TITLE

New *GJA8* variants and phenotypes highlight its critical role in a broad spectrum of eye anomalies

AUTHORS

Fabiola Ceroni[1], Domingo Aguilera-Garcia[2,3], Nicolas Chassaing[4,5], Dorine Arjanne Bax[1], Fiona Blanco-Kelly[2,3,6,7], Patricia Ramos[2,3], Maria Tarilonte[2,3], Cristina Villaverde[2,3], Luciana Rodrigues Jacy da Silva[2,3], Maria Juliana Ballesta-Martínez[8], Maria Jose Sanchez-Soler[8], Richard Holt[1], Lisa Cooper-Charles[9], Jonathan Bruty[9], Yvonne Wallis[9], Dominic McMullan[9], Jonathan Hoffman[10], David Bunyan[11], Alison Stewart[12], Helen Stewart[6], Katherine Lachlan[13,14], DDD Study[15], Alan Fryer[16], Victoria McKay[16], Joëlle Roume[17], Pascal Dureau[18], Anand Saggar[19], Michael Griffiths[9], Patrick Calvas[4,5], Carmen Ayuso[2,3], Marta Corton*[2,3], Nicola Ragge*[1,10]

*joint senior authors

AFFILIATIONS

1 Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, UK;

2 Genetics Service, Instituto de Investigación Sanitaria - Fundación Jiménez Díaz University Hospital- Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain;

3 Centre for Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Madrid, Spain;

4 Service de Génétique Médicale, Hôpital Purpan, CHU Toulouse, France;

5 UMR 1056 Inserm - Université de Toulouse, Toulouse, France;

6 Oxford Centre for Genomic Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, UK;

7 Institute of Ophthalmology, University College London, London, UK;

8 Medical Genetics Department, University Hospital Virgen de la Arrixaca, Murcia, Spain;

9 West Midlands Regional Genetics Laboratory, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK;

10 West Midlands Regional Clinical Genetics Service and Birmingham Health Partners, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK;

11 Wessex Regional Genetics Laboratory, Salisbury NHS Foundation Trust, Salisbury, UK;

12 Sheffield Clinical Genetics Department, Northern General Hospital, Sheffield, UK;

13 Wessex Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Princess Anne Hospital, Southampton, UK;

14 Human Genetics & Genomic Medicine, University of Southampton, Southampton General Hospital, Southampton, UK;

15 Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK;

16 Cheshire and Merseyside Genetics Service, Liverpool Women's NHS Foundation Trust, Liverpool, UK;

17 Department of Clinical Genetics, centre de référence "AnDDI rares", Poissy Hospital GHU PIFO, Poissy, France;

18 Fondation ophtalmologique Adolphe-de-Rothschild, Paris, France;

19 Clinical Genetics Unit, St Georges University of London, London, UK.

Corresponding author: Professor Nicola Ragge, MD DM FRCP FRCOphth FRCPCH

Faculty of Health and Life Sciences, Oxford Brookes University, Gipsy Lane, Oxford, OX3 0BP, UK; West Midlands Regional Clinical Genetics Service and Birmingham Health Partners, Birmingham Women's and Children's NHS Foundation Trust, Mindelsohn Way, Birmingham, B15 2TG, UK.

Tel: +44 1865 484413

Fax: +44 1865 742177

E-mail: nragge@brookes.ac.uk; nicola.ragge@bwnft.nhs.uk

Acknowledgements

We would like to thank the patients and their families for their participation in our study. We are grateful to Suzanne Broadgate for assisting with the coordination of the project. We thank the Centrum Menselijke Erfelijkheid laboratory, Leuven, Belgium, for the diagnostic *PAX2* testing. We also thank Dr Jean-Philippe Bault for the ultrasound examinations performed on family 2. This work was supported by grants from Baillie Gifford, Visually Impaired Children Taking Action (VICTA) (http://www.victa.org.uk/), Microphthalmia, Anophthalmia, Coloboma Support (MACS) (www.macs.org.uk), Oxford Brookes University Central Research Fund, Retina France, Fondation Maladies Rares, Spanish Institute of Health Carlos III (CP12/03256), Spanish Ministry of Economy and Competitiveness (SAF2013-46943-R), Mutua Madrileña Foundation and Cátedra de Patrocinio UAM-IIS-FJD of Genomic Medicine (University Autónoma of Madrid). The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003], see www.ddduk.org/access.html for full acknowledgement.

ABSTRACT

GJA8 encodes connexin 50 (Cx50), a transmembrane protein involved in the formation of lens gap junctions. *GJA8* mutations have been linked to early onset cataracts in humans and animal models. In mice, missense mutations and homozygous *Gja8* deletions lead to smaller lenses and microphthalmia in addition to cataract, suggesting *Gja8* may play a role in both lens development and ocular growth.

Following screening of *GJA8* in a cohort of 426 individuals with severe congenital eye anomalies, primarily anophthalmia, microphthalmia and coloboma, we identified four known (p.(Thr39Arg), p.(Trp45Leu), p.(Asp51Asn) and p.(Gly94Arg)) and two novel (p.(Phe70Leu) and p.(Val97Gly)) likely pathogenic variants in seven families. Five of these co-segregated with cataracts and microphthalmia, whereas the variant p.(Gly94Arg) was identified in an individual with congenital aphakia, sclerocornea, microphthalmia and coloboma. Four missense variants of unknown or unlikely clinical significance were also identified. Furthermore, the screening of *GJA8* structural variants in a subgroup of 188 individuals identified heterozygous 1q21 microdeletions in five families with coloboma and other ocular and/or extraocular findings. However, the exact genotype-phenotype correlation of these structural variants remains to be established.

Our data expand the spectrum of *GJA8* variants and associated phenotypes, confirming the importance of this gene in early eye development.

Key words: Cataract, microphthalmia, coloboma, congenital aphakia, *GJA8,* Cx50

INTRODUCTION

Anophthalmia (absent eye), microphthalmia (small eye) and coloboma (optic fissure closure defects), collectively referred to as AMC, form a spectrum of developmental eye disorders, with an overall estimated incidence of 6-13 per 100,000 births (Shah et al., 2011; Skalicky et al., 2013). AMC can occur alone or in combination with other ocular anomalies, such as early onset cataract and anterior segment dysgenesis (ASD). They are associated with extraocular features in just over half of cases (Shah et al., 2012) and can form part of a syndrome (Slavotinek, 2011). The etiology of AMC is characterised by marked genetic heterogeneity. This reflects the complexity underlying eye morphogenesis, a conserved process that requires a series of highly coordinated events, both at the molecular and the structural level, and is tightly regulated by a network of transcription factors, extracellular signaling molecules, cell-cycle regulators and adhesion proteins (Reis and Semina, 2015).

Connexins (Cxs) are a homogeneous family of transmembrane proteins with a crucial role in intercellular communication. They present a conserved topology, which consists of four transmembrane α-helices (TM1-TM4) joined by two extracellular loops (ECL1 and ECL2) and one cytoplasmic loop (ICL), flanked by a short cytoplasmic N-terminal domain (NT) and a long cytoplasmic and less conserved C-terminal domain (CT). Cxs oligomerise in hexameric complexes called connexons, and allow the transmembrane passage of ions and small solutes (≤1 kDa). Connexons can function independently as hemichannels (HCs) or they can dock with their counterparts on the juxtaposed cell to form a gap junction channel (GJC), enabling the direct exchange of small molecules. Given their role in cell-cell communication and tissue homeostasis, Cxs have been implicated in a variety of biological and pathological processes (Pfenniger et al., 2011; García et al., 2016), including myelin-related diseases (Cx32 and Cx47), heart malformations and arrhythmia (Cx40), hearing loss

and skin disorders (Cx26, Cx30, Cx30.3 and Cx31), oculodentodigital dysplasia, a syndrome also involving microphthalmia, microcornea, cataract and/or spherophakia (Cx43), and early onset cataract (Cx46 and Cx50).

As with AMC, developmental or early onset cataracts are a clinically heterogeneous group of disorders, presenting as isolated anomalies or part of a syndrome. More than 110 genes have been implicated in congenital cataracts (Gillespie et al., 2014), with mutations in Cxs accounting for around 16% of cases with a known genetic cause (Shiels and Hejtmancik, 2017). Since the lens does not have any blood supply, it strongly depends on an extensive network of GJCs for the intercellular communications that are critical for its development and the maintenance of its transparency. The most abundant Cxs in the lens are Cx46 and Cx50, which can also form mixed hexamers. Cx46, encoded by *GJA3,* is expressed only in fiber cells, whereas Cx50, encoded by *GJA8,* is present throughout the lens.

Genetic studies in mice have demonstrated that the homozygous knockout of either *Gja3* or *Gja8* leads to cataracts, but with important phenotypic differences. The deletion of *Gja3* causes severe progressive nuclear cataracts, but does not alter ocular growth (Gong et al., 1997). In contrast, *Gja8*-null mice develop milder nuclear cataracts at an early postnatal age and exhibit significantly smaller lenses and microphthalmia (White et al., 1998; Rong et al., 2002), indicating that the two Cxs have overlapping, but distinct functions. In addition, the targeted replacement of *Gja8* with *Gja3* (Cx50KI46 knockin mice) prevents the loss of crystalline solubility, but not the postnatal growth defect resulting from the *Gja8* deletion (White, 2002), confirming the functional diversity of the two proteins and the involvement of *Gja8* in the control of normal ocular growth. This is also supported by mouse lines carrying missense mutations in *Gja8* (Steele et al., 1998; Graw et al., 2001; Chang et al., 2002; Xia et al., 2012; Berthoud et al., 2013) and by rabbit models with CRISPR-Cas9 mediated *GJA8* knockout (Yuan et al., 2016): both develop cataracts, microphthalmia and smaller lenses.

Moreover, severe cataracts and small lenses have also been observed in transgenic mice overexpressing *Gja8* (Chung et al., 2007), indicating that any significant dysregulation of *Gja8* could be deleterious for eye development.

In humans, missense and frameshift mutations in *GJA8* (OMIM 600897) have been associated with cataracts (Beyer et al., 2013; Yu et al., 2016). Rarely, the phenotype also includes additional ocular abnormalities, mainly microcornea and iris hypoplasia (Devi and Vijayalakshmi, 2006; Hansen et al., 2007; Hu et al., 2010; Sun et al., 2011; Prokudin et al., 2014; Ma et al., 2016), but in a few cases also microphthalmia (Ma et al., 2016) and sclerocornea (Ma et al., 2018). Interestingly, defects in the formation of the lens have also been observed (Ma et al., 2018). The cataracts described in these individuals vary in both their location (e.g., nuclear, zonular, lamellar or total) and appearance (e.g., total, pulverulent or dense). The mutations are predominantly heterozygous and only few homozygous variants have been reported, all in consanguineous families (Ponnam et al., 2007; Schmidt et al., 2008 Ponnam et al., 2013; Ma et al., 2016). These pathogenic variants lead to amino acid alterations distributed throughout the protein (Yu et al., 2016), although mostly localised between the domains TM1 and TM2. They are predicted to affect protein function through various mechanisms, such as by inducing misfolding and/or mislocalisation or by altering channel properties (Beyer et al., 2013).

Copy number variants (CNVs) in the distal region of chromosome 1q21 and including *GJA8* are rare in the general population, but have recurrently been identified in individuals with a broad range of different clinical diagnoses (Brunetti-Pierri et al., 2008; Mefford et al., 2008; Stefansson et al., 2008 Bernier et al., 2016). These primarily include developmental delay, microcephaly and psychiatric disorders, although the enrichment of 1q21 CNVs in individuals with these disorders could partly be related to ascertainment bias. However, some cases also have eye anomalies, such as cataracts (Brunetti-Pierri et al., 2008; Mefford et al.,

2008; Rosenfeld et al., 2012; Bernier et al., 2016; Ha et al., 2016), microphthalmia (Mefford et al., 2008) and coloboma (Brunetti-Pierri et al., 2008). Among the genes affected by these recurrent microdeletions/microduplications, *GJA8* represents a strong candidate for the ocular anomalies described in some of the 1q21 CNV carriers.

To investigate further the importance of *GJA8* in human eye morphogenesis and provide a better understanding of the range of developmental ocular anomalies associated with mutations in this gene, we screened *GJA8* in a cohort of 426 unrelated patients (304 UK, 121 Spanish and 1 large French pedigree) with congenital eye anomalies in the AMC spectrum. Two novel and four known likely pathogenic sequence variants were identified in seven families, with one variant being present in two unrelated families. This expands the catalogue of *GJA8* variants likely to be contributing to eye anomalies and the spectrum of phenotypes associated with this gene. Moreover, we also identified heterozygous 1q21 microdeletions including the gene *GJA8* in five additional families, although the pathogenicity of these variants remains to be established.

MATERIALS AND METHODS

Cohort description.

A cohort of UK, Spanish and French families with AMC (Supplementary Table 1) was analysed for *GJA8* variants. The UK families (n=304) were recruited as part of a national 'Genetics of Eye and Brain anomalies' study, approved by the Cambridge East Ethics Committee (04/Q0104/129) and had not received a genetic diagnosis. Family 5 was also recruited into the Deciphering Developmental Disorders (DDD) Study, which has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC). The UK families consisted of 55 probands with anophthalmia, 205 with microphthalmia and 44 with other anomalies within the AMC spectrum; 168 individuals were bilaterally affected and 160 had extra-ocular anomalies. The Spanish families (n=121) consisted of 6 individuals with anophthalmia, 42 with microphthalmia and 73 with other anomalies within the AMC spectrum; 100 individuals were bilaterally affected and 41 had extra-ocular anomalies. They were consented for genetic studies approved by the Ethics Committee of the University Hospital Fundación Jiménez Díaz (FJD, Madrid, Spain) and according to the tenets of the Declaration of Helsinki. The four-generation French pedigree consisted of 15 individuals with congenital cataracts and microphthalmia and consented for the study during their clinical treatment.

Identification of sequence and structural variants in *GJA8*.

The human gene *GJA8* presents one isoform (NM_005267.4), comprising of two exons, with the coding sequence (CDS) entirely contained within exon 2. Sequence variants in the CDS were detected using a combination of Next-Generation Sequencing (NGS) methods and direct sequencing: 35 patients were screened by whole exome sequencing (WES), 207 patients using different targeted NGS panels of eye development genes including *GJA8*, and 184 patients by Sanger sequencing, which was also used to validate NGS findings and check

family segregation. Additionally, CNV data was available for 188 of these patients: 151 individuals (96 UK and 55 Spanish) had been assessed by array-based Comparative Genomic Hybridization (aCGH), with resolutions ranging from 44 kb to 244 kb, whereas for 37 Spanish individuals, CNVs were detected from NGS data using a read depth comparison approach. A detailed description of the different methods can be found in the Supplementary Materials and Supplementary Table 1. The genomic coordinates of the sequence and structural variants are reported according to Build GRCh37/hg19. The allelic frequencies of the sequence mutations were obtained from the Genome Aggregation Database (gnomAD, http://gnomad.broadinstitute.org/) (Lek et al., 2016). For each variant of interest, amino acid conservation across species was visually inspected using the Vertebrate Multiz Alignment & Conservation (100 Species) track from the UCSC Genome Browser. Three conservation scores were annotated using the database dbNSFP v.3.3 (Liu et al., 2016), specifically the GERP++ Rejected Substitutions (RS) score (Davydov et al., 2010), phyloP 100way_vertebrate score (Siepel et al., 2006) and phastCons 100way_vertebrate score (Siepel et al., 2005). Putative functional effects of amino acid substitutions were evaluated with the *in silico* tools SIFT (Kumar et al., 2009) and PolyPhen-2 (Adzhubei et al., 2010).

Validation of mosaicisms and CNVs.

In order to assess potential mosaicism and independently validate aCGH findings, we developed Digital Droplet PCR (ddPCR) assays for the sequence variant in family 1 and three of the *GJA8* microdeletions identified (families 12, 13 and 14) (Supplementary Materials). ddPCR assays were performed using a ddPCR QX200 System (Bio-Rad Laboratories). Primers and Taqman probes were specifically designed for the *GJA8* variant p.(Thr39Arg) using a custom Applied Biosystems TaqMan SNP Genotyping Assay (Thermo Fisher Scientific). For CNV analysis, commercial Taqman Copy Number assays (Thermo Fisher Scientific) were used for exon 2 of *GJA8* and a reference gene (human RNase P gene).

RESULTS

Point mutations identified in *GJA8*.

Screening of the *GJA8* coding region in our cohort of 426 individuals with AMC detected 10 missense variants in 11 unrelated families (Tables 1 and 2). For each missense variant, the amino acid conservation across species is shown in the Supplementary Fig. 1. Taking into account the segregation patterns, the frequency of the variants in public databases of unaffected individuals (Table 2), *in silico* predictions of functional effects and previous reports from the literature (Table 2), as suggested by (Richards et al., 2015), six of these variants (p.(Thr39Arg), p.(Trp45Leu), p.(Asp51Asn), p.(Phe70Leu), p.(Gly94Arg) and p.(Val97Gly)) were considered likely causative, giving a frequency of 1.6% of independent individuals with AMC conditions (7/426) carrying likely pathogenic *GJA8* sequence variants.

In family 1 (Fig. 1a), the heterozygous variant p.(Thr39Arg) (NM_005267.4:c.116C>G) was identified in the male proband (III:1), who presented with bilateral microphthalmia, sclerocornea, cataracts and nystagmus, left secondary glaucoma and a grossly cupped atrophic disc (Fig. 1b). Extraocular anomalies were not observed. The Sanger sequencing profile was suggestive of mosaicism in his mother (II:2), who was diagnosed with early onset cataracts and right exotropia. Mosaicism was confirmed and quantified by ddPCR in blood samples, with an estimated fractional abundance of 25% for the mutated allele (Fig. 1c). The variant was absent in the maternal grandparents, suggesting that it arose as a *de novo* post-zygotic event in the mother. The substitution of threonine 39, located in the TM1 domain, is predicted deleterious by Polyphen-2 and SIFT. Interestingly, the change p.(Thr39Arg) is absent in dbSNP147 and gnomAD, but has been previously described in a family with congenital cataracts, microcornea and iris hypoplasia (Sun et al., 2011).

In family 2 (Fig. 1d), a four-generation pedigree with autosomal dominant congenital cataracts, we initially identified the variant p.(Trp45Leu) (NM_005267.4:c.134G>T) in the

proband III:2, who had a diagnosis of dense congenital cataracts, microphthalmia and nystagmus. Sanger sequencing was performed on five additional family members (four affected with the same clinical diagnosis and one unaffected) and showed that the variant co-segregated with the ocular phenotype. The same amino acid substitution, predicted deleterious by Polyphen-2 and SIFT, has been previously described in another multi-generation family including eleven individuals with autosomal dominant congenital cataracts (Mohebi et al., 2017). Moreover, a different missense variant affecting the same amino acid, p.(Trp45Ser) (NM_005267.4:c.134G>C, rs864309688), has been reported in a three generation family with bilateral congenital cataracts and microcornea (Vanita et al., 2008), in a sporadic case with bilateral anterior cortical/nuclear cataracts (Ma et al., 2016) and in a three generation family with paediatric cataracts (Javadiyan et al., 2017). Functional experiments showed that p.(Trp45Ser) inhibited the formation of functional intercellular channels or hemichannels and decreased the junctional conductance induced by wild-type Cx50 and Cx46, acting as dominant negative inhibitor (Tong et al., 2011). Tryptophan 45 is an evolutionary conserved residue located in the TM1 domain and its substitution with leucine or serine has not been observed in controls (gnomAD).

In family 3 (Fig. 1e) and family 4 (Fig. 1f), we identified the variant p.(Asp51Asn) (NM_005267.4:c.151G>A; rs864309703), which affects a highly conserved amino acid located in the ECL1 domain. This change, predicted deleterious by Polyphen-2 and SIFT, has been previously reported in a patient with bilateral microphthalmia, congenital cataracts and sclerocornea (Ma et al., 2016; Ma et al., 2018). In family 3, the mutation occurred as a *de novo* event in the male proband, who presented with bilateral microphthalmia with associated cataracts, anterior segment dysgenesis and persistent pupillary membranes. Extraocular anomalies were not observed. In the three-generation family 4, the heterozygous variant was identified in both the proband (III:1) and her affected father (II:4). Head axial computed

tomography scanning of the proband at 29 years old showed borderline bilateral microphthalmia and enophthalmos (posterior displacement of the eye), although her ocular globes had a size of 20mm (right eye) and 18mm (left eye). At 32 years of age, the proband had no light perception on the right and light perception on the left. The right eye was phthisical, with no discernible anterior segment structures; the left eye had a corneal leukoma, cataract and corectopia. The father was diagnosed with bilateral microphthalmia and congenital cataracts. The paternal grandfather (I:1) and one of the paternal uncles (II:3), now deceased, were also affected. The mother (II:5) was affected by congenital glaucoma. However, the proband did not carry a mutation in any known congenital glaucoma-associated genes included in a custom targeted NGS panel containing 121 eye developmental genes, 9 of which are associated with congenital glaucoma. No extraocular anomalies were observed.

In family 5 (Fig. 1g), the novel variant p.(Phe70Leu) (NM_005267.4:c.208T>C) was identified in the proband III:2, diagnosed with bilateral microphthalmia, congenital cataracts and secondary glaucoma. Segregation analysis showed that the mutation was a *de novo* event in the affected mother (II:2), who also had microphthalmia and cataracts. Phenylalanine 70 is a conserved amino acid located in the ECL1 domain, and its substitution is predicted to be deleterious by SIFT and Polyphen-2.

In family 6 (Fig. 1h), we identified a missense variant p.(Gly94Arg) (NM_005267.4:c.280G>A) in a male proband of Chinese ethnicity (II:1) presenting with bilateral congenital aphakia (absence of the lens), corneal opacity, bilateral microphthalmia with iris and optic disc coloboma, and bilateral primary glaucoma. No extraocular anomalies were observed. No details of parental phenotype or DNA were available. Interestingly this change, predicted deleterious by SIFT and Polyphen-2 and located in the TM2 domain, is absent in gnomAD, but has been previously identified as a *de novo* event in a child with bilateral corneal opacification and microcornea, bilateral rudimentary lenses and bilateral

glaucoma (Ma et al., 2018).

In family 7 (Fig. 1i), the female proband (II:1) carried a *de novo* variant p.(Val97Gly) (NM_005267.4:c.290T>G), predicted deleterious by SIFT and Polyphen-2 and located in the TM2 domain. This previously undescribed variant was identified by the DDD study (DECIPHER ID: 259194) and confirmed with Sanger sequencing. She had bilateral microphthalmia, anterior segment dysgenesis and dense cataracts, treated with lensectomies, and right secondary glaucoma, with no extraocular features.

The significance of the other four variants identified in the screening (Supplementary Fig. 2) was considered 'uncertain' (p.(Leu292Gln)) or 'unlikely to be pathogenic' (p.(Leu7Met), p.(Asn220Asp) and p.(Gly333Arg)).

A novel amino acid change, p.(Leu292Gln) (NM_005267.4:c.875T>A) was identified in a proband with bilateral mild cataracts and optic nerve coloboma associated with nystagmus, photophobia and small kidneys (family 8). The substitution of leucine 292, located in the CT domain, is predicted benign by SIFT, but deleterious by Polyphen-2. Sanger sequencing of *PAX2* revealed that the proband II:3 also carried a novel heterozygous frameshift variant in this gene (NM_003987.2:c.529_530ins13, p.(Ala177Glyfs*8)), which introduces a premature stop codon in exon 5. Sanger sequencing excluded the maternal inheritance of both the *GJA8* and the *PAX2* variants; paternal DNA was unavailable for segregation analysis.

The GJA8 variant p.(Leu7Met) (NM_005267.4:c.19C>A; rs150441169), located in the N-terminal domain and predicted deleterious by SIFT and Polyphen-2, was detected in a patient (II:1) with syndromic unilateral microphthalmia, and was inherited from his unaffected father (family 9). The family is of African ethnicity and the minor allele frequency (MAF) for the African/African-American population in gnomAD is 0.28%. Different substitutions of this amino acid have been described before as disease-causative mutations: p.(Leu7Pro) (NM_005267.4:c.20T>C) was identified in a family with inherited cataracts (Mackay et al.,

2014) and p.(Leu7Gln) (NM_153465.1:c.20T>A) in a rat model with nuclear pulverulent cataracts and, in the case of homozygous rat mutants, microphthalmia with hypoplastic lens (Liska et al., 2008). However, in contrast with these previously reported variants, the frequency of the p.(Leu7Met) variant in unaffected individuals, in particular of African/Afro-American ethnicity, suggests that the substitution with a methionine might be tolerated.

The variant p.(Asn220Asp) (NM_005267.4:c.658A>G; rs138140155, gnomAD total-MAF=0.24%) was identified in an individual with bilateral microphthalmia and chorioretinal colobomas involving the optic disc, as well as microcephaly associated with normal development and faltering growth (family 10) and was inherited from her unaffected father. This substitution of asparagine 220, located in the TM4 domain and predicted deleterious by SIFT and Polyphen-2, has been reported before in a proband with congenital cataract and microcornea (Ma et al., 2016) and in a three generation family with congenital cataracts and aphakic glaucoma (Kuo et al., 2017). However, in those families it did not co-segregate with the phenotype and therefore was classified as benign. This was also supported by functional experiments showing that this rare polymorphism did not abolish intercellular channel function (Kuo et al., 2017).

The variant p.(Gly333Arg) (NM_005267.4:c.997G>C; rs587600450, gnomAD total-MAF=0.009%) was observed in a proband with unilateral microphthalmia and chorioretinal coloboma involving the optic disc, and was inherited from her unaffected father (family 11). This heterozygous change occurs in the CT domain and is predicted as tolerated by both SIFT and Polyphen-2.

1q21 copy number variants overlapping with *GJA8*.

GJA8 is part of a complex genomic locus, 1q21.1-q21.2, characterised by the presence of numerous segmental duplications (SDs), which make the region susceptible to recurrent rearrangements. To investigate whether structural variants affecting *GJA8* were present in our

cohort of families with AMC, we examined a subset of 188 unrelated individuals for whom copy number information was available from aCGH and/or NGS data. As such, the samples for which CNV data were generated were not chosen according to any selection criteria applied across the total cohort, and therefore they effectively represented a randomly-selected subset of independent AMC cases. This resulted in the identification of 1q21 microdeletions in five families (Fig. 2, Table 1).

The first microdeletion was identified in a proband (family 12, Fig. 2b) with bilateral coloboma of the iris and choroid, mild dysmorphic features (broad forehead, narrow palpebral fissures, depressed nasal root and low set ears), scoliosis, *genu valgum* and gastroesophageal reflux. She had normal developmental milestones. This CNV, detected from the screening of a custom NGS panel of 121 eye development genes, was further confirmed by both aCGH and ddPCR (Supplementary Fig. 3). It spans approximately 2 Mb (chr1:145388977-147395401**,** Build GRCh37/hg19) and affects 40 RefSeq genes. Segregation analysis revealed that this structural variant arose as a *de novo* event in the proband.

The second microdeletion was identified in a female proband with unilateral chorioretinal coloboma involving the optic disc, band keratopathy, cataract and secondary glaucoma without extraocular anomalies, and was inherited from her unaffected father (family 13, Fig. 2b). The minimal deleted region (chr1:146155983-147824178, Build GRCh37/hg19) spans approximately 1.67 Mb and affects 24 RefSeq genes. To validate the microdeletion and test the hypothesis that the unaffected status of the father could be due to mosaicism, we performed a ddPCR assay. However, this experiment confirmed the full heterozygous status of the microdeletion in both individuals (Supplementary Fig. 3b).

The third microdeletion was found in a female proband with extreme microphthalmia in the right eye and iris, chorioretinal coloboma in the left eye, cleft lip and palate, and neonatal

seizures (family 14, Fig. 2b). The minimal deleted region (chr1:146564743-147735011, Build GRCh37/hg19) spans approximately 1.17 Mb and affects 17 RefSeq genes. The presence of the CNV in the mother was excluded by ddPCR (Supplementary Fig. 3b). The father and other family members were unavailable for phenotypic and segregation analysis.

The fourth microdeletion (chr1:146497694-147825519, Build GRCh37/hg19), spanning approximately 1.33 Mb and affecting 20 RefSeq genes, was identified in two independent families (families 15 and 16, Fig. 2b). In family 15, it occurred as a *de novo* event in a proband with bilateral iris and chorioretinal coloboma involving disc, and nystagmus, without extraocular anomalies. The presence of the CNV in the parents was excluded by aCGH. Clinical re-assessment of the family revealed that the father presented blue dot lens opacities and cavernous disc anomalies with a pit in the right eye and mild cavernous disc anomaly or pronounced optic cup in the left eye. In family 16, a three-generation pedigree with coloboma, the microdeletion was detected by aCGH in the proband (III:6), who showed bilateral chorioretinal coloboma and microphthalmia in the right eye associated with microcephaly and normal development, and in the affected father (II:5), who presented with microphthalmia and coloboma in the right eye. The cousin III:2 was also affected with unilateral iris and chorioretinal coloboma. However, segregation analysis could not be performed on this individual.

The predicted boundaries of these CNVs indicated that these rearrangements belonged to different classes of 1q21 microdeletions. Recurrent 1q21 CNVs occur at four breakpoint regions (BP1-BP4), each corresponding to a large block of highly homologous SDs (Mefford et al., 2008). Further, the locus can be divided into two distinct regions: a proximal region included between BP2 and BP3 and a distal region, flanked by BP3 and BP4, which mediate the most recurrent CNVs of the 1q21 locus. While the microdeletions found in families 13-16 were distal rearrangements occurring between BP3 and BP4 (class I), the microdeletion

detected in family 12 was flanked by the breakpoints BP2 and BP4 and extended from the proximal through to the distal region (class II). Despite their rarity in the general population, both types of 1q21 microdeletions appeared to be enriched in our AMC cohort. We compared the frequency of these CNVs in our cases with control individuals previously reported in the literature (Rosenfeld et al., 2012): BP3-BP4 microdeletions occurred in 4 out of 188 individuals with AMC *versus* 12 out of 65282 controls (2.13% *versus* 0.02%, Fisher's exact test $p=1.17 \times 10^{-7}$), whereas BP2-BP4 microdeletions occurred in 1 out of 188 individuals with AMC *versus* 1 out of 65927 controls (0.532% *versus* 0.002%, Fisher's exact test p=0.0057).

Taking into account all the 1q21 microdeletions identified in our cohort, the minimally deleted region spans approximately 830 kb (chr1:146564743-147395401, Build GRCh37/hg19) and includes 11 genes (*NBPF19, NBPF13P, PRKAB2, CHD1L, PDIA3P1, FMO5, LINC00624, BCL9, ACP6, GJA5, GJA8*). In addition to *Gja8*, a role in eye development has been shown also for *Bcl9,* a co-activator for β-catenin-mediated transcription in Wnt signaling (Bienz, 2005). A recent study has demonstrated that *Bcl9* is also part of the Pax6-dependent regulatory circuit and contributes to mouse lens formation (Cantù et al., 2014). No other genes known to be relevant in eye development are present in the region.

Sequence analysis of the coding region of *GJA8* in the probands carrying 1q21 microdeletions did not reveal any variant on the remaining allele. In family 12, no additional pathogenic variants were identified from the targeted NGS screening of 121 eye developmental genes. In family 13, the NGS targeted sequencing of 351 diagnostic genes for eye developmental anomalies in proband II:2 identified an in-frame deletion of 6bp in *FOXC1* (NM_001453.2:c.1338_1343del, p.(Gly447_Gly448del)), maternally inherited. This rare variant (gnomAD total-MAF=0.06%) is reported as a multi-allelic SNP (rs572346201),

which occurs in a region coding for a poly-Glycine stretch. Although its clinical significance is unknown, due to the repetitive nature of this region, it is likely to represent a natural polymorphism.

DISCUSSION

Mutations in Cx50, encoded by *GJA8*, have been primarily linked to congenital and early onset cataract in humans and also animal models. However, recently in a small number of cases *GJA8* mutations have also been associated with a broader phenotype which can include microphthalmia, sclerocornea and lens abnormalities (Ma et al., 2018).

In this study, we have investigated the role of *GJA8* in a cohort of 426 individuals with a wide range of developmental eye anomalies, and identified 16 families with AMC carrying genetic alterations of *GJA8*. These included six likely pathogenic sequence variants (p.(Thr39Arg), p.(Trp45Leu), p.(Asp51Asn), p.(Phe70Leu), p.(Gly94Arg) and p.(Val97Gly)) detected in seven unrelated families, four missense variants (p.(Leu7Met), p.(Asn220Asp), p.(Leu292Gln) and p.(Gly333Arg)) with uncertain or unlikely clinical significance and four heterozygous 1q21 microdeletions involving *GJA8* detected in five unrelated families of uncertain significance.

Segregation analyses were possible for five out of six likely pathogenic sequence variants and showed that these occurred either *de novo* or co-segregated with the disease in an autosomal dominant fashion. These variants were bioinformatically predicted damaging and have not been reported in unaffected individuals according to public databases. Interestingly, three of these changes (p.(Thr39Arg), p.(Trp45Leu) and p.(Asp51Asn)) have been previously described in families with cataracts (Sun et al., 2011; Ma et al., 2016; Javadiyan et al., 2017; Mohebi et al., 2017; Ma et al., 2018) and a fourth (p.(Gly94Arg)) in a proband with sclerocornea and lens abnormalities (Ma et al., 2018). Given the rarity of these variants, the identification of the same missense changes in unrelated affected individuals strongly

supports a causative relationship between these variants and eye developmental disorders. In particular, p.(Asp51Asn) had been reported as a *de novo* mutation in a patient with bilateral microphthalmia, congenital cataracts and sclerocornea (Ma et al., 2016). In the present study, the same variant was detected in two independent families with a similar phenotype, including microphthalmia, cataracts and other anterior chamber eye anomalies. This emerging genotype-phenotype correlation suggests that this amino acid substitution might have a severe effect on GJA8 function and supports the involvement of this protein in a broader range of eye developmental anomalies.

Our identification of the variant p.(Gly94Arg) in another patient also aids genotype-phenotype correlation for amino acid substitutions of this highly conserved residue. In our cohort, the change was identified in a case with bilateral corneal opacification, congenital aphakia and microphthalmia with iris and optic disc coloboma. The same variant has been previously reported as a *de novo* event in an individual diagnosed with bilateral corneal opacification, glaucoma, and rudimentary lenses (Ma et al., 2018). Interestingly, Ma et al. (2018) also described another variant affecting the same amino acid, p.(Gly94Glu), in a proband with total sclerocornea and cataractous disc-like lenses with microcornea. Mice models expressing heterozygous missense mutations (e.g. Cx50D47A, Cx50S50P, Cx50V64A and Cx50R205G) (Graw et al., 2001; Xia et al., 2006; Xia et al., 2012; Berthoud et al., 2013) or with complete *Gja8* knockout (White et al., 1998; Rong et al., 2002) have shown that Gja8 is important for lens development. Therefore, the identification of glycine 94 mutations in three individuals with lens abnormalities supports the hypothesis that this amino acid is particularly important for GJA8 to perform this role in eye development. Interestingly, the phenotype of bilateral aphakia associated with sclerocornea overlaps with that of individuals with biallelic mutations in *FOXE3* (Iseri et al., 2009). Therefore, when screening patients with this phenotype, it is important to screen for variants in both *FOXE3* and *GJA8*.

Multiple sequence alignment indicated that all likely pathogenic sequence variants identified in our cohort affected conserved residues (Supplementary Fig. 1) and were located within the N-terminal region of the protein (Fig. 3). Our findings are consistent with previous studies, since mutations associated with cataracts tend to cluster between TM1 and TM2 (Yu et al., 2016). The transmembrane domains are thought to play an important role in oligomerisation and pore formation, while the ECL1 domain is important in the docking of two opposing HCs to form the GJCs. While these domains are evolutionarily conserved and present high homology among the members of the Cx family, the CT region is the most isotype-specific domain and contains motifs for regulatory kinases (Liu et al., 2011; Wang et al., 2013). In this region, we identified a novel missense change, p.(Leu292Gln), of unknown clinical significance. The variant was found in a proband (family 8) who also carried an insertion of 13bp in *PAX2* (NM_003987.2:c.529_530ins13, p.(Ala177Glyfs*8)). Heterozygous variants of *PAX2* (MIM 167409) are identified in approximately half of the cases presenting with renal coloboma syndrome (Bower et al., 2012), also known as papillorenal syndrome (OMIM 120330). Therefore, this novel *PAX2* variant is likely to be responsible for optic nerve coloboma and kidney anomalies observed in the patient, but it is possible that the *GJA8* variant might lead to a subtle effect on the protein function and contribute to his mild cataract phenotype.

By contrast, the three additional heterozygous variants p.(Leu7Met) (family 9), p.(Asn220Asp) (family 10) and p.(Gly333Arg) (family 11) were considered as likely benign. These were identified in individuals with AMC, but without cataracts, in unaffected parents either in this or previous studies (Ma et al., 2016; Kuo et al., 2017) and in controls (gnomAD).

Human *GJA8* maps to a structurally complex locus on chromosome 1q21.1-q21.2, with at least four large blocks of highly homologous SDs, which make it prone to nonallelic

499 homologous recombination (NAHR) (Mefford et al., 2008). As with other genomic loci
500 subject to recurrent rearrangements (such as 15q11, 15q13, 16p11.2, 16p12.1, 16p13.11,
501 17q12, 22q11.2) (Girirajan and Eichler, 2010; Stankiewicz and Lupski, 2010), 1q21 CNVs
502 have been associated with a wide range of phenotypes including dysmorphic features,
503 developmental delay, neuropsychiatric disorders, and cardiac and eye anomalies. The
504 reported eye anomalies include cataracts (Brunetti-Pierri et al., 2008; Mefford et al., 2008;
505 Rosenfeld et al., 2012; Bernier et al., 2016; Ha et al., 2016) and in a minority of cases more
506 severe defects such as microphthalmia (Mefford et al., 2008) and coloboma (Brunetti-Pierri et
507 al., 2008).
508 The most common 1q21 CNVs occur between the breakpoints BP3 and BP4 (Fig. 2a),
509 spanning ~1.35 Mb (Mefford et al., 2008). This region contains only ~800 kb of unique (i.e.
510 nonduplicated) DNA sequence (Bernier et al., 2016) and includes at least 11 genes (*NBPF19,*
511 *NBPF13P, PRKAB2, CHD1L, PDIA3P1, FMO5, LINC00624, BCL9, ACP6, GJA5* and
512 *GJA8*), which might contribute to different aspects of the disease manifestations observed.
513 Alternatively, 1q21 CNVs can involve only the proximal region (BP2-BP3) or both the
514 proximal and the distal region (BP1/BP2-BP4). Microdeletions of the proximal region have
515 been reported to be a predisposing factor for Thrombocytopenia-absent radius (TAR)
516 syndrome (Klopocki et al., 2007), together with sequence variants in the *RBM8A* gene.
517 Within the distal region, a potential role in eye development has been shown for two of the
518 genes, *GJA8* and *BCL9.* A recent study has demonstrated that *Bcl9* is a downstream effector
519 of *Pax6* during mouse lens development (Cantù et al., 2014). However, the role of *BCL9* in
520 human eye development has not yet been established. Given the involvement of *GJA8* in both
521 cataractogenesis and ocular growth, as previously described, this gene seems to be a good
522 candidate for the ocular anomalies observed in some of the 1q21 CNV carriers.

In most cases, the 1q21 rearrangements are inherited. Their presence in unaffected parents has brought into question their pathogenic significance, but the analysis of large clinical and population cohorts has shown that 1q21 microdeletions/microduplications occur at significantly higher frequency in individuals with clinical diagnoses compared with controls (Brunetti-Pierri et al., 2008; Mefford et al., 2008; Rosenfeld et al., 2012; Bernier et al., 2016). In particular, the comparison of a large cohort of individuals with developmental delay, intellectual disability, dysmorphic features and congenital anomalies with previously published control cohorts showed that the frequency of BP2-BP4 deletions was 0.024% in cases (11/45744) *versus* 0.002% in controls (1/65927), whereas the frequency of BP3-BP4 deletions was 0.285% in cases (86/30215) *versus* 0.018% in controls (12/65282) (Rosenfeld et al., 2012). This enrichment suggests that these CNVs might increase susceptibility to developmental anomalies with variable expressivity and incomplete penetrance, although the factors underlying their heterogeneous phenotypes remain unexplained. In this study, we identified four microdeletions in five families, three overlapping with the microdeletions/microduplications recurrently found between breakpoints BP3 and BP4 and one larger BP2-BP4 microdeletion encompassing both the proximal and the distal region. Given the rarity of these rearrangements, the presence of 1q21 microdeletions in our AMC cohort, with a frequency of 2.13% in AMC cases for BP3-BP4 microdeletions ($p=1.17 \times 10^{-7}$) and a frequency of 0.53% in AMC cases for BP2-BP4 microdeletions ($p=0.0057$), seems to support their role as a risk factor for developmental disorders, including eye anomalies. However, consistent with previous studies, the segregation pattern in families 13 and 15 indicates that other genetic and/or environmental modifiers are likely to be important for the phenotypic outcome. Therefore, the exact genotype-phenotype correlation remains to be established.

Mouse models have shown that *Gja8* copy number losses and point mutations act through different mechanisms and modes of inheritance. Deletions of the entire coding region of the gene cause cataracts and microphthalmia only when homozygous, indicating a recessive mode of inheritance (White et al., 1998; Rong et al., 2002). In contrast, mouse strains carrying pathogenic *Gja8* missense mutations develop microphthalmia and cataracts in a dominant or semi-dominant fashion (Steele et al., 1998; Graw et al., 2001; Chang et al., 2002; Liska et al., 2008; Xia et al., 2012). Since Cxs function in hexameric complexes which can be homo- or heteromeric, it is possible that the impact of single amino acid substitutions may be more severe than the loss of one functional allele. Mutant Cx subunits can interfere with correct formation of the oligomeric complexes in a dominant negative manner and, since GJCs can be formed by different types of Cx subunits, this effect can also extend to the function of other Cxs. Functional and cellular studies have shown that point mutations can alter the activity of the human GJA8 protein in various ways (Beyer et al., 2013). For instance, pathogenic variants can cause misfolding, improper oligomerisation and/or trafficking defects, leading to a reduced number of functional channels on the membrane. Alternatively, the pathogenic variants could alter some physiological properties of the channels, such as permeability or conductance, or lead to the formation of HCs with new and aberrant functions. Therefore, a single base mutation can affect several aspects of the Cx function. This complexity may explain the phenotypic heterogeneity observed among the carriers of *GJA8* variants, and also the difference in penetrance between sequence and copy number variations.

Intra-familial phenotypic variability was also observed for the sequence variant identified in family 1 (p.(Thr39Arg)) possibly related to mosaicism. While the proband carrying the heterozygous change presented with bilateral cataracts and microphthalmia, his mother, who was 25% mosaic for this variant, had a milder phenotype of early onset cataracts. Therefore,

we hypothesize that the somatic mosaicism detected in the mother may explain her milder phenotype and that lower doses of aberrant GJA8 protein during eye development might be responsible for less severe phenotypic outcomes. In support of this, a recent study has reported a correlation between the severity of developmental eye abnormalities and somatic mosaicisms of *Pax6* mutations in CRISPR/Cas9 genome-edited mouse embryos (Yasue et al., 2017).

In conclusion, this study expands our knowledge of the role of *GJA8* in eye development, highlighting how genetic alterations of this gene are likely to give rise not only to early onset cataracts, but also to other developmental eye anomalies. The screening of *GJA8* in 426 individuals with AMC resulted in the identification of six likely pathogenic variants in seven families. In the six families where segregation analysis was possible, the variants co-segregated with both early onset cataracts and microphthalmia. In one singleton case with aphakia and corneal opacification where no segregation analysis was possible, we identified the variant p.(Gly94Arg). This finding, in combination with two previously reported patients with lens development abnormalities and with variants affecting the same amino acid, highlights the importance of this specific residue in the function of GJA8 and suggests that *GJA8* mutations can be responsible for phenotypes often associated with *FOXE3* variants. The role of *GJA8* microdeletions in AMC remains uncertain: the enrichment of rare 1q21 microdeletions in our cohort seems to support their role as risk factors for developmental eye disorders. However, the incomplete segregation and the phenotypic variability of these variants indicate that other genetic and/or environmental factors might be of importance. In summary, these data expand the spectrum of human phenotypes associated with *GJA8* variants and the identification of specific mutations contributes to our understanding of their genotype-phenotype correlation. Therefore, this study demonstrates the importance of screening *GJA8* in individuals with developmental eye anomalies.

Conflict of Interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods 7(4):248-249.

Balikova I, de Ravel T, Ayuso C, Thienpont B, Casteels I, Villaverde C, Devriendt K, Fryns JP, Vermeesch JR. 2011. High frequency of submicroscopic chromosomal deletions in patients with idiopathic congenital eye malformations. Am J Ophthalmol 151(6):1087-1094.e45.

Bernier R, Steinman KJ, Reilly B, Wallace AS, Sherr EH, Pojman N, Mefford HC, Gerdts J, Earl R, Hanson E and others. 2016. Clinical phenotype of the recurrent 1q21.1 copy-number variant. Genet Med 18(4):341-349.

Berthoud VM, Minogue PJ, Yu H, Schroeder R, Snabb JI, Beyer EC. 2013. Connexin50D47A decreases levels of fiber cell connexins and impairs lens fiber cell differentiation. Invest Ophthalmol Vis Sci 54(12):7614-7622.

Beyer EC, Ebihara L, Berthoud VM. 2013. Connexin mutants and cataracts. Front Pharmacol 4:43.

Bienz M. 2005. beta-Catenin: a pivot between cell adhesion and Wnt signalling. Curr Biol 15(2):R64-67.

Bower M, Salomon R, Allanson J, Antignac C, Benedicenti F, Benetti E, Binenbaum G, Jensen UB, Cochat P, DeCramer S and others. 2012. Update of PAX2 mutations in renal coloboma syndrome and establishment of a locus-specific database. Hum Mutat 33(3):457-466.

Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, Lalani SR, Graham B, Lee B, Shinawi M and others. 2008. Recurrent reciprocal 1q21.1 deletions and

624 duplications associated with microcephaly or macrocephaly and developmental and
625 behavioral abnormalities. Nat Genet 40(12):1466-1471.
626 Cantù C, Zimmerli D, Hausmann G, Valenta T, Moor A, Aguet M, Basler K. 2014. Pax6-
627 dependent, but β-catenin-independent, function of Bcl9 proteins in mouse lens
628 development. Genes Dev 28(17):1879-1884.
629 Chang B, Wang X, Hawes NL, Ojakian R, Davisson MT, Lo WK, Gong X. 2002. A Gja8
630 (Cx50) point mutation causes an alteration of alpha 3 connexin (Cx46) in semi-
631 dominant cataracts of Lop10 mice. Hum Mol Genet 11(5):507-513.
632 Chassaing N, Ragge N, Plaisancié J, Patat O, Geneviève D, Rivier F, Malrieu-Eliaou C,
633 Hamel C, Kaplan J, Calvas P. 2016. Confirmation of TENM3 involvement in
634 autosomal recessive colobomatous microphthalmia. Am J Med Genet A 170(7):1895-
635 1898.
636 Chung J, Berthoud VM, Novak L, Zoltoski R, Heilbrunn B, Minogue PJ, Liu X, Ebihara L,
637 Kuszak J, Beyer EC. 2007. Transgenic overexpression of connexin50 induces
638 cataracts. Exp Eye Res 84(3):513-528.
639 Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. 2010. Identifying a
640 high fraction of the human genome to be under selective constraint using GERP++.
641 PLoS Comput Biol 6(12):e1001025.
642 Devi RR, Vijayalakshmi P. 2006. Novel mutations in GJA8 associated with autosomal
643 dominant congenital cataract and microcornea. Mol Vis 12:190-195.
644 Fernandez-San Jose P, Corton M, Blanco-Kelly F, Avila-Fernandez A, Lopez-Martinez MA,
645 Sanchez-Navarro I, Sanchez-Alcudia R, Perez-Carro R, Zurita O, Sanchez-Bolivar N
646 and others. 2015. Targeted Next-Generation Sequencing Improves the Diagnosis of
647 Autosomal Dominant Retinitis Pigmentosa in Spanish Patients. Invest Ophthalmol Vis
648 Sci 56(4):2173-2182.

García IE, Prado P, Pupo A, Jara O, Rojas-Gómez D, Mujica P, Flores-Muñoz C, González-Casanova J, Soto-Riveros C, Pinto BI and others. 2016. Connexinopathies: a structural and functional glimpse. BMC Cell Biol 17 Suppl 1:17.

Gillespie RL, O'Sullivan J, Ashworth J, Bhaskar S, Williams S, Biswas S, Kehdi E, Ramsden SC, Clayton-Smith J, Black GC and others. 2014. Personalized diagnosis and management of congenital cataract by next-generation sequencing. Ophthalmology 121(11):2124-37.e1-2.

Girirajan S, Eichler EE. 2010. Phenotypic variability and genetic susceptibility to genomic disorders. Hum Mol Genet 19(R2):R176-187.

Gong X, Li E, Klier G, Huang Q, Wu Y, Lei H, Kumar NM, Horwitz J, Gilula NB. 1997. Disruption of alpha3 connexin gene leads to proteolysis and cataractogenesis in mice. Cell 91(6):833-843.

Graw J, Löster J, Soewarto D, Fuchs H, Meyer B, Reis A, Wolf E, Balling R, Hrabé de Angelis M. 2001. Characterization of a mutation in the lens-specific MP70 encoding gene of the mouse leading to a dominant cataract. Exp Eye Res 73(6):867-876.

Ha K, Shen Y, Graves T, Kim CH, Kim HG. 2016. The presence of two rare genomic syndromes, 1q21 deletion and Xq28 duplication, segregating independently in a family with intellectual disability. Mol Cytogenet 9:74.

Hansen L, Yao W, Eiberg H, Kjaer KW, Baggesen K, Hejtmancik JF, Rosenberg T. 2007. Genetic heterogeneity in microcornea-cataract: five novel mutations in CRYAA, CRYGD, and GJA8. Invest Ophthalmol Vis Sci 48(9):3937-3944.

Holt R, Ugur Iseri SA, Wyatt AW, Bax DA, Gold Diaz D, Santos C, Broadgate S, Dunn R, Bruty J, Wallis Y and others. 2017. Identification and functional characterisation of genetic variants in OLFM2 in children with developmental eye disorders. Hum Genet 136(1):119-127.

674 Hu S, Wang B, Zhou Z, Zhou G, Wang J, Ma X, Qi Y. 2010. A novel mutation in GJA8
675 causing congenital cataract-microcornea syndrome in a Chinese pedigree. Mol Vis
676 16:1585-1592.
677 Iseri SU, Osborne RJ, Farrall M, Wyatt AW, Mirza G, Nürnberg G, Kluck C, Herbert H,
678 Martin A, Hussain MS and others. 2009. Seeing clearly: the dominant and recessive
679 nature of FOXE3 in eye developmental anomalies. Hum Mutat 30(10):1378-1386.
680 Javadiyan S, Craig JE, Souzeau E, Sharma S, Lower KM, Mackey DA, Staffieri SE, Elder
681 JE, Taranath D, Straga T and others. 2017. High Throughput Genetic Screening of 51
682 Paediatric Cataract Genes Identifies Causative Mutations in Inherited Paediatric
683 Cataract in South Eastern Australia. G3 (Bethesda) 7(10):3257-3268.
684 Klopocki E, Schulze H, Strauss G, Ott CE, Hall J, Trotier F, Fleischhauer S, Greenhalgh L,
685 Newbury-Ecob RA, Neumann LM and others. 2007. Complex inheritance pattern
686 resembling autosomal recessive inheritance involving a microdeletion in
687 thrombocytopenia-absent radius syndrome. Am J Hum Genet 80(2):232-240.
688 Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous
689 variants on protein function using the SIFT algorithm. Nat Protoc 4(7):1073-1081.
690 Kuo DS, Sokol JT, Minogue PJ, Berthoud VM, Slavotinek AM, Beyer EC, Gould DB. 2017.
691 Characterization of a variant of gap junction protein α8 identified in a family with
692 hereditary cataract. PLoS One 12(8):e0183438.
693 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH,
694 Ware JS, Hill AJ, Cummings BB and others. 2016. Analysis of protein-coding genetic
695 variation in 60,706 humans. Nature 536(7616):285-291.
696 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
697 transform. Bioinformatics 25(14):1754-1760.

698 Liska F, Chylíková B, Martínek J, Kren V. 2008. Microphthalmia and cataract in rats with a
699 novel point mutation in connexin 50 - L7Q. Mol Vis 14:823-828.
700 Liu J, Ek Vitorin JF, Weintraub ST, Gu S, Shi Q, Burt JM, Jiang JX. 2011. Phosphorylation
701 of connexin 50 by protein kinase A enhances gap junction and hemichannel function.
702 J Biol Chem 286(19):16914-16928.
703 Liu X, Wu C, Li C, Boerwinkle E. 2016. dbNSFP v3.0: A One-Stop Database of Functional
704 Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Hum
705 Mutat 37(3):235-241.
706 Ma AS, Grigg JR, Ho G, Prokudin I, Farnsworth E, Holman K, Cheng A, Billson FA, Martin
707 F, Fraser C and others. 2016. Sporadic and Familial Congenital Cataracts: Mutational
708 Spectrum and New Diagnoses Using Next-Generation Sequencing. Hum Mutat
709 37(4):371-384.
710 Ma AS, Grigg JR, Prokudin I, Flaherty M, Bennetts B, Jamieson RV. 2018. New mutations
711 in GJA8 expand the phenotype to include total sclerocornea. Clin Genet 93(1):155-
712 159.
713 Mackay DS, Bennett TM, Culican SM, Shiels A. 2014. Exome sequencing identifies novel
714 and recurrent mutations in GJA8 and CRYGD associated with inherited cataract. Hum
715 Genomics 8:19.
716 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
717 Altshuler D, Gabriel S, Daly M and others. 2010. The Genome Analysis Toolkit: a
718 MapReduce framework for analyzing next-generation DNA sequencing data. Genome
719 Res 20(9):1297-1303.
720 Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, Huang S, Maloney VK, Crolla
721 JA, Baralle D and others. 2008. Recurrent rearrangements of chromosome 1q21.1 and
722 variable pediatric phenotypes. N Engl J Med 359(16):1685-1699.

723 Mohebi M, Chenari S, Akbari A, Ghassemi F, Zarei-Ghanavati M, Fakhraie G, Babaie N,
724 Heidari M. 2017. Mutation analysis of connexin 50 gene among Iranian families with
725 autosomal dominant cataracts. Iran J Basic Med Sci 20(3):288-293.
726 Mose LE, Wilkerson MD, Hayes DN, Perou CM, Parker JS. 2014. ABRA: improved coding
727 indel detection via assembly-based realignment. Bioinformatics 30(19):2813-2815.
728 Pfenniger A, Wohlwend A, Kwak BR. 2011. Mutations in connexin genes and disease. Eur J
729 Clin Invest 41(1):103-116.
730 Ponnam SP, Ramesha K, Tejwani S, Ramamurthy B, Kannabiran C. 2007. Mutation of the
731 gap junction protein alpha 8 (GJA8) gene causes autosomal recessive cataract. J Med
732 Genet 44(7):e85.
733 Ponnam SP, Ramesha K, Matalia J, Tejwani S, Ramamurthy B, Kannabiran C. 2013.
734 Mutational screening of Indian families with hereditary congenital cataract. Mol Vis
735 19:1141-1148.
736 Prokudin I, Simons C, Grigg JR, Storen R, Kumar V, Phua ZY, Smith J, Flaherty M, Davila
737 S, Jamieson RV. 2014. Exome sequencing in developmental eye disease leads to
738 identification of causal variants in GJA8, CRYGC, PAX6 and CYP1B1. Eur J Hum
739 Genet 22(7):907-915.
740 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
741 features. Bioinformatics 26(6):841-842.
742 Reis LM, Semina EV. 2015. Conserved genetic pathways associated with microphthalmia,
743 anophthalmia, and coloboma. Birth Defects Res C Embryo Today 105(2):96-113.
744 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E,
745 Spector E and others. 2015. Standards and guidelines for the interpretation of
746 sequence variants: a joint consensus recommendation of the American College of

747 Medical Genetics and Genomics and the Association for Molecular Pathology. Genet
748 Med 17(5):405-424.
749 Rong P, Wang X, Niesman I, Wu Y, Benedetti LE, Dunia I, Levy E, Gong X. 2002.
750 Disruption of Gja8 (alpha8 connexin) in mice leads to microphthalmia associated with
751 retardation of lens growth and lens fiber maturation. Development 129(1):167-174.
752 Rosenfeld JA, Traylor RN, Schaefer GB, McPherson EW, Ballif BC, Klopocki E, Mundlos S,
753 Shaffer LG, Aylsworth AS, Group qS. 2012. Proximal microdeletions and
754 microduplications of 1q21.1 contribute to variable abnormal phenotypes. Eur J Hum
755 Genet 20(7):754-761.
756 Schmidt W, Klopp N, Illig T, Graw J. 2008. A novel GJA8 mutation causing a recessive
757 triangular cataract. Mol Vis 14:851-856.
758 Shah SP, Taylor AE, Sowden JC, Ragge NK, Russell-Eggitt I, Rahi JS, Gilbert CE, Group
759 SoEAS-USI. 2011. Anophthalmos, microphthalmos, and typical coloboma in the
760 United Kingdom: a prospective study of incidence and risk. Invest Ophthalmol Vis
761 Sci 52(1):558-564.
762 Shah SP, Taylor AE, Sowden JC, Ragge N, Russell-Eggitt I, Rahi JS, Gilbert CE, Group
763 SoEASI. 2012. Anophthalmos, microphthalmos, and Coloboma in the United
764 kingdom: clinical features, results of investigations, and early management.
765 Ophthalmology 119(2):362-368.
766 Shiels A, Hejtmancik JF. 2017. Mutations and mechanisms in congenital and age-related
767 cataracts. Exp Eye Res 156:95-102.
768 Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth
769 J, Hillier LW, Richards S and others. 2005. Evolutionarily conserved elements in
770 vertebrate, insect, worm, and yeast genomes. Genome Res 15(8):1034-1050.

771 Siepel A, Pollard K, Haussler D. 2006. New methods for detecting lineage-specific selection.
772 Research in Computational Molecular Biology: Springer Berlin Heidelberg. p 190–
773 205.
774 Skalicky SE, White AJ, Grigg JR, Martin F, Smith J, Jones M, Donaldson C, Smith JE,
775 Flaherty M, Jamieson RV. 2013. Microphthalmia, anophthalmia, and coloboma and
776 associated ocular and systemic features: understanding the spectrum. JAMA
777 Ophthalmol 131(12):1517-1524.
778 Slavotinek AM. 2011. Eye development genes and known syndromes. Mol Genet Metab
779 104(4):448-456.
780 Stankiewicz P, Lupski JR. 2010. Structural variation in the human genome and its role in
781 disease. Annu Rev Med 61:437-455.
782 Steele EC, Lyon MF, Favor J, Guillot PV, Boyd Y, Church RL. 1998. A mutation in the
783 connexin 50 (Cx50) gene is a candidate for the No2 mouse cataract. Curr Eye Res
784 17(9):883-889.
785 Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, Fossdal R,
786 Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE and others. 2008. Large recurrent
787 microdeletions associated with schizophrenia. Nature 455(7210):232-236.
788 Sun W, Xiao X, Li S, Guo X, Zhang Q. 2011. Mutational screening of six genes in Chinese
789 patients with congenital cataract and microcornea. Mol Vis 17:1508-1513.
790 The 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP,
791 Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA and others. 2015. A
792 global reference for human genetic variation. Nature 526(7571):68-74.
793 Tong JJ, Minogue PJ, Guo W, Chen TL, Beyer EC, Berthoud VM, Ebihara L. 2011. Different
794 consequences of cataract-associated mutations at adjacent positions in the first
795 extracellular boundary of connexin50. Am J Physiol Cell Physiol 300(5):C1055-64.

796 Vanita V, Singh JR, Singh D, Varon R, Sperling K. 2008. A novel mutation in GJA8
797 associated with jellyfish-like cataract in a family of Indian origin. Mol Vis 14:323-
798 326.
799 Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants
800 from high-throughput sequencing data. Nucleic Acids Res 38(16):e164.
801 Wang Z, Han J, David LL, Schey KL. 2013. Proteomics and phosphoproteomics analysis of
802 human lens fiber cell membranes. Invest Ophthalmol Vis Sci 54(2):1135-1143.
803 White TW, Goodenough DA, Paul DL. 1998. Targeted ablation of connexin50 in mice results
804 in microphthalmia and zonular pulverulent cataracts. J Cell Biol 143(3):815-825.
805 White TW. 2002. Unique and redundant connexin contributions to lens development. Science
806 295(5553):319-320.
807 Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, King DA,
808 Ambridge K, Barrett DM, Bayzetinova T and others. 2015. Genetic diagnosis of
809 developmental disorders in the DDD study: a scalable analysis of genome-wide
810 research data. Lancet 385(9975):1305-1314.
811 Xia CH, Liu H, Cheung D, Cheng C, Wang E, Du X, Beutler B, Lo WK, Gong X. 2006.
812 Diverse gap junctions modulate distinct mechanisms for fiber cell formation during
813 lens development and cataractogenesis. Development 133(10):2033-2040.
814 Xia CH, Chang B, Derosa AM, Cheng C, White TW, Gong X. 2012. Cataracts and
815 microphthalmia caused by a Gja8 mutation in extracellular loop 2. PLoS One
816 7(12):e52894.
817 Yasue A, Kono H, Habuta M, Bando T, Sato K, Inoue J, Oyadomari S, Noji S, Tanaka E,
818 Ohuchi H. 2017. Relationship between somatic mosaicism of Pax6 mutation and
819 variable developmental eye abnormalities-an analysis of CRISPR genome-edited
820 mouse embryos. Sci Rep 7(1):53.

Yu Y, Wu M, Chen X, Zhu Y, Gong X, Yao K. 2016. Identification and functional analysis of two novel connexin 50 mutations associated with autosome dominant congenital cataracts. Sci Rep 6:26551.

Yuan L, Sui T, Chen M, Deng J, Huang Y, Zeng J, Lv Q, Song Y, Li Z, Lai L. 2016. CRISPR/Cas9-mediated GJA8 knockout in rabbits recapitulates human congenital cataracts. Sci Rep 6:22024.

FIGURE LEGENDS

Fig. 1 *GJA8* likely pathogenic sequence variants identified in 7 unrelated families with AMC. a. Pedigree of family 1. Sanger sequencing results showing the segregation of the missense variant p.(Thr39Arg) are presented. The chromatogram of individual II:2 is suggestive of mosaicism. **b.** Photographs of the affected individuals of family 1 showing intra-familial phenotypic variability. The mother (II:2, top) presented with a milder phenotype, which included right exotropia and normal sized eyes with bilateral pseudophakia. The proband (III:1, bottom) presented with right microphthalmia and complete corneal opacification on the left. **c.** Absolute quantification of the allele abundance for the variant c.116C>G; p.(Thr39Arg) in family 1. Digital Droplet PCR (ddPCR) assays were performed using a Taqman FAM-labeled probe for genotyping the mutant allele and a VIC-labeled probe to detect the wild-type allele. On the left, 1-D fluorescence amplitude plot of droplets shows mutant allele detection in the FAM channel for the heterozygous carrier (III:1), the putative mosaic mother (II:2), a wild-type homozygous carrier (I:2) and no template control (NTC). FAM-positive droplets (blue), containing the mutant allele, exhibit increased fluorescence compared to negative droplets (grey). On the right, the fractional abundance of the mutated allele, represented in percentage, was calculated for the FAM-positive droplets versus VIC-positive droplets (wild type allele), confirming the mosaicism of this variant in individual II:2. **d.** Pedigree of family 2. On the left, a representative sequence chromatogram shows the heterozygous missense variant p.(Trp45Leu). The genotype of the six individuals tested for the variant is indicated below each symbol. **e-f.** Pedigree of families 3 and 4, both carrying the missense variant p.(Asp51Asn). The sequence chromatograms show that the variant occurred *de novo* in family 3. N/A, genotype not available. In family 4, representative sequence chromatogram showing the p.(Asp51Asn) and pedigree indicating the inheritance of affected status and of the variant. For family 4, fully filled symbols

represent individuals affected with congenital cataracts and microphthalmia, but without glaucoma, quarter filled symbols represent individuals with congenital glaucoma. **g.** Pedigree of family 5 and sequence chromatograms showing the missense variant p.(Phe70Leu). **h.** Pedigree of family 6. The adopted child carries the missense variant p.(Gly94Arg). **i.** Pedigree of family 7. Sanger sequencing results show that the missense variant p.(Val97Gly) arose *de novo* in the child II:1

Fig. 2 *GJA8* structural variants identified in 5 unrelated families with AMC. a. Modified schematic from the UCSC Genome Browser (NCBI Build GRCh37/hg19). Partial ideogram of the chromosome bands 1q21.1-q21.2 and the multiple blocks of highly homologous segmental duplications (SD) present in this region are shown. SD, reported under the UCSC track 'Duplications of >1000 Bases of Non-RepeatMasked Sequence', are stretches of DNA of at least 1 kb in length, sharing a sequence identity of at least 90% with another genomic region on the same or on a different chromosome (inter- or intra-chromosomal SD). The colours indicate different levels of similarity between duplications (grey: 90-98% similarity, yellow: 98-99% similarity, orange: greater than 99% similarity). The breakpoint regions (BP2, BP3 and BP4) overlapping with these SD clusters are represented by green bars. The genomic locations of the 1q21 deletions identified in this study are represented by red bars and indicated with family identifiers. RefSeq Genes are indicated by dark-blue rectangular bars. For genes with multiple isoforms, the bars represent the coordinates of the maximal region among the isoforms. **b**. Pedigrees of the families carrying heterozygous 1q21 deletions

Fig. 3 GJA8 mutation spectrum. Schematic of GJA8 showing the protein domains according to UniProt (entry ID: P48165). Above: previously published mutations are shown. Below: the missense variants identified in our cohort are indicated: red indicates likely pathogenic, blue, likely benign and grey, unknown clinical significance. NT, N-terminal

877 domain; TM, transmembrane domain; ECL, extracellular loop; ICL, cytoplasmic loop; CT,
878 C-terminal domain