Identification of novel coloboma candidate genes through conserved gene expression analyses across four vertebrate species

Violeta Trejo-Reveles[1#] & Nicholas Owen[2#], Brian Ho Ching Chan[1], Maria Toms[2], Jeffrey J Schoenebeck[1], Mariya Moosajee[2,3,4]*, and Joe Rainger[1]*

On behalf of Genomics England Research Consortium

1 Roslin Institute, R(D)SVS, University of Edinburgh, Easter Bush Campus, EH25 9RG, UK.
2 Development, Ageing and Disease, UCL Institute of Ophthalmology, London, UK.
3 Department of Genetics, Moorfields Eye Hospital NHS Foundation Trust, London, UK.
4 Ocular Genomics and Therapeutics, The Francis Crick Institute, London, UK.

# These authors contributed equally
* Correspondence: m.mXXXX@ucl.ac.uk and joeXXX@roslin.ed.ac.uk

Abstract: Ocular coloboma (OC) is a failure of complete optic fissure closure during embryonic development and presents as a tissue defect along the proximal distal axis of the ventral eye. It is classed as part of the clinical spectrum of structural eye malformations with microphthalmia and anophthalmia, collectively abbreviated to MAC. Despite deliberate attempts to identify causative variants in MAC, many patients remain without a genetic diagnosis. To reveal potential candidate genes, we utilised transcriptomes experimentally generated from embryonic eye tissues derived from human, mouse, zebrafish, and chicken at stages coincident with optic fissure closure. Our in-silico analyses found 10 genes with optic fissure specific enriched expression: *ALDH1A3*, *BMPR1B*, *EMX2*, *EPHB3*, *NID1*, *NTN1*, *PAX2*, *SMOC1*, *TENM3*, and *VAX1*. In situ hybridization revealed that all 10 genes were broadly expressed ventrally in the developing eye, but that only *PAX2* and *NTN1* were expressed in cells at the edges of the optic fissure margin. Of these conserved optic fissure genes, *EMX2*, *NID1*, and *EPHB3* have not previously been associated with human MAC cases. Targeted genetic manipulation in zebrafish embryos using CRISPR/Cas9 caused the developmental MAC phenotype for *emx2* and *ephb3*. We scrutined available whole genome sequencing datasets from MAC patients and identified a range of variants with plausible causality. In combination our data suggest that expression of genes involved in ventral eye development are conserved across a range of vertebrate species, and that *EMX2*, *NID1*, and *EPHB3* are candidate loci that should be adopted into clinical diagnostic screens for patients with structural eye malformations.

Keywords: Coloboma; Microphthalmia; Anophthalmia; Optic fissure closure; Eye development.

1. Introduction

Congenital structural eye malformations can cause severe visual impairment and reduced quality of life, with a prevalence ranging up to 19/10,000 newborns[1–3]. However, the genetic and environmental causes remain elusive, leading to lack of clinical management or treatment options. Within these disorders, ocular coloboma (OC) is the most common and considered part of the clinical phenotypic spectrum with anophthalmia (completely missing eye) and microphthalmia (small, underdeveloped eye), together referred to as MAC. The developmental insults that cause MAC are mostly genetic but may also be influenced by environmental factors[4–6]. They occur early in gestation and typically affect the morphogenesis and structural development of the optic vesicle (anophthalmia) or eye cup (microphthalmia and coloboma)[7].

During normal eye development, the anterior neural plate develops a transcriptionally distinct eye field which evaginates bilaterally from the surrounding neuroepithelium as the neural plate folds [8]. These outgrowing neural pouches are referred to as the optic vesicles and each one eventually reaches the overlying surface ectoderm, triggering a mutual invagination of the two tissue to create a bilayered optic cup and lens vesicle [9,10]. The ventral retina remains undifferentiated during optic cup stages, largely through a unique transcriptional signature which enables this region of the eye to undergo morphogenic changes that distinguish it from the dorsal retina [11–15]. The morphogenesis of the ventral eye creates a gap along the proximal-distal axes of each optic cup– the optic fissure. The edges of the optic fissure come closer together as the optic cups continue to grow, displacing neural-crest derived periocular mesenchyme (POM) and vascular endothelia, until they eventually come into direct contact and fuse together to complete the circumferential continuity of the eye. The genetic and physical processes that control the fusion of these opposing epithelial sheets is poorly understood but requires the decoupling and reorientation of cells at the leading edges of the epithelia, similar to a partial epithelial to mesenchymal transition, and coincident with remodeling of overlying basement membrane structures[16–18].

Clinically, OC manifests as an inferonasal gap that could affect one or more tissues including the iris, ciliary body, neural retina, retinal pigmented epithelium (RPE), choroid with/without involvement of the optic disc and macula. It accounts for 10% of childhood blindness and has a variable prevalence, but may affect up to 11.9 per 100,000 births [2,19]. It can manifest with wider systemic features as part of a syndrome (syndromic OC), be associated with other eye defects (complex OC) or occur in isolation (isolated OC, and can affect only one (unilateral) or both (bilateral) eyes. The use of next generation sequencing including targeted gene panels and whole genome sequencing in MAC patients and their families, has led to an increase in genetic diagnosis rates up to 33%[20], with MAC panels now including up

to 86 different genes [21]. However, very few OC loci show recurrence among unrelated families, and recent studies suggest over 80% of cases do not have a genetic diagnosis [7,22,23].

Animal models of optic fissure closure (OFC) have recently been powerful in revealing the transcriptional landscapes associated with ventral eye development and fissure fusion, successfully identifying novel MAC candidate genes[17,24,25]. Using these datasets, we previously generated a pipeline to enable the identification of new coloboma candidate genes based on evolutionary conserved gene expression in the developing optic fissure[26]. This method leveraged microarray data from mouse[24] and RNAseq data from zebrafish[25] to reveal 4 novel coloboma candidate genes and supported the identification of human pathogenic variants therein[26]. Here, we applied a similar strategy, this time exclusively using RNAseq data but also increasing the analysis to four vertebrate species, representing mammals (human, mouse), fish (zebrafish) and avians (chicken). Our analyses revealed three novel candidate genes for roles in eye development with supporting evidence from in vivo work that these are essential for OFC. Based on this, we propose NID1, EMX2 and EPHB3 are now included in targeted gene panels for MAC-patients and families.

2. Materials and Methods

2.1 Chicken eye mRNA sequencing

RNAseq was performed on dissected tissue collected from dorsal and fissure regions of the chicken eye at both fusing (HH St30) and fully fused (HH St34) stages of optic fissure closure (as described in Hardy et al., 2019 [17] and staged according to Hamburger and Hamilton criteria [27]). Three replicates for each sample stage and region were prepared and each sample replicate contained pooled tissue from a minimum of 10 different embryos. Pooled samples were processed for RNA extraction on the day of dissection. Total RNA was extracted using Trizol reagent (Ambion) and treated with Turbo DNaseI (Thermo #AM2238). Libraries were prepared from 500ng of the total-RNA sample using the NEBNEXT Ultra II Directional RNA Library Prep kit (NEB #7760) with Poly-A mRNA magnetic isolation (NEB #E7490) according to the manufacturer's protocol. mRNA libraries were assessed on the Agilent 2100 Electrophoresis Bioanalyser with the DNA High Sensitivity Kit (#5067-4627) for fragment size distribution, and then quantified using the Qubit 2.0 Fluorometer and the Qubit dsDNA HS assay kit. Libraries were run on the Nextseq 550 using a high output flow cell and generated a cluster density of 227K/mm2 with 86% of clusters passing quality filters (PF). This produced 83.1 Gb of data with 89.2% ≥Q30. Coverage for each of the mRNA libraries generated ≥34M PE (2x 75bp) reads (Min: 34M, Max: 43M, Mean: 40M).

95

2.2 Transcript quantification and differential gene expression analysis 96

Sequencing quality was evaluated using FASTQC (version 0.11.4). Sequencing adaptors were trimmed using Cutadapt[28] (version 1.16) and Trimmgalore (version 0.4.1). Trimmed sequences were aligned to the chicken genome (galgal6) using STAR aligner to generate bam files, which were then coordinate sorted and indexed using Samtools[29] (version 1.6). A count matrix of transcripts annotated in Ensembl (version 103) was subsequently generated from bam files using featureCounts part of the Subread package [30,31] (version 2.0.1). Differential gene expression and principal component analysis was performed using DESeq2 [32] in R (version 4.1.0). Multiple test comparisons were corrected using Benjamini and Hochberg correction to obtain the adjusted p -value[33]. Genes with a log2 fold change ≥ 1 and an adjusted p-value ≤0.05 were considered as upregulated or fissure enriched differentially expressed genes (DEGs), while those with a log2 fold change of ≤ -1 and an adjusted P-value of ≤ 0.05 were considered downregulated or dorsally enriched DEGs in the analysis (i.e. Fold change ≥ 2 or ≤ 0.5, respectively). Volcano plots of DEGs were drawn using ggplot2 package in R. Enrichment analysis and functional annotation was performed in DAVID (https://david.ncifcrf.gov/home.jsp) using *Gallus_gallus* as a background list and ENSEMBL gene IDs as inputs. Lists of GO terms were ranked and plotted (-log10(p val)) with cut off values p < 0.001.

2.3 In silico cross species differential gene expression analysis

Sequencing data from Hardy, et.al., Patel et al., and Richardson, et.al., (Refences 13-15) were downloaded using the SRA NCBI Toolkit. All reads were processed as follows: Reads were trimmed using Trimmomatic. Pairwise comparison between optic fissure and dorsal or whole eye (in chick pre-fusion samples) tissue were carried out using the Wald Test in DESeq2 using raw count data generated by featureCounts. Comparisons over time were performed using the LRT test. Ensembl gene ID identifiers were converted to their mouse orthologs using the biological DataBase network tool (https://biodbnet-abcc.ncifcrf.gov/db/dbOrtho.php). All up-regulated genes in the fissure at any stage (absolute LFC ≥ 1.0, adjusted *p* value ≤ 0.05) were selected for further analysis. Using mouse, chicken and zebrafish up-regulated DEGs we identified commonalities across species using Ensembl gene database (version 102) with custom scripts based upon the Bioconductor biomaRt package[34,35] (version 4.2). Briefly, orthologs were identified using bioDBnet biological DataBase network (https://biodbnetbcc.ncifcrf.gov/db/dbOrtho.php) using mouse Ensembl gene IDs as a reference Where orthologues were not automatically identified, manual curation of the list was carried out.

2.4 *In situ* hybridisation gene expression analyses

Chick embryos were staged according to Hamburger and Hamilton[27] and fixed overnight at 4°C in 4 % paraformaldehyde (PFA) in 1.0 M phosphate buffered saline solution (PBS). Embryos were then rinsed in PBS twice and immersed in a 10 % sucrose-PBS solution overnight. The following day eyes were removed and mounted in Neg-50 (Richard Allan Scientific) and snap frozen in iso-pentane (Thermo). Sections were cut at 20 µm on a Cryostat (Leica) onto Superfrost-plus slides (Manufacturer), air dried for 2 hours, and then stored at -80oC. Slides were then thawed for 2h at room temperature and rinsed in PBS, before performing RNAscope following the Manufacturer's protocol (ACD Bio-techne, RNAscope Multiplex Fluorescent Reagent Kit v2 User Manual) and using the target-specific probes (ACD Bio-techne) detailed below in **Table 1**. Slides were then mounted in Fluorsave (Merck #345789) with and counterstained with DAPI (ACD Bio-techne, RNAscope Multiplex Fluorescent Reagent Kit v2 #323100). Images were captured using a Zeiss LSM 880 Confocal microscope with a 20x objective and converted to JPEG file format using FIJI (ImageJ V2.1.0/1.53c) under Creative Commons Licence.

Table 1. RNAscope custom-designed and catalogue (*) probes used.

mRNA target	Name of probe	Channel	Catalogue number
TENM3	Gg-TENM3-C2	C2	1147351-C2
EMX2 (Transcript variant X1)	Gg-EMX2-C2	C2	1060951-C2
PAX2	Gg-PAX2-C2	C2	1147371-C2
EPHB3 (Transcript variant X7)	Gg-EPHB3-C2	C2	458861-C2
NID1	Gg-NID1-C1	C1	1147381-C1
ALDH1A3	Gg-ALDH1A3-C2	C2	1144151-C2
BMPR1B	Gg-BMPR1B-C1	C1	1144161-C1
VAX1	Gg-VAX1-C1	C1	1144181-C1
SMOC1 (Transcript variant X4)	Gg-SMOC1-C2	C2	593601-C2
NTN1	Gg-NTN1-C2	C2	497491-C2*

2.5 Data access at Genomics England, 100,000 genomes project

The 100,000 Genomes Project [36] data has been made available through the secure Genomics England (GEL) research embassy (RE), supported on a high-performance cluster, following information governance and security training, with membership to the Hearing and Sight Genomics England Clinical Interpretation Partnership (GeCIP) in the rare disease programme [37]. Variant call files (VCF) generated through the Illumina Starling pipeline, were passed through quality control filters as previously described (Owen et al 2022). Participant phenotype data was present using Human Phenotype Ontology (HPO) terms[38] and accessed via a LabKey data management portal. HPO terms for anophthalmia,

microphthalmia, coloboma as well as daughter terms were queried using the R LabKey package. The resulting MAC cohort consisted of 346 probands.

2.6 Zebrafish husbandry

Zebrafish (wild-type AB) were bred and maintained according to local UCL and UK Home Office regulations for the care and use of laboratory animals under the Animals Scientific Procedures Act at The Francis Crick Institute aquatics unit. UCL Animal Welfare and Ethical Review Body approved all procedures for experimental protocols, in addition to the UK Home Office (License no. PPL PC916FDE7). All approved standard protocols followed the guidelines of the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research Ethics.

2.7 CRISPR/Cas9 mutagenesis in zebrafish

Guide sequences targeting *emx2*, *ephb3a* and *ephb3b* were selected using the Integrated DNA Technologies (IDT) guide RNA design tool with Refseq release 211 of the zebrafish genome. Custom Alt-R sgRNAs and SpCas9 Nuclease were purchased from IDT. Zebrafish embryos were micro-injected at the one-cell stage with approximately 1 nl of injection solution consisting of: 800 ng/ul Cas9 nuclease, 300 mM KCL, 100 ng/ul of sgRNA (*emx2* only or both *ephb3a* and *ephb3b* sgRNAs). For the non-sgRNA controls, embryos were injected with all components except for sgRNA. To confirm the mutagenesis activity of the sgRNAs, DNA was extracted from 10 injected embryos at 48 hpf and PCRs were used to amplify a region of the gene that included the target site. The amplicons were assessed by Sanger sequencing to verify that mutations were induced at the CRISPR target site. The PCR primer sequences and CRISPR guide sequences used are listed in **Table 2**.

2.8 Phenotypic assessments

For assessing eye size, 3 and 5 dpf injected zebrafish embryos and un-injected wild-type siblings were anaesthetized using 0.2 mg/ml tricaine and imaged using a Zeiss Stereo Discovery V20 microscope. ImageJ (FIJI) was used to measure eye diameter from the images. For each group, the mean ± standard deviation was calculated. Data were compared using either unpaired t-tests or Mann-Whitney tests. For immunofluorescence, whole zebrafish larvae at 3 dpf were fixed in 4% PFA/PBS overnight at 4°C before washing in PBS and incubation in 30% sucrose/PBS overnight at 4°C. The samples were mounted and frozen in TissueTek O.C.T (VWR) using dry ice. 12 µm sagittal sections were cut and collected onto Superfrost PLUS slides (Thermo Fisher Scientific). After air-drying, sections were washed in PBS-0.5%

Triton-X before being blocked for 1 hour with 20% normal goat serum (Sigma-Aldrich) in PBS-0.5% Triton-X. and incubating with anti-laminin (Sigma-Aldrich L9393) diluted 1:50 in antibody solution (2% normal goat serum in PBS-0.5% Triton-X) at 4°C overnight. After washing with PBS-0.5% Triton-X, the sections were incubated with Alexa Fluor 568 nm secondary antibody (Thermo Fisher) diluted 1:500 in antibody solution for 2 hours at room temperature. Finally, the sections were washed and mounted in Prolong Diamond Antifade mountant + DAPI (Thermo Fisher Scientific). The slides were imaged using a Zeiss Invert 880 microscope.

Table 2. Oligonucleotide sequences used for zebrafish gene editing and genotyping

Gene	CRISPR guide sequence (5'-3')	Primer sequences (5'-3')
emx2	CGAGGAGCCCATACGACCAG	Forward: CACGATGTGTTGAGCTGTGC Reverse: CCTTTGCTGGCTTGCGAAAA
ephb3a	ATGACCACTTGAGCCCCATC	Forward: TTACATTCCACCTGCTTACACC Reverse: ACATAAGGATTCTCCCTCCACG
ephb3b	ACACTTGGTACGTTCGGATG	Forward: GCAGTACCTTTGCAGCGTAAC Reverse: GAAGCTCTCATCCGGAGCAA

3. Results

3.1. Transcriptomic analyses during chicken optic fissure closure

We previously used dissected optic fissure tissue from chick retinas to define the transcriptonal landscape during fusion [17]. However, this analysis did not include tissue from fused stages of OFC, and whole eye tissue was used as the non-fusing comparison rather than dorsal retina used in other species OFC transcriptomic studies [25,39]. To enable pan-specific comparisons, we generated transcriptomic data using RNAseq from optic fissure and dorsal retina samples collected from chicken eyes at fusing (HH St30) and fused (HH St34) OFC stages (**Figure 1a**). In both cases, micro-dissected tissue samples were pooled depending on condition and stage (tissue from >10 embryos per condition or stage) prior to total RNA extraction and Illumina paired end sequencing with enrichment for mRNA transcripts. Variance across the samples were assessed by principal component analysis (PCA) and these indicated that the sample replicates clustered well and that groups of samples segregated according to stage (PC1) and tissue region (PC2) (**Supplemental Figure S1**). We then performed differential gene expression (DEG) analyses within the stages to determine fissure-specific gene expression (**Figure 1b**). We found 158 unique upregulated DEGs that only showed increased expression in the fusing (HH St30) fissure, 178 that were increased specifically in fused (HH St34) fissure, and 84 genes that were enriched at both stages. Similarly, for downregulated DEGs (dorsal > fissure) there were 175 genes whose expression was reduced in the fusing fissure compared to dorsal, 83 that were reduced in the fused fissure, and 49 that were reduced at both stages. All gene lists and ontology enrichment outputs can be found in Supplemental Table 1. We then assessed those genes with enriched expression in the fissure at each stage. In accordance with previous studies [17,25,39], at fusing stages (HH.St30) we found *NTN1*, *ALDH1A3*, *VAX1*, *SMOC1*, and *PAX2* were the highest fissure-specific (Fissure > Dorsal) DEGs, whereas *ALDH1A1*, *LYPD6*, *GDF6*, *TBX3* and *TBX5* were the most dorsal-specific DEGs (Dorsal > Fissure) (Figure 1c). *In silico* ontology analysis of the upregulated gene set (Log2FC >1.0, p<0.05; n = 246 genes) in the fusing stage fissure revealed enrichment for GO terms of various developmental processes as well as those implicated in OFC, such as “ECM organisation”, “signal transduction”, “negative regulation of wound healing” and the “BMP signalling pathway”, in addition to multiple terms related to cell-cell adhesion or the regulation of cell-junctional complexes (Figure 1d). In the fused tissue (HH.St34), *NTN1*, *ALDH1A3*, *VAX1*, *SMOC1*, and *PAX2* were also identified as fissure specific DEGs, in addition to *BMPR1B*, *CHRDL1*, and *CD109* (**Figure 1e**). Dorsal-specific DEGs observed at both stages were *TBX5*, *TBX3*, *GDF6* and *LYPD6* (**Figure 1c & e**), consistent with previous transcriptomic analyses of OFC [17,39]. Ontology enrichment analysis for all upregulated DEGs (Log2FC >1.0, p<0.05; n = 262 genes) in the fused fissure (Figure 1e) revealed similar GO terms to those enriched for the fusing fissure related to cell-cell

adhesion and cell-junction assembly, for example: “cell-cell adhesion via plasma-membrane adhesion molecules” and “calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules”. Both stages also included, “inner ear morphogenesis” and “sensory perception of sound”, likely to reflect overlap in epithelial fusion processes involved in inner ear morphogenesis and OFC. The GO term “Axon guidance” was enriched among DEGs in both stages of OFC, and includes genes involved in the ephrin, netrin and semaphorin regulatory pathways. Interestingly, HH.St30 had fewer genes in this GO term classification (9 compared to 27) and, in contrast to HH.St34, did not include members of the UNC5 gene family, as previously shown [17,39]. In combination, these validated the use of these chicken transcriptome datasets for inclusion in subsequent analyses.

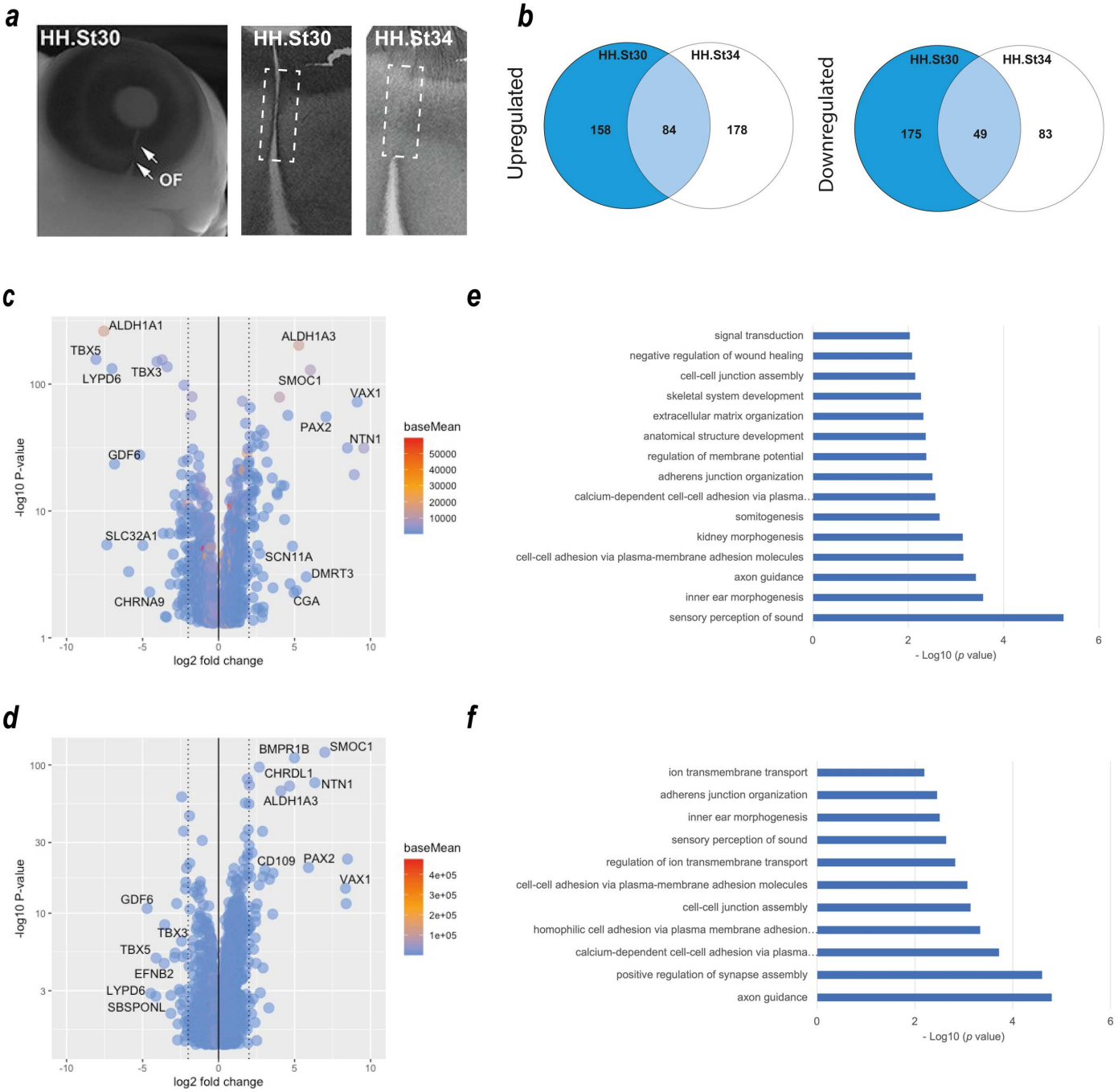

Figure 1. Transcriptome profiling of chicken optic fissure closure. (**a**) Schema for segmental dissection of chick OFM tissue at fusing (HH.St30) and fused (HH.St34) OFC stages. (**b**) Venn diagrams representing the number of DEGs that were (left) significantly upregulated: Fissure expression > Dorsal, or (right) downregulated: Dorsal expression > fissure) between the fusing HH.st30 fissure (blue circle) and fused HH.st34 fissure tissue (clear circle). The number of DEGs that are commonly upregulated/downregulated between the two stages are indicated in the overlapping region. (**c**) Volcano plot showing differentially expressed genes in the fusing stage HH.st30 OFC transcriptomes. (**d**) Volcano plot showing differentially expressed genes in the fused stage HH.st34 transcriptomes. The top 10 genes with the highest and lowest Log2 fold changes are labeled for each graph. (**e-f**) Plots showing DAVID enrichment analysis outputs for fissure specific DEGs at the two developmental stages (**e**, HH.St30; **f**, Hh.St34). Values shown are -Log10 (p value). All accompanying enrichment and gene expression analysis data is in **Supplemental Table 1**.

3.2. Cross-species in silico analysis of optic fissure transcriptomes

To identify evolutionarily conserved gene expression profiles in OFC, whole transcriptome RNAseq datasets from chick (this study and reference [17]), zebrafish[25], mouse and human[39] were compared. Analysis of variance using PCA analyses between species at each of the stages (pre-fusion, fusion, and fused) indicated that the samples clustered broadly according to tissue type (dorsal or fissure) and species (**Figure 2a-c**), although the variance for PC1 ranged from 20-38%. After alignment, differential expression analysis was performed using DESeq2 for successive pairwise condition testing (v1.18.1). Up-regulated genes were defined as those showing a log fold change (Log2FC) of >1.0 and p-adjusted value <0.05. Analysis of upregulated genes at all stages throughout fusion among chicken, mouse, and zebrafish revealed 10 genes that were common to all three species (**Figure 2d**). Among these were the known human MAC genes *TENM3*, *SMOC1*, *PAX2*, *ALDH1A3*, and *BMPR1B* [7,26] , in addition to *NID1*, *VAX1* and *NTN1*, which have been previously identified as MAC or coloboma candidates based on animal studies [13,17,40]. Novel OFC genes not previously associated with MAC in either humans or animal models were *EPHB3* and *EMX2*. We then added human transcriptome data to our analyses and found that those genes upregulated during fusion across all four species were *TEMN3*, *ALDH1A3*, *NTN1*, *SMOC1* and *VAX1* (**Figure 2e**). Within this group, only *NTN1* and *VAX1* have not previously been associated with MAC in humans. Thus, we identified *EMX2* and *EPHB3* as potential new OFC candidates, a core group of 10 genes whose expression was enriched in the fissure during OFC across diverse avian, mammalian, and fish species, and 5 genes whose expression was also enriched in the human fissure during OFC.

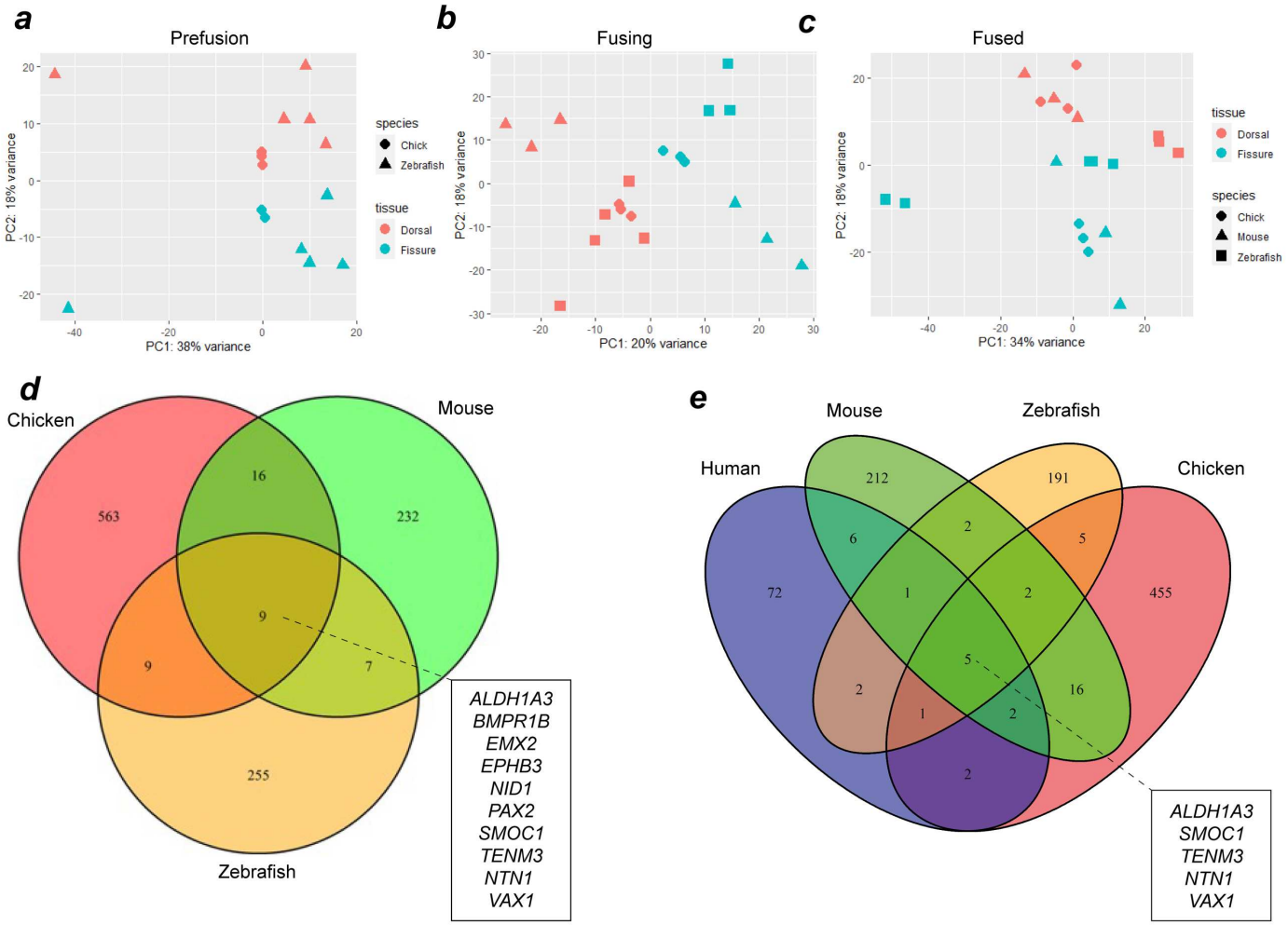

Figure 2. Combined transcriptome analysis across vertebrate species reveals common gene expression signatures during OFC. PCA analyses for aligned transcriptomes for chick and zebrafish (**a**), or for chick, zebrafish, and mouse (**b-c**) at OFC stages prior to fusion (**a**), active fusion (**b**), or in fused fissure stages (**c**). (d) Venn diagram showing intersection analysis for fissure enriched DEGs for all stages of OFC between mouse, chick, and zebrafish. There were 10 fissure specific DEGs shared among all three species: *EPHB3, ALDH1A3, BMPR1B, EMX2, NTN1, PAX2, SMOC1, VAX1, TENM3,* and *NID1*. (e) Analysis of shared DEGs among all four species identified 5 genes with fissure-enriched expression throughout OFC in humans, mouse, chicken, and zebrafish: *SMOC1, NTN1, ALDH1A3, VAX1* and *TENM3*.

3.3. In situ hybridization analysis of enriched gene expression in chicken optic fissures

To qualitatively confirm the transcriptome analyses and to determine if these genes have expression patterns consistent with a role in OFC or broader aspects of ventral eye development, and to identify what cells or ocular tissues these genes were expressed within, we conducted spatial in vivo gene expression analysis using fluorescence in situ

hybridisation on cryosections taken from chick embryonic whole eyes that included both dorsal and ventral regions (**Figure 3**). Eyes were selected at stage HH28 just as fusion is beginning in the chick[17]. This approach revealed that *VAX1*, *BMPR1B, SMOC1, ALDH1A3, NID1*, and *TENM3* were broadly expressed in a ventral-high, dorsal-low pattern in the developing retina and were only observed in the neural retina epithelia. Probes for *SMOC1, NTN1* and *ALDH1A3* gave the strongest signal, whereas *EMX2, NID1*, and *EPHB3* were only weakly detected, broadly in line with the normalised expression values (**Supplemental figure S2**). *NTN1* and *PAX2* expression domains were restricted to the edges of the optic fissure margins, consistent with the localisation of OFC pioneer cells that mediate and directly participate in the fusion process [17,18]. *ALDH1A3, BMPR1B, SMOC1, PAX2, NTN1, VAX1, TENM3* and *NID1* were all expressed in neural retina, with only some expression in the immediate RPE at the fissure edges for these genes. *EMX2* expression was detected in the ventral peri-ocular regions and not detected in RPE or neural retina. As this gene was previously reported as expressed in neuroepithelia, we used whole mount in situ hybridisation with large (500 bp) gene-specific RNA probes to confirm the peri-ocular localisation of *EMX2* expression (*Supplemental figure S3*). *EPHB3* was broadly expressed in a range of cell types in the ventral eye. Of all these genes, expression of only *NID1* was detected in the dorsal retina. Trajectory analyses of expression levels relative to dorsal tissue throughout OFC progression (pre-fusion, fusing, and fused) revealed that no clear temporal gene expression patterns emerged, except that *ALDH1A3* and *EMX2* both decreased after fusion for all species. *BMPR1B*, *SMOC1* and *TENM3* either increased expression in the fissure or it remained constant or slightly decreased after fusion; whereas *NID1*, *NTN1*, *VAX2*, *PAX2* and *EPHB3* all showed decreased expression levels after fusion in 3/4 species (**Supplemental figure S4**).

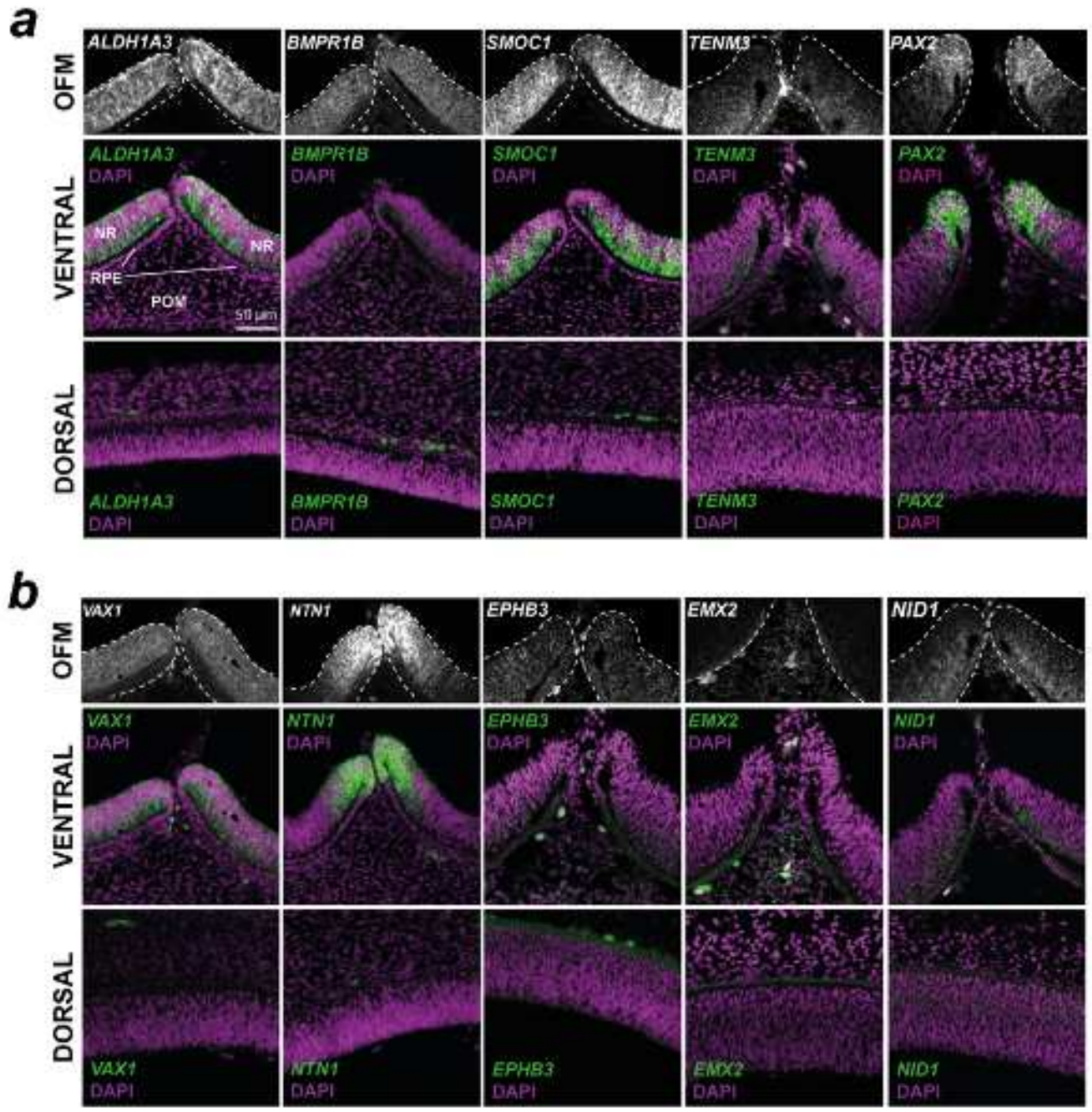

Figure 3. Spatial gene expression analysis for 10 conserved fissure enriched genes in the developing chick eye. Fluorescent in situ hybridisation for (**a**) known coloboma gene targets and (**b**) OFC candidate genes identified in this study counterstained with DAPI nuclear stain on chick eyes at HH.St28, immediately prior to fusion. Note for *EMX2* the OFM region highlights the POM region - expression was not observed in the NR or RPE. In both **a** and **b**, *Top*: grayscale RNAscope probe signal without DAPI highlighting mRNA signal in the immediate optic fissure margin region. *Middle*: mRNA probe and DAPI merged images of broad ventral eye region. *Bottom*: Dorsal eye region with probe and DAPI merged. OFM, optic fissure margin; NR, neural retina; POM, periocular mesenchyme; RPE, retinal pigmented epithelium. Scale = 50 µm.

3.4. Coloboma candidate gene targeting *in vivo*

To determine the requirement of the novel genes identified in this study for eye development and OFC, we generated F0 knock-out zebrafish embryos using CRISPR/Cas9 gene editing (GE) targeting exon DNA immediately downstream of the translation initiation site (ATG) and analysed their eye developmental progress at two stages (**Figure 4**). The zebrafish genome has two duplicated ephb3 genes, ephb3ba and ephb3bb, and therefore we designed guide RNAs to both loci to ensure we generated complete removal of these gene products whose activities could be functionally redundant. Both *emx2* and *ephb3* loss-of-function gene editing events were confirmed by PCR from genomic DNA (**Figure 4a**). In zebrafish, OFC is complete by 56 hours post fertilization [41], and compared to wild type (not injected) and Cas9-only injected embryos, the GE embryos had developed normally at 3 days post fertilisation (dpf) and 5 dpf (**Figure 4b**) although eye sizes were apparently reduced. Measurements of eye diameters revealed smaller eye size in both GE groups compared to controls at 3 dpf and 5 dpf (**Figure 4b,c**). Sections cut from control and GE eyes and stained for laminin showed persistence of the basement membrane surrounding the edges of the optic fissure, consistent with fusion failure and coloboma in both ephb3ab and emx2 targeted embryos (**Figure 4d**). Thus, of the 10 genes identified in this study whose expression in the ventral retina during OFC is conserved across different vertebrates, all can be associated with coloboma or microphthalmia causation in animals or humans when their function is perturbed, and can be considered as MAC candidates for diagnostic screens.

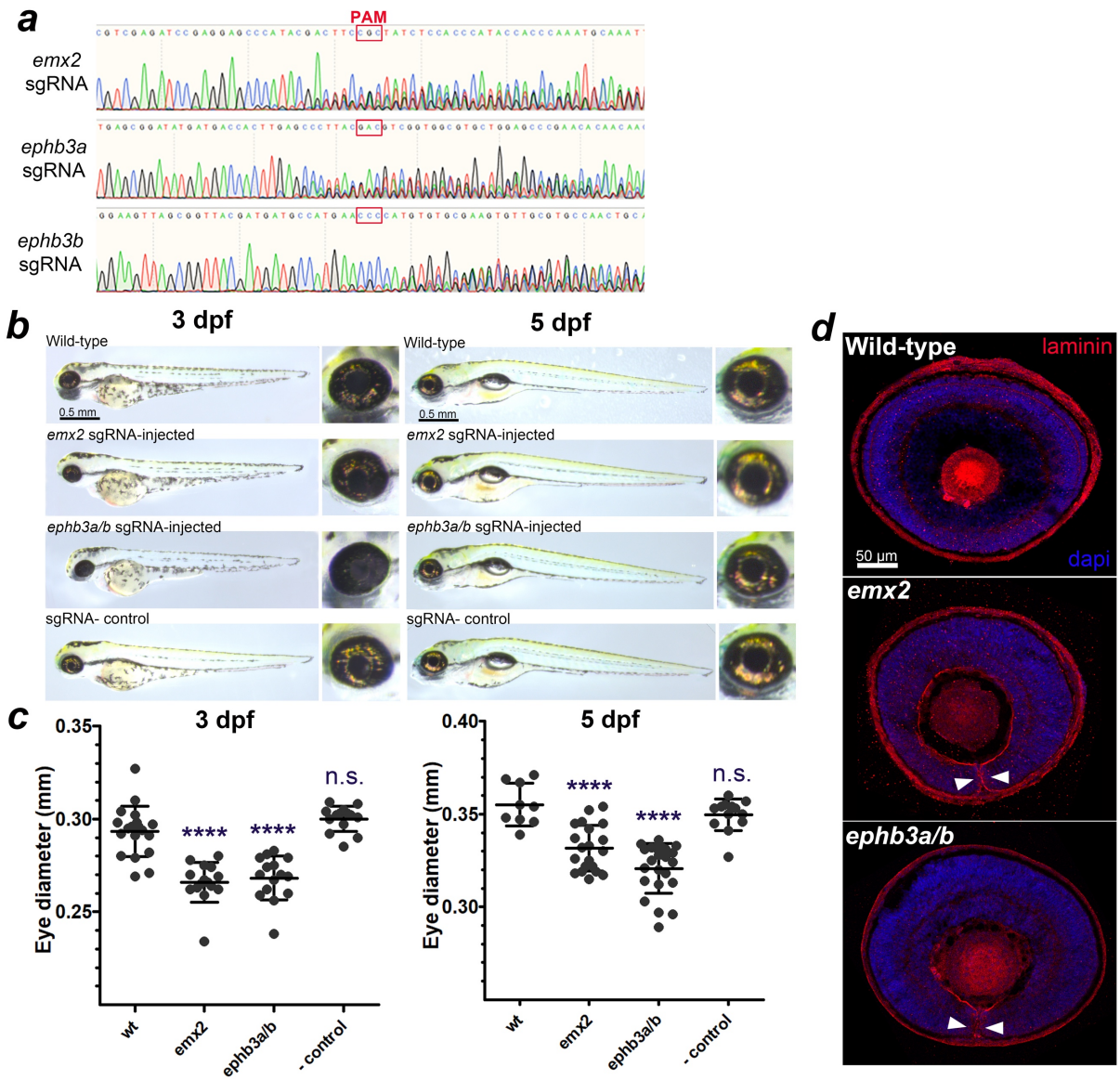

Figure 4. Targeted gene disruption of OFC novel candidates using CRISPR/Cas9 gene targeting in zebrafish embryos. (**a**) Sanger sequence traces for genomic DNA PCR analysis of CRISPR/Cas9 targeted embryos shows successful genetic disruption at the sgRNA site loci for eph3ba, eph3bb, and emx2. PAM sites are indicated. (**b**) Brightfield microscopy analysis of targeted fish at 3 dpf and 5 dpf compared to un-injected controls and Cas9 only controls. (**c**) Eye diameter metrics for fish injected with CRISPR/Cas9 at 3 dpf and 5 dpf (n>10 per target gene; ***p<0.001, ****p<0.0001). (**d**) Confocal analysis of laminin stained cryosections from gene edited and control eyes showing persistence of the basement membrane in the optic fissure margin (arrowheads).

3.5. Screening MAC cohorts for human variants in novel OFC genes.

We then sought to identify any novel disease-causing variants associated with MAC in *EMX2*, *EPHB3*, and *NID1*. A survey of the whole genome sequencing (GS) data of genetically unsolved MAC probands within the Genomics England 100,000 Genomes Project were interogated for variants either assigned as Tier III or variants of uncertain significance (VUS) as per ACMG-AMP guidelines involving the candidate genes. Sequencing data provided a minimum coverage of 15 times for more than 97% of the autosomal genome, and subsequently aligned to either GRCh37 or GRCh38 of the human genome (Isaac, Illumina Inc.). Variants were generated including single-nucleotide variants and indels (insertions or deletions) (Starling, Illumina Inc.), copy number variants (CNV, Canvas, Illumina Inc.), and structural variants (SV, Manta, Illumina Inc.). Genomic coordinates of candidate genes were identified, and used to positionally filter proband variant files, followed by further annotation using dbNSFP (v4.3a). Variants were prioritized in a stepwise manner by filtering using minor allele frequency in publicly available and in house datasets, predicted functional protein consequence, and familial segregation. Surviving variants were further inspected visually using IGV (**Table 3**).

Screening of the proband cohort identified a heterozygous VUS within *EPHB3*, *EPHB3*(NM_004443.4):c.649G>A,(p.Ala217Thr) that several in silico predictions reported a damaging consequence, although some benign (CADD 24.7, FATHMM pathogenic, Polyphen2 damaging, EIGEN pathogenic). The gnomAD genome allele frequency (v3.1.2) for this variant (GRCh38:3-184572969-G-A) reports a total of 11/152182 alleles, although 10/5184 alleles are observed in East Asian population, the probands ethnicity was White. No segregation analysis was undertaken as the sample was a singleton. We identified a second heterozygous VUS, *EPHB3*(NM_004443.4):c.1259G>A (p.Arg420His) in a male proband with HPO terms of autism, microphthalmia, seizures and true anophthalmia. Whilst the population frequency reported was low, in silico prediction of functional consequence was indeterminant.

Variants within the *NID1* gene were also 17rioritized, however no variants survived our filtering. We report a heterozygous variant *NID1* (NM_002508.3):c.2204G>A (p.Arg735His) which has been previously identified as pathogenic in a patient with neural tube defects [42], however the minor allele frequency was too high to enable us to term this as a variant of interest, although it may warrant further investigation (note this has been reported here for completeness). We also report two additional *NID1* coding variants: c.3458C>G(p.Pro1153Arg) and c.2859G>T(p.Lys953Asn) for which in silico predictions for consequence to protein range from uncertain to pathogenic. These alleles were not found in gnomAD. No copy number or structural variants were identified in any of the candidate genes screened.

Family	HPO Terms	Gene	cDNA / GRCh38	Protein	SIFT	Polyphen-2	Mutation Taster	CADD	gnomAD
F1	Optic nerve coloboma, moderate proteinuria, stage 3 chronic kidney disease	*EPHB3*	NM_004443.4 c.649 G>A 3:184572969:G:T	p.(Ala217Thr)	Tolerated	Probably Damaging	Disease Causing	24.7	0.001%
F2	Microphthalmia, true anophthalmia, intellectual disability, autism, seizures	*EPHB3*	NM_004443.4 c.1259 G>A 3:184577088:G:A	p.(Arg420His)	Benign	Probably Damaging	Disease Causing	22.6	$3x10^{-5}$
F3	Chorioretinal coloboma, macrocephaly, unilateral cleft lip, sleep apnoea, abnormal aggressive impulsive or violent behaviour	*NID1*	NM_002508.3 c.2204 G>A 1:236017198:C:T	p.(Arg735His)	Deleterious	Possibly Damaging	Disease Causing	24.04	0.016
F4	Bilateral retinal coloboma, mild global developmental delay, abnormality of the vertebrae, bulbar palsy, Sprengel anomaly, abnormality of the ear, skeletal system, nervous system	*NID1*	NM_002508.3 c.3458 C>G 1:235979873:G:C	p.(Pro1153Arg)	Pathogenic	Probably Damaging	Uncertain	25.9	NF
F5	Microphthalmia, anterior segment dysgenesis, iris coloboma, chorioretinal coloboma,	*NID1*	NM_002508.3 c.2859 G>T 1:235990955:C:A	p.(Lys953Asn)	Uncertain	Probably Damaging	Uncertain	24.4	NF

Table 3. Summary of in silico analysis of single nucleotide variants (SNVs) identified within MAC patient data. Screening of probands enrolled in the 100,000 Genomes Project presenting with HPO terms including microphthalmia, anophthalmia or coloboma for in silico predicted damaging variants results in 5 families. cDNA position of the SNV, subsequent amino acid change along with the genomic location for the human reference genome GRCh38 are presented. All variants identified were heterozygous. Outcomes of predictive algorithms SIFT, Polyphen-2 and MutationTaster are reported, with respective scaled CADD scores and allele frequency as reported in gnomAD (accessed Oct 2022, NF – not found).

4. Discussion

The main purpose of this study was to identify novel MAC candidates, and in this endeavor, we found two new genes, *EMX2* and *EPHB3*, that should now be associated with eye development and optic fissure closure defects. Our cross-species meta-analysis also provided evidence that genes previously associated with coloboma causation in model organisms, such as *VAX1*, *NTN1*, and *NID1*, should be more widely recognized as human MAC candidates.

The identification of novel MAC genes is challenging, largely due to the poor recurrence rates among non-related affected families, genetic heterogeneity, limits in sample sizes, and human phenotype ontology inconsistencies [21,23]. Single variants, even if they are predicted to be damaging or are shown in animal models as deleterious and phenocopy the human phenotype, still do not meet the strict criteria to be considered as causative [43]. Here, we present novel variants and report new genes associated with MAC but were unable to unambiguously confirm their causality. Identifying coloboma candidate genes is even more challenging, largely due to the difficulties in defining the transcriptional networks that govern tissue fusion itself [16]. This is a highly complex process, involving small populations of cells from both the edges of the retinal epithelium, and the peri-ocular mesenchyme (POM) that directly, or indirectly, mediate fusion in the eye, respectively [17,18,44]. We hypothesized that performing in silico data mining of vertebrate OFC transcriptomes may yield novel genes that are expressed in the pioneer cells or in the adjacent POM. This suffered the inherent limitations that OFC in these diverse vertebrates are not identical, for example the timing and duration of the OFC process and the overall size of the eye [16]. In addition, there are inter-species differences to the presence or absence of apoptotic foci in pioneer cells [17,39], and the intercalation of optic nerve cells and astrocytes at the proximal but not medial region of the fissure [45]. We aimed to overcome this second difference by carefully dissecting the medial and not proximal OF from our chicken retinas. Despite these differences, and similar to a previous study [26], we successfully showed conservation of gene expression across these species during OFC, and found that the majority of these genes had already been associated with MAC phenotypes, further validating our approach.

We identified POM specific expression of *EMX2* during OFC and showed that its targeted disruption led to MAC phenotypes in zebrafish, however we were unable to identify any novel pioneer cell markers. We propose that the identity of pioneer cells is highly transient and due to the rapid changes in cell behavior required for fusion, their unique molecular signatures may only be revealed through single-cell transcriptomic analyses rather than through

bulk RNAseq approaches. However, the generation of a pioneer cell reporter line in any of these species could enable the selective isolation of these cells, and their subsequent characterization throughout the OFC process, at multiple 'omics levels. Our study shows that the gene expression conservation we observed in OFC among different vertebrates would permit either of these model species and approaches to be used for such experiments, and we propose these will be the logical next steps for OFC research and OC candidate gene identification.

EMX2 is a homeobox transcription factor required for central nervous system and urogenital development [46,47]. It is a homologue of the Drosophila empty spiracles (ems) gene that is essential for anterior head development and brain segmentation. Mice lacking functional Emx2 die at birth and have no kidneys or reproductive organs [46]. In humans, heterozygous *EMX2* mutations have been associated with Schizencephaly [OMIM #269160], but no MAC or other ocular malformations have previously been reported associated with mutations in this gene. Our study provides the first example for genetic targeting of *EMX2* causing developmental eye disorders, with microphthalmia and coloboma observed in Crispr/Cas9-generated mutant *emx2* zebrafish. On review of the available literature, an apparent coloboma and ventral eye defect in a single image panel of an *Emx2*$^{-/-}$ mouse embryo at gestational stage E12.5 was noted, when OFC should be complete [48]. As this was unreported by the authors, without further data it is unclear what the penetrance or underpinning mechanism of this phenotype is, however this appears to support our data indicating the necessity for EMX2 in eye development. Further investigation is required to understand its precise role in OFC.

NID1 encodes nidogen, a component of the basement membrane that overlies the edges of the fissures as they become apposed prior to fusion. A previous study [40] found that the zebrafish orthologue *nid1* is one of several paralogous nidogens expressed in the fissure margin, but whose expression is downregulated specifically at the onset of fusion, presumably helping to mediate the remodeling of the basement membrane (BM) to enable fusion through either reducing the integrity of the BM or by exposing proteolytic sites in the remaining components of the OFM BM. We found *NID1* among the 3-species list (chicken, zebrafish, and mouse) of OFM enriched genes and found that its expression was reduced in fused versus fusing transcriptomes. We also showed its expression was localized to the apical-most cells of the chicken ventral neural retina. Morpholino-based knock-down of *nid1* in the previously mentioned study caused coloboma [40], and therefore we did not generate mutants for this gene. However, given the

zebrafish phenotype, and the conservation of OFM NID1 expression across the range of species shown here, we suggest this gene should be considered in the genetic analysis of MAC patients.

Ephrin type-B receptor 3 is a transmembrane tyrosine kinase protein with affinity for the ephrin-B family of ligands and is encoded by the *EPHB3* gene. Ephrin receptor-ligand interactions at cell surfaces mediate numerous dynamic developmental processes, such as migration, proliferation, and cell fate determination, mediated through changes to cytoskeletal dynamics [49], and can also act as tumor suppressor genes [50]. A role for Ephrin signaling in eye development is also emerging: dominant mutations in *EPHA2* cause isolated congenital cataracts, and with its ligand *EFNA5* are important for normal lens development [51–54]. Heterozygous pathogenic *EPHA2* variants have also been reported in two families presenting with congenital cataract and non-syndromic microphthalmia, widening the phenotype associated with this gene [55]. Even more recently, biallelic mutations in *EPHA2* have been found in a syndromic form of microphthalmia with anterior segment dysgenesis [56]. Mutations in *EPHB3* have not yet been associated with defects in eye development, but the zebrafish CRISPR/Cas9 mutant phenotype data presented here together with its developmental expression in the ventral chicken eye tissue suggests this gene has an important role in normal oculogenesis and OFC. Inclusion of *EPHB3* on targeted gene panels for structural eye disease may help to improve diagnostic rates for MAC patients.

The ventral vertebrate retina remains undifferentiated during optic cup development largely through its unique transcriptional signature. The gene expression data presented in this study suggests this is largely conserved with *VAX1*, *ALDH1A3, SMOC1, NTN1* and *TENM3* all detected as ventrally expressed genes across human, mouse, chicken, and zebrafish species. Interestingly, *PAX2*, which when mutated is known to cause ocular coloboma in humans and zebrafish [57–59], was upregulated in mouse, chicken, and zebrafish fissures, and was highly specific to the pioneer cell region of the retina. However, *PAX2* was not included as a conserved OFC enriched gene when we added human expression data. This is likely due to the limitations of the datasets used within our in-silico approach, as the human dataset from Patel et al[39] showed variance in expression levels between dorsal and fissure tissues across the human samples analysed (n=3). Furthermore, *PAX2* was not featured in their list of significantly differentially expressed genes from the human and mouse fissure margins [39]. This highlights the difficulty in obtaining high-quality human samples and reinforces the utility of multi-species approaches for OFC research.

In summary, we applied an in-silico pipeline combining new and existing RNAseq datasets to reveal genes whose expression is enriched in the ventral eye during optic fissure closure stages of development across a range of diverse species. This approach was coupled to in vivo validation by spatial analysis of gene expression and revealed a core group of genes that are expressed during OFC; 5 were already known human MAC genes (*ALDH1A3, BMPR1B, PAX2, TENM3, SMOC1*) and 3 *(NID1, NTN1, VAX1*) were previously associated with structural eye defects in animal models. All our conserved OFC genes including the two novel MAC candidates, *EMX2* and *EPHB3*, should be considered for molecular screening based on their gene expression profiles during eye development and the presence of coloboma and microphthalmia phenotypes in animal models. This study expands the use of developmental expression analyses in model systems to provide supportive evidence towards identifying candidate genes for human structural eye malformations.

Author Contributions: Conceptualization, VTR, NO, MM and JR; methodology, BC, VTR, NO, MT, JJS, JR; software, BC, VTR, NO, JJS; investigation, BC, VTR, NO, MT, JR; formal analysis, BC, VTR, NO, MT; Validation, MT, JJS, JR; resources, MM, JR; data curation, BC, VTR, NO; writing—original draft preparation, JR; writing—review and editing, VTR, MT, NO, MM,; funding acquisition, JR, VTR, MM. All authors have read and agreed to the published version of the manuscript.

Funding: JR was supported by a UKRI Future leaders Fellowship (MR/S033165/1). BC and JR were supported by the Biotechnology and Biological Sciences Research Council (BBS/E/D/10002071). VTR's doctoral studentship was awarded by the Mexican National Council of Science and Technology (CONACyT). MM was funded by the Wellcome Trust (205174/Z/16/Z), Moorfields Eye Charity and BIPOSA Research Award.

Ethics statement: Zebrafish and chickens were maintained according to institutional regulations for the care and use of laboratory animals under the UK Animals Scientific Procedures Act. All chicken embryology work was carried out on embryos prior to 2/3 gestation and no regulated procedures were performed. For zebrafish, all work was conducted under the UCL Animal Welfare and Ethical Review Body (License no. PPL PC916FDE7). All approved standard protocols followed the guidelines of the Association for Research in Vision and Ophthalmology Statement for the Use of Animals in Ophthalmic and Vision Research Ethics.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Data Availability Statement: All chicken RNAseq data files generated for this study are submitted to the NCBI Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo) with the accession number XXX [updated after review].

Acknowledgments: We would like to thank the staff at the Greenwood Building (Roslin Institute) for chicken husbandry, Aara Patel and Jane Sowden (UCL) for sharing RNAseq data, and James Prendergast (Roslin Institute) for help with Bioinformatics. This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support.

Supplementary Materials.

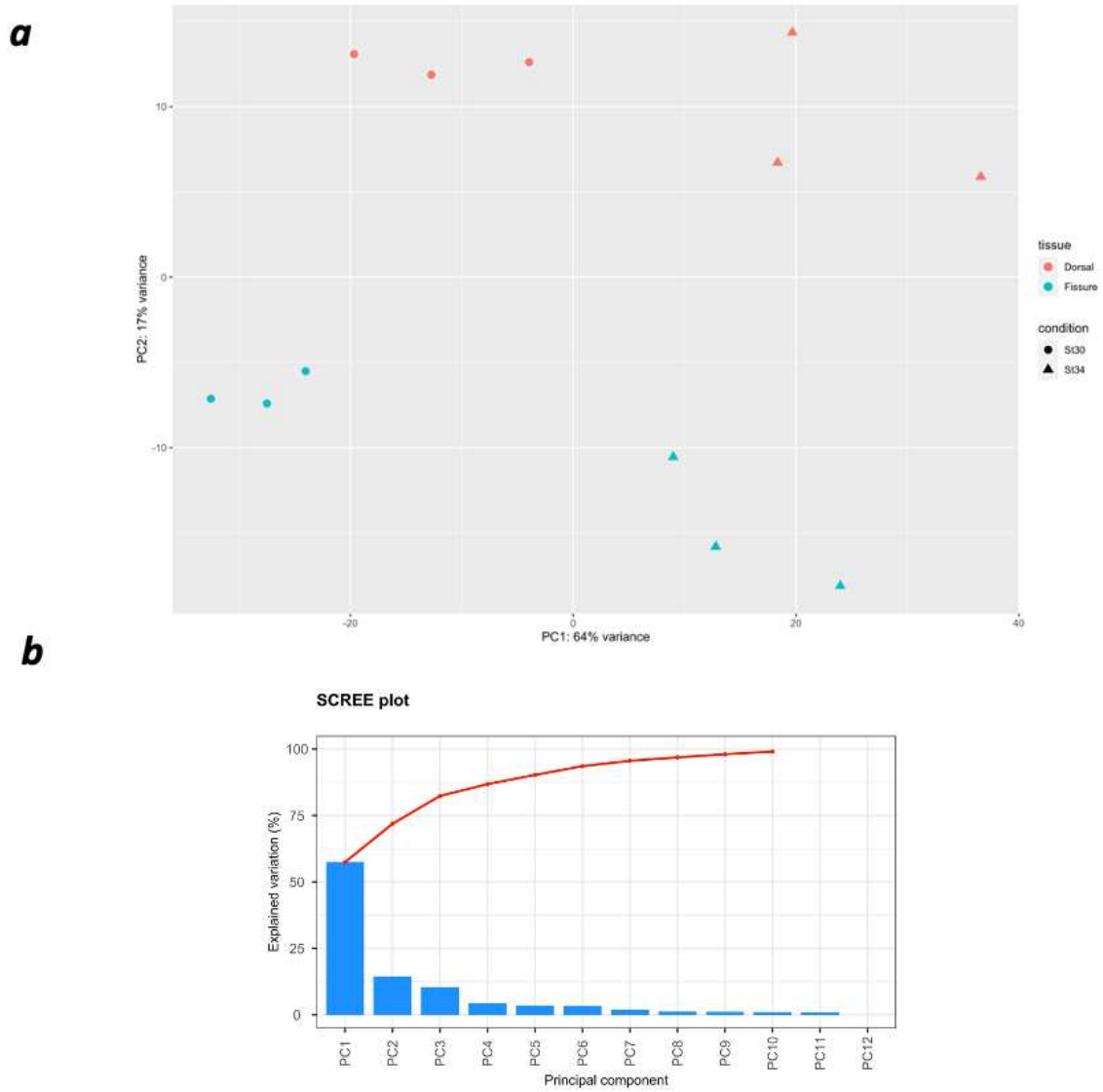

Figure S1. Validation of chicken RNAseq data. (a) Colorimetric Principal-Component Analysis (PCA) plot for samples labeled according to tissue (red: dorsal; blue: fissure) and condition (circle: HH St30; triangles: HH St34). PC1 is plotted against PC2 (n=3 for each tissue region and condition; total = 12 samples). (b) SCREE plot showing PC1 markedly reports difference between HH.St 30 versus HH.St34 samples, whereas PC2 separates fissure and dorsal tissue samples.

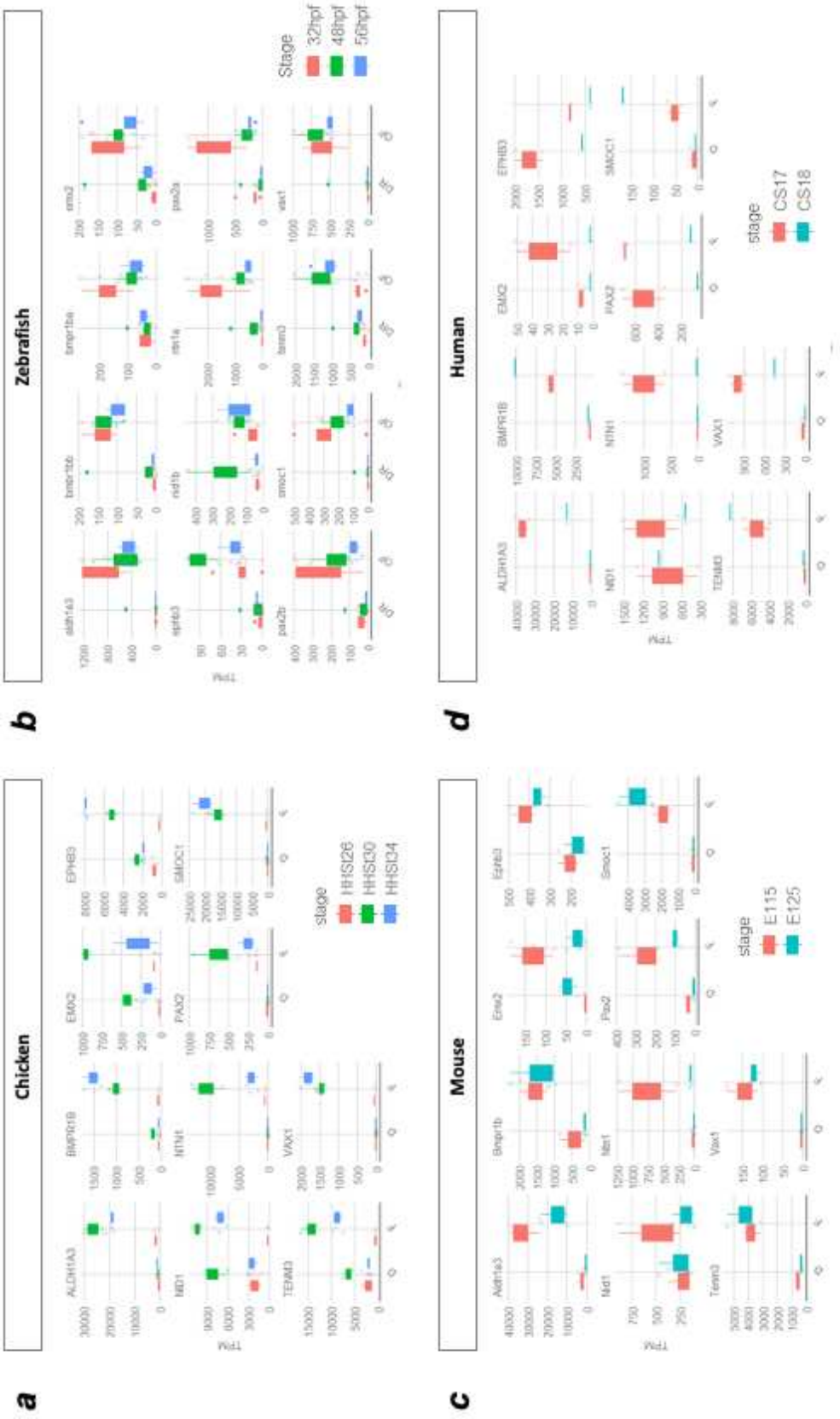

Figure S2: Expression values for all species and stages in the OFM and dorsal retina. TPM, transcripts per million; D, dorsal; F, fissure.

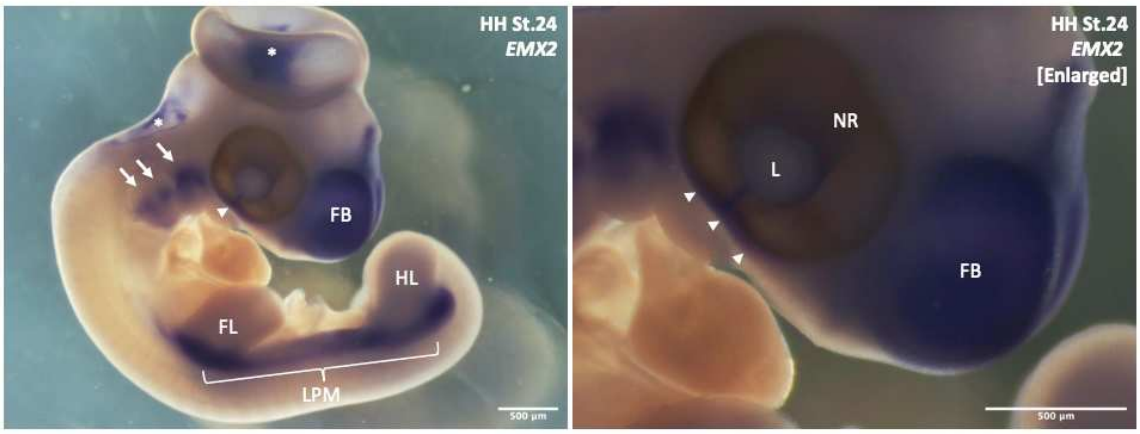

Figure S3. Colorimetric whole mount in situ hybridization for EMX2 expression in HH St24 chicken embryo. Arrowheads indicate periocular mesenchyme region. Arrows indicate pharyngeal arch regions. FB, forebrain; HL, hindlimb; FL, forelimb, LPM, lateral plate mesoderm region; L, lens; NR, neural retina. Asterisks indicate artifactual staining from substrate trapping. Scale bar = 500 µm.

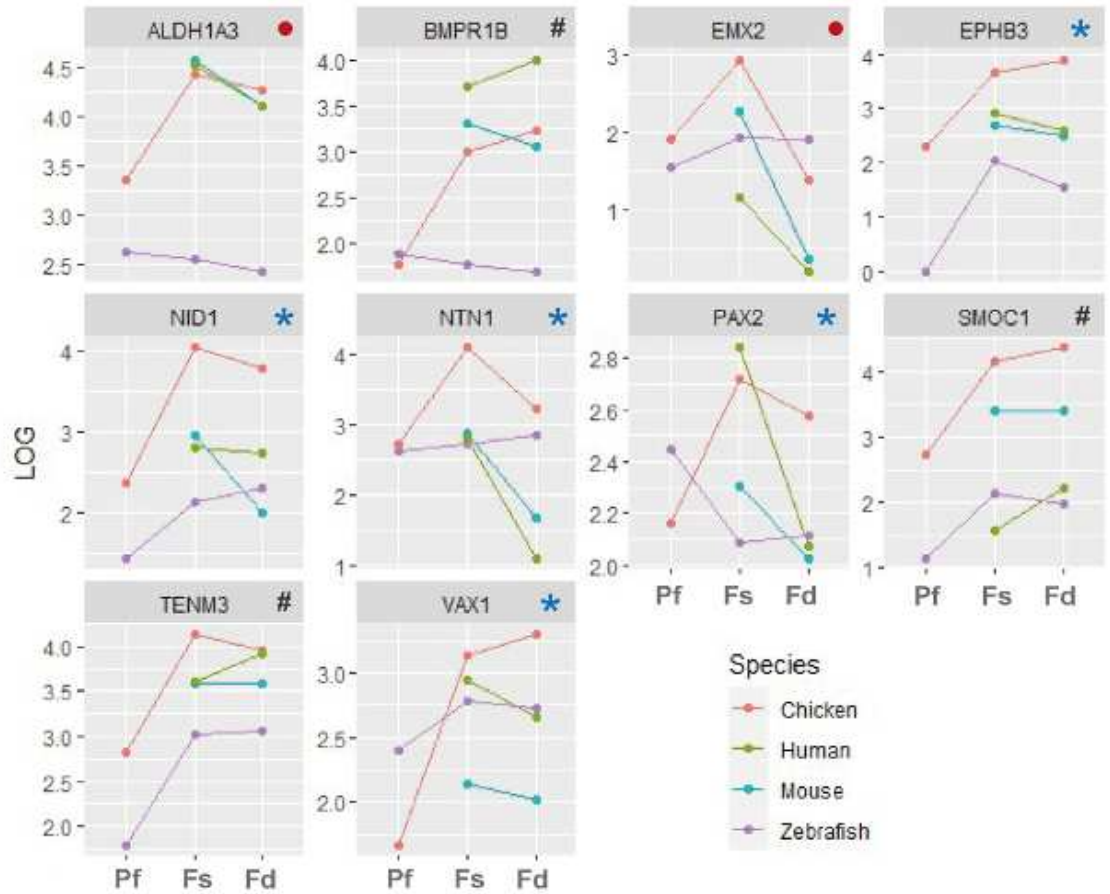

Figure S4. Expression trajectories in the OFM based on RNAseq expression data for humans, mouse, chicken, and zebrafish. Red dots indicate genes with reduced fissure expression post-fusion. Hatching indicates genes with increased or constant levels of expression post-fusion, and blue asterisks are genes with decreased fissure expression in the majority (3 of 4) of species.

References

1. Stoll, C.; Alembik, Y.; Dott, B.; Roth, M.P. Epidemiology of Congenital Eye Malformations in 131,760 Consecutive Births. Ophthalmic Paediatr. Genet. 1992, 13, 179–186, doi:10.3109/13816819209046487.

2. Morrison, D.; FitzPatrick, D.; Hanson, I.; Williamson, K.; van Heyningen, V.; Fleck, B.; Jones, I.; Chalmers, J.; Campbell, H. National Study of Microphthalmia, Anophthalmia, and Coloboma (MAC) in Scotland: Investigation of Genetic Aetiology. J. Med. Genet. 2002, 39, 16–22, doi:10.1136/jmg.39.1.16.

3. Shah, S.P.; Taylor, A.E.; Sowden, J.C.; Ragge, N.; Russell-Eggitt, I.; Rahi, J.S.; Gilbert, C.E. Anophthalmos, Microphthalmos, and Coloboma in the United Kingdom: Clinical Features, Results of Investigations, and Early Management. Ophthalmology 2012, 119, 362–368, doi:10.1016/j.ophtha.2011.07.039.

4. Williamson, K.A.; FitzPatrick, D.R. The Genetic Architecture of Microphthalmia, Anophthalmia and Coloboma. Eur. J. Med. Genet. 2014, 57, 369–380, doi:10.1016/j.ejmg.2014.05.002.

5. Selzer, E.B.; Blain, D.; Hufnagel, R.B.; Lupo, P.J.; Mitchell, L.E.; Brooks, B.P. Review Article Review of Evidence for Environmental Causes of Uveal Coloboma. Surv. Ophthalmol. 2021, 67, 1031–1047, doi:10.1016/j.survophthal.2021.12.008.

6. Harding, P.; Moosajee, M. The Molecular Basis of Human Anophthalmia and Microphthalmia; 2019; Vol. 7; ISBN 4420760869.

7. Patel, A.; Sowden, J.C. Genes and Pathways in Optic Fissure Closure. Semin. Cell Dev. Biol. 2017, doi:10.1016/j.semcdb.2017.10.010.

8. Sinn, R.; Wittbrodt, J. An Eye on Eye Development. Mech. Dev. 2013, 130, 347–358.

9. Fuhrmann, S. Eye Morphogenesis and Patterning of the Optic Vesicle. Curr Top Dev Biol. 2010, 93, 61–84, doi:10.1016/B978-0-12-385044-7.00003-5.Eye.

10. Chow, R.L.; Lang, R.A. Early Eye Development in Vertebrates. Annu. Rev. Cell Dev. Biol. 2001.

11. Zhang, X.M.; Yang, X.J. Temporal and Spatial Effects of Sonic Hedgehog Signaling in Chick Eye Morphogenesis. Dev. Biol. 2001, 233, 271–290, doi:10.1006/dbio.2000.0195.

12. Barbieri, A.M.; Lupo, G.; Bulfone, A.; Andreazzoli, M.; Mariani, M.; Fougerousse, F.; Consalez, G.G.; Borsani, G.; Beckmann, J.S.; Barsacchi, G.; et al. A Homeobox Gene, Vax2, Controls the Patterning of the Eye Dorsoventral Axis. Proc. Natl. Acad. Sci. U. S. A. 1999, doi:10.1073/pnas.96.19.10729.

13. Mui, S.H.; Kim, J.W.; Lemke, G.; Bertuzzi, S. Vax Genes Ventralize the Embryonic Eye. Genes Dev. 2005, 19, 1249–1259, doi:10.1101/gad.1276605.

14. Adler, R., & Belecky-Adams, T.L. The Role of Bone Morphogenetic Proteins in the Differentiation of the Ventral Optic Cup. Development 2002, 129, 3161–3171, doi:https://doi.org/10.1242/dev.129.13.3161.

15. Sehgal, R.; Karcavich, R.; Carlson, S.; Belecky-Adams, T.L. Ectopic Pax2 Expression in Chick Ventral Optic Cup Phenocopies Loss of Pax2 Expression. Dev. Biol. 2008, 319, 23–33, doi:10.1016/j.ydbio.2008.03.041.

16. Chan, B.H.C.; Moosajee, M.; Rainger, J. Closing the Gap: Mechanisms of Epithelial Fusion During Optic Fissure Closure. Front. Cell Dev. Biol. 2021, 8.

17. Hardy, H.; Prendergast, J.G.; Patel, A.; Dutta, S.; Trejo-Reveles, V.; Kroeger, H.; Yung, A.R.; Goodrich, L. V.; Brooks, B.; Sowden, J.C.; et al. Detailed Analysis of Chick Optic Fissure Closure Reveals Netrin-1 as an Essential Mediator of Epithelial Fusion. Elife 2019, 8, doi:10.7554/eLife.43877.

18. Eckert, P.; Knickmeyer, M.D.; Heermann, S. In Vivo Analysis of Optic Fissure Fusion in Zebrafish: Pioneer Cells, Basal Lamina, Hyaloid Vessels, and How Fissure Fusion Is Affected by Bmp. Int. J. Mol. Sci. 2020, 21, doi:10.3390/ijms21082760.

19. Shah, S.P.; Taylor, A.E.; Sowden, J.C.; Ragge, N.K.; Russell-Eggitt, I.; Rahi, J.S.; Gilbert, C.E. Anophthalmos, Microphthalmos, and Typical Coloboma in the United Kingdom: A Prospective Study of Incidence and Risk. Investig. Ophthalmol. Vis. Sci. 2011, 52, 558–564, doi:10.1167/iovs.10-5263.

20. Harding, P.; Gore, S.; Malka, S.; Rajkumar, J.; Oluonye, N.; Moosajee, M. Real-World Clinical and Molecular Management of 50 Prospective Patients with Microphthalmia, Anophthalmia and/or Ocular Coloboma. Br. J. Ophthalmol. 2022, bjo-2022-321991, doi:10.1136/bjo-2022-321991.

21. Patel, A.; Hayward, J.D.; Tailor, V.; Nyanhete, R.; Ahlfors, H.; Gabriel, C.; Jannini, T.B.; Abbou-Rayyah, Y.; Henderson, R.; Nischal, K.K.; et al. The Oculome Panel Test: Next-Generation Sequencing to Diagnose a Diverse Range of Genetic Developmental Eye Disorders. Ophthalmology 2019, 126, 888–907, doi:10.1016/j.ophtha.2018.12.050.

22. Rainger, J.; Williamson, K.A.; Soares, D.C.; Truch, J.; Kurian, D.; Gillessen-Kaesbach, G.; Seawright, A.; Prendergast, J.; Halachev, M.; Wheeler, A.; et al. A Recurrent de Novo Mutation in ACTG1 Causes Isolated Ocular Coloboma. Hum. Mutat. 2017, 38, 942–946, doi:10.1002/humu.23246.

23. Jackson, D.; Malka, S.; Harding, P.; Palma, J.; Dunbar, H.; Moosajee, M. Molecular Diagnostic Challenges for Non-Retinal Developmental Eye Disorders in the United Kingdom. Am. J. Med. Genet. Part C Semin. Med. Genet. 2020, doi:10.1002/ajmg.c.31837.

24. Brown, J.D.; Dutta, S.; Bharti, K.; Bonner, R.F.; Munson, P.J.; Dawid, I.B.; Akhtar, A.L.; Onojafe, I.F.; Alur, R.P.; Gross, J.M.; et al. Expression Profiling during Ocular Development Identifies 2 Nlz Genes with a Critical Role in Optic Fissure Closure. Proc. Natl. Acad. Sci. U. S. A. 2009, 106, 1462–1467, doi:10.1073/pnas.0812017106.

25. Richardson, R.; Owen, N.; Young, R.M.; Tracey-White, D.; Moosajee, M.; Toms, M. Transcriptome Profiling of Zebrafish Optic Fissure Fusion. Sci. Rep. 2019, 9, 1–12, doi:10.1038/s41598-018-38379-5.

26. Owen, N.M.M. Identification of Four Novel Human Ocular Coloboma Genes - Suppl. Data. Genet. Med. 2022, 1–12.

27. Hamburger, V.; Hamilton, H.L. A Series of Normal Stages in the Development of the Chick Embryo. J. Morphol. 1951, 88, 49–92, doi:10.1002/jmor.1050880104.

28. Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. Mol. Biol. Netw. 2011, 17, 10–12.

29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map Format and SAMtools. Bioinformatics 2009, 25, 2078–2079.

30. Liao, Y., Smyth, G. K., & Shi, W. The Subread Aligner: Fast, Accurate and Scalable Read Mapping by Seed-and-Vote. Nucleic Acids Res. 2013, 41, doi:https://doi.org/10.1093/nar/gkt214.

31. Liao, Y., Smyth, G. K., & Shi, W. No Title. Bioinformatics 2014, 30, 923–930, doi:https://doi.org/10.1093/bioinformatics/btt656.

32. Love, M. I., Huber, W., & Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. Genome Biol. 2014, 15, 1–21, doi:https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8.

33. Benjamini, Y. and Hochberg, Y. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. 1995, 57, 289–300, doi:DOI: 10.2307/2346101.

34. Durinck, S.; Moreau, Y.; Kasprzyk, A.; Davis, S.; De Moor, B.; Brazma, A.; Huber, W. BioMart and Bioconductor: A Powerful Link between Biological Databases and Microarray Data Analysis. Bioinformatics 2005, 21, 3439–3440, doi:10.1093/bioinformatics/bti525.

35. Durinck S, Spellman P, Birney E, H.W. Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor Package BiomaRt. Nat. Protoc. 2009, 4, 1184–1191.

36. Https://Www.Genomicsengland.Co.Uk/Initiatives/100000-Genomes-Project.

37. Genomics England Research Consortium 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. N. Engl. J. Med. 2021, 385, 1868–1880.

38. Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J. O. B., Danis, D., Gourdine, J. P., Gargano, M., Harris, N. L., Matentzoglu, N., McMurry, J. A., Osumi-Sutherland, D., Cipriani, V., Balhoff, J. P., Conlin, T., Blau, H., Baynam, G., Palmer, R., Gratian, P.N. Expansion of the Human Phenotype Ontology (HPO) Knowledge Base and Resources. Nucleic Acids Res. 2019, 8, D1018-1027, doi:10.1093/nar/gky1105.

39. Patel, A.; Anderson, G.; Galea, G.; Balys, M.; Sowden, J.C. A Molecular and Cellular Analysis of Human Embryonic Optic Fissure Closure Related to the Eye Malformation Coloboma. Development 2020, 44, dev.193649, doi:10.1242/dev.193649.

40. Carrara, N.; Weaver, M.; Piedade, W.P.; Vöcking, O.; Famulski, J.K. Temporal Characterization of Optic Fissure Basement Membrane Composition Suggests Nidogen May Be an Initial Target of Remodeling. Dev. Biol. 2019, 452, 43–54, doi:10.1016/j.ydbio.2019.04.012.

41. Knickmeyer, M.D.; Mateo, J.L.; Eckert, P.; Roussa, E.; Rahhal, B.; Zuniga, A.; Krieglstein, K.; Wittbrodt, J.; Heermann, S. TGFb-Facilitated Optic Fissure Fusion and the Role of Bone Morphogenetic Protein Antagonism. Open Biol. 2018, 8, doi:10.1098/rsob.170134.

42. Ishida M, Cullup T, Boustred C, James C, Docker J, English C; GOSgene, Lench N, Copp AJ, Moore GE, Greene NDE, S.P. A Targeted Sequencing Panel Identifies Rare Damaging Variants in Multiple Genes in the Cranial Neural Tube Defect, Anencephaly. Clin. Genet. 2018, 93, 870–879, doi:10.1111/cge.13189.

43. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 2015, 17, 405–424, doi:10.1038/gim.2015.30.

44. Gestri, G.; Bazin-Lopez, N.; Scholes, C.; Wilson, S.W. Cell Behaviors during Closure of the Choroid Fissure in the Developing Eye. Front. Cell. Neurosci. 2018, 12, 1–12, doi:10.3389/fncel.2018.00042.

45. Bernstein, C.S.; Anderson, M.T.; Gohel, C.; Slater, K.; Gross, J.M.; Agarwala, S. The Cellular Bases of Choroid Fissure Formation and Closure. Dev. Biol. 2018, 440, 137–151, doi:10.1016/j.ydbio.2018.05.010.

46. Yoshida, M.; Suda, Y.; Matsuo, I.; Miyamoto, N.; Takeda, N.; Kuratani, S. Yoshida-1997 Emx1 and Emx2 Functions in Development of Dorsal Telencephalon.Pdf. 1997, 111, 101–111.

47. Suda, Y.; Hossain, Z.M.; Kobayashi, C.; Hatano, O.; Yoshida, M.; Matsuo, I.; Aizawa, S. Emx2 Directs the Development of Diencephalon in Cooperation with Otx2. Development 2001, 128, 2433–2450, doi:10.1242/dev.128.13.2433.

48. Cecchi, C. Emx2: A Gene Responsible for Cortical Development, Regionalization and Area Specification. Gene 2002, 291, 1–9, doi:10.1016/S0378-1119(02)00623-6.

49. Klein, R. Eph/Ephrin Signalling during Development. Dev. 2012, 139, 4105–4109, doi:10.1242/dev.074997.

50. Zhao K, He J, Wang YF, Jin SD, Fan Y, Fang N, Qian J, Xu TP, G.R. EZH2-Mediated Epigenetic Suppression of EphB3 Inhibits Gastric Cancer Proliferation and Metastasis by Affecting E-Cadherin and Vimentin Expression. Gene 2019, 686, 118–124, doi:10.1016/j.gene.2018.11.015.

51. Shi, Y.; De Maria, A.; Bennett, T.; Shiels, A.; Bassnett, S. A Role for Epha2 in Cell Migration and Refractive Organization of the Ocular Lens. Invest. Ophthalmol. Vis. Sci. 2012, 53, 551–559, doi:10.1167/iovs.11-8568.

52. Jun, G.; Guo, H.; Klein, B.E.K.; Klein, R.; Jie, J.W.; Mitchell, P.; Miao, H.; Lee, K.E.; Joshi, T.; Buck, M.; et al. EPHA2 Is Associated with Age-Related Cortical Cataract in Mice and Humans. PLoS Genet. 2009, 5, doi:10.1371/journal.pgen.1000584.

53. Lee, J.; Lee, B.K.; Gross, J.M. Bcl6a Function Is Required during Optic Cup Formation to Prevent P53-Dependent Apoptosis and Colobomata. Hum. Mol. Genet. 2013, 22, 3568–3582, doi:10.1093/hmg/ddt211.

54. Coopera, M.A.; Sona, A.I.; Komlosb, D.; Suna, Y.; Kleimanc, N.J.; Zhoua, R. Loss of Ephrin-A5 Function Disrupts Lens Fiber Cell Packing and Leads to Cataract. Proc. Natl. Acad. Sci. U. S. A. 2008, 105, 16620–16625, doi:10.1073/pnas.0808987105.

55. Harding, P.; Toms, M.; Schiff, E.; Owen, N.; Bell, S.; Lloyd, I.C.; Moosajee, M. EPHA2 Segregates with Microphthalmia and Congenital Cataracts in Two Unrelated Families. Int. J. Mol. Sci. 2021, 22, 2190, doi:https://doi.org/10.3390/ijms22042190.

56. Courdier, C.; Gemahling, A.; Guindolet, D.; Barjol, A.; Scaramouche, C.; Bouneau, L.; Calvas, P.; Martin, G.; Chassaing, N.; Plaisancié, J. EPHA2 Biallelic Disruption Causes Syndromic Complex Microphthalmia with Iris Hypoplasia. Eur. J. Med. Genet. 2022, 65, 104574, doi:https://doi.org/10.1016/j.ejmg.2022.104574.

57. Benetti, E.; Artifoni, L.; Salviati, L.; Pinello, L.; Perrotta, S.; Zuffardi, O.; Zacchello, G.; Murer, L. Renal Hypoplasia without Optic Coloboma Associated with PAX2 Gene Deletion. Nephrol. Dial. Transplant. 2007, 22, 2076–2078, doi:10.1093/ndt/gfm187.

58. Gregory-Evans, C.Y.; Williams, M.J.; Halford, S.; Gregory-Evans, K. Ocular Coloboma: A Reassessment in the Age of Molecular Neuroscience. J. Med. Genet. 2004, 41, 881–891, doi:10.1136/jmg.2004.025494.

59. Sanyanusin P, Schimmenti LA, McNoe LA, Ward TA, Pierpont ME, Sullivan MJ, Dobyns WB, E.M. Mutation of the PAX2 Gene in a Family with Optic Nerve Colobomas, Renal Anomalies and Vesicoureteral Reflux. Nat. Genet. 1995, 9, 358–364, doi:10.1038/ng0495-358.