Multi-omics analysis in human retina uncovers ultraconserved *cis*-regulatory elements at rare eye disease loci

Victor Lopez Soriano[*1,2], Alfredo Dueñas Rey[*1,2], Rajarshi Mukherjee[3], Genomics England Research Consortium[4], Frauke Coppieters[1,2,5], Miriam Bauwens[1,2], Andy Willaert[1,2], Elfride De Baere[1,2]

1- Department of Biomolecular Medicine, Ghent University, Ghent, Belgium.
2- Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium.
3- Department of Ophthalmology, St James's University Hospital, Leeds, UK.
4- Genomics England, London, UK.
5- Department of Pharmaceutics, Ghent University, Ghent, Belgium.

*equal contribution

Corresponding author: Elfride De Baere (elfride.debaere@ugent.be)

Running title: Identification of ultraconserved CREs in retina

ABSTRACT

Cross-species genome comparisons have revealed a substantial number of ultraconserved non-coding elements (UCNEs). Several of these elements have proved to be essential tissue- and cell type-specific *cis*-regulators of developmental gene expression. Here, we characterized a set of UCNEs, as candidate CREs (cCREs) during retinal development and evaluate the contribution of their genomic variation to rare eye diseases, for which pathogenic non-coding variants are emerging. Integration of bulk and single-cell retinal multi-omics data revealed 594 genes under potential *cis*-regulatory control of UCNEs, of which 45 are implicated in rare eye disease. Mining of candidate *cis*-regulatory UCNEs in WGS data derived from the rare eye disease cohort of Genomics England revealed 178 ultrarare variants within 84 UCNEs associated with 29 disease genes. Overall, we provide a comprehensive annotation of ultraconserved non-coding regions acting as cCREs during retinal development which can be targets of non-coding variation underlying rare eye diseases.

KEYWORDS

Ultraconserved non-coding elements (UCNEs), retina, single-cell, multi-omics, enhancer, *cis*-regulatory elements (CREs), genomic variation, rare eye disease.

31 **INTRODUCTION**

32 The 'dark matter of the genome' harbors functional *cis*-regulatory elements (CREs) such as promoters,
33 enhancers, silencers, and insulators, whose orchestrated activity is essential to provide spatial and
34 temporal patterns of gene expression that ensure proper tissue development and homeostasis (Field
35 & Adelman, 2020; Spielmann & Mundlos, 2016). The influence of these regulatory elements on their
36 target genes spans within architectural chromatin units known as topologically associating domains
37 (TADs), demarcated by boundaries enriched in CTCF and cohesin proteins (Dixon et al., 2012). Due to
38 their high context specificity, the annotation of candidate CREs (cCREs) poses an attractive yet
39 challenging task. In this regard, the Encyclopedia of DNA Elements (ENCODE) (Abascal et al., 2020;
40 Dunham et al., 2012) represents a powerful tool to characterize functional cCREs, as it provides a
41 robust inventory of well-defined cCREs supported by epigenetic data derived from a wide variety of
42 both human tissues and cell types.

43 Over the past decades, comparative genomics have been made progress to map functional cCREs by
44 establishing cross-comparison between species. As a result, several databases of non-coding
45 sequences that exhibit extremely high conservation have been generated, including UCEs (Bejerano et
46 al., 2004; Lomonaco et al., 2014), UCNEs (Dimitrieva & Bucher, 2013), ANCORA (Engström et al., 2008)
47 and CONDOR (Woolfe et al., 2007). One of the most comprehensive resources is UCNEbase, which
48 comprises exclusively >200bp-long non-coding genomic regions that exhibit ≥95% sequence identity
49 between human and chicken. In total, 4,135 regions across the genome comply with the criteria and
50 are listed as ultraconserved non-coding elements (UCNEs). The vast majority of these highly
51 constrained elements cluster around key developmental genes, such as *PAX6*, and have been long
52 hypothesized to act as tissue-specific transcriptional regulators (Dimitrieva & Bucher, 2013;
53 Polychronopoulos et al., 2017). Ultraconserved CREs have been shown to be consistently depleted of
54 common variants (Drake et al., 2006; Snetkova, Ypsilanti, et al., 2021), indicating that purifying
55 selection has shifted ultraconserved CRE-derived allele frequencies towards magnitudes similar to

56 those observed for missense variants (Drake et al., 2006), hence reinforcing that variation within these
57 regions is more likely to have functional consequences.

58 Due to the wide implementation of whole genome sequencing (WGS) in human genetic studies and
59 ambitious initiatives such as the 100,000 Genome Project (100kGP) launched by Genomics England
60 (GEL) (Smedley et al., 2021), previously overlooked regions within the vast non-coding fraction of the
61 genome have been investigated in patients with rare diseases and have shown an emerging role of
62 non-coding variants in disease pathogenesis. Two remarkable examples of functional evidence of
63 disease driven by disruption of conserved CREs can be found precisely in the rare eye disease field, in
64 which mutations located within highly conserved non-coding regions have been linked to
65 developmental defects. A first example is a *de novo* single nucleotide variant (SNV) in an
66 ultraconserved enhancer (SIMO element) 150 kb downstream of an intact *PAX6* transcriptional unit,
67 found in a case with aniridia (Bhatia et al., 2013). Second, tandem duplications within a gene desert
68 downstream the *IRXA* cluster (Cipriani et al., 2017; Small et al., 2016) have been found in patients
69 affected by North Carolina Macular Dystrophy (NCMD); the shared duplicated region harbors an
70 enhancer element defined as UCNE that could act as a *cis*-regulator during retinal development (Van
71 de Sompele et al., 2022).

72 The retina is a heterogeneous tissue composed of neuronal (rod and cone photoreceptors, bipolar
73 cells, ganglion cells, horizontal cells, and amacrine cells) and non-neuronal cell types (astrocytes,
74 Müller cells and resident microglia) that arise from the common pool of retinal progenitor cells (RPCs)
75 (Turner et al., 1990; Turner & Cepko, 1988). This cellular complexity is the result of spatiotemporally
76 controlled gene expression programs during retinal development, requiring concerted action of
77 thousands of CREs. In the wide spectrum of retinopathies, inherited retinal diseases (IRD) represent a
78 leading cause of early-onset vision impairment affecting over 2 million people worldwide. Despite
79 targeted panel-based sequencing and whole exome sequencing (WES)-based genetic testing 30-50%
80 of IRD cases remain unsolved (Ellingford et al., 2016; Haer-Wigman et al., 2017), raising the hypothesis

81 that disease-causing mutations are found in the non-coding genome that are mostly not covered by
82 standard genetic testing.

83 In this study we set out to annotate UCNEs as cCREs that modulate gene expression during retinal
84 development and that can play a role in retinopathies such as IRD. Firstly, we performed a
85 comprehensive integration of publicly available multi-omics datasets derived from human retina based
86 on biochemical features associated with regulation of gene expression. Secondly, UCNEs co-localizing
87 with genomic regions showing regulatory activity were linked to target genes that exhibit expression
88 in the retina. Finally, we evaluated the contribution of genomic variation within these regions to rare
89 eye diseases. This allowed us to identify an ultrarare SNV in a candidate *cis*-regulatory UCNE located
90 upstream *PAX6* in a family displaying autosomal dominant foveal abnormalities. Overall, this work has
91 improved the functional annotation of UCNEs in human retina representing understudied targets of
92 non-coding variation that may explain missing heritability in rare eye diseases.

93 RESULTS

94 Integration of bulk and single-cell multi-omics data enables the identification of 1,487 95 UCNEs with a candidate *cis*-regulatory role in human retina

96 The wealth of publicly available data allowed the integration of transcriptomic (bulk and scRNA-seq)
97 and epigenomic (DNase-seq, scATAC-seq, ChIP-seq) data to evaluate the regulatory capacity of UCNEs
98 over the major developmental stages of human retina (Fig. 1A). More specifically, to identify which
99 UCNEs might have a potential *cis*-regulatory role, an intersection was performed with genomic regions
100 characterized by an open chromatin context in the retina at different developmental stages, as
101 supported either by DNase (ENCODE) or scATAC-seq (Thomas et al., 2022) peaks. This resulted in a
102 total of 1,487 UCNEs, i.e., approximately one third of UCNEs display a retinal open context (Fig. 1B,
103 Supplementary Table 1). Approximately 80% of the DNase-seq peaks were also supported by scATAC-
104 seq, hence defining high-confidence candidate *cis*-regulatory UCNEs (Fig. 1B). Nevertheless, to account
105 for differences in time points between datasets, all 1,487 UCNEs were included in downstream

106 analyses. Apart from the display of a retinal open chromatin context, we also evaluated the overlap
107 between UCNEs and genomic regions featured by markers associated with regulation of gene
108 expression. A total of 834 UCNEs were found to display at least one of the assessed features in the
109 interrogated retinal developmental stages; more specifically 111 UCNEs were identified to display the
110 active enhancer mark H3K27ac, of which most (95%) also exhibit other histone marks related to active
111 enhancers (H3K4me1 and H3K4me2) (Fig. 1B). Table 1 and Supplementary Table 2 include summarized
112 and detailed overviews of the number of identified UCNEs per marker and stage, respectively.

113 To investigate the functional relevance of the putative regulatory UCNEs, we performed a correlation
114 with the 1,942 experimentally validated human non-coding fragments of the VISTA Enhancer Browser
115 (Fig. 1C). Out of these 1,942 elements, approximately 50% (998/1,942) display enhancer activity. When
116 overlapped with all 4,135 UCNEs, a slight increase of positive elements (505/922) is observed;
117 however, when only the identified retinal accessible UCNEs were considered, roughly 68% were found
118 to be positive (272/402), thereby reinforcing their predicted regulatory role (Fig. 1C). Interestingly,
119 evaluation of the anatomical description of the reporter gene expression patterns of these 272
120 elements revealed a potential enrichment in eye (Fig. 1D). An illustrative example of one of these
121 elements and its corresponding multi-omics-based characterization is shown in Fig. 1E-F.

122 **Table 1. Overview of the number of events based on the peak identification for each marker at the**
123 **different stages of retinal differentiation.** Abbreviations: FW: fetal weeks.

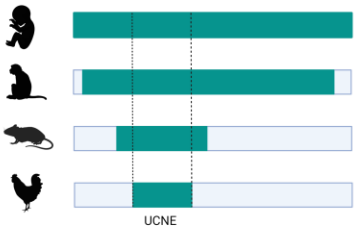
124

	Stage				
ChIP-seq	**FW13/14**	**FW15/16**	**FW18/20**	**FW23/24**	**Total Unique**
H3K27ac	73	45	74	94	127
H3K4me1	235	490	48	323	582
H3K4me2	347	354	256	225	465
H3K4me3	94	124	102	97	136
H3K27me3	119	54	128	126	154
H3K9/14Ac	51	76	36	31	94
H3K9me3	0	1	1	29	29
PolII	17	66	59	55	106
CTCF	51	69	22	37	71

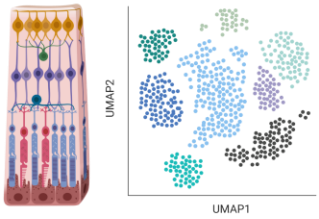
A

Annotation of ultraconserved non-coding regions acting as candidate *cis*-regulatory elements

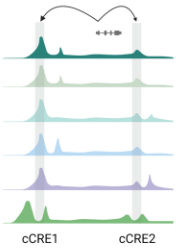
Retrieval of ultraconserved non-coding elements



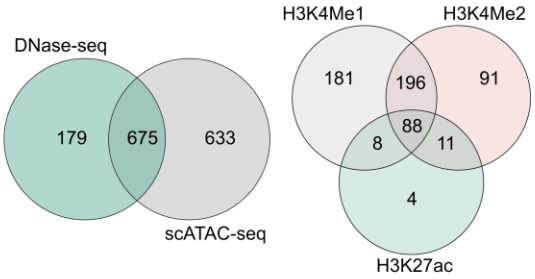
Retinal bulk & single-cell multi-omics dataset integration



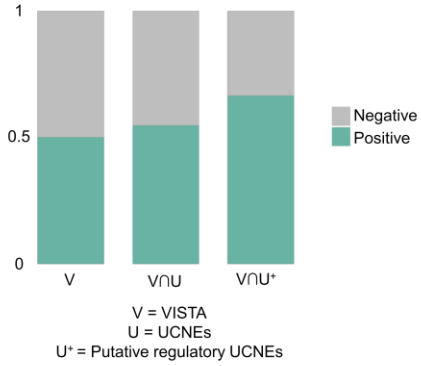
Prediction of UCNEs with potential *cis*-regulatory role



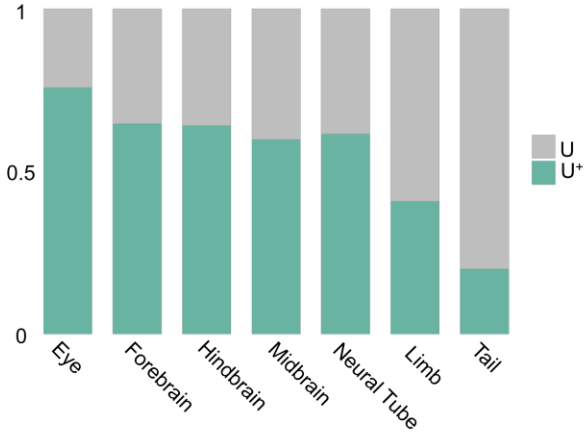
B



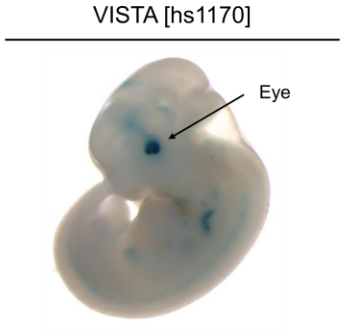
C



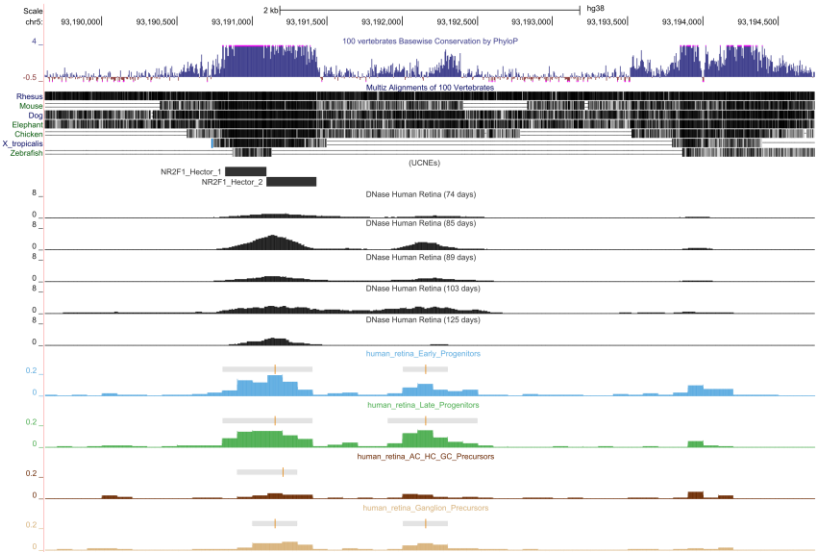
D



E



F



125

126 **Figure 1. Integration of publicly available datasets for the characterization of the ultraconserved**
127 **non-coding elements (UCNEs) library. A)** Overview of the integrative multi-omics analysis for the
128 prediction of UCNEs with potential *cis*-regulatory role in human retina. **B)** Venn diagrams illustrating
129 the CREs displaying open chromatin features based on scATAC-seq and DNase-seq (left) and their
130 overlap of active enhancer marks (H3K27ac, H3K4me1 and H3K4me2). **C)** Barplot showing the
131 proportions of elements from the VISTA enhancer browser (V), UCNEs (U), and putative regulatory
132 UCNEs (characterized by retinal datasets) (U^+) displaying reporter expression (positive). **D)**
133 Proportional stacked barplot showing the distribution of tissues (eye, forebrain, hindbrain, midbrain,
134 neural tube, limb and tail) in which the putative regulatory UCNEs display reporter expression. **E-F)**
135 Illustration of one of the characterized UCNEs (NR2F1_Hector_1/2) displaying open chromatin
136 supported by DNase-seq (embryonic day 74-85, 89 and 103-125) and scATAC-seq (AC/HC/GC
137 Precursors Cells, Early Progenitor Cells, Ganglion Precursor Cells, Late Retinal Progenitor Cells) and
138 enhancer reporter expression in the eye. Figures obtained from the UCSC VISTA (hs1170) enhancer (E)
139 and genome (F) browsers. Abbreviations: FW: fetal weeks; V: VISTA enhancer elements; U: UCNEs; U^+:
140 Putative active regulatory UCNE; ∩: intersection.

141

142 ***In silico* prediction and integration of retinal chromatin conformation data identify putative**

143 **target genes under UCNE *cis*-regulation**

144 We annotated the identified active UCNEs to assign them potential target genes and thus assess their
145 association with active gene expression in the retina. A total of 724 target genes were assigned to the
146 initially identified 1,487 UCNEs displaying candidate *cis*-regulatory activity. The vast majority (74.5%)
147 of queried regions were assigned as putatively regulating two genes based on the used association rule
148 (Fig. 2A). Furthermore, most of these regions were found to be located at 50-500kb from the TSS of
149 the regulated target genes regardless of their orientation (Fig. 2B). Out of these 724 target genes, 594
150 (81.9%) are expressed in the retina in at least one of the interrogated developmental stages (52 to 136
151 days post-conception). To evaluate further whether chromatin contacts between the UCNEs and the
152 promoters of their target genes can occur in retina, we integrated the 2,948 retinal TADs identified by
153 Marchal et al. (Marchal et al., 2022). A large majority (393/594) of the target genes were found to be
154 in TADs also harboring their associated UCNE (Supplementary Table 1). These genes were then used
155 as input for Gene Ontology analysis, which revealed an overrepresentation of terms related to
156 regulation of gene expression and differentiation, thereby confirming the expected roles of UCNEs as
157 tissue-specific CREs during development, particularly of the nervous system (Fig. 2C).

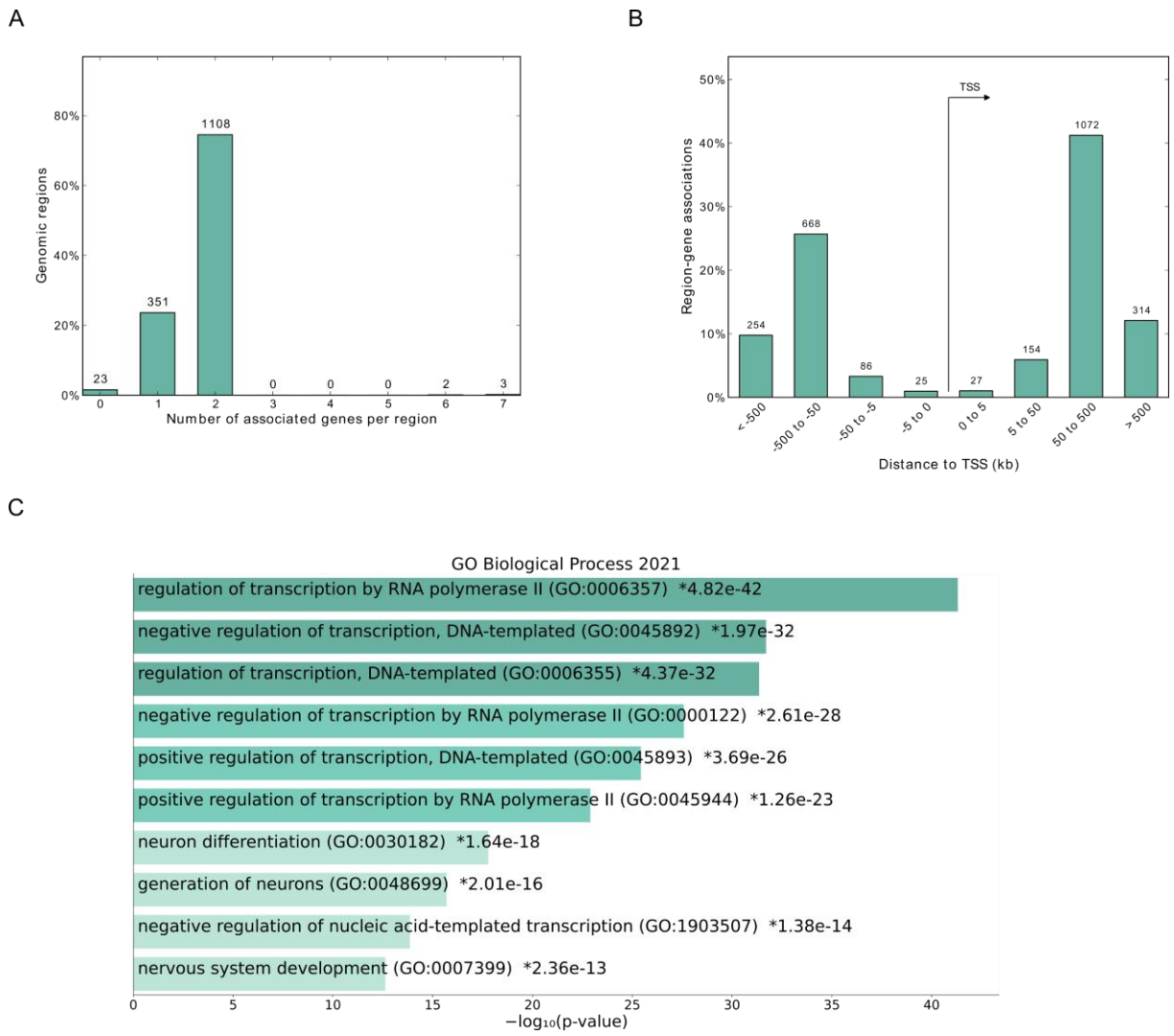


Figure 2. Target genes under putative regulation of characterized UCNEs in retina. A) Bar plot showing the number of associated genes per retinal UCNE. **B)** Distance distribution from the TSS and its associated UCNE. **C)** Gene ontology for the UCNE-associated target genes. Abbreviations: GO: Gene Ontology; TSS: transcription start site.

As a last layer of characterization of these active UCNEs and their candidate target genes, we assigned to each target gene cell-type specificity information based on expression data. Out of the 1,487 active UCNEs, 808 (54.3%) displayed a cell-type specific open chromatin context consistent with their target gene(s) expression signature(s) (Supplementary Table 1).

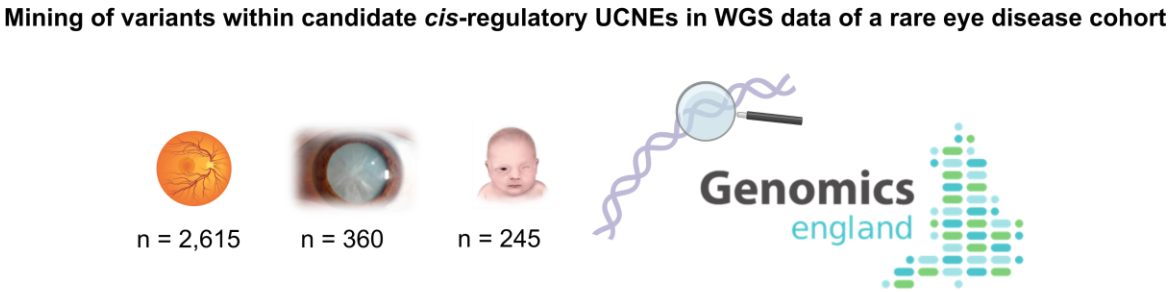
Mining of WGS data from a rare eye disease cohort reveals rare variants within active UCNEs

Out of the 594 putative target genes under UCNE *cis*-regulation, 45 were found to be previously linked to a rare eye disease phenotype, of which 19 are also associated with other developmental disorder

171 phenotypes (Supplementary Table 4). Considering the extreme selective constraints of UCNEs, we
172 hypothesized that genetic changes in these regions could contribute to disease. Therefore, we
173 evaluated the genomic variation within the active UCNEs associated with these 45 disease genes in a
174 sub-cohort of individuals affected by rare eye diseases (n = 3,220, Supplementary Table 5) from the
175 100,000 Genomes Project (Fig. 3A).

176 As expected, a depletion of common variation (MAF>1%) was observed across these regions (Fig. 3B).
177 A total of 431 (426 SNVs and 5 SVs) ultrarare variants, i.e., absent from reference population public
178 databases, were identified in 199 active UCNEs linked to these 45 genes (Supplementary Table 6).
179 Notably, out of the 5 identified SVs, one corresponds to the known shared duplicated region
180 downstream of *IRX1*, located within the NCMD-linked *MCDR3* locus [MIM: 608850]; in particular, this
181 duplication, identified in 8 affected individuals of 4 different families, involves a UCNE exhibiting
182 chromatin accessibility in developing horizontal cells. Finally, out of these ultrarare variants, 178 are
183 located within 84 UCNEs displaying histone modification marks (in at least one of the interrogated
184 stages), associated with 29 genes. This set defined our primary search space for variants with potential
185 functional effects and further assessment.

A



B

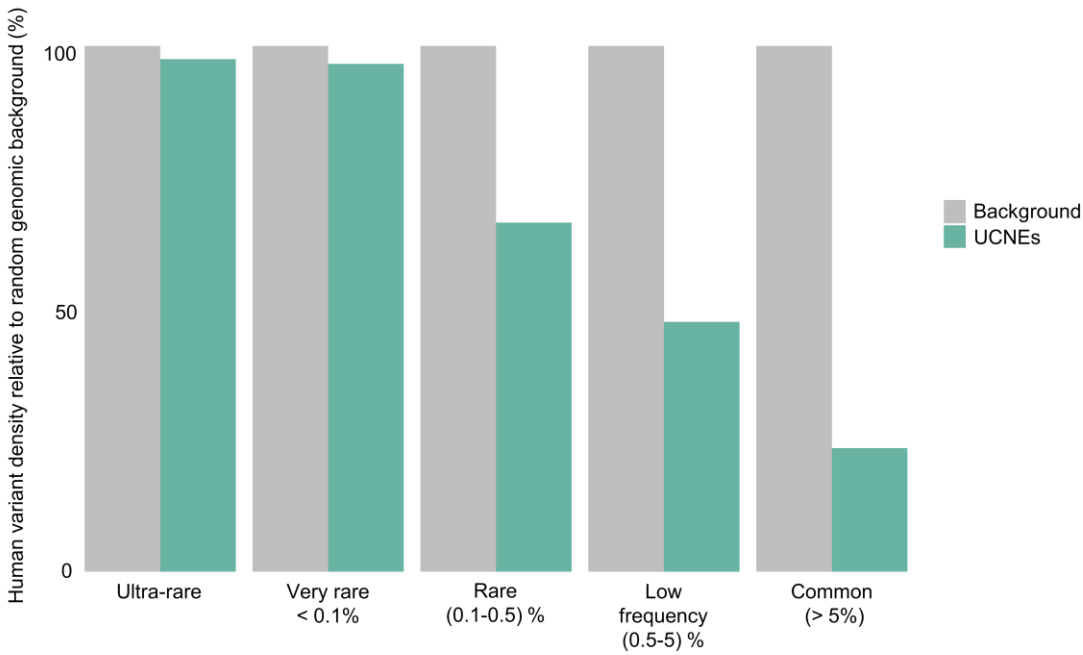


Figure 3. Contribution of UCNE genomic variation to missing heritability in rare eye diseases. A) Overview of a sub-cohort comprising n=3,220 participants of the 100,000 Genomes Project affected by posterior segment abnormalities (n=2,615), anterior segment abnormalities (n=360), and ocular malformations (n=245). **B)** Variant population frequencies within putative retinal UCNEs normalized to a background composed of randomly selected sequences (see Methods).

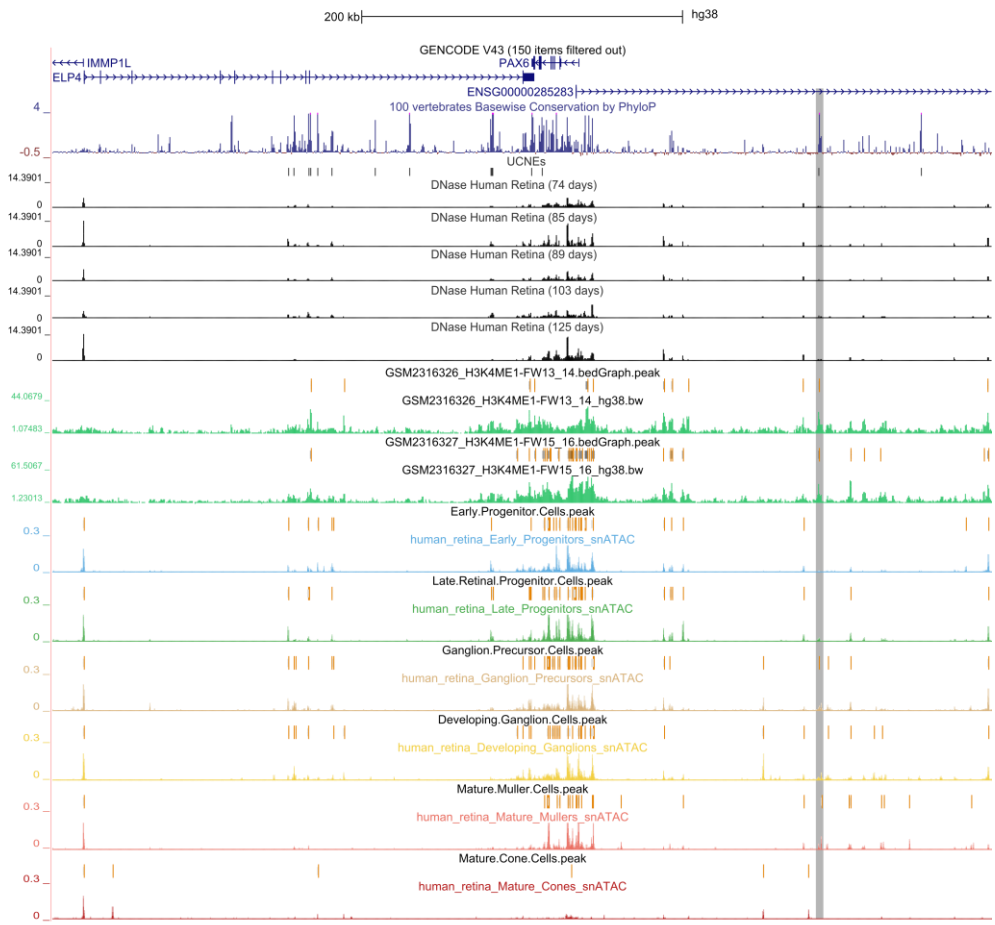
An ultrarare SNV in an active UCNE upstream of *PAX6* found in a family segregating autosomal dominant foveal anomalies

We identified an ultrarare SNV (chr11:31968001T>C) within a candidate *cis*-regulatory UCNE located ~150kb upstream of *PAX6*. This variant was found in four individuals of a family segregating autosomal dominant foveal abnormalities (Fig. S1A-B; Supplementary Table 7). A *CFH* missense variant (c.1187A>G, p.Asn396Ser) was initially reported in the affected individuals but could not explain the

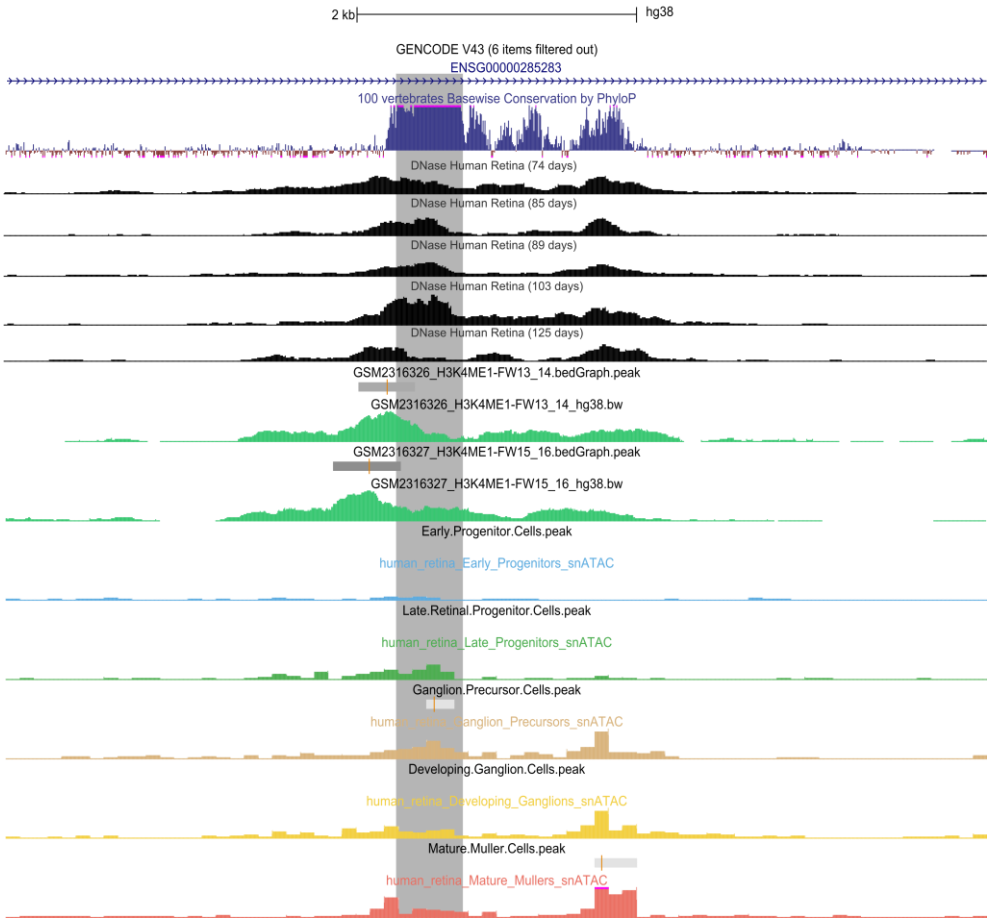
200 foveal anomaly. Given the phenotype, a variant screening was performed with a particular focus on
201 the *PRDM13* and *IRX1* NCMD loci, both associated with foveal or macular mis-development. A total of
202 three and two SNVs within the *IRX1* and *PRDM13* loci respectively were found to segregate with
203 disease. However, these SNVs are present in individuals from reference population databases and are
204 not located in genomic regions displaying *cis*-regulatory features (Supplementary Table 8). Equally, loci
205 associated with foveal hypoplasia, nystagmus and hypopigmentation (*AHR*, *FRMD7*, *GPR143*, and
206 *SLC38A8*) were assessed for SNV or SV segregating with disease. No other (likely) pathogenic variants
207 in these loci or any other known IRD/rare eye disease gene were identified in the affected cases
208 (Supplementary Table 8).

209 The identified chr11:31968001T>C variant affects a nucleotide residue that is conserved for at least
210 360 million years of evolution that separate humans from *X. tropicalis*. *In silico* assessment of this
211 variant and flanking sequence pointed to a likely deleterious effect and revealed a predicted disruption
212 of several TF binding motifs (Supplementary Table 8_2). More specifically, this variant is located within
213 a UCNE that is catalogued as a cCRE in ENCODE and featured by distal enhancer-like signatures in
214 bipolar neurons (EH38E1530321). Regarding its retinal context, this UCNE displays accessible
215 chromatin in the early stages of retinal development (7/8 gestational weeks), and in particular in
216 ganglion cell precursors (Fig. 4A-B). Importantly *PAX6* is expressed within this pool of cells and,
217 interestingly, this expression appears to be enriched within a specific sub-population, as observed from
218 the distribution of expression values across all the cells composing this cluster (Fig. S2). Additionally,
219 this UCNE was found to display the active enhancer mark H3K4Me1 at the earliest time points available
220 for this dataset (13/14 and 15/16 gestational weeks).

A


200 kb
hg38
GENCODE V43 (150 items filtered out)
IMMP1L
PAX6
ELP4
ENSG00000285283
100 vertebrates Basewise Conservation by PhyloP
UCNEs
DNase Human Retina (74 days)
DNase Human Retina (85 days)
DNase Human Retina (89 days)
DNase Human Retina (103 days)
DNase Human Retina (125 days)
GSM2316326_H3K4ME1-FW13_14.bedGraph.peak
GSM2316326_H3K4ME1-FW13_14_hg38.bw
GSM2316327_H3K4ME1-FW15_16.bedGraph.peak
GSM2316327_H3K4ME1-FW15_16_hg38.bw
Early.Progenitor.Cells.peak
human_retina_Early_Progenitors_snATAC
Late.Retinal.Progenitor.Cells.peak
human_retina_Late_Progenitors_snATAC
Ganglion.Precursor.Cells.peak
human_retina_Ganglion_Precursors_snATAC
Developing.Ganglion.Cells.peak
human_retina_Developing_Ganglions_snATAC
Mature.Muller.Cells.peak
human_retina_Mature_Mullers_snATAC
Mature.Cone.Cells.peak
human_retina_Mature_Cones_snATAC

B


2 kb
hg38
GENCODE V43 (6 items filtered out)
ENSG00000285283
100 vertebrates Basewise Conservation by PhyloP
DNase Human Retina (74 days)
DNase Human Retina (85 days)
DNase Human Retina (89 days)
DNase Human Retina (103 days)
DNase Human Retina (125 days)
GSM2316326_H3K4ME1-FW13_14.bedGraph.peak
GSM2316326_H3K4ME1-FW13_14_hg38.bw
GSM2316327_H3K4ME1-FW15_16.bedGraph.peak
GSM2316327_H3K4ME1-FW15_16_hg38.bw
Early.Progenitor.Cells.peak
human_retina_Early_Progenitors_snATAC
Late.Retinal.Progenitor.Cells.peak
human_retina_Late_Progenitors_snATAC
Ganglion.Precursor.Cells.peak
Developing.Ganglion.Cells.peak
Mature.Muller.Cells.peak
human_retina_Mature_Mullers_snATAC

221

222 **Figure 4. Characterization of the *cis*-regulatory landscape of *PAX6* in human retina. A)** Visualization
223 of the epigenomic context of the identified *PAX6*-associated UCNE (highlighted in *gray*) in relation to
224 the *PAX6* locus (*top;* chr11:31,490,261-32,075,591) and **B)** zoomed-in of the identified element
225 (*bottom*; chr11:31,965,000-31,972,000). Figures are obtained from the UCSC genome browser.

226

227 This region has also been functionally validated in transgenic mouse assays (VISTA element hs855),
228 which revealed gene enhancer activity mainly in the forebrain and, to some extent, in the retina (3/6
229 embryos) (Fig. 5A). A closer assessment of this UCNE with regard to its proximity to the *PAX6* promoter
230 in other species revealed its syntenic conservation up to zebrafish; interestingly, this element localized
231 in closer proximity to the *PAX6* promoter in species like *X. tropicalis* (60kb distance from the TSS) or
232 *Danio rerio* (15kb distance from the *pax6a* TSS) (Fig. S3).

233 *In vivo* enhancer assays in zebrafish display retinal activity for the active UCNE upstream of
234 *PAX*6

235 Given the putative retinal activity suggested by the VISTA enhancer assay for this UCNE, we performed
236 zebrafish enhancer assays to gain further insights into its regulatory range of action. We found
237 consistent reporter expression in several regions, including the forebrain, hypothalamus/otic vesicle,
238 somites, nosepit and eye. Importantly, we observed that the GFP expression in the eye initiated in a
239 small fraction of the embryos at 2 dpf (4/22 GFP^+ embryos) increasing to a larger proportion of the
240 embryos by3 dpf (21/22 GFP^+ embryos) and eventually disappeared from 4dpf (Fig. 5B). A detailed
241 overview of the results of these assays can be found in Supplementary Table 10. Altogether, these
242 observations are consistent with the epigenomic characterization based on retinal datasets and
243 provides further evidence of a potential *cis*-regulatory role of this UCNE on *PAX6* expression.

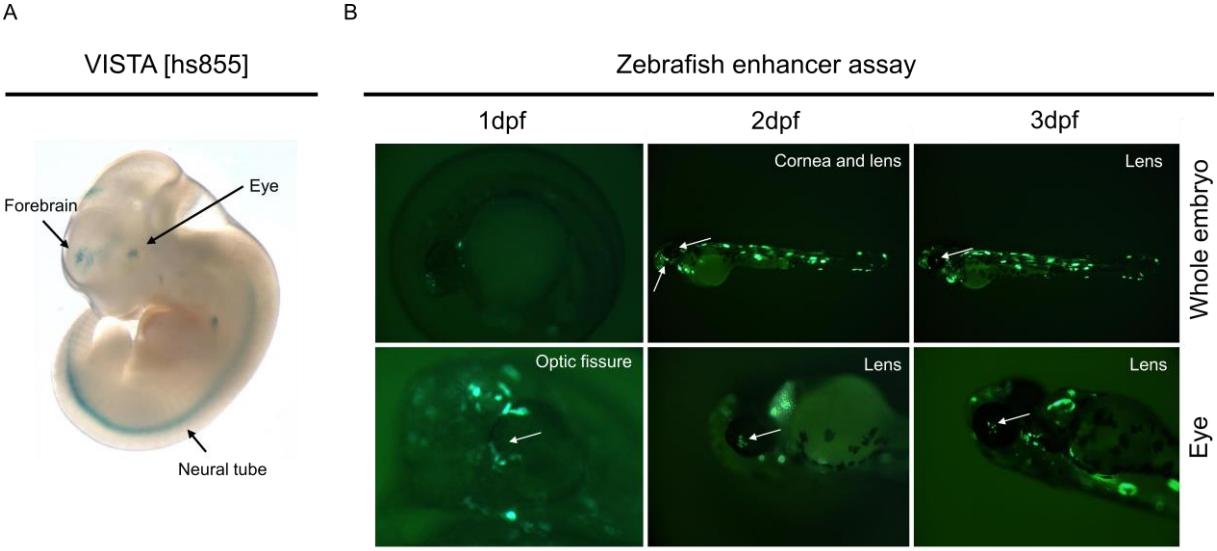


Figure 5. *In vivo* evaluation of the *PAX6*-associated UCNE. A) This UCNE displays reporter activity in the human embryonic forebrain (6/9 embryos) as well as in other structures including eye (3/6 embryos) and neural tube. Figure obtained from the VISTA enhancer browser (hs855). **B)** Zebrafish enhancer assays at 3 different time points (1 dpf, 2 dpf, 3 dpf) showing GFP-positive tissues (optic fissure -1 dpf-, lens cells -2,3 dpf- and forebrain -2,3 dpf-; *white arrows*). Abbreviations. dpf: days post-fertilization.

DISCUSSION

Since the landmark study of Bejerano et al. (Bejerano et al., 2004) almost twenty years ago, ultraconservation of non-coding elements and their paradoxical functional (in)dispensability (Ahituv et al., 2007; Chiang et al., 2008; Dickel et al., 2018; Drake et al., 2006; Snetkova, Pennacchio, et al., 2021) have equally stirred controversy and fascination among scientists from diverse disciplines. A recent study demonstrated that UCNEs with robust enhancer activity during embryonic development appear to be unexpectedly resilient to mutation (Snetkova, Pennacchio, et al., 2021). There are, however, clear instances in which variation within ultraconserved CREs can be a driver of human disease (Benko et al., 2009; Bhatia et al., 2013; Ghiasvand et al., 2011; Klopocki et al., 2008; Kvon et al., 2020; Lettice et al., 2003; Martínez et al., 2010; Plaisancié et al., 2018; Wieczorek et al., 2010). Despite the substantial body of literature on these intriguing elements, thus far there are no comprehensive studies based on the integration of multi-omics to map the regulatory capacity of UCNEs in a specific tissue or cells, and on human genetic data to assess their contribution to disease. Apart from a few well-known examples,

265 variants in CREs are an underrepresented cause of Mendelian diseases. A major obstacle hampering
266 their identification is the need to define the search space of the CREs they affect in a tissue- and even
267 cell-type specific manner. Here we set out to address this challenge in human retina, of which the *cis*-
268 regulatory architecture has been well-studied (Cherry et al., 2020; Marchal et al., 2022; Thomas et al.,
269 2022) and for which there is emerging *cis*-regulatory variation implicated in disease (Bhatia et al., 2013;
270 Ghiasvand et al., 2011; Plaisancié et al., 2018; Van de Sompele et al., 2022).

271 Although ENCODE provides a well-annotated inventory of cCREs, its retinal datasets do not include the
272 earliest stages of RPC differentiation. To overcome this and to incorporate cell-type specificity
273 information, we integrated scATAC-seq, which allowed us to identify 1,487 UCNEs characterized by
274 accessible chromatin in at least one major stage of retinal development. We also interrogated the
275 epigenetic landscape of these elements by analyzing histone modification patterns associated with e.g.
276 active/poised enhancers (H3K27ac, H3K4me1, H3K4me2) (Creyghton et al., 2010; Rada-Iglesias et al.,
277 2011; Santos-Rosa et al., 2002) and highly packed chromatin (H3K27me2 and H3K27me3) (Juan et al.,
278 2016). This analysis provided a useful approach to trace the activity of elements that are active at a
279 specific stage during retinal development. In total, 111 UCNEs were found to display active enhancer
280 marks during retinal development. Similarly to ENCODE, the analyzed ChIP-seq datasets do not cover
281 the earliest stages and hence these 111 elements are more likely to be related to stages of
282 differentiation corresponding to later-born cell types (i.e., rod photoreceptors, bipolar cells, Müller
283 glial, and some varieties of amacrine cells) (Rapaport et al., 2004). Overall, our data integration spans
284 all the major stages of retinogenesis and thus provides a comprehensive framework for the systematic
285 characterization of ultraconserved retinal CREs.

286 Establishing an association between cCREs and their putative target genes is essential for the
287 interpretation of genomic variation that could disrupt cell type-specific binding sites of TFs and/or long-
288 range chromatin contacts that can lead to ectopic expression of genes (de Bruijn et al., 2020; Lupiáñez
289 et al., 2015). Here we combined GREAT (McLean et al., 2010) with retinal expression and Hi-C derived
290 retinal data to assign a range of action to candidate *cis*-regulatory UCNEs. In total, 594 retina-expressed

291 genes can be under UCNE *cis*-regulation; interestingly most of the UCNEs were found to be distributed
292 in distal position from the TSS of their assigned target genes, as observed for enhancers promoting
293 expression of developmental genes (Benko et al., 2009; Long et al., 2020; Pachano et al., 2022). In this
294 regard, a large fraction of the identified ultraconserved putative enhancers clusters around key
295 developmental genes (e.g. *MAB21L2*, *OTX2*, *PAX6*, *SOX2*, *ZEB2*) known to be controlled by complex
296 regulatory landscapes (Polychronopoulos et al., 2017). Besides, 45 of these genes are implicated in
297 rare eye diseases. Interestingly, coding variants in these genes lead in most cases to syndromic
298 developmental phenotypes with other manifestations apart from the ocular ones as collated in the
299 G2P database [23]. As has been shown before, the phenotype caused by a coding mutation of a
300 developmental gene can be different from the phenotype caused by a mutation in a CRE controlling
301 spatiotemporal expression of this gene. This is exemplified by the *PRDM13* locus, for which bi-allelic
302 coding *PRDM13* variants result in hypogonadotropic hypogonadism and perinatal brainstem
303 dysfunction in combination with cerebellar hypoplasia (Whittaker et al., 2021) while *cis*-regulatory
304 variants in retinal CREs leads to NCMD, a developmental macular disease (Van de Sompele et al., 2022).
305 From our analyses we could identify a cCRE for *PRDM13* that displays an open chromatin context in
306 early RPCs and developing amacrine cells, which are precisely the retinal cell types in which *PRDM13*
307 is expressed (Goodson et al., 2018; Watanabe et al., 2015). Although no disease associations have been
308 established thus far for other genes for which we identified candidate *cis*-regulatory UCNEs, some of
309 them are known to play important roles in retina development, e.g., *NFIB* (Clark et al., 2019), and
310 therefore represent potential target regions for non-coding variants contributing to missing
311 heritability.

312 A limitation of this analysis is the usage of a pre-defined set of genomic regions, in particular UCNEs,
313 as this database is static and rarely updated (Polychronopoulos et al., 2017). Our approach could thus
314 be expanded using more flexible criteria, thus including other less conserved, albeit disease-relevant
315 loci (Ghiasvand et al., 2011). Nonetheless, the advantage of working with highly conserved genomic
316 regions is the availability of already validated experimental data (Visel et al., 2007); additionally, as

317 shown here, the integration of tissue-specific epigenomic data provides a robust framework to identify
318 potential *bona fide* enhancers and their corresponding target genes, and clinically relevant variants
319 therein.

320 To further investigate the contribution of genomic variation within candidate *cis*-regulatory UCNEs to
321 disease, we mined WGS data from a sub-cohort of 100,000 Genome Project Genomics England
322 participants affected by different rare eye diseases. As reported previously (Katzman et al., 2007;
323 Snetkova, Pennacchio, et al., 2021), we observed a substantial depletion of common variation across
324 these regions when compared to a background comprising randomly selected genomic regions. This
325 apparently strong purifying selection has been proven difficult to reconcile with recent findings
326 demonstrating remarkable resilience of UCNEs to variation (Snetkova, Ypsilanti, et al., 2021).
327 Moreover, it has been suggested that weaker but uniform levels of purifying selection across hundreds
328 of bases and different species could bring together these otherwise contradictory observations and
329 explain why rare variants are not significantly depleted within UCNEs (Dukler et al., 2022). Our primary
330 search space for variants with potential functional effects comprised 178 ultrarare variants located
331 within 84 retina-active UCNEs associated with 29 genes. Interestingly, most of them were predicted to
332 have a likely functional and/or deleterious effect, which could be explained by a skewed cumulative
333 importance towards evolutionary conservation-related features in the predictive models (Rogers et
334 al., 2018; Schwarz et al., 2019; Shi et al., 2021; Zhou & Troyanskaya, 2015).

335 Out of all variants found, one ultrarare variant (chr11:31968001T>C) within a UCNE displaying open
336 chromatin in progenitor ganglion cells and linked to *PAX6* was further dissected. This variant was
337 exclusively found within the studied family, displaying isolated foveal abnormalities, and initially solved
338 with an ultrarare missense variant in *CFH*, which could not fully explain the clinical presentation
339 (Raychaudhuri et al., 2011; Taylor et al., 2019). As no other likely pathogenic variants were found
340 within relevant loci (Ehrenberg et al., 2021; Kuht et al., 2020, 2022; Small et al., 2016; Van de Sompele
341 et al., 2022), and the observed phenotypes of the affected individuals match within the *PAX6* disease
342 spectrum, mis-regulation of *PAX6* expression cannot be excluded as a pathogenetic mechanism.

343 Genetic defects of *PAX6* have been found in aniridia, a pan-ocular disorder characterized by the
344 absence or hypoplasia of the iris, nystagmus and foveal hypoplasia (Cunha et al., 2019), the latter
345 comprising thinning of macular inner and outer retinal layers consistent with misdirected foveal
346 development (Pedersen et al., 2020). Recently, a phenotype characterized by isolated foveal
347 hypoplasia with nystagmus has been also linked to *PAX6* variation (Lima Cunha et al., 2021). Given the
348 known inter- and intra-familial phenotypic variability observed in *PAX6*-associated disorders (Dubey et
349 al., 2015; Yokoi et al., 2016), variable expressivity cannot be discarded for the identified variant. Thus
350 far, in terms of regulatory variation in CREs, only a single SNV located within the ultraconserved SIMO
351 element has been associated with *PAX6*-disease. This subtle change (SNV), identified *de novo* in an
352 individual with aniridia and foveal hypoplasia, was found to disrupt an autoregulatory PAX6 binding
353 site (Bhatia et al., 2013). Importantly, we could establish *in vivo*, using the zebrafish as animal model,
354 a developmental expression pattern in the eye driven by this UCNE in tissues for which *PAX6* expression
355 is relevant (Ashery-Padan et al., 2000; Collinson et al., 2003; Georgala et al., 2011; Kimura et al., 2005;
356 Pituello et al., 1999) and within a time window consistent with the period in which the zebrafish retina
357 has become fully laminated (Morris & Fadool, 2005; Schmitt & Dowling, 1999). However, to confirm
358 the association of this UCNE and impact of the identified variant on *PAX6* expression, further functional
359 assays are needed. The phenotypic consequence of the variant itself, nevertheless, constitutes a
360 limitation for modelling since currently no suitable models, including retinal organoids, fully
361 recapitulate foveal patterning (Hussey et al., 2022).

362 Overall, our work is exemplar for how the wealth of publicly available multi-omics data can be used to
363 exploit the regulatory capacity of UCNEs in a tissue- and cell type-specific way. As demonstrated here,
364 UCNEs can represent understudied regions of non-coding variation underlying missing heritability in
365 Mendelian diseases. With the increasing implementation of WGS in rare disease research and
366 diagnosis, the delineation of tissue and cell type-specific CREs will be a prerequisite to identify and fully
367 interpret the pathogenic nature of non-coding variants. With this study, we have illustrated how the
368 creation of a comprehensive set of functionally annotated UCNEs in a target tissue can represent a

369 powerful initial strategy to narrow down the variant search space, particularly for developmental
370 phenotypes.

371 # METHODS

372 **Integration of UCNEs with bulk and single-cell epigenomic, regulatory and transcriptional**
373 **datasets from human developing and adult retina**

374 The 4,135 genomic regions defined as UCNEs (Dimitrieva & Bucher, 2013) were used as the basis for
375 the integration of multiple publicly available multi-omics datasets derived from embryonic and adult
376 human retina. More specifically, to evaluate the potential function of UCNEs as *cis*-regulatory modules,
377 we made use of DNase-seq (Abascal et al., 2020), ChIP-seq of histone modifications (Aldiri et al., 2017),
378 and single-cell (sc) ATAC-seq (Thomas et al., 2022) derived from retinal tissue collected at distinct
379 stages of development. Furthermore, to correlate potential *cis*-regulatory activity with gene
380 expression, we mined both bulk (Hoshino et al., 2017) and scRNA-seq (Thomas et al., 2022) generated
381 at stages overlapping or extending the ones of the epigenomic datasets. Additionally, as a last layer of
382 functional characterization, we integrated into our analyses the experimental data of the VISTA
383 Enhancer Browser (Visel et al., 2007), which allows the classification of candidate regulatory elements
384 based on *in vivo* enhancer reporter assays tested in transgenic mice at embryonic day 11.5 (Pennacchio
385 et al., 2006).

386 **Identification of candidate *cis*-regulatory UCNEs in retina**

387 Data generated by scATAC-seq of embryonic (53, 59, 74, 78, 113, and 132 days) and adult (25, 50, and
388 54 years old) human retinal cells were obtained (GSE183684) and imported into R (v4.0.5). Count
389 matrices were processed using the *ArchR* single-cell analysis package (v1.0.1) (Granja et al., 2021) as
390 described in Thomas et al., 2022(Thomas et al., 2022) with minor modifications. Briefly, the total
391 number of cells after filtering out doublets was 61,313. Single-nucleus RNA-seq of the same tissue and
392 timepoints (GSE183684) were integrated using the unconstrained integration method. Both datasets

393 were then used to assign retinal cell class identities to the different clusters and subsequent peak
394 calling. BigWig files from each annotated cell cluster were extracted and converted into bedGraph files
395 using *bigWigtoBedGraph* UCSC utility; peak detection was performed using *bdgpeakcall* (MACS2.2.7.1)
396 (Zhang et al., 2008) with default parameters and a value of 0.4 as cut-off. Similarly, bigWig files
397 corresponding to histone modification patterns (H3K27Ac, H3K27me3, H3K36me3, H3K4me1,
398 H3K4me2, H3K4me3, H3K9/14Ac, H3K9me3, PolII, CTCF) were retrieved (GSE87064) and converted
399 into bedGraph files; in this case, the cut-off value for peak calling by *bdgpeakcall* was set at 20.
400 Additionally, the subset of ENCODE DNase hypersensitivity sites (rDHSs) identified in embryonic retina
401 (74-85, 89, 103-105 days) were obtained (ENCSR786VSQ, ENCSR666FML, ENCSR632UXV) and elements
402 featured by *Low-DNase* filtered out. Finally, to identify putative *cis*-regulatory UCNEs, we screened for
403 overlaps between UCNEs and the retrieved sc-ATAC/ChIP/DNase-seq peaks using *bedtools intersect*
404 (*BEDTools* (2.30.0)(Quinlan & Hall, 2010) with default parameters; of note *bedtools window* including
405 a ±250bp-long window was used when overlapping the ChIP-seq peaks to characterize more broadly
406 the chromatin status in the vicinity of the accessible UCNEs (Fig. S3).

407 **Identification and characterization of target genes under active UCNE regulation**

408 To assign potential target genes to the identified active UCNEs we used the Genomic Regions
409 Enrichment of Annotations Tool (*GREAT*) (McLean et al., 2010). Briefly, each gene in the genome was
410 assigned a basal regulatory domain (5kb and 1kb upstream and downstream of the transcription start
411 site {TSS}, respectively) with an extension of up to 1Mb to the nearest gene's basal domain. Curated
412 regulatory domains were also included. Each active UCNE was then associated with all genes whose
413 regulatory region it overlaps.

414 Potential target genes were subsequently filtered based on retinal expression. To do so, we retrieved
415 RNA-seq paired-end FASTQ files derived from fetal retina samples (52 to 136 days post-conception)
416 characterized in Hoshino et al. (Hoshino et al., 2017). Transcripts were quantified through pseudo-
417 alignment by *Kallisto* (v.0.46.1)(Bray et al., 2016) using default parameters for both index build (derived

418 from the Ensembl human release 101) and transcript abundance estimation. A custom R script was
419 then used to filter out target genes exhibiting no expression (TPM<0.5) at any of the interrogated
420 stages. Additionally, to evaluate the reliability of the regulatory domains assigned by *GREAT* to these
421 target genes, we integrated the 2,948 retinal TADs described by Marchal et al. (Marchal et al., 2022).
422 To characterize further the putatively regulated target genes, we made use of the integration of the
423 scATAC-seq gene score and the scRNA-seq gene expression matrices generated previously. For each
424 target gene, its expression was ranked by percentile of expression across all clusters. A gene expression
425 signature was then assigned by retrieving the cluster identities exhibiting an expression value above
426 the 80th percentile threshold.

427 Finally, to identify overrepresented Gene Ontology (GO) terms and infer possible enriched pathways
428 among the potential target genes under UCNE regulation, GO enrichment analyses were performed
429 using *Enrichr* (Xie et al., 2021).

430 **Interrogation of WGS data in a rare eye disease cohort**

431 Active UCNEs associated with genes expressed in retina were filtered further based on their implication
432 in disease as retrieved from the comprehensive *gene-disease pairs and attributes* list provided by G2P
433 (Eye and Developmental Disease –DD– Panels; 2022-03-17) (Lenassi et al., 2021; Thormann et al., 2019)
434 extended with the NCMD-associated *IRX1* locus(Small et al., 2016). To assess the contribution of
435 genomic variation within these loci to disease, an analysis was performed to detect small variants
436 (SNVs, and indels < 50bp), large structural variants (SVs) including copy number variants (CNVs)
437 overlapping these disease-gene associated UCNE loci through query of a sub-cohort of participants
438 with rare eye disease phenotypes (n = 3,220) from the 100,000 Genomes Project (100KGP, Genomics
439 England).

440 Retrieved variants were annotated with VEP (v.107). As a first filter, only variants with filtering allele
441 frequency <0.5% and no observed homozygotes in gnomAD v3.1 were considered for further
442 assessment. Variants were evaluated for their potential pathogenic/modifying effect using *in silico*

443 prediction tools focusing on transcription factor (TF) binding site disruption -QBiC-Pred- and chromatin
444 state effects -DeepSea, CARMEN, FATHMM-XF, RegulationSpotter. Furthermore, we annotated these
445 variants with our integrated analyses in order to evaluate them within their regulatory context and
446 potential target genes.

447 For each candidate variant, we compared the similarities between the participant phenotype (HPO
448 terms) and the ones known for its target gene through literature search and clinical assessment by the
449 recruiting clinician when possible. Finally, for each candidate variant identified in participants whose
450 cases were not solved through 100KGP, a variant screening of known retinal disease genes (Genomics
451 England Retinal Disorder panel)(Martin et al., 2019) was performed to discard (likely) pathogenic
452 variants, both SNVs and SVs, that could provide an alternative molecular diagnosis. For each instance
453 for which only the UCNE variant remained as candidate, we placed a clinical collaboration request with
454 Genomics England.

455 **Targeted sequencing and reverse phenotyping**

456 We performed segregation testing of two ultrarare SNVs initially identified in three affected individuals
457 of a 3-generation family displaying autosomal dominant foveal abnormalities, namely an SNV in a
458 candidate *cis*-regulatory UCNE located upstream of *PAX6* and a missense *CFH* variant. Genomic DNA
459 was extracted using Oragene-DNA saliva kits (OG-500, DNA Genotek) according to manufacturer's
460 instructions. Targeted sequencing of the variants was performed on genomic DNA by PCR amplification
461 followed by Sanger sequencing using the BigDye Terminator v3.1 kit (Life Technologies). Primer
462 sequences can be found in Supplementary Table 9.

463 Following their genetic assessment, four members of this 3-generation family were clinically re-
464 evaluated. The examination included visual acuity assessment, slit-lamp examination for anterior and
465 posterior segment anomalies, and intraocular pressure measurement. Detailed imaging involving
466 ultra-wide field fundus photography, ultra-wide field autofluorescence imaging and optical coherence
467 tomography (OCT) was performed.

468 **Generation of *in vivo* reporter constructs and functional characterization of candidate *cis*-**
469 **regulatory UCNE upstream of *PAX6* using enhancer assays in zebrafish embryos**

470 Primers were designed to amplify the sequence of the candidate *cis*-regulatory UCNE located upstream
471 of *PAX6* from human genomic DNA (Roche). The PCR product was then cloned into the E1b-GFP-Tol2
472 enhancer assay vector containing an E1b minimal promoter followed by the Green Fluorescent Protein
473 (GFP) reporter gene(Li et al., 2010) by restriction-ligation cloning. The primer sequences can be found
474 in Supplementary Table 9. The recombinant vector containing the cCRE-*PAX6*-UCNE was then amplified
475 in One Shot TOP10 Chemically Competent E. coli cells (Invitrogen) and purified using the QIAprep Spin
476 Miniprep Kit (Qiagen). The sequence of the insert was confirmed by Sanger sequencing using the
477 BigDye Terminator v3.1 kit (Life Technologies). The constructs were then microinjected into the yolk
478 of at least 70 wild-type zebrafish embryos at single-cell stage using the Tol2 transposase system for
479 germline integration of the transgene according to Bessa *et al*. (Bessa et al., 2009) with minor
480 modifications. As readout, GFP fluorescence was observed and photographed with a Leica M165 FC
481 Fluorescent Stereo Microscope (Leica Microsystems) and its localization annotated at 1, 2 and 3 days
482 post fertilization (dpf) to evaluate enhancer activity.

483

484

485

486

487

488

489

490

491

492 **COMPETING INTEREST STATEMENT**

493 The authors declare that they have no competing interests.

494 **FUNDING**

495 This work was supported by the Ghent University Special Research Fund (BOF20/GOA/023) (EDB);
496 H2020 MSCA ITN grant (No. 813490 StarT) (EDB, FC, MB), EJPRD19-234 Solve-RET (EDB). EDB is a Senior
497 Clinical Investigator (1802220N) of the Research Foundation-Flanders (FWO); VLS and ADR are an Early
498 Starting Researcher of StarT (grant No. 813490). EDB is member of ERN-EYE (Framework Partnership
499 Agreement No 739534-ERN-EYE).

500 **ACKNOWLEDGEMENTS**

501 This research was made possible through access to the data and findings generated by the 100,000
502 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly
503 owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded
504 by the National Institute for Health Research and NHS England. The Welcome Trust, Cancer Research
505 UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes
506 Project uses data provided by patients and collected by the National Health Service as part of their
507 care and support. We acknowledge Chris Inglehearn (Leeds Institute of Medical Research, School of
508 Medicine, University of Leeds, Leeds, UK) for his helpful advice and comments for the manuscript. We
509 also would like to thank the Zebrafish Facility Ghent (ZFG) Core at Ghent University. Hanna De Saffel,
510 Quinten Mahieu, Angelika Jürgens and Lies Vantomme are thanked for their excellent technical
511 assistance.

512 **ETHICS APPROVAL AND CONSENT TO PARTICIPATE**

513 The 100,000 Genomes Project Protocol has ethical approval from the HRA Committee East of England
514 – Cambridge South (REC Ref 14/EE/1112). This study was registered with Genomics England under
515 Research Registry Projects 465.

AUTHORS' CONTRIBUTIONS

V.L.S.: Conception and project design, acquisition of data, analysis and interpretation of data, drafting and revising the manuscript.

A.D.R.: Conception and project design, acquisition of data, analysis and interpretation of data, drafting and revising the manuscript.

R.M: Acquisition of data, analysis and interpretation of data, revising the manuscript.

G.E: Acquisition of data, revising the manuscript.

F.C: Acquisition of data, revising the manuscript.

M.B: Project supervision, acquisition of data, revising the manuscript.

A.W: Acquisition of data, analysis and interpretation of data, revising the manuscript.

E.D.B.: Conception and project supervision, acquisition of data, analysis and interpretation of data, drafting and revising the manuscript.

CONSENT FOR PUBLICATION

Not applicable. We present only de-identified data.

AVAILABILITY OF DATA AND MATERIALS

The data that support the findings of this study are available within the Genomics England (protected) Research Environment but restrictions apply to the availability of these data, as access to the Research Environment is limited to protect the privacy and confidentiality of participants. De-identified data as well as analysis scripts are, however, available from the authors upon reasonable request. Extended data generated in this study are available in the supplementary materials.

536 **REFERENCES**

537 Abascal, F., Acosta, R., Addleman, N. J., Adrian, J., Afzal, V., Aken, B., Akiyama, J. A., Jammal, O. Al, …
538 Zimmerman, J. (2020). Expanded encyclopaedias of DNA elements in the human and mouse
539 genomes. *Nature*, *583*(7818), 699–710. https://doi.org/10.1038/s41586-020-2493-4

540 Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L. A., & Rubin, E. M. (2007). Deletion of
541 ultraconserved elements yields viable mice. *PLoS Biology*, *5*(9), 1906–1911.
542 https://doi.org/10.1371/journal.pbio.0050234

543 Aldiri, I., Xu, B., Wang, L., Chen, X., Hiler, D., Griffiths, L., Valentine, M., Shirinifard, A., … Dyer, M. A.
544 (2017). The Dynamic Epigenetic Landscape of the Retina During Development, Reprogramming,
545 and Tumorigenesis. *Neuron*, *94*(3), 550-568.e10. https://doi.org/10.1016/j.neuron.2017.04.022

546 Ashery-Padan, R., Marquardt, T., Zhou, X., & Gruss, P. (2000). Pax6 activity in the lens primordium is
547 required for lens formation and for correct placement of a single retina in the eye. *Genes and*
548 *Development*, *14*(21), 2701–2711. https://doi.org/10.1101/gad.184000

549 Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004).
550 Ultraconserved elements in the human genome. *Science*, *304*(5675), 1321–1325.
551 https://doi.org/10.1126/science.1098119

552 Benko, S., Fantes, J. A., Amiel, J., Kleinjan, D. J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., …
553 Lyonnet, S. (2009). Highly conserved non-coding elements on either side of SOX9 associated
554 with Pierre Robin sequence. *Nature Genetics*, *41*(3), 359–364. https://doi.org/10.1038/ng.329

555 Bessa, J., Tena, J. J., De La Calle-Mustienes, E., Fernández-Miñán, A., Naranjo, S., Fernández, A.,
556 Montoliu, L., Akalin, A., … Gómez-Skarmeta, J. L. (2009). Zebrafish Enhancer Detection (ZED)
557 vector: A new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions
558 in zebrafish. *Developmental Dynamics*, *238*(9), 2409–2417. https://doi.org/10.1002/dvdy.22051

559 Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M. T., De Marco, R., Damante, G., Grainger, R., …
560 Kleinjan, D. A. (2013). Disruption of autoregulatory feedback by a mutation in a remote,
561 ultraconserved PAX6 enhancer causes aniridia. *American Journal of Human Genetics*, *93*(6),
562 1126–1134. https://doi.org/10.1016/j.ajhg.2013.10.028

563 Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq
564 quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

565 Cherry, T. J., Yang, M. G., Harmin, D. A., Tao, P., Timms, A. E., Bauwens, M., Allikmets, R., Jones, E. M.,
566 … Greenberg, M. E. (2020). Mapping the cis-regulatory architecture of the human retina reveals
567 noncoding genetic variation in disease. *Proceedings of the National Academy of Sciences of the*
568 *United States of America*, *117*(16), 9001–9012. https://doi.org/10.1073/pnas.1922501117

569 Chiang, C. W. K., Derti, A., Schwartz, D., Chou, M. F., Hirschhorn, J. N., & Wu, C. T. (2008).
570 Ultraconserved elements: Analyses of dosage sensitivity, motifs and boundaries. *Genetics*,
571 *180*(4), 2277–2293. https://doi.org/10.1534/genetics.108.096537

572 Cipriani, V., Silva, R. S., Arno, G., Pontikos, N., Kalhoro, A., Valeina, S., Inashkina, I., Audere, M., …
573 Moore, A. T. (2017). Duplication events downstream of IRX1 cause North Carolina macular
574 dystrophy at the MCDR3 locus. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-
575 017-06387-6

576 Clark, B. S., Stein-O'Brien, G. L., Shiau, F., Cannon, G. H., Davis-Marcisak, E., Sherman, T., Santiago, C.
577 P., Hoang, T. V., … Blackshaw, S. (2019). Single-Cell RNA-Seq Analysis of Retinal Development
578 Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. *Neuron*, *102*(6),
579 1111-1126.e5. https://doi.org/10.1016/j.neuron.2019.04.010

580 Collinson, J. M., Quinn, J. C., Hill, R. E., & West, J. D. (2003). The roles of Pax6 in the cornea, retina,
581 and olfactory epithelium of the developing mouse embryo. *Developmental Biology*, *255*(2),
582 303–312. https://doi.org/10.1016/S0012-1606(02)00095-7

583 Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J.,
584 Lodato, M. A., … Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers
585 and predicts developmental state. *Proceedings of the National Academy of Sciences of the*
586 *United States of America*, *107*(50), 21931–21936. https://doi.org/10.1073/pnas.1016071107

587 Cunha, D. L., Arno, G., Corton, M., & Moosajee, M. (2019). The spectrum of PAX6 mutations and
588 genotype-phenotype correlations in the eye. *Genes*, *10*(12).
589 https://doi.org/10.3390/genes10121050

590 de Bruijn, S. E., Fiorentino, A., Ottaviani, D., Fanucchi, S., Melo, U. S., Corral-Serrano, J. C., Mulders,
591 T., Georgiou, M., … Hardcastle, A. J. (2020). Structural Variants Create New Topological-
592 Associated Domains and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis
593 Pigmentosa. *American Journal of Human Genetics*, *107*(5), 802–814.
594 https://doi.org/10.1016/j.ajhg.2020.09.002

595 Dickel, D. E., Ypsilanti, A. R., Pla, R., Zhu, Y., Barozzi, I., Mannion, B. J., Khin, Y. S., Fukuda-Yuzawa, Y.,
596 … Visel, A. (2018). Ultraconserved Enhancers Are Required for Normal Development. *Cell*,
597 *172*(3), 491-499.e15. https://doi.org/10.1016/j.cell.2017.12.017

598 Dimitrieva, S., & Bucher, P. (2013). UCNEbase - A database of ultraconserved non-coding elements
599 and genomic regulatory blocks. *Nucleic Acids Research*, *41*(D1), 101–109.
600 https://doi.org/10.1093/nar/gks1092

601 Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological
602 domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*,
603 *485*(7398), 376–380. https://doi.org/10.1038/nature11082

604 Drake, J. A., Bird, C., Nemesh, J., Thomas, D. J., Newton-Cheh, C., Reymond, A., Excoffier, L., Attar, H.,
605 … Hirschhorn, J. N. (2006). Conserved noncoding sequences are selectively constrained and not
606 mutation cold spots. *Nature Genetics*, *38*(2), 223–227. https://doi.org/10.1038/ng1710

607 Dubey, S. K., Mahalaxmi, N., Vijayalakshmi, P., & Sundaresan, P. (2015). Mutational analysis and
608 genotype-phenotype correlations in southern Indian patients with sporadic and familial aniridia.
609 *Molecular Vision*, *21*(January), 88–97.

610 Dukler, N., Mughal, M. R., Ramani, R., Huang, Y. F., & Siepel, A. (2022). Extreme purifying selection
611 against point mutations in the human genome. *Nature Communications*, *13*(1).
612 https://doi.org/10.1038/s41467-022-31872-6

613 Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., …
614 Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome.
615 *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

616 Ehrenberg, M., Bagdonite-Bejarano, L., Fulton, A. B., Orenstein, N., & Yahalom, C. (2021). Genetic
617 causes of nystagmus, foveal hypoplasia and subnormal visual acuity- other than albinism.
618 *Ophthalmic Genetics*, *42*(3), 243–251. https://doi.org/10.1080/13816810.2021.1888128

619 Ellingford, J. M., Barton, S., Bhaskar, S., O'Sullivan, J., Williams, S. G., Lamb, J. A., Panda, B.,
620 Sergouniotis, P. I., … Black, G. C. M. (2016). Molecular findings from 537 individuals with
621 inherited retinal disease. *Journal of Medical Genetics*, *53*(11), 761–767.
622 https://doi.org/10.1136/jmedgenet-2016-103837

623 Engström, P. G., Fredman, D., & Lenhard, B. (2008). Ancora: A web resource for exploring highly

624 conserved noncoding elements and their association with developmental regulatory genes.
625 *Genome Biology*, *9*(2), 8–11. https://doi.org/10.1186/gb-2008-9-2-r34

626 Field, A., & Adelman, K. (2020). Evaluating Enhancer Function and Transcription. *Annual Review of*
627 *Biochemistry*, *89*, 213–234. https://doi.org/10.1146/annurev-biochem-011420-095916

628 Georgala, P. A., Carr, C. B., & Price, D. J. (2011). The role of Pax6 in forebrain development.
629 *Developmental Neurobiology*, *71*(8), 690–709. https://doi.org/10.1002/dneu.20895

630 Ghiasvand, N. M., Rudolph, D. D., Mashayekhi, M., Brzezinski, J. A., Goldman, D., & Glaser, T. (2011).
631 Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA
632 disease. *Nature Neuroscience*, *14*(5), 578–588. https://doi.org/10.1038/nn.2798

633 Goodson, N. B., Nahreini, J., Randazzo, G., Uruena, A., Johnson, J. E., & Brzezinski, J. A. (2018).
634 Prdm13 is required for Ebf3+ amacrine cell formation in the retina. *Developmental Biology*,
635 *434*(1), 149–163. https://doi.org/10.1016/j.ydbio.2017.12.003

636 Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y., & Greenleaf, W. J.
637 (2021). Author Correction: ArchR is a scalable software package for integrative single-cell
638 chromatin accessibility analysis (Nature Genetics, (2021), 53, 3, (403-411), 10.1038/s41588-021-
639 00790-6). *Nature Genetics*, *53*(6), 935. https://doi.org/10.1038/s41588-021-00850-x

640 Haer-Wigman, L., Van Zelst-Stams, W. A. G., Pfundt, R., Van Den Born, L. I., Klaver, C. C. W., Verheij, J.
641 B. G. M., Hoyng, C. B., Breuning, M. H., … Yntema, H. G. (2017). Diagnostic exome sequencing in
642 266 Dutch patients with visual impairment. *European Journal of Human Genetics*, *25*(5), 591–
643 599. https://doi.org/10.1038/ejhg.2017.9

644 Hoshino, A., Ratnapriya, R., Brooks, M. J., Chaitankar, V., Wilken, M. S., Zhang, C., Starostik, M. R.,
645 Gieser, L., … Reh, T. A. (2017). Molecular Anatomy of the Developing Human Retina.
646 *Developmental Cell*, *43*(6), 763-779.e4. https://doi.org/10.1016/j.devcel.2017.10.029

647 Hussey, K. A., Hadyniak, S. E., & Johnston, R. J. (2022). Patterning and Development of
648 Photoreceptors in the Human Retina. *Frontiers in Cell and Developmental Biology*, *10*(April), 1–
649 21. https://doi.org/10.3389/fcell.2022.878350

650 Juan, A. H., Wang, S., Ko, K. D., Zare, H., Tsai, P. F., Feng, X., Vivanco, K. O., Ascoli, A. M., … Sartorelli,
651 V. (2016). Roles of H3K27me2 and H3K27me3 Examined during Fate Specification of Embryonic
652 Stem Cells. *Cell Reports*, *17*(5), 1369–1382. https://doi.org/10.1016/j.celrep.2016.09.087

653 Katzman, S., Kern, A. D., Bejerano, G., Fewell, G., Fulton, L., Wilson, R. K., Salama, S. R., & Haussler, D.
654 (2007). Human genome ultraconserved elements are ultraselected. *Science*, *317*(5840), 915.
655 https://doi.org/10.1126/science.1142430

656 Kimura, J., Suda, Y., Kurokawa, D., Hossain, Z. M., Nakamura, M., Takahashi, M., Hara, A., & Aizawa, S.
657 (2005). Emx2 and Pax6 function in cooperation with Otx2 and Otx1 to develop caudal forebrain
658 primordium that includes future archipallium. *Journal of Neuroscience*, *25*(21), 5097–5108.
659 https://doi.org/10.1523/JNEUROSCI.0239-05.2005

660 Klopocki, E., Ott, C. E., Benatar, N., Ullmann, R., Mundlos, S., & Lehmann, K. (2008). A
661 microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal
662 thumb-polysyndactyly syndrome. *Journal of Medical Genetics*, *45*(6), 370–375.
663 https://doi.org/10.1136/jmg.2007.055699

664 Kuht, H. J., Han, J., Maconachie, G. D. E., Park, S. E., Lee, S. T., McLean, R., Sheth, V., Hisaund, M., …
665 Thomas, M. G. (2020). SLC38A8 mutations result in arrested retinal development with loss of
666 cone photoreceptor specialization. *Human Molecular Genetics*, *29*(18), 2989–3002.
667 https://doi.org/10.1093/hmg/ddaa166

Kuht, H. J., Maconachie, G. D. E., Han, J., Kessel, L., van Genderen, M. M., McLean, R. J., Hisaund, M., Tu, Z., … Thomas, M. G. (2022). Genotypic and Phenotypic Spectrum of Foveal Hypoplasia: A Multicenter Study. *Ophthalmology*, *129*(6), 708–718. https://doi.org/10.1016/j.ophtha.2022.02.010

Kvon, E. Z., Zhu, Y., Kelman, G., Novak, C. S., Plajzer-Frick, I., Kato, M., Garvin, T. H., Pham, Q., … Pennacchio, L. A. (2020). Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. *Cell*, *180*(6), 1262-1271.e15. https://doi.org/10.1016/j.cell.2020.02.031

Lenassi, E., Carvalho, A., Thormann, A., Fletcher, T., Hardcastle, C., Hunt, S. E., Sergouniotis, P. I., Michaelides, M., … Ellingford, J. M. (2021). EyeG2P: an automated variant filtering approach improves efficiency of diagnostic genomic testing for inherited ophthalmic disorders. *MedRxiv*, 2021.07.23.21261017. https://www.medrxiv.org/content/10.1101/2021.07.23.21261017v1%0Ahttps://www.medrxiv.org/content/10.1101/2021.07.23.21261017v1.abstract

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., … de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, *12*(14), 1725–1735. https://doi.org/10.1093/hmg/ddg180

Li, Q., Ritter, D., Yang, N., Dong, Z., Li, H., Chuang, J. H., & Guo, S. (2010). A systematic approach to identify functional motifs within vertebrate developmental enhancers. *Developmental Biology*, *337*(2), 484–495. https://doi.org/10.1016/j.ydbio.2009.10.019

Lima Cunha, D., Owen, N., Tailor, V., Corton, M., Theodorou, M., & Moosajee, M. (2021). PAX6 missense variants in two families with isolated foveal hypoplasia and nystagmus: evidence of paternal postzygotic mosaicism. *European Journal of Human Genetics*, *29*(2), 349–355. https://doi.org/10.1038/s41431-020-00737-1

Lomonaco, V., Martoglia, R., Mandreoli, F., Anderlucci, L., Emmett, W., Bicciato, S., & Taccioli, C. (2014). UCbase 2.0: Ultraconserved sequences database (2014 update). *Database*, *2014*, 1–8. https://doi.org/10.1093/database/bau062

Long, H. K., Osterwalder, M., Welsh, I. C., Hansen, K., Davies, J. O. J., Liu, Y. E., Koska, M., Adams, A. T., … Wysocka, J. (2020). Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. *Cell Stem Cell*, *27*(5), 765-783.e14. https://doi.org/10.1016/j.stem.2020.09.001

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., … Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, *161*(5), 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004

Marchal, C., Singh, N., Batz, Z., Advani, J., Jaeger, C., Corso-Díaz, X., & Swaroop, A. (2022). High-resolution genome topology of human retina uncovers super enhancer-promoter interactions at tissue-specific and multifactorial disease loci. *Nature Communications*, *13*(1), 1–16. https://doi.org/10.1038/s41467-022-33427-1

Martin, A. R., Williams, E., Foulger, R. E., Leigh, S., Daugherty, L. C., Niblock, O., Leong, I. U. S., Smith, K. R., … McDonagh, E. M. (2019). PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nature Genetics*, *51*(11), 1560–1565. https://doi.org/10.1038/s41588-019-0528-2

Martínez, F., Monfort, S., Rosellá, M., Oltra, S., Blesa, D., Quiroga, R., Mayo, S., & Orellana, C. (2010). Enrichment of ultraconserved elements among genomic imbalances causing mental delay and

713 congenital anomalies. *BMC Medical Genomics*, *3*. https://doi.org/10.1186/1755-8794-3-54

714 McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., &
715 Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature*
716 *Biotechnology*, *28*(5), 495–501. https://doi.org/10.1038/nbt.1630

717 Morris, A. C., & Fadool, J. M. (2005). Studying rod photoreceptor development in zebrafish.
718 *Physiology and Behavior*, *86*(3), 306–313. https://doi.org/10.1016/j.physbeh.2005.08.020

719 Pachano, T., Haro, E., & Rada-Iglesias, A. (2022). Enhancer-gene specificity in development and
720 disease. *Development (Cambridge, England)*, *149*(11). https://doi.org/10.1242/dev.186536

721 Pedersen, H. R., Baraas, R. C., Landsend, E. C. S., Utheim, O. A., Utheim, T. P., Gilson, S. J., & Neitz, M.
722 (2020). PAX6 genotypic and retinal phenotypic characterization in congenital aniridia.
723 *Investigative Ophthalmology and Visual Science*, *61*(5). https://doi.org/10.1167/IOVS.61.5.14

724 Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky,
725 S., Dubchak, I., … Rubin, E. M. (2006). In vivo enhancer analysis of human conserved non-coding
726 sequences. *Nature*, *444*(7118), 499–502. https://doi.org/10.1038/nature05295

727 Pituello, F., Medevielle, F., Foulquier, F., & Duprat, A. M. (1999). Activation of Pax6 depends on
728 somitogenesis in the chick embryo cervical spinal cord. *Development*, *126*(3), 587–596.
729 https://doi.org/10.1242/dev.126.3.587

730 Plaisancié, J., Tarilonte, M., Ramos, P., Jeanton-Scaramouche, C., Gaston, V., Dollfus, H., Aguilera, D.,
731 Kaplan, J., … Corton, M. (2018). Implication of non-coding PAX6 mutations in aniridia. *Human*
732 *Genetics*, *137*(10), 831–846. https://doi.org/10.1007/s00439-018-1940-x

733 Polychronopoulos, D., King, J. W. D., Nash, A. J., Tan, G., & Lenhard, B. (2017). Conserved non-coding
734 elements: Developmental gene regulation meets genome organization. *Nucleic Acids Research*,
735 *45*(22), 12611–12624. https://doi.org/10.1093/nar/gkx1074

736 Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic
737 features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

738 Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique
739 chromatin signature uncovers early developmental enhancers in humans. *Nature*, *470*(7333),
740 279–285. https://doi.org/10.1038/nature09692

741 Rapaport, D. H., Wong, L. L., Wood, E. D., Yasumura, D., & Lavail, M. M. (2004). Timing and
742 topography of cell genesis in the rat retina. *Journal of Comparative Neurology*, *474*(2), 304–324.
743 https://doi.org/10.1002/cne.20134

744 Raychaudhuri, S., Iartchouk, O., Chin, K., Tan, P. L., Tai, A. K., Ripke, S., Gowrisankar, S., Vemuri, S., …
745 Seddon, J. M. (2011). A rare penetrant mutation in CFH confers high risk of age-related macular
746 degeneration. *Nature Genetics*, *43*(12), 1232–1236. https://doi.org/10.1038/ng.976

747 Rogers, M. F., Shihab, H. A., Mort, M., Cooper, D. N., Gaunt, T. R., & Campbell, C. (2018). FATHMM-
748 XF: Accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*,
749 *34*(3), 511–513. https://doi.org/10.1093/bioinformatics/btx536

750 Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S.
751 L., Mellor, J., & Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3.
752 *Nature*, *419*(6905), 407–411. https://doi.org/10.1038/nature01080

753 Schmitt, E. A., & Dowling, J. E. (1999). Early retinal development in the Zebrafish, Danio rerio: Light
754 and electron microscopic analyses. *Journal of Comparative Neurology*, *404*(4), 515–536.
755 https://doi.org/10.1002/(SICI)1096-9861(19990222)404:4<515::AID-CNE8>3.0.CO;2-A

Schwarz, J. M., Hombach, D., Köhler, S., Cooper, D. N., Schuelke, M., & Seelow, D. (2019). RegulationSpotter: Annotation and interpretation of extratranscriptic DNA variants. *Nucleic Acids Research*, *47*(W1), W106–W113. https://doi.org/10.1093/nar/gkz327

Shi, F.-Y., Wang, Y., Huang, D., Liang, Y., Liang, N., Chen, X.-W., & Gao, G. (2021). Computational Assessment of the Expression-modulating Potential for Noncoding Variants. *Genomics, Proteomics & Bioinformatics*, *86*. https://doi.org/10.1016/j.gpb.2021.10.003

Small, K. W., DeLuca, A. P., Whitmore, S. S., Rosenberg, T., Silva-Garcia, R., Udar, N., Puech, B., Garcia, C. A., … Stone, E. M. (2016). North Carolina Macular Dystrophy Is Caused by Dysregulation of the Retinal Transcription Factor PRDM13. *Ophthalmology*, *123*(1), 9–18. https://doi.org/10.1016/j.ophtha.2015.10.006

Smedley, D., Smith, K. R., Martin, A., Thomas, E. A., McDonagh, E. M., Cipriani, V., Ellingford, J. M., Arno, G., … Caulfield, M. (2021). 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. *New England Journal of Medicine*, *385*(20), 1868–1880. https://doi.org/10.1056/nejmoa2035790

Snetkova, V., Pennacchio, L. A., Visel, A., & Dickel, D. E. (2021). Perfect and imperfect views of ultraconserved sequences. *Nature Reviews Genetics*, *0123456789*. https://doi.org/10.1038/s41576-021-00424-x

Snetkova, V., Ypsilanti, A. R., Akiyama, J. A., Mannion, B. J., Plajzer-Frick, I., Novak, C. S., Harrington, A. N., Pham, Q. T., … Dickel, D. E. (2021). Ultraconserved enhancer function does not require perfect sequence conservation. *Nature Genetics*, *53*(4), 521–528. https://doi.org/10.1038/s41588-021-00812-3

Spielmann, M., & Mundlos, S. (2016). Looking beyond the genes: The role of non-coding variants in human disease. *Human Molecular Genetics*, *25*(R2), R157–R165. https://doi.org/10.1093/hmg/ddw205

Taylor, R. L., Poulter, J. A., Downes, S. M., McKibbin, M., Khan, K., Inglehearn, C. F., Webster, A. R., Hardcastle, A. J., … van Heyningen, V. (2019). Loss-of-Function Mutations in the CFH Gene Affecting Alternatively Encoded Factor H-like 1 Protein Cause Dominant Early-Onset Macular Drusen. *Ophthalmology*, *126*(10), 1410–1421. https://doi.org/10.1016/j.ophtha.2019.03.013

Thomas, E. D., Timms, A. E., Giles, S., Harkins-Perry, S., Lyu, P., Hoang, T., Qian, J., Jackson, V. E., … Cherry, T. J. (2022). Cell-specific cis-regulatory elements and mechanisms of non-coding genetic disease in human retina and retinal organoids. *Developmental Cell*, *57*(6), 820-836.e6. https://doi.org/10.1016/j.devcel.2022.02.018

Thormann, A., Halachev, M., McLaren, W., Moore, D. J., Svinti, V., Campbell, A., Kerr, S. M., Tischkowitz, M., … FitzPatrick, D. R. (2019). Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-10016-3

Turner, D. L., & Cepko, C. L. (1988). A common progenitor for neurons and glia persists in rat retina late in development. *Nature*, *328*(6126), 131–136. https://doi.org/10.1038/328131a0

Turner, D. L., Snyder, E. Y., & Cepko, C. L. (1990). Lineage-independent determination of cell type in the embryonic mouse retina. *Neuron*, *4*(6), 833–845. https://doi.org/10.1016/0896-6273(90)90136-4

Van de Sompele, S., Small, K. W., Cicekdal, M. B., Soriano, V. L., D'haene, E., Shaya, F. S., Agemy, S., Van der Snickt, T., … De Baere, E. (2022). Multi-omics approach dissects cis-regulatory mechanisms underlying North Carolina macular dystrophy, a retinal enhanceropathy. *American*

Journal of Human Genetics, *109*(11), 2029–2048. https://doi.org/10.1016/j.ajhg.2022.09.013

Visel, A., Minovitsky, S., Dubchak, I., & Pennacchio, L. A. (2007). VISTA Enhancer Browser - A database of tissue-specific human enhancers. *Nucleic Acids Research*, *35*(SUPPL. 1), 88–92. https://doi.org/10.1093/nar/gkl822

Watanabe, S., Sanuki, R., Sugita, Y., Imai, W., Yamazaki, R., Kozuka, T., Ohsuga, M., & Furukawa, T. (2015). Prdm13 regulates subtype specification of retinal amacrine interneurons and modulates visual sensitivity. *Journal of Neuroscience*, *35*(20), 8004–8020. https://doi.org/10.1523/JNEUROSCI.0089-15.2015

Whittaker, D. E., Oleari, R., Gregory, L. C., Le Quesne-Stabej, P., Williams, H. J., Torpiano, J. G., Formosa, N., Cachia, M. J., … Dattani, M. T. (2021). A recessive PRDM13 mutation results in congenital hypogonadotropic hypogonadism and cerebellar hypoplasia. *Journal of Clinical Investigation*, *131*(24). https://doi.org/10.1172/JCI141587

Wieczorek, D., Pawlik, B., Li, Y., Akarsu, N. A., Caliebe, A., May, K. J. W., Schweiger, B., Vargas, F. R., … Wollnik, B. (2010). A specific mutation in the distant sonic hedgehog (SHH) cis-regulator (ZRS) causes Werner Mesomelic Syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. *Human Mutation*, *31*(1), 81–89. https://doi.org/10.1002/humu.21142

Woolfe, A., Goode, D. K., Cooke, J., Callaway, H., Smith, S., Snell, P., McEwen, G. K., & Elgar, G. (2007). CONDOR: A database resource of developmentally associated conserved non-coding elements. *BMC Developmental Biology*, *7*, 1–11. https://doi.org/10.1186/1471-213X-7-100

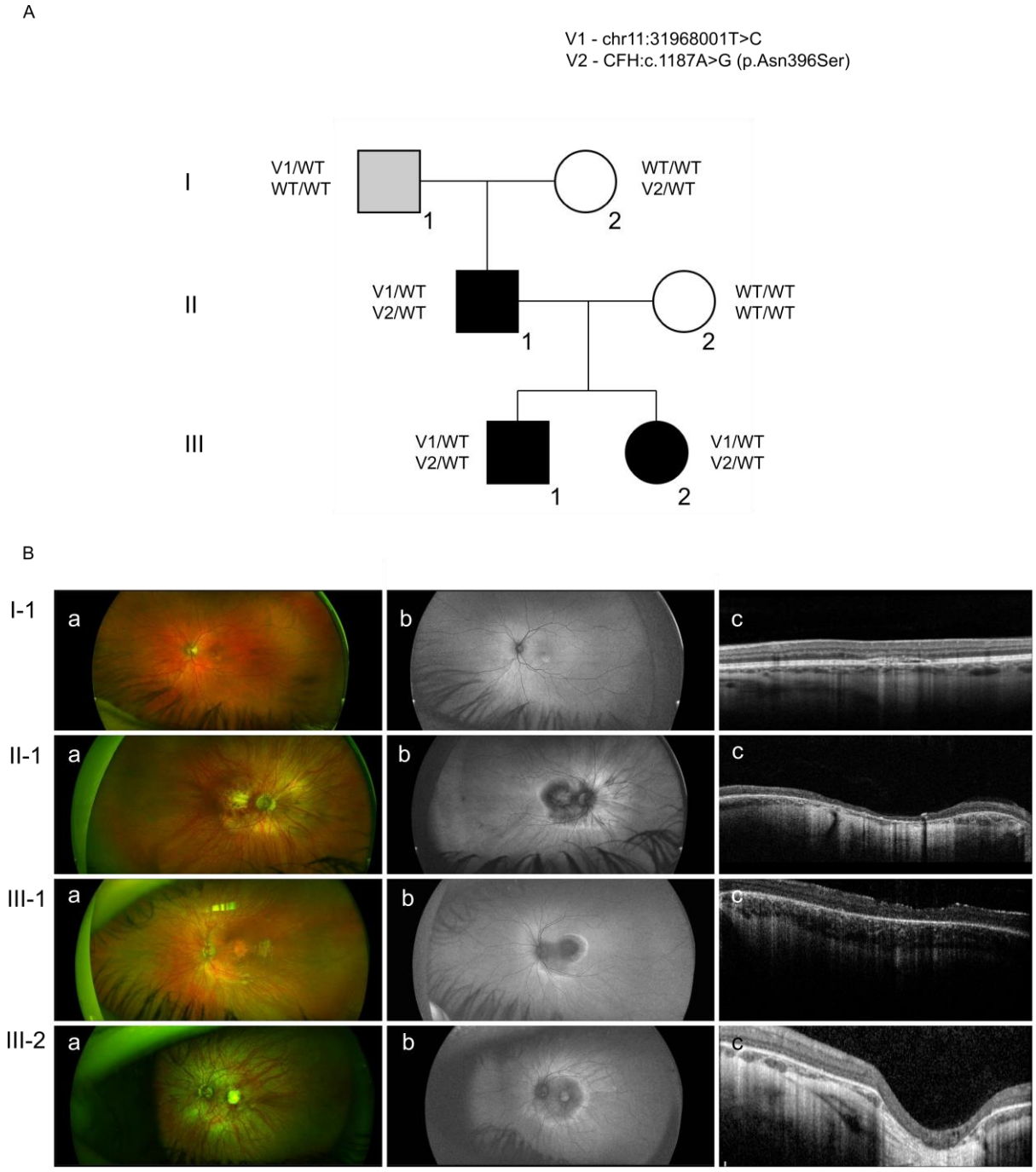
Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., … Ma'ayan, A. (2021). Gene Set Knowledge Discovery with Enrichr. *Current Protocols*, *1*(3), 1–51. https://doi.org/10.1002/cpz1.90

Yokoi, T., Nishina, S., Fukami, M., Ogata, T., Hosono, K., Hotta, Y., & Azuma, N. (2016). Genotype–phenotype correlation of PAX6 gene mutations in aniridia. *Human Genome Variation*, *3*(1), 1–5. https://doi.org/10.1038/HGV.2015.52

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., … Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9). https://doi.org/10.1186/gb-2008-9-9-r137

Zhou, J., & Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. *Nature Methods*, *12*(10), 931–934. https://doi.org/10.1038/nmeth.3547

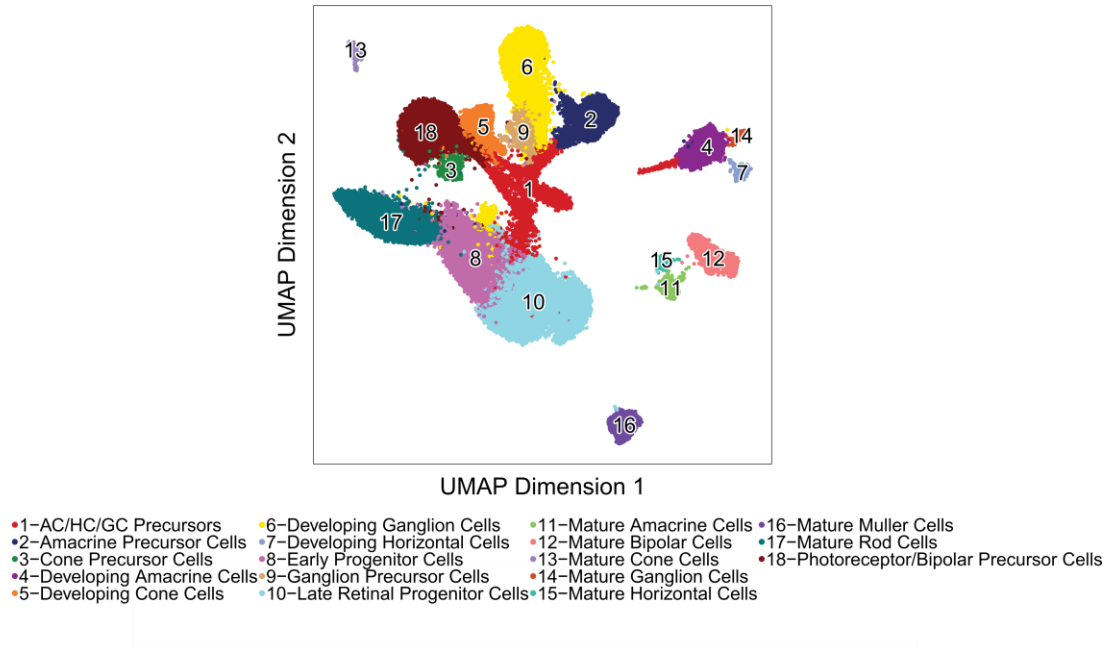
833 **SUPPLEMENTARY MATERIAL**



834

835 **Figure S1. *PAX6*-associated variant exclusively found in a family presenting foveal anomalies. A)**
836 Pedigree of the family, indicating segregation of the *PAX6*-associated UCNE variant (V1) and of the *CFH*
837 variant (V2). **B)** Ophthalmological assessment (fundus examination, FAF, and OCT) revealed tessellated
838 fundi with atrophic areas at the macula involving the fovea in II-1, III-1 and III-2, while I-1 displayed
839 only an area of pallor inferior to the left fovea and corresponding hyper-autofluorescence (see Table
840 S7).

A



B

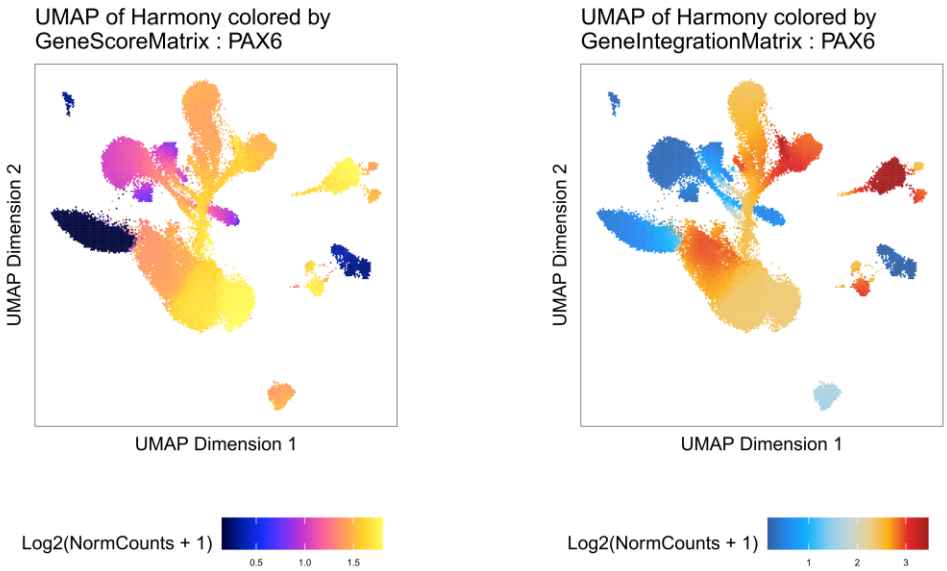


Figure S2. Single-cell characterization of the *PAX6* gene. A) UMAP of the analyzed single cell dataset (Thomas et al., 2022).**B)** Feature plots (scATAC-seq and scRNA-seq) for the *PAX6* locus.

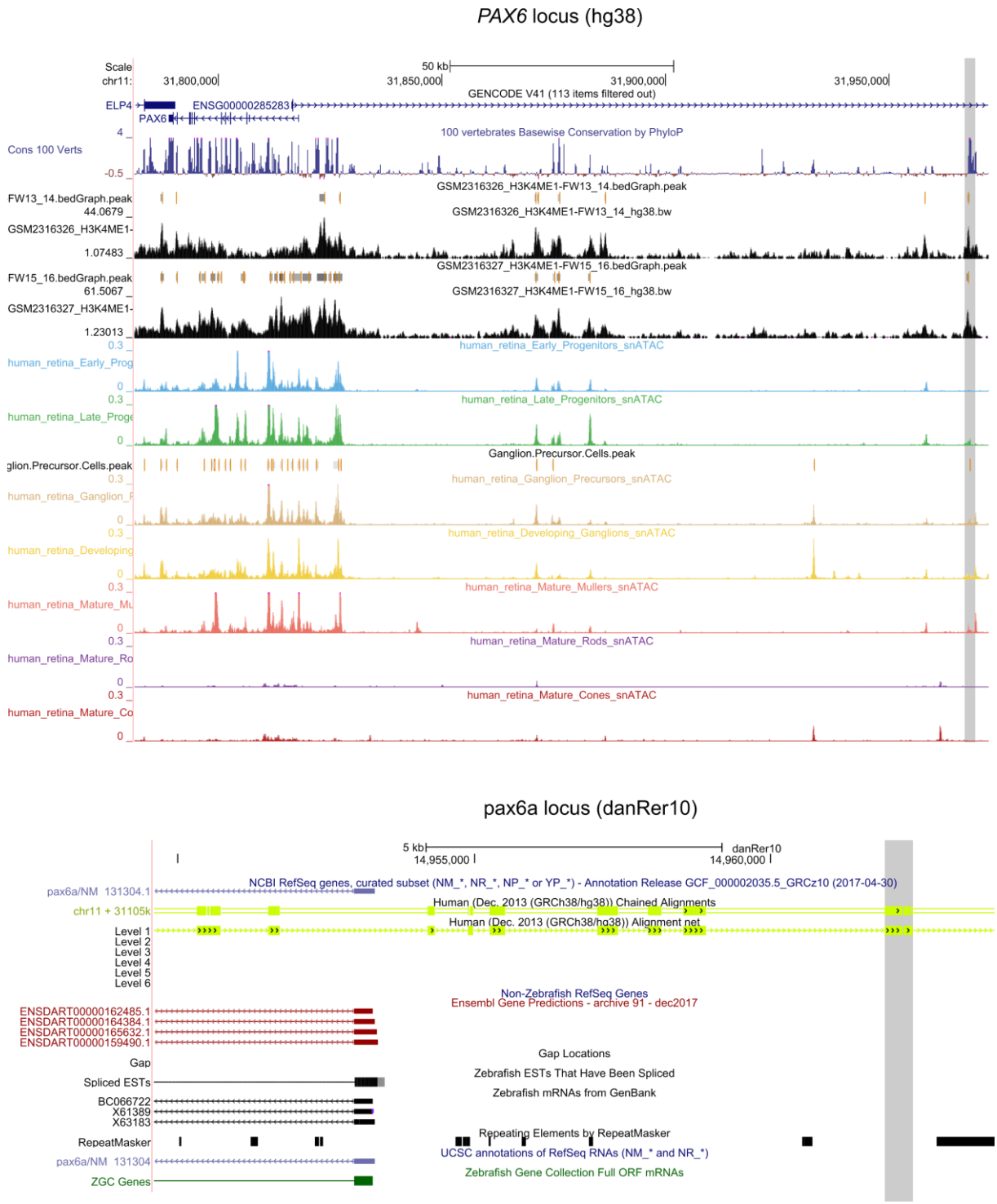
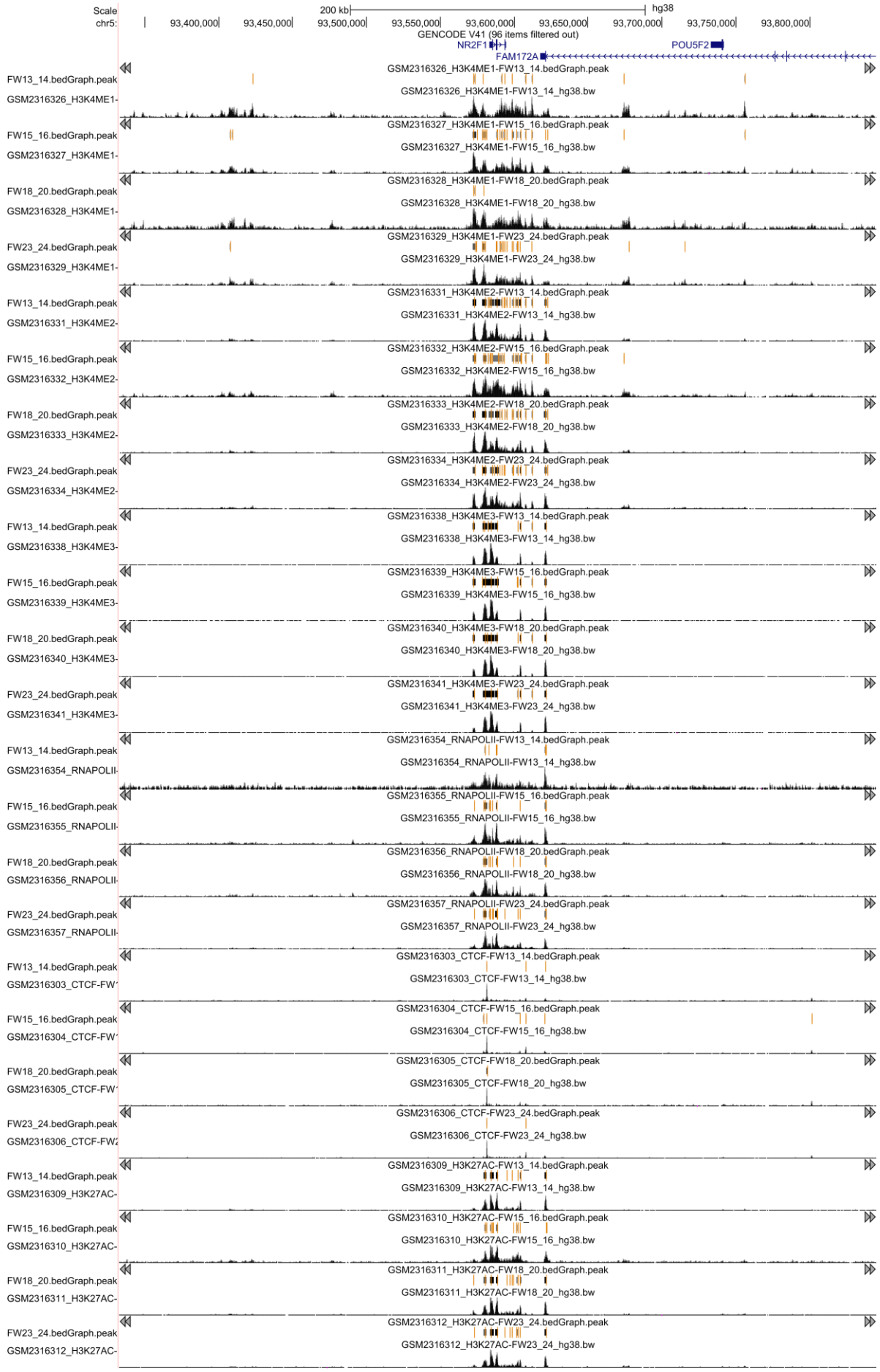


Figure S3. Syntenic region in zebrafish of the human *PAX6* locus. Location of the characterized UCNE (in gray) linked to *PAX6*, found exclusively in the *pax6a* locus of zebrafish at a shorter distance (≈15kb) when compared to the human *PAX6* locus (≈150kb) with respect to the *PAX6* transcription start site.


Scale
chr5:
200 kb
hg38
93,400,000
93,450,000
93,500,000
93,550,000
93,600,000
93,650,000
93,700,000
93,750,000
93,800,000
GENCODE V41 (96 items filtered out)
NR2F1
POU5F2
FAM172A
GSM2316326_H3K4ME1-FW13_14.bedGraph.peak
GSM2316326_H3K4ME1-FW13_14_hg38.bw
GSM2316327_H3K4ME1-FW15_16.bedGraph.peak
GSM2316327_H3K4ME1-FW15_16_hg38.bw
GSM2316328_H3K4ME1-FW18_20.bedGraph.peak
GSM2316328_H3K4ME1-FW18_20_hg38.bw
GSM2316329_H3K4ME1-FW23_24.bedGraph.peak
GSM2316329_H3K4ME1-FW23_24_hg38.bw
GSM2316331_H3K4ME2-FW13_14.bedGraph.peak
GSM2316331_H3K4ME2-FW13_14_hg38.bw
GSM2316332_H3K4ME2-FW15_16.bedGraph.peak
GSM2316332_H3K4ME2-FW15_16_hg38.bw
GSM2316333_H3K4ME2-FW18_20.bedGraph.peak
GSM2316333_H3K4ME2-FW18_20_hg38.bw
GSM2316334_H3K4ME2-FW23_24.bedGraph.peak
GSM2316334_H3K4ME2-FW23_24_hg38.bw
GSM2316338_H3K4ME3-FW13_14.bedGraph.peak
GSM2316338_H3K4ME3-FW13_14_hg38.bw
GSM2316339_H3K4ME3-FW15_16.bedGraph.peak
GSM2316339_H3K4ME3-FW15_16_hg38.bw
GSM2316340_H3K4ME3-FW18_20.bedGraph.peak
GSM2316340_H3K4ME3-FW18_20_hg38.bw
GSM2316341_H3K4ME3-FW23_24.bedGraph.peak
GSM2316341_H3K4ME3-FW23_24_hg38.bw
GSM2316354_RNAPOLII-FW13_14.bedGraph.peak
GSM2316354_RNAPOLII-FW13_14_hg38.bw
GSM2316355_RNAPOLII-FW15_16.bedGraph.peak
GSM2316355_RNAPOLII-FW15_16_hg38.bw
GSM2316356_RNAPOLII-FW18_20.bedGraph.peak
GSM2316356_RNAPOLII-FW18_20_hg38.bw
GSM2316357_RNAPOLII-FW23_24.bedGraph.peak
GSM2316357_RNAPOLII-FW23_24_hg38.bw
GSM2316303_CTCF-FW13_14.bedGraph.peak
GSM2316303_CTCF-FW13_14_hg38.bw
GSM2316304_CTCF-FW15_16.bedGraph.peak
GSM2316304_CTCF-FW15_16_hg38.bw
GSM2316305_CTCF-FW18_20.bedGraph.peak
GSM2316305_CTCF-FW18_20_hg38.bw
GSM2316306_CTCF-FW23_24.bedGraph.peak
GSM2316306_CTCF-FW23_24_hg38.bw
GSM2316309_H3K27AC-FW13_14.bedGraph.peak
GSM2316309_H3K27AC-FW13_14_hg38.bw
GSM2316310_H3K27AC-FW15_16.bedGraph.peak
GSM2316310_H3K27AC-FW15_16_hg38.bw
GSM2316311_H3K27AC-FW18_20.bedGraph.peak
GSM2316311_H3K27AC-FW18_20_hg38.bw
GSM2316312_H3K27AC-FW23_24.bedGraph.peak
GSM2316312_H3K27AC-FW23_24_hg38.bw

848

Figure S4. Illustration of the peak identification of scATAC-seq and ChIP-seq data. A) Peak identification within the *NR2F1* locus for the following marks: H3K4me1, H3K4me2, H3K4me3, CTCF, RNAPolII and H3K27ac across the different time points.

Table S1. Characterization of UCNEs based on: UCNE ID – UCNE name extracted from the original library elaborated by Dimitrieva et al., 2013; **g.coordinates (hg38) –** genomic coordinate of the UCNE; **Target gene –** gene association extracted from GREAT analysis; **Bulk RNA-seq (target gene mean expression) –** expression value (TPM) obtained from the RNA-seq based on samples related to the human retinal development from Hoshino et al., 2018; **Bulk RNA-seq (target gene expression rank) –** expression rank obtained from the RNA-seq based on samples related to the human retinal development from Hoshino et al., 2018; **Bulk RNA-seq (target gene maximum expression) –** maximum values of gene expression (TPM) obtained from the RNA-seq based on samples related to the human retinal development from Hoshino et al., 2018; **Bulk RNA-seq (stage of maximum target gene expression)** – stage where the maximum levels of gene expression where observed, obtained from the RNA-seq based on samples related to the human retinal development from Hoshino et al., 2018; **scRNA-seq target gene expression –** characterization of the cell identity where the expression pattern of the target gene is expected, to perform correlation between enhancer-target gene, obtained from the scRNA-seq based on samples related to human retinas from Thomas et al., 2022; **DNase-seq (stage)** – specific stages where the open chromatin context was identified, obtained from the DNase-seq experiments related to the retinal development from ENCODE; **scATAC-seq** – cell clusters where the peak identification was retrieved; **Retinal TAD Support** – qualitative estimation of the association UCNE-target gene; **ChIP-seq** - epigenomic marks, CTCFs and PolII peaks observed within the genomic context of the characterized UCNE; **VISTA (Element ID)** – assessment of the inclusion of the UCNE element within the VISTA enhancer browser and its unique ID; **VISTA (Assay result)** – reporter assay result from the VISTA enhancer browser; **VISTA (Expression pattern) –** includes the tissues where the expression pattern for the reporter assay was observed.

Table S2. Characterization of UCNEs based on chromatin accessibility. UCNE_id – Ultraconserved non-coding element name; **Stage** – retinal development stages where a peak from DNase is retrieved; **#Stages** – number of stages for the specific element where it was retrieved; **scATAC-seq** – cell cluster where the peak identification was retrieved; **#Peaks** – total number of cell clusters where the peak identification was retrieved.

Table S3. Characterization of UCNEs based on epigenomic marks.

Table S4. Disease association of UCNE target genes. Gene; UCNE_id; g.coordinates; #OMIM disease name; confidence category; allelic requirement; mutation consequence; phenotypes; organ specificity list; PMIDs.

Table S5. Overview of the eye disease sub-cohort of 100,000 Genomes Project (Genomics England). Normalized Disease Group; Normalized Disease Sub Group; Normalized Specific Disease; Participant Count.

Table S6. Variants retrieved within the UCNEs that are linked to an eye or retinal disease phenotype. It includes the **gene**, the **UCNE_id** or name and the retrieved **variant** (hg38).

Table S7. Phenotypic description of the studied family segregating autosomal dominant foveal abnormalities. P**articipant information** (ID, origin and sex), the **molecular findings** (carriers of the V1 (chr11:31968001T>C) and/or V2 (CFH, c.1187A>G (p.Asn396Ser)), **diagnosis and age of onset** (Clinical diagnosis, Age at diagnosis, Age of onset visual field loss, Age of onset BCVA loss) and **clinical findings** (Age at examination, Slit-lamp examination, BCVA OD/OS (LogMAR), Fundus Examination, Visual field, FAF, OCT, Other findings).

Table S8. Analysis of the TFBS motif disruption potentially exerted by the chr11:31968001T>C variant (qBiC-PRED) and retrieved variants within genes associated with macular developmental defects and

896 **foveal hypoplasia (*IRX1*, *PRMD13*, *SLC38A8*, *GPR143*, *FRMD7* and *AHR*).** All coordinates are in the
897 hg38 assembly.

898 **Table S9. Set of primers used in this study for cloning and segregation.**

899 **Table S10. Overview of the transgenic enhancer assays in zebrafish for the *PAX6*-associated UCNE.**