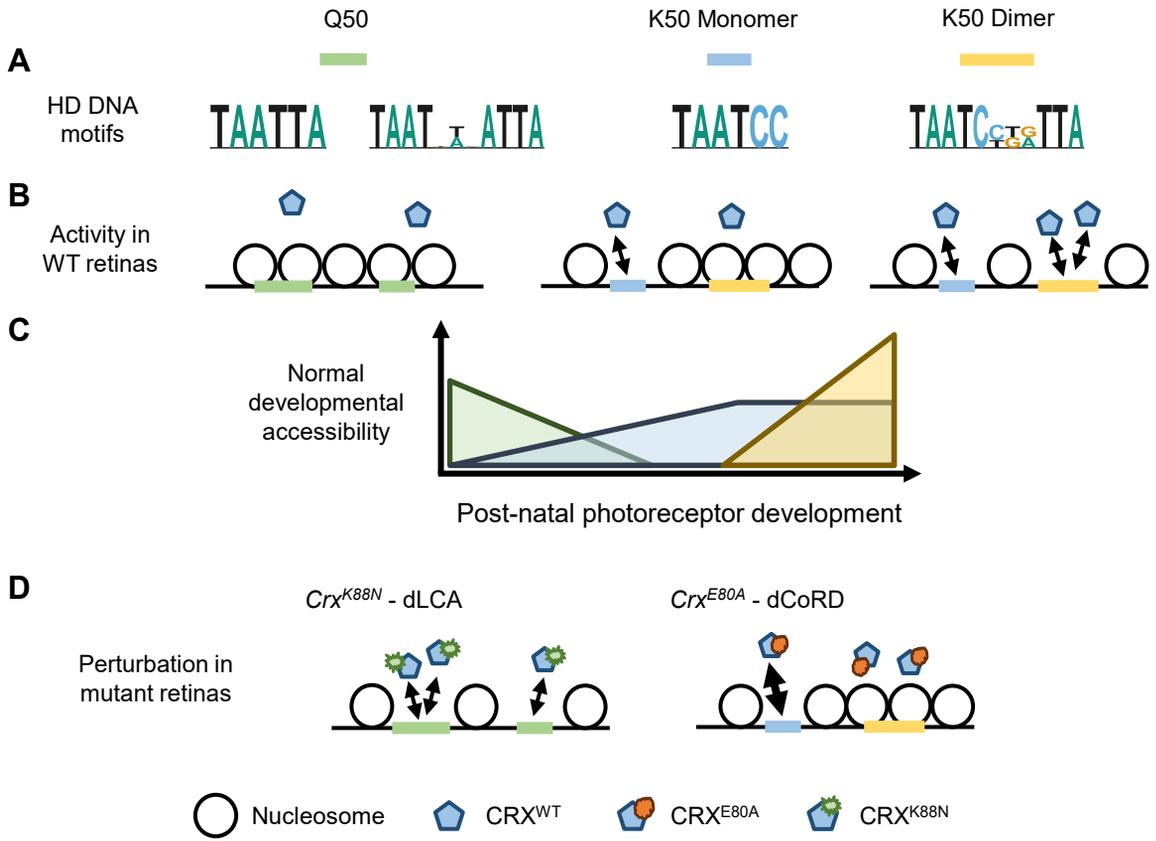


Figure 8



Aberrant homeodomain-DNA cooperative dimerization underlies distinct developmental defects in two dominant *CRX* retinopathy models

Yiqiao Zheng, Gary D. Stormo and Shiming Chen

Genome Res. published online December 23, 2024
Access the most recent version at doi:[10.1101/gr.279340.124](https://doi.org/10.1101/gr.279340.124)

Supplemental Material <http://genome.cshlp.org/content/suppl/2025/01/31/gr.279340.124.DC1>

P<P Published online December 23, 2024 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This manuscript is Open Access. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International license), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>
