# The utility of massively parallel sequencing for posterior polymorphous corneal dystrophy type 3 molecular diagnosis

Dudakova Lubica[a], Evans Cerys J.[b], Pontikos Nikolas[b], Hafford-Tear Nathaniel J.[b], Malinka Frantisek[a,c], Skalicka Pavlina[a,d], Horinek Ales[e,f], Francis L. Munier[g], Nathalie Voide[g], Studeny Pavel[h], Vanikova Lucia[i], Kubena Tomas[j], Rojas Lopez Karla E.[b], Davidson Alice E.[b], Hardcastle Alison J.[b], Tuft Stephen J.[k], Liskova Petra[a,b,d]

[a] Research Unit for Rare Diseases, Department of Paediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University and General University Hospital in Prague, Ke Karlovu 2, 128 08 Prague 2, Czech Republic

[b] UCL Institute of Ophthalmology, 11-43 Bath Street, EC1V 9EL, London, United Kingdom

[c] Department of Computer Science, Czech Technical University in Prague, Karlovo namesti 13, 128 08 Prague, Czech Republic

[d] Department of Ophthalmology, First Faculty of Medicine, Charles University and General University Hospital in Prague, U Nemocnice 2, 128 08, Prague, Czech Republic

[e] 3rd Department of Medicine - Department of Endocrinology and Metabolism, First Faculty of Medicine, Charles University and General University Hospital in Prague, U Nemocnice 1, 128 08, Prague 2, Czech Republic

[f] Institute of Biology and Human Genetics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Albertov 4, 128 00, Prague, Czech Republic

[g] Jules-Gonin Eye Hospital, Fondation Asile des Aveugles, University of Lausanne, Avenue de France 15, 1004 Lausanne, Switzerland

[h] Ophthalmology Department, Third Faculty of Medicine, Charles University and Teaching Hospital Kralovske Vinohrady, Srobarova 1150/50, 100 34, Prague, Czech Republic.

[i] Center for Eye Microsurgery, Gagarinova 7/B, 821 03, Bratislava, Slovakia

[j] Ophthalmology Clinic of Dr. Tomas Kubena, U Zimniho stadionu 1759, 760 00, Zlin, Czech Republic

[k] Moorfields Eye Hospital, 162 City Road, EC1V 2PD, London, United Kingdom

Correspondence to:
Petra Liskova, MD, PhD
Research Unit for Rare Diseases, Department of Paediatrics and Adolescent Medicine
First Faculty of Medicine, Charles University and General University Hospital in Prague
Ke Karlovu 2, 128 08, Prague, Czech Republic
Tel: +420 224 967 139, email: petra.liskova@lf1.cuni.cz

**ABSTRACT**

The aim of this study was to identify the molecular genetic cause of disease in posterior polymorphous corneal dystrophy (PPCD) probands of diverse origin and to assess the utility of massively parallel sequencing in the detection of *ZEB1* mutations. We investigated a total of 12 families (five British, four Czech, one Slovak and two Swiss). Ten novel and two recurrent disease-causing mutations in *ZEB1*, were identified in probands by Sanger (n=5), exome (n=4) and genome (n=3) sequencing. Sanger sequencing was used to confirm the mutations detected by massively parallel sequencing, and to perform segregation analysis. Genome sequencing revealed that one proband harboured a novel ~0.34 Mb heterozygous *de novo* deletion spanning exons 1-7 and part of exon 8. Transcript analysis confirmed that the *ZEB1* transcript is detectable in blood-derived RNA samples and that the disease-associated variant c.482-2A>G leads to aberrant pre-mRNA splicing. *De novo* mutations, which are a feature of PPCD3, were found in the current study with an incidence rate of at least 16.6%. In general, massively parallel sequencing is a time-efficient way to detect PPCD3-associated mutations and, importantly, genome sequencing enables the identification of full or partial heterozygous *ZEB1* deletions that can evade detection by both Sanger and exome sequencing. These findings contribute to our understanding of PPCD3, for which currently, 49 pathogenic variants have been identified, all of which are predicted to be null alleles.

**Key words:** *ZEB1;* posterior polymorphous corneal dystrophy type 3; massively parallel sequencing; exome; genome; aberrant splicing; breakpoint mapping

# 1. INTRODUCTION

Posterior polymorphous corneal dystrophy (PPCD) is a genetically heterogeneous, autosomal dominant disorder, characterised by vesicular lesions, bands and opacities at the level of Descemet membrane and the corneal endothelium (Cibis et al., 1977). Decreased visual acuity can result from amblyopia, irregular astigmatism from corneal steepening, secondary glaucoma and/or corneal edema (Krachmer, 1985; Liskova et al., 2010; Liskova et al., 2013). To date, variants within three genes *ZEB1*, *OVOL2* and *GRHL2* have been identified to be disease-causing (Krafchak et al., 2005; Davidson et al., 2016; Le et al., 2016; Liskova et al., 2018).

Mutations in the zinc finger E-box binding homeobox 1 (*ZEB1)* cause PPCD type 3 (PPCD3; OMIM #609141) (Krafchak et al., 2005). In addition to ocular features, it has been reported that patients with PPCD3 have an increased incidence of hernia and hydrocele (Krafchak et al., 2005; Liskova et al., 2013). Three cases with agenesis or hypoplasia of the corpus callosum have also been documented (Jang et al., 2014; Chaudhry et al., 2017). Individuals with PPCD3 can be asymptomatic, and there are rare reports of incomplete penetrance (Krafchak et al., 2005; Liskova et al., 2013).

*ZEB1* encodes a zinc finger transcription factor. The protein plays a role in epithelial-mesenchymal transition (EMT) by activating the transcription of mesenchymal genes while repressing epithelial genes (Park et al., 2008). EMT is important for embryonic development, wound healing, fibrosis and cancer progression (Chen et al., 2017). Importantly, it is hypothesised that dysregulation of this pathway underpins the pathophysiology of PPCD (Davidson et al., 2016; Liskova et al., 2018). PPCD3 is attributed to ZEB1 haploinsufficiency, (Liskova et al., 2016; Chung et al., 2017) and the reduced levels of ZEB1 induce an altered cellular response to apoptotic stimuli and cell barrier function (Zakharevich et al., 2017).

In this study, we demonstrate the utility of massively parallel sequencing for the identification of *ZEB1* mutations. Furthermore, we establish that it is possible to determine the effects of *ZEB1* variants on pre-mRNA splicing using blood-derived RNA. Finally, we present a

comprehensive summary of *ZEB1* mutations that have been reported to cause PPCD3 to date.

## 2. METHODS

### 2.1. Participants, clinical examination and samples

The study was approved by the Ethics committee of the General University Hospital in Prague (reference no. 151/11 S-IV) or Moorfields Eye Hospital (REC references 13/LO/1084 and 09/H0724/25) and adhered to the tenets of the Helsinki Declaration. All participants signed informed consent. Twelve families were investigated; four white Czech (C31, C32, C34, C36), four white British (B9, B10, B11 and B12), one South Asian British (B13), one white Slovak (SK1) and two white Swiss (SW1, SW2). A diagnosis of PPCD was based on established clinical criteria (Cibis et al., 1977; Liskova et al., 2010; Liskova et al., 2013). Best corrected visual acuity (BCVA) was measured using Snellen charts and converted to decimal values. Participating individuals were asked to provide information on their general health status.

Genomic DNA from probands and any additional available family members was extracted from venous blood using a Gentra Puregene blood kit (Qiagen, Hilden, Germany) or from saliva using an Oragene kit (Oragene OG-300, DNA Genotek, Canada). In one affected individual total RNA was isolated from venous blood using a QIAamp RNA Blood Mini Kit (Qiagen), transcribed into cDNA with the SuperScript III Reverse Transcriptase (Thermo Fisher, Waltham, Massachusetts).

### 2.2. Sanger sequencing

We first excluded the presence of pathogenic variants in regulatory regions of *OVOL2* and *GRHL2* in all probands by conventional Sanger sequencing (Davidson et al., 2016; Liskova et al., 2018). In probands from families C31, C32, C34 and B9-B11 (Figure 1) the entire *ZEB1* coding region and flanking intronic sequences (NM_030751.5, NG_017048.1) were

amplified by PCR and directly sequenced (Evans et al., 2015). Sequence variants were described according to the Human Genome Variation Society guidelines (http://varnomen.hgvs.org/) (den Dunnen et al., 2016). Direct sequencing was also used to confirm the likely disease-causing variants found by massively parallel sequencing which included a large deletion. Furthermore, direct sequencing was also used for segregation analysis within the families.

## 2.3. Massively parallel sequencing

DNA from the probands from families C31, B12, B13, SW1 and SW2 (Figure 1) were subjected to exome sequencing. Sequencing libraries were generated using a SureSelect Human All Exome V6 capture kit (Agilent, Santa Clara, California). Libraries were sequenced on HiSeq4000 sequencer (Illumina). Probands from families C31, C36 and SK1 (Figure 1-3) were analysed by genome sequencing using a TruSeq Nano DNA library preparation kit and a HiSeq X Ten sequencer (Illumina, San Diego, California). Reads were aligned to the GRCh37/hg19 human reference sequence with NovoAlign V3.02.08 (Novocraft, Malaysia). Variant calling was performed with Genome Analysis Tool Kit (GATK) HaplotypeCaller (version 4.0.1.2) (McKenna et al., 2010). Variants were annotated using the Variant Effect Predictor (McLaren et al., 2016).

Both coding and non-coding rare variants with a minor allele frequency (MAF) ≤ 0.005 as per gnomAD (Lek et al., 2016) in genes known to be implicated in the pathogenesis of corneal endothelial dystrophies (*ZEB1*, *GRHL2*, *OVOL2*, *COL8A2*, *SLC4A11*, *TCF4*) (Biswas et al., 2001; Krafchak et al., 2005; Vithana et al., 2006; Baratz et a., 2010; Davidson et al., 2016; Le et al., 2016; Liskova et al., 2018) were further investigated for potential pathogenicity. The frequency of detected variants was also determined in 2,430 exomes generated from Czech individuals available through projects of the National Centre for Medical Genomics (http://ncmg.cz/en). Larger structural variations were compared to entries in the Database of Genomic Variants (MacDonald et al., 2014). The exome and genome sequencing data,

aligned to the human reference GRCh37/hg19, were visualized with the Integrated Genomics Viewer (IGV) (Broad institute, California, USA) (Thorvaldsdottir et al., 2013).

In one patient we did not detect a pathogenic variant and we therefore manually searched aligned genome sequencing reads for stretches of homozygosity around the *ZEB1* locus indicative of a heterozygous deletion. Importantly, we used the soft clip function of the IGV software to visualise the likely breakpoints. Primers were designed at each end (forward 5´-CTTTCTGCTGAGGCCATTTC-3´ and reverse 5´-TGGGTGACTAGAGCCAGACC-3´ primer) and the PCR product was Sanger sequenced to confirm the breakpoints. Control data on copy number variations in the *ZEB1* genomic region were mined from the Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/home; accessed December 2018).

A schematic diagram of the pipeline used to process genomic DNA samples in our study is shown in Supplementary Figure S1.

## 2.4. Paternity testing

Paternity testing was performed in families C31 and C32 to prove *de novo* origin of the detected mutations using an established set of 16 forensic markers (Ensenberger et al., 2010).

## 2.5. Functional assessment of pre-mRNA splicing

To assess the effect of a selected variant on pre-mRNA splicing, cDNA of an affected individual II:2 from family C36 was amplified and analysed by direct sequencing, spanning a region of *ZEB1* exons 4-7 using forward 5´-CTGAGGCACCTGAAGAGGAC-3´ and reverse 5´-TTGCAGTTTGGGCATTCATA-3´ primer.

## 2.6. Review of mutations associated with PPCD3

We performed a comprehensive and manually curated review of all publicly available literature reporting *ZEB1* PPCD-associated variants. All variants listed in this article are

annotated in accordance with transcript NM_030751.5 and GRCh37/hg19 human genome assembly.

## 3. RESULTS

The clinical, demographic and sequencing data of 19 affected individuals from 12 families with PPCD3 are summarized in Table 1.

Slit-lamp examination showed characteristic bilateral corneal signs in all affected subjects. Only 9 eyes of 5 individuals (26.3%) (mean age 27±11.9; range 8-40) had a BCVA of 1.0, while moderate or severe visual impairment (i. e. BCVA <0.33) was present in 10 eyes of 7 individuals (36.8%) (mean age 39±16,5, range 20-59). Two individuals had nystagmus, including subject II:2 from family C34 who was noted to have cloudy corneas from 6 weeks of age. Four individuals (21%) had a keratoplasty in at least one eye. None of the participants in this study had been diagnosed with glaucoma. Two individuals out of 19 reported that they had undergone surgery for hernia and one for hydrocele. None of the probands had signs of cognitive deficiency consistent with agenesis or hypoplasia of corpus callosum (Jang et al., 2014; Chaudhry et al., 2017).

In total, we identified ten novel heterozygous variants in *ZEB1* that were evaluated as pathogenic; c.454del, c.482-2A>G, c.646G>T, c.1045G>T, c.1498_1501del, c.1669C>T, c.1918C>T, c.2577del, c.2650C>T (Figure 1, 2) and a large ~0.34 Mb deletion (Figure 3). Two known *ZEB1* mutations were also found; c.689_690del and c.1576dup (Figure 1). In concordance with previous observations (Krafchak et al., 2005; Evans et al., 2015; Liskova et al., 2016) all affected individuals had a heterozygous loss-of-function allele. The c.1576dup mutation was observed in an individual who also had a unique c.469C>G; p.(His157Asp) variant, predicted to be benign by four out of the five software tools we used (Supplementary Table 1). Unfortunately, familial samples were not able available to determine the phase of these variants in the affected proband. In samples analysed by massively parallel

sequencing no potentially disease-associated variants were found in other genes known to be associated with corneal endothelial dystrophies.

Of note, the proband from family C31 with sporadic disease underwent screening by Sanger sequencing, followed by exome sequencing and genome sequencing. Visualising the aligned genome sequencing reads in IGV using a soft clip function enabled us to identify likely deletion breakpoints (Supplementary Figure S2). Sanger sequencing across the region confirmed the presence of the deletion encompassing *ZEB1* exons 1-7 and part of exon 8 (chr10.hg19:g.31,476,838_31,812,958del) (Figure 3).

In samples that underwent exome sequencing, the sequencing depth of the *ZEB1* gene (NM_030751.5) exceeded 20x in at least 97.7% of the coding region across all samples analysed (n=5). In samples subjected to genome sequencing, the sequencing depth of the entire *ZEB1* region (NG_017048.1) exceeded 20x in at least 97.9% of the region across all samples analysed (n=3).

In family C36, analysis of the *ZEB1* transcript in the blood of an individual who had a splice site variant, c.482-2A>G, demonstrated that the mutation alters pre-mRNA splicing, resulting in exon 5 skipping and insertion of a premature termination codon, described as r.482_684del, p.Gly161Glufs*9 (Figure 2).

None of the novel pathogenic mutations located in *ZEB1* coding region or the unique variant p.(His157Asp) were found in gnomAD, Database of Genomic Variants or in data specific to the Czech population. Segregation analysis within the families also supported their pathogenicity (Figure 1-3). As parents in families C31 and C32 were clinically unaffected and did not carry *ZEB1* mutations detected in their children, we suspected *de novo* origin of the mutations, which was subsequently confirmed by paternity testing (Figure 1,3).

A schematic representation of all PPCD3-associated *ZEB1* variants that have been reported to date, including this study, is provided in Figure 4. In total, 49 mutations have been identified in 54 families of various ethnic backgrounds. All are predicted to be null alleles and hence support the hypothesis that PPCD3 arises due to *ZEB1* haploinsufficiency. Further details on the identified mutations are shown in Supplementary Table 2.

8

**4. DISCUSSION**

In this study we report ten novel and two recurrent mutations in *ZEB1,* as well as the first genetic study of PPCD3 patients from Slovakia and Switzerland. Importantly, we show that massively parallel sequencing is a useful tool, particularly genome sequencing, to genetically diagnose the condition. We also demonstrate for the first time that it is possible to experimentally determine the effect of *ZEB1* variants on pre-mRNA splicing using whole-blood derived RNA. In addition, we put our work in context by providing a comprehensive review of all disease-associated *ZEB1* variants reported to date.

All *ZEB1* reported pathogenic variants either introduce premature termination codons or comprise large deletions predicted to completely disrupt the gene transcription (Krafchak et al., 2005; Aldave et al, 2007; Liskova et al., 2007; Vincent et al., 2009; Liskova et al., 2010; Nguyen et al., 2010; Bakhtiare et al., 2013; Liskova et al., 2013; Lechner et al., 2013; Jang et al., 2014; Evans et al., 2015; Liskova et al., 2016). Although it has been anticipated that aberrant *ZEB1* transcripts are likely eliminated by the mRNA nonsense mediated decay surveillance pathway, cellular studies have shown that *ZEB1* disease-causing variants may, in addition to the reduced protein production, also result in impaired nuclear localization of the encoded protein (Chung et al., 2014).

Of the 49 *ZEB1* pathogenic variants that have been described, five (10%) were large deletions (Liskova et al., 2016; Chaudhry et al., 2017). Their size varies considerably (0.071-3.3 Mb), and importantly they can evade detection by exome sequencing, qPCR or SNP arrays, depending on primer or SNP locations. Genome sequencing overcomes these limitations. In proband from family 31 it has allowed us to detect the fifth unique deletion in association with PPCD3 and to precisely visualize and map its breakpoints. On this basis we decided to use genome sequencing as the first line approach in two other probands. Massively parallel sequencing also enabled us to screen multiple genes associated with corneal endothelial disease, which could be useful in determining the genetic cause in cases of bilateral corneal edema of uncertain aetiology (Bakhtiari et al., 2013; Evans et al., 2015).

We also conducted transcript analysis using blood-derived RNA to show for the first time that a novel splicing mutation c.482-2A>G causes aberrant pre-mRNA splicing.

There have only been two recurrent *ZEB1* mutations reported. In this study we have documented the *de novo* origin of c.689_690del; p.(His230Arg*7), previously found in an American family (Bakhtiari et al., 2013). We have also identified c.1576dup; p.(Val526Glyfs*3) in a patient of South Asian ethnicity, previously reported in three families from North America (Krafchak et al., 2005; Bakhtiari et al., 2013; Lechner et al, 2013) and one British family (Evans et al., 2015). Interestingly, this patient (B13-II:1) also harboured an additional unique missense variant in *ZEB1* p.(His157Asp). Unfortunately, we were not able to determine the phase of this variant with respect to c.1576dup to assess its contribution to the phenotype. However, *in silico* analysis suggests that it is likely benign and, if not, we hypothesise that it is located on the same allele as the loss-of-function variant given that bi-allelic loss-of-function *ZEB1* alleles are embryonic lethal (Takagi et al., 1998; Liu et al., 2008).

The *de novo* mutation rate amongst PPCD3 probands is unknown, partly because individuals with PPCD3 can be asymptomatic, and some subjects even show incomplete penetrance for the disease, and thus a reported family history can be unreliable (Krafchak et al., 2005; Liskova et al., 2007; Liskova et al., 2010; Bakhtiari et al., 2013; Evans et al., 2015). Herein, we identified *de novo* mutations in two of 12 families studied, resulting in an incidence rate of 16.6% in this study. However, this figure may be an underestimate because although there was no family history reported in families B10-B12 and SW2, the parents of the probands were unavailable for examination or testing.

*ZEB1* null alleles are found at an extremely low frequency in large-scale public databases such as gnomAD and the Database of Genomic Variants. As they lack individual-level data without details of ophthalmic examination, the presence of such variants in these datasets does not exclude their potential pathogenic role (Cooper et al., 2011; Coe et al., 2014).

In summary, given that *ZEB1* deletions of various sizes as well as non-coding pathogenic variants in *GRHL2* and *OVOL2* are present in a significant number of PPCD cases, we

consider genome sequencing to be a time and cost-effective approach to identify the

molecular genetic cause of PPCD.

**ACKNOWLEDGEMENTS**

## REFERENCES

Aldave, A.J., Yellore, V.S., Yu, F., Bourla, N., Sonmez, B., Salem, A.K., Rayner, S.A., Sampat K.M., Krafchak, C.M., Richards, J.E., 2007. Posterior polymorphous corneal dystrophy is associated with *TCF8* gene mutations and abdominal hernia. Am J Med Genet A. 143A, 2549-2556.

Bakhtiari, P., Frausto, R.F., Roldan, A.N., Wang, C., Yu, F., Aldave, A.J., 2013. Exclusion of pathogenic promoter region variants and identification of novel nonsense mutations in the zinc finger E-box binding homeobox 1 gene in posterior polymorphous corneal dystrophy. Mol Vis. 19, 575-580.

Baratz, K.H., Tosakulwong, N., Ryu, E., Brown, W.L., Branham, K., Chen, W., Tran, K.D., Schmid-Kubista, K.E., Heckenlively, J.R., Swaroop, A., Abecasis, G., Bailey, K.R., Edwards, A.O., 2010. E2-2 protein and Fuchs's corneal dystrophy. N Engl J Med. 363, 1016-1024.

Biswas, S., Munier, F.L., Yardley, J., Hart-Holden, N., Perveen, R., Cousin, P., Sutphin, J.E., Noble, B., Batterbury, M., Kielty, C., Hackett, A., Bonshek, R., Ridgway, A., McLeod, D., Sheffield, V.C., Stone, E.M., Schorderet, D.F., Black, G.C., 2001. Missense mutations in COL8A2, the gene encoding the alpha2 chain of type VIII collagen, cause two forms of corneal endothelial dystrophy. Hum Mol Genet. 10, 2415-2423.

Cibis, G.W., Krachmer, J.A., Phelps, C.D., Weingeist, T.A., 1977. The clinical spectrum of posterior polymorphous dystrophy. Arch Ophthalmol. 95, 1529-1537.

Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W., Vulto-van Silfhout, A.T., Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E., Schuurs-Hoeijmakers, J.H., Hoischen, A., Pfundt, R., Krumm, N., Carvill, G.L., Li, D., Amaral, D., Brown, N., Lockhart, P.J., Scheffer, I.E., Alberti, A., Shaw, M., Pettinato, R., Tervo, R., de Leeuw, N., Reijnders, M.R., Torchia, B.S., Peeters, H., O'Roak, B.J., Fichera, M., Hehir-Kwa, J.Y., Shendure, J., Mefford, H.C., Haan, E., Gécz, J., de Vries, B.B., Romano, C., Eichler, E.E., 2014. Refining analyses of copy number variations identifies specific genes associated with developmental delay. Nat Genet. 46, 1063-1071.

Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., Abdel-Hamid, H., Bader, P., McCracken, E., Niyazov, D., Leppig, K., Thiese, H., Hummel, M., Alexander, N., Gorski, J., Kussmann, J., Shashi, V., Johnson, K., Rehder, C., Ballif, B.C., Shaffer, L.G., Eichler, E.E., 2011. A copy number variation morbidity map of developmental delay. Nat Genet. 43, 838-846.

Chaudhry, A., Chung, B.H., Stavropoulos, D.J., Araya, M.P., Ali, A., Heuon, E., Chitayat, D., 2017. Agenesis of the corpus callosum, developmental delay, autism spectrum disorder, facial dysmorphism, and posterior polymorphous corneal dystrophy associated with *ZEB1* gene deletion. Am J Med Genet A. 173, 2467-2471.

Chen, T., You, Y., Jiang, H., Wang, Z.Z., 2017. Epithelial-mesenchymal transition (EMT): A biological process in the development, stem cell differentiation, and tumorigenesis. J Cell Physiol. 232, 3261-3272.

Chung, D.W., Frausto, R.F., Ann, L.B., Jang, M.S., Aldave, A.J., 2014. Functional impact of *ZEB1* mutations associated with posterior polymorphous and Fuchs´ endothelial dystrophies. Invest Ophthalmol Vis Sci. 55, 6159-6166.

Chung, D.D., Frausto, R.F., Lin, B.R., Hanser, E.M., Cohen, Z., Aldave, A.J., 2017. Transcriptomic Profiling of Posterior Polymorphous Corneal Dystrophy. Invest Ophthalmol Vis Sci. 58, 3202-3214.

Davidson, A.E., Liskova, P., Evans, C.J., Dudakova, L., Noskova, L., Pontikos, N., Hartmannova, H., Hodanova, K., Stranecky, V., Kozmik, Z., Levis, H.J., Idigo, N., Sasai, N., Mahler, G.J., Bellingham, J., Veli, N., Ebenzer, N.D., Cheetham, M.E., Daniels, J.T., Thaung, C.M., Jirsova, K., Plagnol, V., Filipec, M., Kmoch, S., Tuft, S.J., Hardastle, A.J., 2016. Autosomal-Dominant Corneal Endothelial Dystrophies CHED1 and PPCD1 Are Allelic Disorders Caused by Non-coding Mutations in the Promoter of *OVOL2*. Am J Hum Genet. 98,75-89.

den Dunnen, J.T., Dalgleish, R., Maglott, D.R., Hart, R.K., Greenblatt, M.S., McGowan-Jordan, J., Roux, A.F., Smith, T., Antonarakis, S.E., Taschner, P.E., 2016. HGVS

Recommendations for the Description of Sequence Variants: 2016 Update. Hum

Mutat. 37, 564-569.

Ensenberger, M.G., Thompson, J., Hill, B., Homick, K., Kearney, V., Mayntz-Press, K.A.,

Mazur, P., McGuckian, A., Myers, J., Raley, K., Raley, S.G., Rothove, R., Wilson, J.,

Wieczorek, D., Fulmer, P.M., Storts, D.R., Krenke, B.E., 2010. Developmental

validation of the PowerPlex 16 HS System: an improved 16-locus fluorescent STR

multiplex. Forensic Sci Int Genet. 4, 257-264.

Evans, C.J., Liskova, P., Dudakova, L., Hrabcikova, P., Horinek, A.,Jirsova, K., Filipec, M.,

Hardcastle, A.J., Davidson, A.E., Tuft, S.J.,  2015. Identification of six novel mutations

in *ZEB1* and description of the associated phenotypes in patients with posterior

polymorphous corneal dystrophy 3. Ann Hum Genet. 79, 1-9.

Gwilliam, R., Liskova, P., Filipec, M., Kmoch, S., Jirsova, K., Huckle, E.J., Stables, C.L.,

Bhattacharya, S.S., Hardcastle, A.J., Deloukas, P., Ebenzer, N.D., 2005. Posterior

polymorphous corneal dystrophy in Czech families maps to chromosome 20 and

excludes the *VSX1* gene. Invest Ophthalmol Vis Sci. 46, 4480-4484.

Heon, E., Mathers, W.D., Alward, W.L., Weisenthal, R.W., Sunden, S.L., Fishbaugh, J.A.,

Taylor, C.M., Krachmer, J.H., Sheffield, V.C., Stone, E.M., 1995. Linkage of posterior

polymorphous corneal dystrophy to 20q11. Hum Mol Genet. 4, 485-488.

Jang, M.S., Roldan, A.N., Frausto, R.F., Aldave, A.J., 2014. Posterior polymorphous corneal

dystrophy 3 is associated with agenesis and hypoplasia of the corpus callosum.

Vision Res. 100, 88-92.

Krachmer, J.H., 1985 Posterior polymorphous corneal dystrophy: a disease characterized by

epithelial-like endothelial cells which influence management and prognosis. Trans Am

Ophthalmol Soc. 83, 413-475.

Krafchak, C.M., Pawar, H., Moroi, S.E., Sugar, A., Lichter, P.R., Mackey, D.A., Mian, S.,

Nairus, T., Elner, V., Schteingart, M.T., Downs, C.A., Kijek, T.G., Johnson, J.M.,

Trager, E.H., Rozsa, F.W., Mandal, M.N., Epstein, M.P., Vollrath, D., Ayyagari, R.,

Boehnke, M., Richards, J.E., 2005. Mutations in *TCF8* cause posterior polymorphous

corneal dystrophy and ectopic expression of *COL4A3* by corneal endothelial cells. Am J Hum Genet. 77, 694-708.

Le, D.J., Chung, D.W., Frausto, R.F., Kim, M.J., Aldave, A.J., 2016. Identification of potentially pathogenic variants in the posterior polymorphous corneal dystrophy 1 locus. Plos One. 11(6), e0158467.

Lechner, J., Dash, D.P., Muszynska, D., Hosseini, M.,  Segev, F., George, S., Frazer, D.G., Moore, J.E., Kaye, S.B., Young, T., Simpson, D.A., Churchill, A.J., Heon, E., Willoughby, C.E., 2013. Mutational spectrum of the *ZEB1* gene in corneal dystrophies supports a genotype-phenotype correlation. Invest Ophthalmol Vis Sci. 54,3215-3223.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E, Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., Exome Aggregation Consortium, 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 536, 285-291.

Liskova, P., Tuft, S.J., Gwilliam, R., Ebenezer, N.D., Jirsova, K., Prescott, Q., Martincova, R., Pretorius, M., Sinclair, N., Boase, D.L., Jeffrey, M.J., Deloukas, P., Hardcastle, A.J., Filipec, M., Bhattacharya, S.S., 2007. Novel mutations in the *ZEB1* gene identified in

Czech and British patients with posterior polymorphous corneal dystrophy. Hum Mutat. 28, 638.

Liskova, P., Filipec, M., Merjava, S., Jirsova, K., Tuft, S.J., 2010. Variable ocular phenotypes of posterior polymorphous corneal dystrophy caused by mutations in the *ZEB1* gene. Ophthalmic Genet. 31, 230-234.

Liskova, P., Palos, M., Hardcastle, A.J., Vincent, A.L., 2013. Further genetic and clinical insights of posterior polymorphous corneal dystrophy 3. JAMA Ophthalmol. 131, 1296-1303.

Liskova, P., Evans, C.J., Davidson, A.E., Zaliova, M., Dudakova, L., Trkova, M., Stranecky, V., Carnt, N., Plagnol, V., Vincent, A.L., Tuft, S.J., Hardcastle, A.J., 2016. Heterozygous deletions at the *ZEB1* locus verify haploinsufficiency as the mechanism of disease for posterior polymorphous corneal dystrophy type 3. Eur J Hum Genet. 24, 985-991.

Liskova, P., Dudakova, L., Evans, C.J., Rojas Lopez, K.E., Pontikos, N., Athanasiou, D., Jama, H., Sach, J., Skalicka, P., Stranecky, V., Kmoch, S., Thaung, C., Filipec, M., Cheetham, M.E., Davidson, A.E., Tuft, S.J., Hardcastle, A.J., 2018. Ectopic *GRHL2* Expression Due to Non-coding Mutations Promotes Cell State Transition and Causes Posterior Polymorphous Corneal Dystrophy 4. Am J Hum Genet. 102, 447-459.

Liu, Y., Peng, X., Tan, J., Darling, D.S., Kaplan, H.J., Dean, D.C., 2008. Zeb1 mutant mice as a model of posterior corneal dystrophy. Invest Ophthalmol Vis Sci. 49, 1843-1849.

MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., Scherer, S.W., 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 42, D986-992.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297-1303.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P.,

Cunningham, F., 2016. The Ensembl Variant Effect Predictor. Genome Biol. 17, 122.

Nguyen, D.Q., Hosseini, M., Billingsley, G., Heon, E., Churchill, A.J., 2010. Clinical

phenotype of posterior polymorphous corneal dystrophy in a family with a novel *ZEB1*

mutation. Acta Ophthalmol. 88, 695-699.

Park, S.M., Gaur, A.B., Lengyel, E., Peter, M.E., 2008. The miR-200 family determines the

epithelial phenotype of cancer cells by targeting the E-cadherin repressors *ZEB1* and

*ZEB2*. Genes Dev. 22, 894-907.

Takagi, T., Moribe, H., Kondoh, H., Higashi, Y., 1998. DeltaEF1, a zinc finger and

homeodomain transcription factor, is required for skeleton patterning in multiple

lineages. Development. 125, 21-31.

Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV):

high-performance genomics data visualization and exploration. Brief Bioinform. 14,

178-192.

Vincent, A.L., Niederer, R.L., Richards, A., Karolyi, B., Patel, D.V., McGhee, C.N., 2009.

Phenotypic characterisation and *ZEB1* mutational analysis in posterior polymorphous

corneal dystrophy in a New Zealand population. Mol Vis. 15, 2544-2553.

Vithana, E.N., Morgan, P., Sundaresan, P., Ebenezer, N.D., Tan, D.T., Mohamed, M.D.,

Anand, S., Khine, K.O., Venkataraman, D., Yong, V.H., Salto-Tellez, M.,

Venkatraman, A., Guo, K., Hemadevi, B., Srinivasan, M., Prajna, V., Khine, M.,

Casey, J.R., Inglehearn, C.F., Aung, T., 2006. Mutations in sodium-borate

cotransporter SLC4A11 cause recessive congenital hereditary endothelial dystrophy

(CHED2). Nat Genet. 38, 755-757.

Zakharevich, M., Kattan, J.M., Chen, J.L., Lin, B.R., Cervantes, A.E., Chung, D.D., Frausto,

R.F., Alvade, A.J., 2017. Elucidating the molecular basis of PPCD: Effects of

decreased *ZEB1* expression on conreal endothelial cell function. Mol Vis. 23, 740-

752.

**Figure 1: Disease-causing *ZEB1* point mutations and small heterozygous deletions identified in ten families.** Affected individuals are shown as black symbols. Individuals who tested positive for a particular mutation in each family are labelled with +/- and those who were negative with -/-. The proband from family C32 harboured a known mutation c.689_690del which arose *de novo*; three affected individuals from family C34 carried a novel mutation c.646G>T; in family SK1 three individuals had novel mutation c.1045G>T; in family B9 two affected individuals had a novel mutation c.2577del and in families B10, B11 and B12 the probands were found to each harbour a different novel mutation c.1918C>T, c.1669C>T and c.2650C>T, respectively. The proband from family B13 carried a recurrent mutation c.1576dup. Two Swiss families SW1 and SW2 carried novel c.1498_1501del and c.454del, respectively.

**Figure 2: Assessment of the c.482-2A>G *ZEB1* variant on pre-mRNA splicing in family C36.** Pedigree (A), a novel heterozygous mutation located in a canonical splice site (B). PCR of blood derived cDNA (C) from a control individual (line 1 marked as C) and the patient sample (Line 2, marked as P) showing two different products. The smaller fragment was found to represent skipping of exon 5 (D) due to aberrant pre-mRNA splicing (r.482_684del) (NM_030751.5, NG_017048.1). Direct sequencing of the PCR amplified products confirming expected splicing patterns in the control sample (E) and one wild type and one aberrant transcript resulting in exon 5 skipping in the patient sample (F).

**Figure 3: A schematic representation of the *de novo* partial heterozygous *ZEB1* deletion identified in family C31.** The breakpoints of the deletion were identified, and the corresponding sequence chromatogram is shown.

**Figure 4**: **Schematic diagram of the *ZEB1* gene showing 49 heterozygous pathogenic mutations identified to date in patients with posterior polymorphous corneal dystrophy type 3.** Larger deletions are shown as a bar with coordinates of the deleted region (GRCh37/hg19). Mutations identified in the current study are highlighted in red, novel variants are underlined. Description of coding and splicing mutations are based on reference sequence NM_030751.5, NG_017048.1 and numbering starts at the translation initiation site.