


# Exome sequencing identifies genetic variants in anophthalmia and microphthalmia

Jingjing Li<sup>1</sup> | Wei Yang<sup>2</sup> | Yuejun Jessie Wang<sup>1</sup> | Chen Ma<sup>2</sup> |  
Cynthia J. Curry<sup>3</sup> | Daniel McGoldrick<sup>4</sup> | Deborah A. Nickerson<sup>4,5</sup> |  
Jessica X. Chong<sup>5,6</sup> | Elizabeth E. Blue<sup>5,7</sup> | James C. Mullikin<sup>8</sup> | Jennita Reefhuis<sup>9</sup> |  
Wendy N. Nembhard<sup>10</sup> | Paul A. Romitti<sup>11</sup> | Martha M. Werler<sup>12</sup> |  
Marilyn L. Browne<sup>13,14</sup> | Andrew F. Olshan<sup>15</sup> | Richard H. Finnell<sup>16,17,18</sup> |  
Marcia L. Feldkamp<sup>19</sup> | Faith Pangilinan<sup>8</sup> | Lynn M. Almli<sup>9</sup> | Mike J. Bamshad<sup>4,5,6</sup> |  
Lawrence C. Brody<sup>8</sup> | Mary M. Jenkins<sup>9</sup> | Gary M. Shaw<sup>2</sup>  | NISC Comparative  
Sequencing Program<sup>20</sup> | University of Washington Center for Mendelian Genomics<sup>4</sup> |  
the National Birth Defects Prevention Study

<sup>1</sup>Department of Neurology School of Medicine, The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, The Bakar Computational Health Sciences Institute, The Parker Institute for Cancer Immunotherapy, University of California San Francisco, San Francisco, California, USA

<sup>2</sup>Department of Pediatrics, Stanford University School of Medicine, Stanford, California, USA

<sup>3</sup>Genetic Medicine, Department of Pediatrics, University of California, San Francisco, California, USA

<sup>4</sup>Department of Genome Sciences, University of Washington, Seattle, Washington, USA

<sup>5</sup>Brotman Baty Institute for Precision Medicine, Seattle, Washington, USA

<sup>6</sup>Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, Washington, USA

<sup>7</sup>Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington, USA

<sup>8</sup>National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA

<sup>9</sup>National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

<sup>10</sup>Department of Epidemiology, Fay W. Boozman College of Public Health, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

<sup>11</sup>Department of Epidemiology, University of Iowa College of Public Health, Iowa City, Iowa, USA

<sup>12</sup>Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA

<sup>13</sup>Birth Defects Registry, New York State Department of Health, Albany, New York, USA

<sup>14</sup>Department of Epidemiology and Biostatistics, School of Public Health, University at Albany, Rensselaer, New York, USA

<sup>15</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

<sup>16</sup>Department of Molecular and Cellular Biology, Center for Precision Environmental Health, Baylor College of Medicine, Houston, Texas, USA

<sup>17</sup>Department of Molecular and Human Genetics, Center for Precision Environmental Health, Baylor College of Medicine, Houston, Texas, USA

<sup>18</sup>Department of Medicine, Center for Precision Environmental Health, Baylor College of Medicine, Houston, Texas, USA

<sup>19</sup>Division of Medical Genetics, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, Utah, USA

<sup>20</sup>NIH Intramural Sequencing Center, National Human Genome Research Institute, Bethesda, Maryland, USA

## Correspondence

Gary M. Shaw, Department of Pediatrics,  
Stanford University School of Medicine,  
453 Quarry Road Palo Alto, 94304 CA,  
Stanford, CA, USA.  
Email: [gmsshaw@stanford.edu](mailto:gmsshaw@stanford.edu)

Jingjing Li, Department of Neurology School of  
Medicine, The Eli and Edythe Broad Center of  
Regeneration Medicine and Stem Cell  
Research, The Bakar Computational Health  
Sciences Institute, The Parker Institute for  
Cancer Immunotherapy, University of  
California San Francisco, 35 Medical Center  
Way, San Francisco, 94143 CA, USA.  
Email: [jingjing.li@ucsf.edu](mailto:jingjing.li@ucsf.edu)

## Funding information

Centers for Disease Control and Prevention,  
Grant/Award Numbers: FOA DD09-001, PA  
02081, PA 96043; California Department of  
Public Health; National Institutes of Health;  
National Heart, Lung, and Blood Institute,  
Grant/Award Number: U24 HG008956;  
National Human Genome Research Institute,  
Grant/Award Number: UM1 HG006493;  
University of Washington

## Abstract

Anophthalmia and microphthalmia (A/M) are rare birth defects affecting up to 2 per 10,000 live births. These conditions are manifested by the absence of an eye or reduced eye volumes within the orbit leading to vision loss. Although clinical case series suggest a strong genetic component in A/M, few systematic investigations have been conducted on potential genetic contributions owing to low population prevalence. To overcome this challenge, we utilized DNA samples and data collected as part of the National Birth Defects Prevention Study (NBDPS). The NBDPS employed multi-center ascertainment of infants affected by A/M. We performed exome sequencing on 67 family trios and identified numerous genes affected by rare deleterious nonsense and missense variants in this cohort, including *de novo* variants. We identified 9 nonsense changes and 86 missense variants that are absent from the reference human population (Genome Aggregation Database), and we suggest that these are high priority candidate genes for A/M. We also performed literature curation, single cell transcriptome comparisons, and molecular pathway analysis on the candidate genes and performed protein structure modeling to determine the potential pathogenic variant consequences on PAX6 in this disease.

## KEYWORDS

congenital abnormalities, genetic epidemiology, newborn eye abnormalities

## 1 | INTRODUCTION

Anophthalmia and microphthalmia (A/M) are complex eye malformations (Dong et al., 2015; Skalicky et al., 2013; Verma & Fitzpatrick, 2007). Human eye development is initiated at 3 weeks gestation from the anterior neural plate (Sinn & Wittbrodt, 2013; Zagozewski, Zhang, & Eisenstat, 2014). Anophthalmia is defined as a total absence of the eye tissue or any structures associated with the eye (Plaisancie et al., 2019). True anophthalmia is an abnormality of eye development that occurs at the time of developing optic vesicle at 3–4 weeks of gestation leading to absence of the eye, optic nerve and chiasm (Harding & Moosajee, 2019). Clinical anophthalmia is considered in the absence of ocular structures with histologically detected remnants (Plaisancie et al., 2019). Microphthalmia is believed to arise early in pregnancy as well, but at least one report has suggested identification of microphthalmia development in midpregnancy (Blazer, Zimmer, Mezer, & Bronshtein, 2006). Microphthalmia is defined as reduction in the volume of the eye, usually characterized by corneal diameters less than 10 mm or anteroposterior globe diameter less than 20 mm, which is part of the spectrum of clinical anophthalmia and includes coloboma, anterior segment abnormalities and cataract (Skalicky et al., 2013; Verma & Fitzpatrick, 2007). A/M are rare congenital malformations with a prevalence of up to 2 per 10,000 live births (Mai et al., 2019; Shaw et al., 2005).

Current understanding of suspected A/M risk factors has been largely derived from clinical case series and experimental models

(Graw, 2019; Kumar, Sandell, Trainor, Koentgen, & Duester, 2012; Richardson, Tracey-White, Webster, & Moosajee, 2017). A significant genetic component of A/M has emerged from such clinical series. Several chromosome abnormalities have been implicated, such as trisomies 9, 13, and 18, 14q22.1-q23.2 deletion and 3q, 4p, and 10q duplications (Verma & Fitzpatrick, 2007).

The genetic basis of these conditions is further supported by identification of several candidate genes through clinical genomic sequencing. Approximately 20 genes are associated with A/M, but collectively these only explain a very small proportion of A/M cases (Bardakjian & Schneider, 2011; Williamson & FitzPatrick, 2014). Among these genes, one of the most extensively studied is PAX6 given its primary role in regulating eye development (Harding, Brooks, FitzPatrick, & Moosajee, 2020; Patel & Sowden, 2019; Verma & Fitzpatrick, 2007). Compound heterozygous nonsense variants in PAX6 were previously associated with anophthalmia (Glaser et al., 1994). However, the prevalence of PAX6 nonsense variants in A/M cases is extremely rare, and PAX6 variants are more commonly seen in many other human ocular phenotypes, such as aniridia (Patel & Sowden, 2019; Verma & Fitzpatrick, 2007). SOX2 has also been implicated in A/M in many studies; but may only be associated in approximately 10% of A/M cases (Dash et al., 2020; Fantès et al., 2003; Matias-Perez et al., 2018). Thus, known genetic risk factors explain a very small fraction of A/M cases. An important observation made from previous genetic analyses is the genetic heterogeneity in A/M, where different genes contribute to the molecular etiologies

of the A/M phenotype and showed that different individuals with A/M have different sets of variants.

Such genetic heterogeneity poses a significant challenge for classical family segregation analysis for Mendelian traits and candidate gene studies, and suggests the need for an unbiased genomic screen if we are to identify genetic underpinnings in A/M. What has made the investigation more challenging is the rarity of A/M cases, with existing knowledge mostly derived from nonsystematic small collections of affected probands. Exome sequencing analyses have been previously attempted in A/M cases but were limited to small patient cohorts (<30) or to a small set of preselected genes with previously suggestive evidence of microphthalmia, anophthalmia, and coloboma (Deml et al., 2016; Haug et al., 2021). Here we are able to capitalize on the systematic data collection of the National Birth Defects Prevention Study (NBDPS), the largest birth defects study conducted in the United States. We performed exome sequencing on 67 family trios who participated in the NBDPS and were affected by A/M. We integrated exome analysis, single-cell transcriptome comparison, molecular pathway analysis, protein structure modeling and literature curation, resulting in the identification of numerous genes involved in A/M, and specific retinal cell types potentially implicated in the pathogenesis of A/M.

## 2 | METHODS AND MATERIALS

### 2.1 | Study populations

The NBDPS is a multi-center case-control study of over 30 major birth defects. This study included data on pregnancies with estimated due dates between October 1997 and December 2011. Complete study methods have been described previously (Reefhuis et al., 2015). Cases and controls were ascertained from 10 birth defects surveillance centers: Arkansas, California, Georgia, Iowa, Massachusetts, New Jersey, New York, North Carolina, Texas, and Utah (Reefhuis et al., 2015). The institutional review boards of each study center provided approval for the NBDPS.

Infants and fetuses diagnosed with A/M had their medical records reviewed for eligibility by clinical geneticists. All included cases required diagnosis by an ophthalmologist or confirmation by surgical pathology or autopsy. Infants were ineligible if their only diagnosis was “small eyes” or “small palpebral fissures.” Infants were also ineligible if their A/M was associated with anterior encephalocele, anencephaly, holoprosencephaly, or amniotic band syndrome. Cases strongly suspected to have a chromosomal abnormality or single-gene condition were also excluded (Rasmussen et al., 2003). Cases were classified as “isolated” or “with accompanying malformations” based on the presence and type of accompanying malformations. Cases with no other major malformations, i.e., only minor anomalies or related anomalies, were considered isolated. Cases with at least one additional, believed to be unrelated major malformation, were considered A/M with accompanying malformations. The term “unrelated” refers

to defects in different body parts or systems and not a part of sequence.

Following a computer-assisted telephone interview, each woman was mailed cytobrushes to collect buccal cell specimens for herself, her child (if living), and her child's biological father (if available). Further details about cytobrush collection are found in Jenkins et al. 2019. Of the 235 A/M cases with completed maternal interviews, there were only 76 trios (proband, mother, father) with DNA available from cytobrushes, and eight trios were removed from the 76 because at least one family member had DNA amounts less than 200 ng. Of the remaining 68 trios, another proband was removed due to poor sample quality, leading to exclusion of the entire family. Therefore, we had 67 probands for sequence analysis as part of a trio, and they served as the analytic base for the current work.

### 2.2 | Exome sequencing

Details are found in Jenkins et al. 2019. Briefly, exome sequencing was performed at the National Institutes of Health Intramural Sequencing Center (NISC, Rockville, MD; <https://www.nisc.nih.gov/>). We used SeqCap<sup>®</sup> EZ Human Exome + UTR kit v3.0 (Roche NimbleGen, Madison, WI) for exome capture. Sequencing was performed using Illumina's HiSeq 2500 system (Illumina, San Diego, CA), which generated 126 bp paired-end reads. Illumina Genome Analyzer Pipeline software (version 1.18.64.0) was used for downstream analysis. Illumina's Efficient Large-scale Alignment of Nucleotide Databases (ELAND) was used to map the paired-end reads onto the reference human genome (hg19). We used the Burrows-Wheeler Aligner (BWA, v0.7.10) for read alignment (Li & Durbin, 2009). We followed the Genome Analysis Toolkit (GATK) Best Practice protocol for variant call involving “duplicate removal” (Picard MarkDuplicates; v1.111), indel realignment (GATK IndelRealigner; v3.2-2), and base quality recalibration (GATK BaseRecalibrator; v3.2-2). Raw Variant Call Format files were generated by GATK HaplotypeCaller (v3.2). Called raw variants were further refined by the variant quality score recalibration (VQSR) procedure. We only considered 117,608 high confidence single nucleotide variants (SNVs) with “PASS” tags assigned by GATK VQSR.

Data quality control (QC) included an assessment of (1) total reads (minimum of 50 million PE50 reads); (2) library complexity; (3) capture efficiency; (4) coverage distribution: 90% at 8X required for completion; (5) capture uniformity; (6) raw error rates; (7) transition/transversion ratio (Ti/Tv); (8) distribution of known and novel variants relative to dbSNP (typically <7%); (9) fingerprint concordance >99%; (10) sample homozygosity and heterozygosity; and (11) sample contamination validation. Exome completion was defined as having >90% of the exome target at >8× coverage and >80% of the exome target at >20× coverage. Typically, this required mean coverage of the target at 50–60×. Variant annotation was based on Annovar v2017Jul17 (Wang, Li, & Hakonarson, 2010), which returns annotations that include dbSNP rsIDs, gene names, and predicted functional effects (e.g., nonsynonymous, stopgain, or stoploss).

## 2.3 | Functional genomic analysis

To identify genes affected by nonsense variants, we required variants to be absent in the Genome Aggregation Database (gnomAD) v2.1.1 and the probability of being loss-of-function intolerant (pLI) score (Lek et al., 2016) to be >0.9, as a predictor of extreme haplo-insufficiency. In analyses of missense variants, our criteria called for variants to be absent in the gnomAD database v2.1.1, a Phred-scaled Combined Annotation Dependent Depletion (CADD) score (Kircher et al., 2014) > 20, Variant Effect Scoring Tool 3.0 (VEST3) score (Carter, Douville, Stenson, Cooper, & Karchin, 2013)  $\geq 0.7$  and Meta Likelihood ratio (LR) score (Dong et al., 2015)  $\geq 0.7$ , as predictors of extreme deleteriousness. We used Enrichr (Kuleshov et al., 2016) for functional enrichment analysis. We analyzed single-cell transcriptomic data across retina developmental stages from gestational week 9 to 27 (Lu et al., 2020). We first downloaded preprocessed data from GEO with the accession number GSE138002. Then we followed cell type definitions (Lu et al., 2020) and standardized gene expression in each cell type by total number of sequenced reads. We then performed quantile normalization across all cell types. For each gene, we averaged its expression across all cells in a given cell type, and for each cell type, we compared expression differences between the candidate genes and control genes (background) using a Wilcoxon rank-sum test, followed by Benjamini-Hochberg correction. For protein structure analysis, we implemented Site Directed Mutator (Pandurangan, Ochoa-Montano, Ascher, & Blundell, 2017) and DynaMut (Rodrigues, Pires, & Ascher, 2018) to compute physicochemical properties associated with any residues in a given protein structure.

## 2.4 | Sanger sequencing

Sanger sequencing was performed by ACGT, Inc. (<http://www.acgtinc.com/>) to validate a select group of the rare nonsense (loss-of-function) variants ( $n = 9$ ) detected during exome sequencing. The primer set was synthesized based on the reference sequence. Each locus of interest underwent PCR amplification of genomic DNA (>100 ng/reaction) and bidirectional sequencing. Sequence data from each locus were aligned, and the variants of interest were independently called by at least two individuals. Among the nine variants, Sanger sequencing was attempted in five probands and two trios, i.e., total of seven probands and parents of two of them with sufficient DNA quantities remaining.

## 3 | RESULTS

Among the 67 probands with A/M, over 90% had microphthalmia, 82% were considered isolated in their phenotype, and 72% occurred unilaterally (Table 1). Demographic characteristics of the probands and their parents are also shown in Table 1. Although no parent of any proband self-reported to have A/M, there were three probands with first degree relatives who self-reported eye conditions. These

**TABLE 1** Descriptive characteristics (percentages)<sup>a</sup> of sequenced anophthalmia/microphthalmia probands, National Birth Defects Prevention Study, 1997–2011

|  | Total (N = 67) |
|--|----------------|
| Infant sex                               |                |
| Male                                     | 36 (53.7)      |
| Female                                   | 31 (46.3)      |
| Gestational age (weeks)                  |                |
| Preterm (<37)                            | 15 (22.4)      |
| Term ( $\geq 37$ )                       | 52 (77.6)      |
| Plurality                                |                |
| Singleton                                | 63 (94.0)      |
| Multiple                                 | 4 (6.0)        |
| Phenotypes                               |                |
| Both                                     | 1 (1.5)        |
| Anophthalmia                             | 5 (7.5)        |
| Microphthalmia                           | 61 (91.0)      |
| Isolated/with accompanying malformations |                |
| Isolated                                 | 55 (82.1)      |
| With accompanying malformations          | 12 (17.9)      |
| Laterality                               |                |
| Unilateral                               | 48 (71.6)      |
| Bilateral                                | 19 (28.4)      |
| Maternal age at delivery (years)         |                |
| < 25                                     | 22 (32.8)      |
| 25–29                                    | 22 (32.8)      |
| $\geq 30$                                | 23 (34.3)      |
| Maternal race/ethnicity                  |                |
| Non-Hispanic White                       | 41 (61.2)      |
| Non-Hispanic Black                       | 3 (4.5)        |
| Hispanic                                 | 19 (28.4)      |
| Other                                    | 4 (6.0)        |
| Paternal age at delivery (years)         |                |
| <25                                      | 14 (20.9)      |
| 25–29                                    | 20 (29.9)      |
| $\geq 30$                                | 32 (47.8)      |
| Paternal race/ethnicity                  |                |
| Non-Hispanic White                       | 40 (59.7)      |
| Non-Hispanic Black                       | 4 (6.0)        |
| Hispanic                                 | 18 (26.9)      |
| Other                                    | 3 (4.5)        |

<sup>a</sup>Sums of percentages may not equal 100 owing to rounding.

conditions could be part of the A/M spectrum. The first proband's father was reported as having coloboma, the second proband's mother and sibling were reported to be blind at birth, and the third proband's mother reported Duane syndrome (data not shown).

We performed exome sequencing on 67 trios, generating high quality sequence data for each of the 201 individuals. Subsequent QC

analysis confirmed high quality of the called variants (Figure S1). We identified 102,793 SNVs in coding sequences among the 67 probands. We had less confidence on indel and structural variant calls, therefore, we focused our analyses on SNVs.

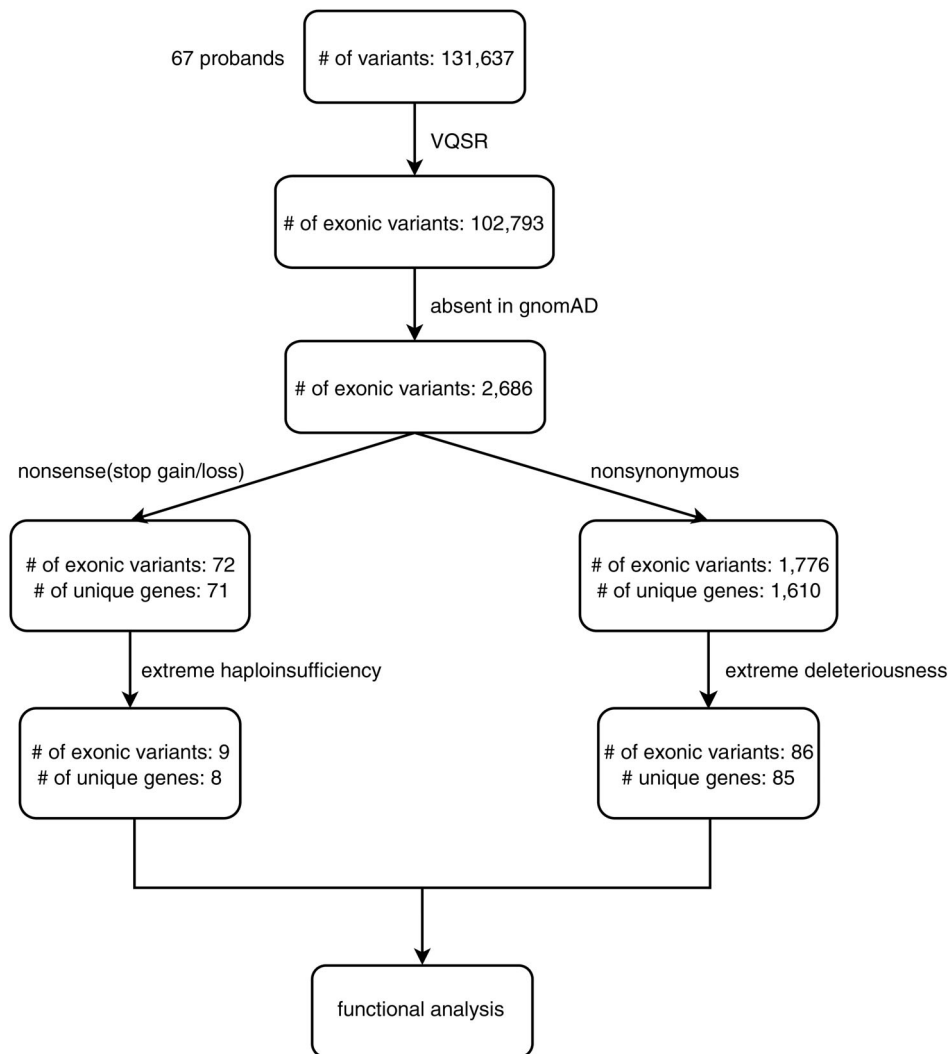
Given the low population prevalence of A/M (2/10,000 live births), we followed the rare disease, rare variant model and studied the role of rare variants in A/M. These rare variants also included *de novo* variants in the analyses. Studying all rare variants, instead of only focusing on *de novo* variants, enabled us to study private variants in affected families. For exonic variants identified in the 67 probands with A/M, we excluded those observed in the global human population represented by exome and genome sequences from 125,748 and 15,708 individuals, respectively (the gnomAD database v2.1.1), resulting in an entire collection of 2686 rare exonic variants identified in the probands with A/M. Thus, these variants represent those rare in the human population.

Mapping these rare variants onto their respective protein coding sequences, we identified 72 nonsense (loss-of-function) and 1776 missense variants. For loss-of-function variants, we considered them to have deleterious effects if the affected proteins were intolerant to

copy loss, thereby receiving extreme probability of loss of function intolerance (pLI) scores (Lek et al., 2016) (pLI >0.9). For missense variants, we implemented a set of machine learning algorithms to jointly predict their mutational consequences and only considered those variants receiving consensus predicted mutational deleteriousness. Following the procedures outlined in Figure 1, we identified 9 and 86 rare deleterious nonsense and missense variants respectively.

We first analyzed the nine rare nonsense variants affecting eight genes (Table 2). Gene *CHD7* was implicated in two unrelated probands. We performed literature curation for each of the genes in Table 2 and found evidence that each is involved in eye development (Bergman et al., 2012; Du et al., 2015; Ergun, Akay, Ergun, & Percin, 2017; Li, Jin, & Zhong, 2019; Meijer, 1993; Papakostas et al., 2018). For example, *LRP5*, *NUMA1* and *TUBGCP3* have been identified to be involved in retinal development and *KRT31* to be involved with eye morphological alterations (Du et al., 2015; Ergun et al., 2017; Li et al., 2019). Thus, this agnostic analysis of rare nonsense variants successfully identified candidate genes in A/M.

Previous work has identified the retina as an important tissue of origin in A/M (Chao et al., 2010; Garcia-Llorca, Aspelund,



**FIGURE 1** Analysis schematic. Overview of the approach to identifying candidate genes from 67 probands. Following GATK best practice, we applied the variant quality score recalibration (VQSR) procedure to refine the raw variants from the initial genotype calls. To define rare variants, we considered those not seen in the genome Aggregation database (gnomAD)

**TABLE 2** Nonsense variants (stopgain) in probands with anophthalmia/microphthalmia identified by exome sequencing

| Gene                      | [Chromosome: position] | [REF/ALT] | Predicted protein change | PMID                                       | Eye development function   | Origin         | Family ID | Phenotype   |
|---------------------------|------------------------|-----------|--------------------------|--|--|----------------|-----------|---|
| <b>VPS13D<sup>a</sup></b> | 1:12557600             | C/T       | p.Q4212X                 | 30,789,691                                 | Eye movement abnormalities   | <i>De novo</i> | #46       | Unilateral microphthalmia with accompanying malformations |
| <i>XKR4</i>               | 8:56015062             | C/G       | p.S5X                    | 29,625,478                                 | Mutation showed increase risk of poor vision                               | Unknown        | #54       | Unilateral microphthalmia, isolated                       |
| <b>CHD7<sup>b</sup></b>   | 8:61735063             | C/T       | p.R987X                  | 22,539,353                                 | Known in A/M   | <i>De novo</i> | #55       | Unilateral microphthalmia, isolated                       |
| <i>CHD7</i>               | 8:61750761             | C/T       | p.R1494X                 | 22,539,353                                 | Known in A/M   | Unknown        | #20       | Unilateral microphthalmia with accompanying malformations |
| <i>LRP5</i>               | 11:68191161            | C/T       | p.R497X                  | 28,111,184                                 | Retinal development  | Paternal       | #17       | Bilateral microphthalmia, isolated                        |
| <i>NUMA1</i>              | 11:71726895            | G/A       | p.Q552X                  | 25,593,321                                 | Eye tissue development   | Paternal       | #45       | Bilateral microphthalmia, isolated                        |
| <i>TUBGCP3</i>            | 13:113181728           | G/T       | p.S459X                  | 31,178,691                                 | Retinal progenitor cell proliferation; Knockout showed small eye phenotype | Paternal       | #16       | Unilateral microphthalmia, isolated                       |
| <i>FAM192A</i>            | 16:57206739            | G/A       | p.Q59X                   | The Human Protein Atlas                    | Expression in the retina   | Maternal       | #64       | Unilateral microphthalmia, isolated                       |
| <i>KRT31</i>              | 17:39553461            | C/A       | p.E111X                  | International Mouse Phenotyping Consortium | Eye morphology   | Maternal       | #61       | Unilateral microphthalmia, isolated                       |

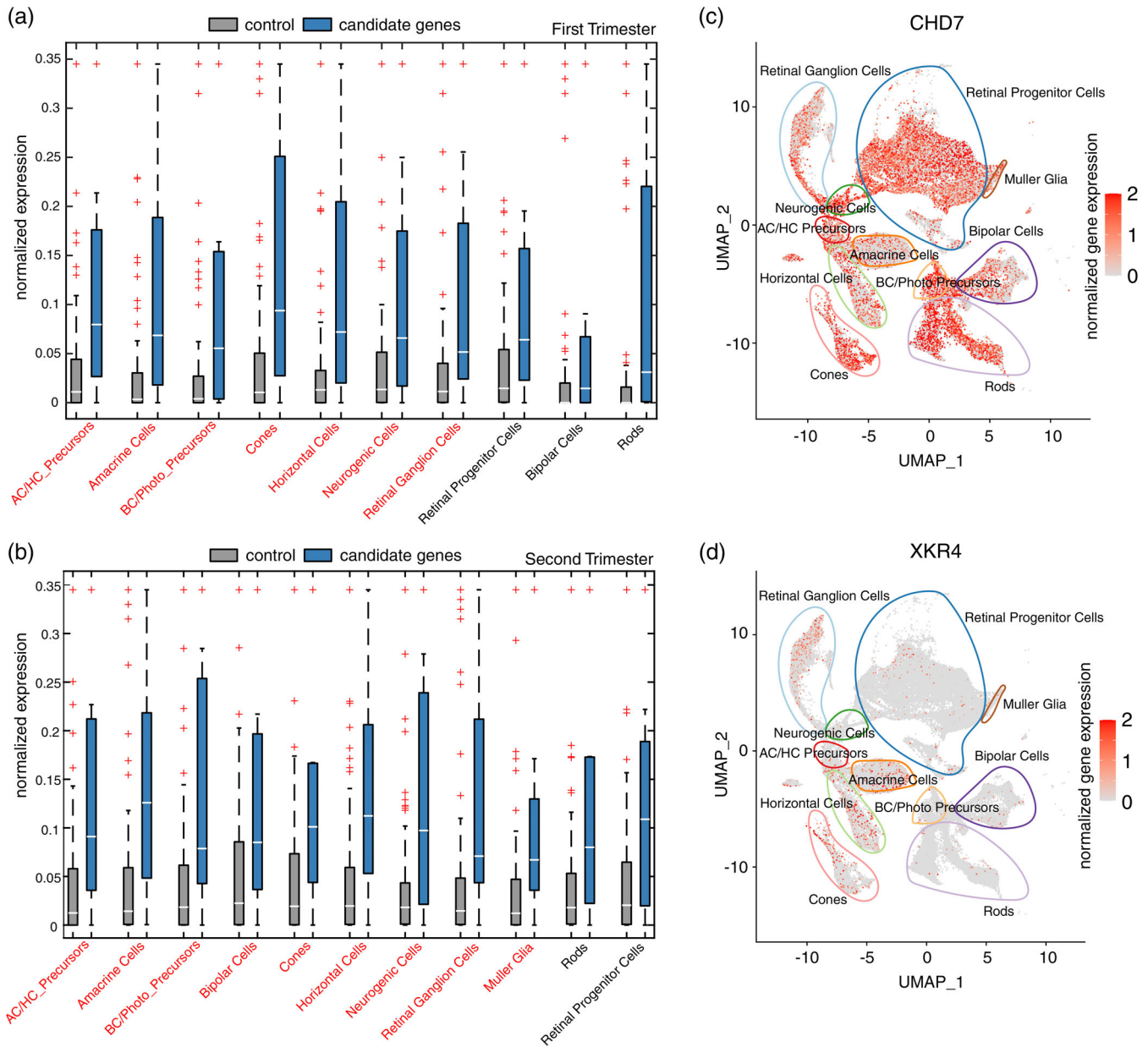
Abbreviations: ALT, alternate; A/M, anophthalmia/microphthalmia; PMID, PubMed unique identifier; REF, reference.

<sup>a</sup>*De novo* not confirmed by Sanger sequencing (proband result was confirmed, but parent samples were not available).

<sup>b</sup>*De novo* confirmed by Sanger sequencing.

Ogmundsdottir, Steingrímsson, & Eysteinnsson, 2019). To derive mechanistic insights, we analyzed single-cell transcriptome data across retina developmental stages (Lu et al., 2020), where the global transcriptomes were profiled in each of the 10 cell types in the developing retina from gestational weeks 9 to 27. As a background control set, we identified a list of 71 genes following the same procedure in the probands (Figure 1), which were affected by rare nonsense variants, without considering their tolerance to gene copy loss. In this way, predicted mutational pathogenicity could be directly associated with functional specificity. We observed that across 7 of the 10 retinal cell types, the identified eight genes in Table 2 displayed significantly increased expression in aggregate relative to the background control genes (false discovery rates, FDRs < 0.05, Wilcoxon rank-sum test, followed by Benjamini-Hochberg correction), and the pattern was

consistent in both the first and second trimesters (Figure 2a and b). Therefore, the single-cell analysis confirmed the increased aggregate expression of the identified genes in retinal development, and the comparisons revealed potential cell types that may be involved in development of A/M. We used the detected genes, *CHD7* and *XKR4*, for example, to reveal potential cell types implicated in A/M. *CHD7* is more ubiquitously expressed across all retinal cell types, despite a stronger pattern in retinal progenitor cells, cones and bipolar cell /photoreceptor precursors (Figure 2c). However, *XKR4* displayed strong specificities in the developing retina in only a few cell types, including retinal ganglion cells, amacrine, horizontal cells, and cones (Figure 2d). Future functional experiments seem warranted to examine the role of *XKR4* in retinal development through its cell-type-specific expression.



**FIGURE 2** Single-cell transcriptome data analysis of candidate genes across retina developmental stages. (a) In the first trimester, the identified eight genes displayed significantly increased expression relative to the background control genes in AC/HC\_Precursors ( $N = 695$ ), amacrine cells ( $N = 253$ ), BC/Photo\_Precursors, cones ( $N = 689$ ), horizontal cells ( $N = 1883$ ), neurogenic cells ( $N = 1123$ ), and retinal ganglion cells ( $N = 6297$ ). The  $p$  values were calculated by Wilcoxon rank-sum test, followed by Benjamini-Hochberg correction. Cell types in red font along the x-axis had a false discovery rate (FDR)  $< 0.05$ . AC/HC, amacrine cell/horizontal cell; BC/Photo, bipolar cell/photoreceptor. (b) In the second trimester, the identified eight genes displayed significantly increased expression relative to the background control genes in AC/HC\_Precursors ( $N = 1039$ ), amacrine cells ( $N = 10,706$ ), BC/Photo\_Precursors ( $N = 1825$ ), bipolar cells ( $N = 5949$ ), cones ( $N = 3985$ ), horizontal cells ( $N = 4618$ ), neurogenic cells ( $N = 2197$ ), retinal ganglion cells ( $N = 2662$ ), and Muller glia ( $N = 343$ ). The  $p$  values were calculated by Wilcoxon rank-sum test, followed by Benjamini-Hochberg correction. Cell types in red font along the x-axis had a false discovery rate (FDR)  $< 0.05$ . (a and b) The bottom and top of the boxes denote the first and third quartiles, respectively. The whiskers indicate the minimal value within 1.5 interquartile range (IQR) of the lower quartile and the maximum value within 1.5 IQR of the upper quartile. The plus symbols represent outliers. The black dashed lines indicate the lower and upper limits of the regions with regular scale. Outliers outside of the black dashed lines are visualized with compressed scale in the regions surrounded by gray lines for better visualization. (c) *CHD7* is ubiquitously expressed across all retinal cell types, despite a stronger pattern in retinal progenitor cells, cones and BC/photo precursors. Red and gray colors represent high and low read number in *CHD7* expression, respectively. Uniform manifold approximation and projection (UMAP) is an efficient dimension reduction algorithm commonly used in single-cell RNA sequencing analysis. (d) *XKR4* displayed strong specificity in the developing retina only in very few cell types, including retinal ganglion cells, amacrine cells, horizontal cells, and cones. Red and gray colors represent high and low *XKR4* expression, respectively.

We investigated the parent of origin for these identified nonsense variants (Table 2). We identified two putative *de novo* variants *VPS13D* (c.C12634T, p.Q4212X) and *CHD7* (c.C2959T, p.R987X). While additional DNA samples were unavailable for verifying the *de novo* status of the *VPS13D* (c.C12634T, p.Q4212X) variant (Family #46), we were able to use Sanger sequencing and successfully confirmed the *de novo* variant c.C2959T, p.R987X in *CHD7* (Family #55, Figure S2). In addition to these *de novo* variants, other variants were observed to be either paternally or maternally transmitted. We additionally performed Sanger sequencing and confirmed the presence of four identified loss-of-function variants in the corresponding probands, including variants in *CHD7* (c.C4480T, p.R1494X), *NUMA1* (c.C1654T, p.Q552X), *TUBGCP3* (c.C1376A, p.S459X), and *FAM192A* (c.C175T, p.Q59X). The *LRP5* (c.C1489T, p.R497X) variant in the proband was not confirmed due to failed PCR. We were not able to perform Sanger validation on *XKR4* (c.C14G, p.S5X) and *KRT31* (c.G331T, p.E111X) due to insufficient DNA remaining post exome sequencing studies.

We implemented three machine learning algorithms to predict potential missense variant effects (Carter et al., 2013; Dong et al., 2015; Kircher et al., 2014). We considered a subset of 86 missense variants receiving consensus deleteriousness predictions from all three algorithms, which represented rare deleterious missense variants of high confidence in the A/M proband cohort (Figure 1). These 86 deleterious variants affected 85 genes. We performed pathway analysis and observed functional enrichment for eye physiologies among these affected genes (Table 3). Mouse mutants of these genes displayed abnormal horizontal cell morphology, and blindness (e.g., *UPS38* and *WNK4*, Table 4). In humans, these 85 identified genes exhibited phenotypic enrichment in hypoplasia of the fovea, the macula and the retina, as well as in opacification of the corneal stroma and optic atrophy (Table 3). We performed the same analyses on genes affected by rare potential missense variant without considering mutational consequences, and the same functional enrichment could not be observed, thereby correlating predicted mutational pathogenicity with functional specificities to eye physiologies. Investigating rare missense variants captured additional genes that could be contributing to A/M. Among these rare deleterious missense variants, seven were likely *de novo* variants, 4 were unknown in origin, and 75 were parentally transmitted (Table S1b) further indicating either the variants were not pathologic or the potential heterogeneity in the mode of inheritance underlying A/M.

We also manually examined each of the identified 85 unique genes (Table S1b) and our literature curation uncovered their overall functions in eye development. Some exemplar genes are shown in Table 4, which were not included as member genes in the above enriched functional ontology categories (Table 3). We particularly note *UPS38* (c.C2281A, p.L761I) and *WNK4* (c.A151G, p.T51A), whose mutants in model organisms displayed a loss of eyes (Table 4) (Shimizu, Goto, Sato, & Shibuya, 2013; Tsai et al., 2019), replicating the expected A/M traits in humans.

**TABLE 3** Functional enrichment summary for missense variants

| Function   | FDR (candidate genes) | FDR (control) |
|--|-----------------------|---------------|
| Detection of visible light (GO:0009584)              | 8.82e−3               | 0.25          |
| Visual perception (GO:0007601)                       | 8.66e−3               | 0.28          |
| Abnormal horizontal cell morphology (MP:0006068)     | 0.05                  | 0.72          |
| Blindness (MP:0002001)                               | 0.048                 | 0.76          |
| Hypoplasia of the fovea (HP:0007750)                 | 7.2e−4                | 0.29          |
| Aplasia/hypoplasia of the macula (HP:0008059)        | 1.51e−3               | 0.41          |
| Aplasia/hypoplasia affecting the retina (HP:0008061) | 1.53e−3               | 0.44          |
| Opacification of the corneal stroma (HP:0007759)     | 1.53e−3               | 0.68          |
| Optic atrophy (HP:0000648)                           | 1.59e−3               | 0.55          |

Abbreviations: FDR, false discovery rate; GO, gene ontology; HP, human phenotype; MP, mammalian phenotype.

Among all identified genes, *PAX6* was of prime interest given its role as a key transcription factor driving eye development (Glaser et al., 1994). In our analysis, a potential deleterious missense variant in the N-terminus of the *PAX6* protein sequence was detected in one A/M proband (potential pathogenic variant was absent from the reference human population). As shown in Figure 3a, this potential pathogenic variant (c.G56A, p.G19E) [refseq ID: NM\_001368911.2] altered the wildtype amino acid from a glycine (G) residue into a glutamic acid (E), localized in the paired DNA binding domain of the transcription factor *PAX6*. To gain mechanistic insights into the mutational impact on this main eye development driver protein, we modeled the 3D structure of *PAX6* (PDB ID: 6PAX) and mapped the p.G19E potential pathogenic variant onto the *PAX6* crystal structure. Thermodynamic analysis of the p.G19E potential pathogenic variant revealed that the mutant allele is predicted to substantially reduce the solvent accessibility associated with the wildtype allele from 113.9% to 66.7%, leading to increased packing density in this local position. As such, the residue-occluded packing density at this position displayed about a 40% increase from 0.16 to 0.22. By burying this mutant amino acid into the protein interior, as opposed to the wildtype residue, which is exposed to the solvent, this potential pathogenic variant in fact destabilized the *PAX6* structure with  $\Delta\Delta G = -3.34$  kcal/mol. We further computed the alteration of vibrational entropy in the *PAX6* crystal structure that resulted from this amino acid change. As shown in Figure 3b, the mutated residue had a significant effect on inducing a rigidification (loss of molecular flexibility) of the *PAX6* structure in the N-terminus.

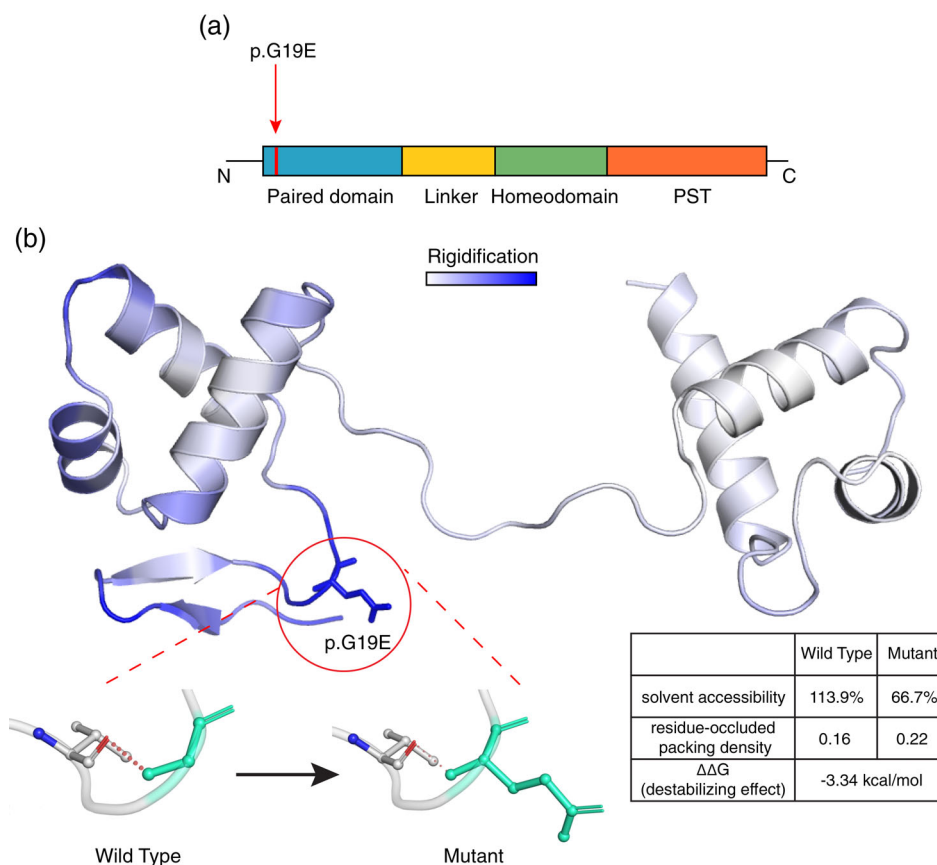
For the three probands with first degree relatives reported as having eye malformations, there were identified missense variants inherited from their affected parents (Table S1b).



**TABLE 4** Selected missense variants in probands with anophthalmia/microphthalmia identified by exome sequencing

| Gene   | [chromosome: position] | [REF/ALT] | Predicted protein change | PMID                      | Eye development function  | Family ID | Origin   | Phenotype   |
|--------|------------------------|-----------|--------------------------|---------------------------|---|-----------|----------|---|
| AGRN   | 1:985283               | G/C       | p.G1582A                 | 17,196,957;<br>28,221,305 | Overexpression showed eye development defects; mutant showed restrictive eye. | #42       | Maternal | Unilateral microphthalmia, isolated                       |
| USP38  | 4:144135410            | C/A       | p.L761I                  | 31,723,061                | Mutant showed loss of eyes  | #60       | Paternal | Unilateral microphthalmia with accompanying malformations |
| WASHC5 | 8:126068971            | C/A       | p.R507L                  | 31,971,710                | Eye anomalies in Ritscher-Schinzel syndrome                                   | #62       | Paternal | Unilateral anophthalmia with accompanying malformations   |
| PAX6   | 11:31824340            | C/T       | p.G19E                   | 7,951,315                 | A key regulator in eye development  | #59       | Maternal | Bilateral microphthalmia, isolated                        |
| COL4A1 | 13:110829240           | C/T       | p.G954E                  | 24,628,545                | Known in A/M  | #36       | De novo  | Bilateral microphthalmia with accompanying malformations  |
| WNK4   | 17:40936497            | A/G       | p.T51A                   | 23,517,227                | Mutant showed loss of eyes in <i>Xenopus</i> embryos                          | #11       | Maternal | Unilateral microphthalmia, isolated                       |

Abbreviations: ALT, alternate; A/M, anophthalmia/microphthalmia; PMID, PubMed unique identifier; REF, reference.



**FIGURE 3** Protein structure analysis of a potential deleterious missense variant in *PAX6*. (a) The gene structure of transcription factor *PAX6*. The potential pathogenic variant (p.G19E) altered the wildtype amino acid from a glycine (G) residue into a glutamic acid (E) in the paired DNA binding domain of *PAX6*. PST represents a carboxyl-terminal transactivation domain rich of proline (P), serine (S), and threonine (T). (b) The p.G19E potential pathogenic variant of *PAX6* had a significant effect in loss of molecular flexibility in the N-terminus. 3D structure of mutated *PAX6*, blue and white colors stand for high and low rigidification. The zoomed in region indicates the differences between wild type and mutant in mutated residue location, red represents hydrogen bonds and light-green represents residues. The table shows the prediction outcomes of interatomic interactions.

Finally, we investigated whether various factors reported during the maternal interview appeared more or less frequently among case infants with versus without the identified gene variants. That is, we compared each of the three groups, 9 with *de novo*, 9 with nonsense, and 48 with nonsense or missense variants to the group of 19 case infants without identified variants (some of the infants overlapped between the three groups). More than 200 factors (such as maternal/paternal demographic information, maternal pregnancy history, lifestyle factors, multivitamin/folic acid intake, and nutrient intake from food) were investigated. The sample sizes were expectedly sparse, with none of the assessed factors observed to occur with higher frequency in one group vs. another (data not shown).

## 4 | DISCUSSION

We employed an agnostic search strategy based on exome sequences in 201 individuals from 67 trio families where the proband had A/M. This search strategy confirmed a few known genes in A/M and also identified novel candidate genes of interest. Contributions of the novel genes identified were orthogonally supported by single-cell transcriptomic data (Lu et al., 2020), molecular pathway analysis, protein structure modeling, and curated literature reports.

We identified 9 nonsense and 86 deleterious missense variants that are absent from gnomAD. Our functional analysis validated their physiological roles in regulating eye development, thereby providing potential mechanistic insights for the findings. We observed recurrent disruptions of *CHD7* in two independent probands, the gene for CHARGE syndrome, in which colobomatous microphthalmia is a component of the syndrome (Haug et al., 2021). For these two probands with a *CHD7* variant, the phenotype of one also included microtia and unilateral cleft lip and palate, and the phenotype of the other did not include any additional malformations. For other candidate genes, two have been implicated in model organisms, such as the loss of eye phenotypes associated with *UPS38* and *WNK4* mutants in mouse or other model organisms. To gain further possible mechanistic insights, we analyzed single-cell transcriptomic data across retinal developmental stages and identified several cell types displaying aggregate expression enrichment of the identified candidate genes, thereby suggesting vulnerable cell types specific to A/M etiologies. However, it is important to note that in addition to the retina, many other eye components are also implicated in A/M, and future inquiry is required to determine A/M-associated cell types in different eye components.

We especially focused on the analysis of a potential pathogenic missense variant in *PAX6*, a key transcription factor driving eye development (Glaser et al., 1994). The association of *PAX6* with A/M was previously established by a case study identifying loss-of-function variants in *PAX6*, and our work leveraging protein structure modeling now provides evidence for the involvement of missense variants in *PAX6* in A/M. Intriguingly, at the same position, mutating the wildtype glycine residue into two different amino acids resulted in two

different eye development disorders: the glycine (G) to tryptophan (W) alteration resulted in aniridia (Wolf et al., 1998), whereas the glycine (G) to glutamic acid (E) was associated with microphthalmia (Figure 3). We postulated that this phenotypic disparity likely resulted from the different side chain structures of the tryptophan and the glutamic acid residues, leading to varying degrees of perturbing the *PAX6* protein structure.

The mutated residue, p.G19E, destabilized the local structure in the paired DNA binding domain of *PAX6*, thereby likely perturbing the *PAX6*-mediated gene regulatory network and, thus, leading to eye developmental disorders. Intriguingly, a previous report also identified several missense variants in *PAX6* associated with bilateral microphthalmia (Williamson et al., 2020), which, despite different locations from our observation, all resided in the paired DNA binding domain of *PAX6* (Figure 3a). Our observation, in conjunction with previous work, indicates that the *PAX6* paired DNA binding domain is a potential pathogenic variant hotspot in bilateral microphthalmia. We also observed that the mutated residue in the proband in our study was maternally transmitted. Further review of maternal interview data indicated blindness in both the proband's mother and a sibling.

In addition, we conducted exploratory analyses of indels that could potentially affect splice sites. However, none of the identified splicing events would be considered reportable based on ACMG practice guidelines (Miller et al., 2021; Richards et al., 2015). While this study is the largest sequencing effort for A/M and the data derived from population-based case ascertainment, it is still of relatively small size. Such limited sample size obviated a robust exploration of possible pregnancy exposures that may, in combination with gene variants, modify risks of A/M. Further, the analysis was based on exome sequence data and not whole genome sequence data rendering us unable to explore potential contributions from noncoding regions of the genome. Therefore, future work with larger studies will help confirm the candidate genes identified here and pursue some of the hypothetical directions offered by our observations.

## AUTHOR CONTRIBUTIONS

Gary M. Shaw supervised the study. Lynn M. Almli, Marilyn L. Browne, Cynthia J. Curry, Marcia L. Feldkamp, Richard H. Finnell, Mary M. Jenkins, Wendy N. Nembhard, Andrew F. Olshan, Jennita Reefhuis, Paul A. Romitti, Gary M. Shaw, Martha M. Werler, and NBDPS, contributed to the NBDPS study design and data and specimen acquisition. Mike Bamshad, Elizabeth E. Blue, Lawrence C. Brody, Jessica X. Chong, Daniel McGoldrick, James Mullikin, Deborah A. Nickerson, Faith Pangilinan, NISC Comparative Sequencing Program, and UWCMG conducted experimental studies. Jingjing Li, Chen Ma, Yuejun Jessie Wang, and Wei Yang analyzed the data, and Mike Bamshad, Elizabeth E. Blue, Lawrence C. Brody, Jessica X. Chong, Jingjing Li, Chen Ma, Daniel McGoldrick, Deborah A. Nickerson, Gary M. Shaw, Yuejun Jessie Wang, Wei Yang, and UWCMG interpreted the data. Jingjing Li, Yuejun Jessie Wang, and Wei Yang created the

initial draft of the article and reviewed the literature and Lynn M. Almlı, Cynthia J. Curry, Mary M. Jenkins, Faith Pangilinan, Paul A. Romitti, and Gary M. Shaw critically revised the article. All authors read and approved the final article.

## ACKNOWLEDGMENTS

This work was supported through cooperative agreements under PA 96043, PA 02081, and FOA DD09-001 from the Centers for Disease Control and Prevention to the Centers for Birth Defects Research and Prevention participating in the NBDPS. Sequencing data was reprocessed and analyzed by the University of Washington Center for Mendelian Genomics and was funded by the National Human Genome Research Institute (NHGRI) and National Heart, Lung, and Blood Institute grants UM1 HG006493 and U24 HG008956. This work was supported by the Division of Intramural Research of NHGRI, National Institutes of Health. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention, the National Institutes of Health, or the California Department of Public Health.

## CONFLICT OF INTEREST

Dr. Richard Finnell formerly held a leadership position in TeratOmic Consulting LLC, a now defunct organization. Further, he serves on the editorial board of the journal Reproductive and Developmental Medicine and receives travel funding to attend quarterly editorial board meetings.

## DATA AVAILABILITY STATEMENT

The data are from a CDC-funded multi-site study and not publicly available at this time due to privacy or ethical restrictions.

## ORCID

Gary M. Shaw  <https://orcid.org/0000-0001-7438-4914>

## REFERENCES

- Bardakjian, T. M., & Schneider, A. (2011). The genetics of anophthalmia and microphthalmia. *Current Opinion in Ophthalmology*, 22(5), 309–313. <https://doi.org/10.1097/ICU.0b013e328349b004>
- Bergman, J. E., Janssen, N., van der Sloot, A. M., de Walle, H. E., Schoots, J., Rendtorff, N. D., ... Hofstra, R. M. (2012). A novel classification system to predict the pathogenic effects of CHD7 missense variants in CHARGE syndrome. *Human Mutation*, 33(8), 1251–1260. <https://doi.org/10.1002/humu.22106>
- Blazer, S., Zimmer, E. Z., Mezer, E., & Bronshtein, M. (2006). Early and late onset fetal microphthalmia. *American Journal of Obstetrics and Gynecology*, 194(5), 1354–1359. <https://doi.org/10.1016/j.ajog.2005.11.010>
- Carter, H., Douville, C., Stenson, P. D., Cooper, D. N., & Karchin, R. (2013). Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics*, 14(Suppl 3), S3. <https://doi.org/10.1186/1471-2164-14-S3-S3>
- Chao, R., Nevin, L., Agarwal, P., Riemer, J., Bai, X., Delaney, A., ... Slavotinek, A. (2010). A male with unilateral microphthalmia reveals a role for TMX3 in eye development. *PLoS One*, 5(5), e10565. <https://doi.org/10.1371/journal.pone.0010565>
- Dash, S., Brastrom, L. K., Patel, S. D., Scott, C. A., Slusarski, D. C., & Lachke, S. A. (2020). The master transcription factor SOX2, mutated in anophthalmia/microphthalmia, is post-transcriptionally regulated by the conserved RNA-binding protein RBM24 in vertebrate eye development. *Human Molecular Genetics*, 29(4), 591–604. <https://doi.org/10.1093/hmg/ddz278>
- Deml, B., Reis, L. M., Lemyre, E., Clark, R. D., Kariminejad, A., & Semina, E. V. (2016). Novel mutations in PAX6, OTX2 and NDP in anophthalmia, microphthalmia and coloboma. *European Journal of Human Genetics*, 24(4), 535–541. <https://doi.org/10.1038/ejhg.2015.155>
- Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., & Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human Molecular Genetics*, 24(8), 2125–2137. <https://doi.org/10.1093/hmg/ddu733>
- Du, J., Aleff, R. A., Soragni, E., Kalari, K., Nie, J., Tang, X., ... Wieben, E. D. (2015). RNA toxicity and missplicing in the common eye disease fuchs endothelial corneal dystrophy. *Journal of Biological Chemistry*, 290(10), 5979–5990. <https://doi.org/10.1074/jbc.M114.621607>
- Ergun, S. G., Akay, G. G., Ergun, M. A., & Percin, E. F. (2017). LRP5-linked osteoporosis-pseudoglioma syndrome mimicking isolated microphthalmia. *European Journal of Medical Genetics*, 60(3), 200–204. <https://doi.org/10.1016/j.ejmg.2017.01.007>
- Fantes, J., Ragge, N. K., Lynch, S. A., McGill, N. I., Collin, J. R., Howard-Peebles, P. N., ... FitzPatrick, D. R. (2003). Mutations in SOX2 cause anophthalmia. *Nature Genetics*, 33(4), 461–463. <https://doi.org/10.1038/ng1120>
- García-Llorca, A., Aspelund, S. G., Ogmundsdottir, M. H., Steingrımsson, E., & Eysteınsson, T. (2019). The microphthalmia-associated transcription factor (Mitf) gene and its role in regulating eye function. *Scientific Reports*, 9(1), 15386. <https://doi.org/10.1038/s41598-019-51819-0>
- Glaser, T., Jepeal, L., Edwards, J. G., Young, S. R., Favor, J., & Maas, R. L. (1994). PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nature Genetics*, 7(4), 463–471. <https://doi.org/10.1038/ng0894-463>
- Graw, J. (2019). Mouse models for microphthalmia, anophthalmia and cataracts. *Human Genetics*, 138(8–9), 1007–1018. <https://doi.org/10.1007/s00439-019-01995-w>
- Harding, P., Brooks, B. P., FitzPatrick, D., & Moosajee, M. (2020). Anophthalmia including next-generation sequencing-based approaches. *European Journal of Human Genetics*, 28(3), 388–398. <https://doi.org/10.1038/s41431-019-0479-1>
- Harding, P., & Moosajee, M. (2019). The molecular basis of human anophthalmia and microphthalmia. *Journal of Developmental Biology*, 7(3), 16. <https://doi.org/10.3390/jdb7030016>
- Haug, P., Koller, S., Maggi, J., Lang, E., Feil, S., Włodarczyk, A., ... Berger, W. (2021). Whole exome sequencing in coloboma/microphthalmia: Identification of novel and recurrent variants in seven genes. *Genes (Basel)*, 12(1), 65. <https://doi.org/10.3390/genes12010065>
- Jenkins, M. M., Almlı, L. M., Pangilinan, F., Chong, J. X., Blue, E. E., Shapira, S. K., ... National Birth Defects Prevention Study. (2019). Exome sequencing of family trios from the National Birth Defects Prevention Study: Tapping into a rich resource of genetic and environmental data. *Birth Defects Research*, 111(20), 1618–1632. <https://doi.org/10.1002/bdr2.1554>
- Kircher, M., Witten, D. M., Jain, P., O’Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative

- pathogenicity of human genetic variants. *Nature Genetics*, 46(3), 310–315. <https://doi.org/10.1038/ng.2892>
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., ... Ma'ayan, A. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1), W90–W97. <https://doi.org/10.1093/nar/gkw377>
- Kumar, S., Sandell, L. L., Trainor, P. A., Koentgen, F., & Duester, G. (2012). Alcohol and aldehyde dehydrogenases: Retinoid metabolic effects in mouse knockout models. *Biochimica et Biophysica Acta*, 1821(1), 198–205. <https://doi.org/10.1016/j.bbali.2011.04.004>
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., ... Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285–291. <https://doi.org/10.1038/nature19057>
- Li, G., Jin, D., & Zhong, T. P. (2019). Tubgcp3 is required for retinal progenitor cell proliferation during zebrafish development. *Frontiers in Molecular Neuroscience*, 12, 126. <https://doi.org/10.3389/fnmol.2019.00126>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Lu, Y., Shiau, F., Yi, W., Lu, S., Wu, Q., Pearson, J. D., ... Clark, B. S. (2020). Single-cell analysis of human retina identifies evolutionarily conserved and species-specific mechanisms controlling development. *Developmental Cell*, 53(4), 473–491 e479. <https://doi.org/10.1016/j.devcel.2020.04.009>
- Mai, C. T., Isenburg, J. L., Canfield, M. A., Meyer, R. E., Correa, A., Alverson, C. J., ... National Birth Defects Prevention Network. (2019). National population-based estimates for major birth defects, 2010–2014. *Birth Defects Research*, 111(18), 1420–1435. <https://doi.org/10.1002/bdr2.1589>
- Matias-Perez, D., Garcia-Montano, L. A., Cruz-Aguilar, M., Garcia-Montalvo, I. A., Nava-Valdez, J., Barragan-Arevalo, T., ... Zenteno, J. C. (2018). Identification of novel pathogenic variants and novel genotype-phenotype correlations in Mexican subjects with microphthalmia and/or anophthalmia by next-generation sequencing. *Journal of Human Genetics*, 63(11), 1169–1180. <https://doi.org/10.1038/s10038-018-0504-1>
- Meijer, I. A. (1993). VPS13D movement disorder. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, G. Mirzaa, & A. Amemiya (Eds.), *GeneReviews*(R). Seattle (WA): University of Washington, Seattle.
- Miller, D. T., Lee, K., Chung, W. K., Gordon, A. S., Herman, G. E., Klein, T. E., ... Group, A. S. F. W. (2021). ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 23(8), 1381–1390. <https://doi.org/10.1038/s41436-021-01172-3>
- Pandurangan, A. P., Ochoa-Montano, B., Ascher, D. B., & Blundell, T. L. (2017). SDM: A server for predicting effects of mutations on protein stability. *Nucleic Acids Research*, 45(W1), W229–W235. <https://doi.org/10.1093/nar/gkx439>
- Papakostas, T. D., Morrison, M. A., Lane, A. M., Awh, C., DeAngelis, M. M., Gragoudas, E. S., & Kim, I. K. (2018). Genetic risk factors for radiation vasculopathy. *Investigative Ophthalmology and Visual Science*, 59(3), 1547–1553. <https://doi.org/10.1167/iovs.17-22791>
- Patel, A., & Sowden, J. C. (2019). Genes and pathways in optic fissure closure. *Seminars in Cell and Developmental Biology*, 91, 55–65. <https://doi.org/10.1016/j.semcdb.2017.10.010>
- Plaisancie, J., Ceroni, F., Holt, R., Zazo Seco, C., Calvas, P., Chassaing, N., & Ragge, N. K. (2019). Genetics of anophthalmia and microphthalmia. Part 1: Non-syndromic anophthalmia/microphthalmia. *Human Genetics*, 138(8–9), 799–830. <https://doi.org/10.1007/s00439-019-01977-y>
- Rasmussen, S. A., Olney, R. S., Holmes, L. B., Lin, A. E., Keppler-Noreuil, K. M., Moore, C. A., & National Birth Defects Prevention Study. (2003). Guidelines for case classification for the National Birth Defects Prevention Study. *Birth Defects Research. Part A: Clinical and Molecular Teratology*, 67(3), 193–201. <https://doi.org/10.1002/bdra.10012>
- Reefhuis, J., Gilboa, S. M., Anderka, M., Browne, M. L., Feldkamp, M. L., Hobbs, C. A., ... National Birth Defects Prevention Study. (2015). The national birth defects prevention study: A review of the methods. *Birth Defects Research. Part A: Clinical and Molecular Teratology*, 103(8), 656–669. <https://doi.org/10.1002/bdra.23384>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... Committee, A. L. Q. A. (2015). 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. *Genetics in Medicine*, 17(5), 405–424. <https://doi.org/10.1038/gim.2015.30>
- Richardson, R., Tracey-White, D., Webster, A., & Moosajee, M. (2017). The zebrafish eye—a paradigm for investigating human ocular genetics. *Eye (London, England)*, 31(1), 68–86. <https://doi.org/10.1038/eye.2016.198>
- Rodrigues, C. H., Pires, D. E., & Ascher, D. B. (2018). DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Research*, 46(W1), W350–W355. <https://doi.org/10.1093/nar/gky300>
- Shaw, G. M., Carmichael, S. L., Yang, W., Harris, J. A., Finnell, R. H., & Lammer, E. J. (2005). Epidemiologic characteristics of anophthalmia and bilateral microphthalmia among 2.5 million births in California 1989–1997. *American Journal of Medical Genetics Part A*, 137(1), 36–40. <https://doi.org/10.1002/ajmg.a.30840>
- Shimizu, M., Goto, T., Sato, A., & Shibuya, H. (2013). WNK4 is an essential effector of anterior formation in FGF signaling. *Genes to Cells*, 18(6), 442–449. <https://doi.org/10.1111/gtc.12048>
- Sinn, R., & Wittbrodt, J. (2013). An eye on eye development. *Mechanisms of Development*, 130(6–8), 347–358. <https://doi.org/10.1016/j.mod.2013.05.001>
- Skalicky, S. E., White, A. J., Grigg, J. R., Martin, F., Smith, J., Jones, M., ... Jamieson, R. V. (2013). Microphthalmia, anophthalmia, and coloboma and associated ocular and systemic features: Understanding the spectrum. *JAMA Ophthalmol*, 131(12), 1517–1524. <https://doi.org/10.1001/jamaophthalmol.2013.5305>
- Tsai, I. C., Adams, K. A., Tzeng, J. A., Shennib, O., Tan, P. L., & Katsanis, N. (2019). Genome-wide suppressor screen identifies USP35/USP38 as therapeutic candidates for ciliopathies. *JCI Insight*, 4(22), e130516. <https://doi.org/10.1172/jci.insight.130516>
- Verma, A. S., & Fitzpatrick, D. R. (2007). Anophthalmia and microphthalmia. *Orphanet Journal of Rare Diseases*, 2, 47. <https://doi.org/10.1186/1750-1172-2-47>
- Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16), e164. <https://doi.org/10.1093/nar/gkq603>
- Williamson, K. A., & FitzPatrick, D. R. (2014). The genetic architecture of microphthalmia, anophthalmia and coloboma. *European Journal of Medical Genetics*, 57(8), 369–380. <https://doi.org/10.1016/j.ejmg.2014.05.002>
- Williamson, K. A., Hall, H. N., Owen, L. J., Livesey, B. J., Hanson, I. M., Adams, G. G. W., ... FitzPatrick, D. R. (2020). Recurrent heterozygous PAX6 missense variants cause severe bilateral microphthalmia via predictable effects on DNA-protein interaction. *Genetics in*

*Medicine*, 22(3), 598–609. <https://doi.org/10.1038/s41436-019-0685-9>

Wolf, M. T., Lorenz, B., Winterpacht, A., Drechsler, M., Schumacher, V., Royer-Pokora, B., ... Wildhardt, G. (1998). Ten novel mutations found in aniridia. *Human Mutation*, 12(5), 304–313. [https://doi.org/10.1002/\(SICI\)1098-1004\(1998\)12:5<304::AID-HUMU3>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1098-1004(1998)12:5<304::AID-HUMU3>3.0.CO;2-D)

Zagozewski, J. L., Zhang, Q., & Eisenstat, D. D. (2014). Genetic regulation of vertebrate eye development. *Clinical Genetics*, 86(5), 453–460. <https://doi.org/10.1111/cge.12493>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Li, J., Yang, W., Wang, Y. J., Ma, C., Curry, C. J., McGoldrick, D., Nickerson, D. A., Chong, J. X., Blue, E. E., Mullikin, J. C., Reefhuis, J., Nembhard, W. N., Romitti, P. A., Werler, M. M., Browne, M. L., Olshan, A. F., Finnell, R. H., Feldkamp, M. L., Pangilinan, F., Almli, L. M., Bamshad, M. J., Brody, L. C., Jenkins, M. M., Shaw, G. M., NISC Comparative Sequencing Program University of Washington Center for Mendelian Genomics the National Birth Defects Prevention Study (2022). Exome sequencing identifies genetic variants in anophthalmia and microphthalmia. *American Journal of Medical Genetics Part A*, 188A:2376–2388. <https://doi.org/10.1002/ajmg.a.62874>