Journal of
MEDICAL Genetics

Biallelic variants in Plexin B2 (*PLXNB2*) cause amelogenesis imperfecta, hearing loss and intellectual disability.

Journal:	*Journal of Medical Genetics*
Manuscript ID	jmg-2023-109728.R2
Article Type:	Original research
Date Submitted by the Author:	n/a
Complete List of Authors:	Smith, Claire; University of Leeds Faculty of Medicine and Health, Institute of Medical Research, St James's University Hospital Laugel-Haushalter, Virginie; Université de Strasbourg, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258, CRNS-UMR7104 Hany, Ummey; University of Leeds Faculty of Medicine and Health, Institute of Medical Research, St James's University Hospital Best, Sunayna; University of Leeds Faculty of Medicine and Health, Institute of Medical Research, St James's University Hospital; Leeds Teaching Hospitals NHS Trust, Yorkshire Regional Genetics Service Taylor, Rachel; Manchester University NHS Foundation Trust, Manchester Centre for Genomic Medicine, St Mary's Hospital; EMQN CIC Poulter, James; University of Leeds Faculty of Medicine and Health, Leeds Institute of Medical Research, St James's University Hospital Wortmann, Saskia; Paracelsus Medical University Salzburg, Department of Pediatrics, University Children's Hospital; Radboudumc, Amalia Children´s Hospital Feichtinger, Rene; Paracelsus Medical University Salzburg, Department of Pediatrics, University Children's Hospital Mayr, Johannes; Paracelsus Medical University Salzburg, Department of Paediatrics, University Children's Hospital Al Bahlani, Suhaila; Government of Oman Ministry of Health, Dental & OMFS Clinic, Al Nahdha Hospital Nikolopoulos, Georgios; BSRC Alexander Fleming, Institute for Fundamental Biomedical Research Rigby, Alice; University of Leeds Faculty of Medicine and Health, Institute of Medical Research; University of Leeds Faculty of Medicine and Health, School of Dentistry Black, Graeme; University of Manchester Institute of Science and Technology Department of Biomolecular Science, Division of Evolution and Genomic Sciences, Manchester Academic Health Science Centre; Manchester University NHS Foundation Trust, Manchester Centre for Genomic Medicine, St Mary's Hospital Watson, Christopher; University of Leeds Faculty of Medicine and Health, Institute of Medical Research, St James's University Hospital; Leeds Teaching Hospitals NHS Trust, North East and Yorkshire Genomic Laboratory Hub, Central Lab, St James's University Hospital Mansour, Sahar; University of London, Lymphovascular Research Unit, Molecular and Clinical Sciences Research Institute, St George's Hospital;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	St George's University Hospitals NHS Foundation Trust, SW Thames Regional Centre for Genomics Inglehearn, Chris; University of Leeds Faculty of Medicine and Health, Institute of Medical Research, St James's University Hospital Mighell, Alan; University of Leeds Faculty of Medicine and Health, School of Dentistry Bloch-Zupan, Agnès; University of Strasbourg, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258, CNRS-UMR7104; University of Strasbourg Faculty of Medicine
Keywords:	Dentistry, Dentistry, Genetic Research, Genetics, Genetics, Medical

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

BMJ

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Biallelic variants in Plexin B2 (*PLXNB2*) cause amelogenesis imperfecta, hearing loss and intellectual disability.

Claire E L Smith[1*], Virginie Laugel-Haushalter[2*], Ummey Hany[1], Sunayna Best[1,3], Rachel L Taylor[4,5,6], James A. Poulter[1], The UK Inherited Retinal Disease Consortium, Genomics England Research Consortium, Saskia B Wortmann[7,8], René G Feichtinger[7], Johannes A Mayr[7], Suhaila Al Bahlani[9], Georgios Nikolopoulos[10], Alice Rigby[1,11], Graeme C Black[4,5], Christopher M. Watson[1,12], Sahar Mansour[13,14], Chris F Inglehearn[1], Alan J Mighell[11*] and Agnès Bloch-Zupan[2,15,16*]

[1]Leeds Institute of Medical Research, St James's University Hospital, University of Leeds, Beckett Street, Leeds. LS9 7TF. UK.

[2] Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258, CNRS-UMR7104, Université de Strasbourg, Illkirch, France.

[3]Yorkshire Regional Genetics Service, Leeds Teaching Hospitals NHS Trust, Leeds. LS7 4SA. UK.

[4]Division of Evolution and Genomic Sciences, Manchester Academic Health Science Centre, Faculty of Biology, Medicine and Health, School of Biological Sciences, University of Manchester, Manchester, M13 9NT. UK.

[5]Manchester Centre for Genomic Medicine, St. Mary's Hospital, Manchester University NHS Foundation Trust, Manchester. M13 9WL. UK.

[6]EMQN CIC, Pencroft Way, Manchester Science Park, Manchester. M15 6SE. UK.

[7]University Children's Hospital, Salzburger Landesklinken (SALK) and Paracelsus Medical University (PMU), Müllner Hauptstrasse 48, 5020 Salzburg, Austria.

[8]Amalia Children's Hospital, Radboud University Medical Center, Geert Grooteplein Zuid 10, 6525 GA Nijmegen, the Netherlands.

[9]Dental & O.M.F.S Clinic, Al Nahdha Hospital, Ministry of Health, Muscat. HGQJ+GV6. Oman.

[10]Institute for Fundamental Biomedical Research, BSRC Alexander Fleming, 16672 Vari, Attica, Greece.

[11]School of Dentistry, University of Leeds. LS2 9JT. UK.

[12]North East and Yorkshire Genomic Laboratory Hub, Central Lab, St. James's University Hospital, Leeds. LS9 7TF. UK.

[13]Lymphovascular Research Unit, Molecular and Clinical Sciences Research Institute, University of London St George's, London. SW17 0QT. UK.

[14]SW Thames Regional Centre for Genomics, St George's University Hospitals NHS Foundation Trust, London. SW17 0QT. UK.

[15]Faculté de Chirurgie Dentaire, Université de Strasbourg, 8 Rue Sainte Elisabeth, F-67000 Strasbourg, France.

[16]Centre de référence des maladies rares orales et dentaires O-Rares, Filière Santé Maladies rares TETE COU, European Reference Network CRANIO, Pôle de Médecine et Chirurgie Bucco-dentaires, Hôpital Civil, Hôpitaux Universitaires de Strasbourg (HUS), 1 place l'hôpital, 67000 Strasbourg, France.

* These authors contributed equally to the study.

Correspondence to:

Dr Claire E L Smith email C.E.L.Smith@leeds.ac.uk

Telephone (+44) 113 343 8445

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Background:

Plexins are large transmembrane receptors for the semaphorin family of signalling proteins. Semaphorin-plexin signalling controls cellular interactions that are critical during development as well as in adult life stages. Nine plexin genes have been identified in humans, but despite the apparent importance of plexins in development, only biallelic *PLXND1* and *PLXNA1* variants have so far been associated with Mendelian genetic disease.

Methods:

Eight individuals from six families presented with a recessively inherited variable clinical condition, with core features of amelogenesis imperfecta (AI) and sensorineural hearing loss (SNHL), with variable intellectual disability. Probands were investigated by exome or genome sequencing. Common variants and those unlikely to affect function were excluded. Variants consistent with autosomal recessive inheritance were prioritised. Variant segregation analysis was performed by Sanger sequencing. RNA expression analysis was conducted in C57Bl6 mice.

Results:

Rare biallelic pathogenic variants in plexin B2 (*PLXNB2*), a large transmembrane semaphorin receptor protein, were found to segregate with disease in all six families. The variants identified include missense, nonsense, splicing changes and a multi-exon deletion. *Plxnb2* expression was detected in differentiating ameloblasts.

Conclusion:

We identify rare biallelic pathogenic variants in *PLXNB2* as a cause of a new autosomal recessive, phenotypically diverse syndrome with AI and SNHL as core features. Intellectual disability, ocular disease, ear developmental abnormalities and lymphoedema were also present in multiple cases. The variable syndromic human phenotype overlaps with that seen in *Plxnb2* knockout mice, and, together with the rarity of human *PLXNB2* variants, may explain why pathogenic variants in *PLXNB2* have not been reported previously.

What is already known on this topic

Plexins are large transmembrane proteins that act as receptors for the semaphorin family of signalling proteins. Semaphorin-plexin signalling controls cellular interactions that are critical during development as well as in adult life stages. Nine plexin genes have been identified in humans, but despite the apparent importance of plexins in development, only biallelic *PLXND1* and *PLXNA1* variants have so far been associated with Mendelian genetic disease.

What this study adds

We identify rare biallelic pathogenic variants in *PLXNB2* as a cause of a new autosomal recessive, phenotypically diverse syndrome with amelogenesis imperfecta and sensorineural hearing loss as core features. Intellectual disability, ocular disease, ear developmental abnormalities and lymphoedema were also present in multiple cases. The variable syndromic human phenotype overlaps with that seen in *Plxnb2* knockout mice, and, together with the rarity of human *PLXNB2* variants, may explain why pathogenic variants in *PLXNB2* have not been reported previously.

How this study might affect research, practice or policy

Individuals presenting with amelogenesis imperfecta and sensorineural hearing loss should be screened for mutations in *PLXNB2* and tested for other features of the syndrome. *PLXNB2*

should be added to amelogenesis imperfecta, deafness and intellectual disability gene panels

to improve mutation detection rates. Affected families should receive appropriate genetic

counselling.

Introduction

Development is a cascade of highly dynamic, time-critical, complex processes involving many signalling molecules and receptors that act to regulate cell proliferation, migration, adhesion and differentiation. This results in the formation of complex tissues, which further organise to effect organogenesis. Plexins are large transmembrane proteins that act as receptors for the semaphorin family of signalling proteins. Semaphorin-plexin signalling controls cellular interactions that are critical during development as well as in adult life stages (reviewed by Perala et al. 2012).[1] Semaphorin signalling modulates changes to both actin and microtubule organisation and therefore to the overall cytoskeleton, cell morphology, cell adhesion and cell motility (reviewed in Alto & Terman[2]).

The plexin gene family was originally identified in humans, and members were grouped according to the domain structure of the encoded proteins.[3, 4] Nine genes have been identified in both humans and mice, with class A plexins consisting of 4 genes (A1-A4), class B of three (B1-B3) and classes C (C1) and D (D1) of one each.[1, 4] Plexin family members share a common structure, with extracellular and intracellular portions. The extracellular portions contain a sema domain that binds with semaphorin ligands to activate signalling, as well as two or three PSI (plexin, semaphorin and integrin) domains and two or three glycine-proline rich IPT (immunoglobulin, plexin and transcription factor) domains. The intracellular portions are highly conserved[5] and contain two R-Ras GAP motifs and one set of Plexin Rho-GTPase Association Motifs (PRAMs).

Class B plexins have an additional intracellular C-terminal PDZ interaction domain[6] and an extracellular cleavage site for proprotein convertases.[1] Plexin B2 (PLXNB2) participates in axonal guidance and cell migration.[7] It is expressed widely but it also demonstrates a specific temporospatial pattern of expression throughout development, and its expression is

distinct from that of other plexins, suggesting non-redundancy.[8] RNA transcripts are detectable in mice from early fetal stages to adulthood within the brain.[9] In situ hybridisation in E14 mouse embryos revealed high expression within several regions of the central nervous system, including many regions of the brain and retina.[8] Expression was also high in the developing tooth bud, oral epithelium and in cartilage, with lower expression also detected in the cochlea, lung, kidney, epidermis and intestine.[8] The exact role of PLXNB2 in tooth development is currently unknown, but semaphorins and plexin B1 have been found to be important for innervation of tooth buds[10] and for dental stem cell migration ex vivo.[11]

Plxnb2 knockout mice (*Plxnb2*[-/-]) vary in phenotype, depending upon their genetic background. *Plxnb2*[-/-] mice produced on inbred backgrounds did not survive gestation.[12, 13] The majority developed exencephaly, reflecting the importance of plexin B2 activity for neural tube closure and potentially also its influence on the actin cytoskeleton. Other defects noted included abnormal development of the dentate gyrus, defects in cerebellar foliation and lamination, retarded development of the olfactory bulb and impaired neuronal proliferation. In contrast, when the same pathogenic variant was introduced into outbred CD1 mice, neural tube closure defects were less common, and after 4 generations, around 30% of *Plxnb2*[-/-] mice were viable and fertile. Despite the knockout mice having no obvious behavioural or motor defects, their cerebella were smaller and major brain foliation defects were still present.[12] Heterozygous *Plxnb2*[+/-] mice had no apparent abnormalities.

Human *PLXNB2* variants (MIM*604293) and aberrant PLXNB2 expression have been associated with lung cancer,[14, 15] acute myeloid leukaemia,[16] amyotrophic lateral sclerosis,[17] glioblastoma,[18] autism spectrum disorders with regression,[19] psoriasis,[20] and first trimester euploid miscarriage.[21] In contrast, despite the apparent importance of plexins in development, only biallelic *PLXND1* (MIM*620282) and *PLXNA1* (MIM*601055)

variants have so far been associated with Mendelian genetic disease in humans. *PLXND1* variants cause multiple types of congenital heart defects (MIM#620294).[22] *PLXNA1* variants cause Dworschak-Punetha neurodevelopmental syndrome which includes speech regression, autistic features and hyperactivity, variable sensorineural hearing loss (SNHL) and ocular, brain, facial and skin abnormalities (MIM#619955).[23] The same authors also suggested that pathogenic variants in other plexins may be embryonic lethal or may cause a range of phenotypes that have not yet been recognised as part of one syndrome.[23]

Here we describe six families with probands carrying rare biallelic *PLXNB2* variants. Affected individuals manifest a complex, variable syndromic phenotype, the core features of which appear to be SNHL and amelogenesis imperfecta (AI), with intellectual disability also present in most cases.

Materials and Methods

Patients

Affected individuals and family members were recruited in accordance with the principles outlined by the declaration of Helsinki, with local ethical approval (REC 13/YH/0028). Clinical evaluation captured disease features as part of routine patient care. Genomic DNA was obtained from venous blood samples using a salt-based extraction protocol, or from saliva using Oragene DNA Sample Collection kits (DNA Genotek, Ottawa, ON, Canada) as detailed in the manufacturer's instructions.

Sequencing and analysis

Individuals were recruited and genomic DNA was subjected to SNP genotyping, exome or genome sequencing at different institutions. Sequencing and analysis methods for each family can be found in the Supplementary materials and methods. In summary, variants identified in next generation short-read sequencing data were filtered to exclude all changes other than missense, frameshift or stop variants, exonic insertion/deletions or variants located at splice consensus sites (up to 8 bp within introns or 3 bp within exons away from splice junctions). Synonymous variants outside of the splice region were discarded. Variants in the Genome Aggregation Database (gnomAD) (v2.2.1)[24] were excluded if present at a global minor allele frequency of 1% or higher. Variants were also filtered based on the mode of inheritance. In families known to be consanguineous, homozygous variants were prioritised. Population specific high frequency variants and platform artefacts were excluded by removing variants also present in exomes of individuals of the same ethnicity without dental disease that had been sequenced using the same reagents and platform. Splicing prediction analysis was carried out using NetGene2 (v2.4.2)[25] and Splice AI.[26] CADD v1.6,[27]

REVEL,[28] Polyphen-2 (HumVar model)[29] and SIFT[30] were used to assess each variant's potential to be disease causing. Copy number variants were identified using ExomeDepth (v1.0.0).[31] Variants were confirmed and segregation analysis was performed for all available family members by Sanger sequencing. Primer sequences used are shown in Supplementary Table 1. Genomic coordinates are based on the GRCh37 human reference genome, the reference gene sequence used for *PLXNB2* is MANE Select transcript NM_012401.4 (ENST00000359337.9) and protein variant nomenclature for PLXNB2 is based on RefSeq protein NP_036533.2 (ENSP00000352288.4). The corresponding references used for *CRYBB3* are MANE Select transcript NM_004076.5 (ENST00000215855.7) and for CRYBB3 RefSeq protein NP_004067.1 (ENSP00000215855.2). All variants identified as part of this study were uploaded to ClinVar: SCV002822954 - SCV002822961. *In silico* modelling of the effect of the variants on the PLXNB2 protein tertiary structure was completed using iTASSER MTD[32] using the default parameters. The protein structures were visualised with UCSF Chimera.[33]

Mouse tissue preparation

All animals were maintained in accordance with the French Ministry of Agriculture guidelines for the use of laboratory animals under study SC67-218-37- IGBMC and APAFIS#3957-2016020516359388v1 and in accordance with the NIH guidelines provided in the Guide for the Care and Use of Laboratory Animals. All methods and experimental procedures were reviewed and approved by an institutional safety committee.

Mouse embryos/fetuses were collected at E14.5, E16.5, E19.5 or on the day of birth and analyzed as detailed in supplementary materials and methods.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Results

Initially we recruited two consanguineous families, one Turkish (Family 1) and one Omani
(Family 2) (Figure 1). In Family 1, two male double first cousins have bilateral SNHL,
intellectual disability, AI and severe myopia (Supplementary Figure 1). In addition to the
shared phenotype, one (II:3) also has bilateral cataracts and the other (II:6) has unilateral
renal agenesis and pyloric stenosis. Analysis of GeneChip® Human Mapping 250K *Nsp* SNP
data from affected individuals II:3 and II:6 showed two homozygous regions encompassing
chr7:34,080,354-41,760,000 and chr22:49,714,781-51,1756,26. Exome sequencing of II:3
and II:6 revealed a shared homozygous variant in *PLXNB2* within the region on chromosome
22, c.2413A>T, p.(Ile805Phe) (Table 1) which segregates with disease in the family
(Supplementary Figure 2). It affects the extracellular portion of the protein, specifically the
first cell surface receptor IPT domain and is predicted to be damaging by all pathogenicity
prediction software tested (Supplementary Table 2). The residue affected is highly conserved
(Supplementary Figure 3) and the variant is absent from gnomAD. Individual II:3 was also
found to carry variant c.388G>A, p.(Glu130Lys) in crystallin beta 3 (*CRYBB3*;
MIM*123630) (Supplementary Figure 4), which is likely to explain the bilateral cataracts
observed in him (MIM#609741).[34] Variants that passed population and pathogenicity
prediction filters but were not investigated further are detailed in Supplementary Table 3 for
each family.

Patient (sex/age range)	Zygosity and variants (NM_012401.4, NP_036533.2)	Phenotype				
		Auditory	Dental	Developmental /neurological	Vision	Other
Family 1 II:3 (M/12-18)	Homozygous c.2413A>T: p.(Ile805Phe); c.2413A>T: p.(Ile805Phe)	SNHL	AI, conical permanent incisors	Global developmental delay, moderate intellectual disability	Myopia, horizontal nystagmus, (congenital cataract due to *CRYBB3* variant)	Ear lobe skin blind-ended tracts
Family 1 II:6 (M/19-21)	Homozygous c.2413A>T: p.(Ile805Phe); c.2413A>T: p.(Ile805Phe)	SNHL; labyrinthine malformation	AI, conical permanent incisors	Global developmental delay, epilepsy, moderate intellectual disability	Severe myopia with scattered papillae, horizontal nystagmus, macular atrophy	Ear lobe skin blind-ended tracts, unilateral renal agenesis, pyloric stenosis, asthma, recurrent bronchitis, intrauterine growth retardation, finger pads, watch glass toe nails, overweight
Family 2 IV:2 (M/12-18)	Homozygous c.2248G>A: p.(Asp750Asn); c.2248G>A: p.(Asp750Asn)	SNHL	AI	Intellectual disability	No obvious abnormality, not examined	
Family 3 II:1 (M/6-11)	Compound heterozygous c.750C>A: p.(Cys250*); c.3117G>A: p.(Thr1039=)	SNHL (mild) 500-4000 Hz,	AI with hypoplasia	Normal	Developmental macular abnormality with pale fundus, attenuated blood vessels, high myopia, nystagmus, microcornea	Ear lobe skin blind-ended tracts, cleft palate, hypertelorism, keratopathy
Family 4 II:1 (F/19-21)	Compound heterozygous c.2606del: p.(Phe869Serfs*45); c.3982_3986del: p.(Phe1328His*65)	SNHL	AI; missing upper permanent lateral incisors	Normal	No obvious abnormality, not examined	Ear lobe skin blind-ended tracts, Bilateral primary lower limb lymphoedema (onset aged 3), nevus, cellulitis
Family 5 II:1 (F/50-59)	Homozygous c.5197-337_5310del: p.(Asp1733_Arg1779del); c.5197-337_5310del: p.(Asp1733_Arg1779del)	SNHL	AI	Mild/moderate intellectual disability	No obvious abnormality, not examined	Bilateral primary lower limb lymphoedema

Patient (sex/age range)	Zygosity and variants (NM_012401.4, NP_036533.2)	Phenotype				
		Auditory	Dental	Developmental /neurological	Vision	Other
Family 5 II:2 (M/50-59)	Homozygous c.5197-337_5310del: p.(Asp1733_Met1770del) ; c.5197-337_5310del: p.(Asp1733_Met1770del)	SNHL	AI	Intellectual disability	No obvious abnormality, not examined	Unilateral lymphoedema of one foot
Family 6 III:1 (M/2-5)	Homozygous c.4609G>A: p.(Gly1537Ser); c.4609G>A: p.(Gly1537Ser)	Could not be assessed; no current indication of hearing loss	Clinical tooth failure (cause unclear); Could not be assessed further	Profound intellectual disability, non-verbal, autistic features, hyperactive behaviour	Strabismus, no other obvious abnormality, could not be assessed.	Mild generalised muscular hypotonia

Table 1. The variants in *PLXNB2* identified in six families and the disease features observed. Splicing prediction tools suggest that exon 35 is

entirely skipped leading to p.(Asp1733_Arg1779del) instead of p.(Asp1733_Met1770del) as would be predicted from the proportion of the gene

deleted. Age ranges are shown for individuals to maintain anonymity. Age of onset was from birth unless otherwise stated. Abbreviations used:

AI amelogenesis imperfecta; SNHL sensorineural hearing loss. Variants are based on genome build GRCh37 and *PLXNB2* transcript

ENST00000359337.9, NM_012401.4 and PLXNB2 protein ENSP00000352288.4, NP_036533.2.

In Family 2, two Omani brothers born of a first cousin union, were found to have bilateral SNHL, intellectual disability and AI (Table 1, Figure 1 and Supplementary Figure 5). Exome sequencing of one affected brother revealed a homozygous missense variant in *PLXNB2*, c.2248G>A, p.(Asp750Asn). This replaces a charged residue with an uncharged one, affecting a highly conserved residue in the extracellular portion of the protein (Supplementary Figure 3). SIFT predicts the variant to be deleterious and it was not identified in gnomAD (Supplementary Table 2).

Family 3 was identified independently via the UK Inherited Retinal Dystrophy Consortium. They are a white British family with an affected male child born with facial clefting, who also presented with nystagmus shortly after birth. In middle childhood, he was diagnosed with retinal dystrophy, high myopia, microcorneas and mild keratopathy. He also has mild bilateral SNHL for sounds ranging from 500-4000 Hz, and AI (Table 1, Figures 1, 2 and Supplementary Figure 6). Exome sequencing revealed biallelic compound heterozygous variants in *PLXNB2*, one a nonsense variant, c.750C>A, p.(Cys250*), and one synonymous variant altering the final base of the splice donor site of exon 19, predicted to affect splicing, c.3117G>A, p.(Thr1039=). Analysis with splice prediction tools Splice AI[26] and NetGene2[25] predicted loss of the donor site, suggesting that some of intron 19 may be retained in the mature transcript (Supplementary Table 4, Supplementary Figure 7). Neither variant was identified in gnomAD (Supplementary Table 2). A second fetus, who presented with clefting, was electively aborted. Genotyping showed that the fetus was heterozygous for the nonsense *PLXNB2* variant only.

Next, we searched the UK 100,000 Genomes dataset[35] for patients with biallelic variants in *PLXNB2* and a similar phenotype. We identified one affected white British female (Family 4) with SNHL, AI, lower limb lymphoedema and cellulitis (Table 1, Figures 1, 2, Supplementary Figure 8). She carries biallelic compound heterozygous frameshift variants

c.2606delT, p.(Phe869Serfs*45) and c.3982_3986delCTTT, p.(Phe1328Hisfs*65) in

PLXNB2 (Supplementary Figure 9), both predicted to produce transcripts that are subject to

nonsense mediated decay.[36] Variant c.2606delT, p.(Phe869Serfs*45) has previously been

identified in gnomAD as a heterozygous variant in one individual, suggesting an allele

frequency of 4.024 x10^{-6}. Variant c.3982_3986delCTTT, p.(Phe1328Hisfs*65) was not

present in gnomAD.

With increased understanding of the clinical presentation associated with biallelic *PLXNB2*

variants, we identified Family 5 through further collaboration. Two affected siblings of

Pakistani origin, born of a consanguineous union (Figure 1) presented with deafness, AI,

intellectual disability and lower limb lymphoedema (Figure 2, Table 1, Supplementary

Figures 10, 11). Exome sequencing of II:1 and subsequent Exome Depth analysis revealed a

homozygous deletion spanning exons 34 and 35 (Reads ratio 0.0291, Bayes factor 18.3;

Supplementary Table 2). The breakpoints predicted in the exome sequence (Supplementary

Figure 12), chr22:50,715,085 and chr22:50,715,672, were confirmed by PCR, which also

confirmed the deletion was present in II:2. This deletion is in-frame and is predicted to delete

at least 38 amino acids (aa) (p.(Asp1733_Met1770del)) from the 1838aa protein, including all

of exon 34 and part of exon 35. Splice prediction tool NetGene2 (v2.4.2) predicts that exon

35 will be skipped entirely (Supplementary Table 5), resulting in an in-frame deletion of

47aa, p.(Asp1733_Arg1779del), suggesting that an abnormal PLXNB2 protein may be

produced. The deleted region is part of the Rho-GAP catalytic domain critical to the protein's

function, but the deletion would leave the catalytic arginine residues at 1395, 1396 and 1691

intact.

Using GeneMatcher,[37] we identified one further patient with biallelic *PLXNB2* variants and

an overlapping disease phenotype. Family 6 are of Iraqi origin. One affected male was born

of first cousin consanguineous parents (Figure 1). In early childhood, he has severe

developmental delay and autistic features, and his tooth enamel shows evidence of severe damage from a limited visual inspection (Table 1, Supplementary Figure 13). However, it was not possible to carry out a detailed dental examination or to obtain dental radiographs. The presence of AI, as opposed to severe caries could therefore not be confirmed, and testing for SNHL and eye disease was not possible. Upon exome sequencing, the affected individual was found to carry a homozygous *PLXNB2* variant c.4609G>A, p.(Gly1537Ser), which affects the highly conserved Rho-GAP domain that lies within the intracellular portion of PLXNB2 and changes the residue from a non-polar to a polar residue (Table 1, Supplementary Table 2, Supplementary Figure 3). The residue is conserved in all species examined and this variant was not present in gnomAD.

We next used iTASSER MTD to try to assess the effect of each of the variants on the overall predicted structure of PLXNB2 (Supplementary Figure 14). *In silico* predictions of WT PLXNB2 structure (panel A) and of these variants are likely to be of limited use due the small percentage of PLXNB2 covered by known crystal structures and the lack of appropriate homologous protein structures upon which to base the WT structure. The local structural changes for the missense variants (Asp750Asn, Ile805Phe and Gly1537Ser) are shown in panels B, C and D. There were minor changes in the solvent accessibility for Asp750Asn and Ile805Phe. For Ile805Phe, the mutant structure is characterized as undefined, in comparison to the WT strand structure. The model for the deletion Asp1733_Arg1779del shows extensive structural differences to the WT protein.

To gain insight into *Plxnb2* expression in mice, we performed *in situ* hybridization using a digoxigenin-labeled antisense riboprobe generated from the same DNA template as previously used by the EURExpress consortium (www.eurexpress.org) (Supplementary Figure 15). Mouse embryos were analyzed at E14.5, E16.5 and E19.5. *Plxnb2* expression was detected in the kidney and lung (Fig. 3A and B). Discrete expression was observed in the

developing inner ear (cochlea: Fig. 3C). Expression was also detected in olfactory epithelium

and retina (Fig. 3D and F), and in the small and large intestine (Fig. 3E).

We also interrogated the GTEx portal[38] (https://gtexportal.org/) to determine the

expression of *PLXNB2* in human tissues. This showed that there was relatively high

expression of *PLXNB2* in the cerebellum and cerebellar hemisphere compared to other

regions of the brain. There was also high expression in the kidney, salivary gland, ovary,

prostate, spleen and thyroid, amongst other tissues (Supplementary Figure 16).

Since all individuals studied had AI or were suspected to have AI, we investigated the

expression of *Plxnb2* in the mouse during various stages of tooth development (Fig. 3G, I and

K lower molar and H, J and L lower incisor). *Plxnb2* transcripts were observed at E14.5 in

the epithelial tissues of the developing teeth, at E16.5 in the epithelial and mesenchymal

compartments of the incisor and in the epithelial part of the molar. At E19.5 labeling was

observed in differentiating ameloblasts.

Discussion

Here we describe biallelic pathogenic variants in *PLXNB2* in six families of diverse ethnicity with individuals affected by a variable syndromic phenotype. Four of these are consanguineous families with affected individuals homozygous for an extremely rare variant that has almost certainly been passed down both branches of the family, and one includes affected cousins, demonstrating significant co-segregation of the disease with biallelic *PLXNB2* variants. This syndrome has AI and SNHL as core symptoms, and intellectual disability, lower limb lymphoedema, ocular abnormalities and a variety of other conditions are also seen in some, but not all, cases. Given the varied and complex roles of PLXNB2 in development,[1] it seems unsurprising that biallelic variants cause a syndromic disease phenotype. The variants all have pathogenicity scores indicating that they are predicted to be deleterious (Supplementary Table 2). The case series we have accumulated, together with published evidence of an overlapping condition in *Plxnb2* knockout mice, provides compelling evidence for biallelic variants in *PLXNB2* as the cause of this recessively inherited condition.

All variants implicated are extremely rare, with all but one absent in gnomAD (v2.1.1), which contains approximately 124,000 individuals with good quality sequence across *PLXNB2*. The rarity of these variants, together with the constraint metrics available in gnomAD (o/e = 0.19, pLI = 0.99),[39] suggest that *PLXNB2* loss-of-function variants are not well tolerated and affect viability. However, the identification of an individual who may entirely lack PLXNB2 (Family 4 II:1) seems to contradict this. This is consistent with observations in *Plxnb2* knockout mice, where homozygosity for the knockout allele was lethal on one genetic background but viable on another.[12, 13] The *PLXNB2* variants identified herein include missense, frameshift and splice variants, a premature termination

codon and a deletion spanning two exons, which are all observed to be protein damaging variants.

One possible interpretation of these findings is that all the observed human variants have the effect of being functional knockouts, and that, as observed in mouse models, the viability of such embryos is determined by the genetic background. Genetic background might also explain the highly variable phenotypes observed in different cases, with only AI and SNHL as consistent features. Pathogenic variants in *PLXNA1* cause an overlapping and similarly varying range of phenotypes to pathogenic variants in *PLXNB2*.[23]. The impact of genetic background on variation in disease phenotype, severity and survival, has been noted for one *Plxnb2*^{-/-} mouse model,[12] which suggests that other cosegregating variants may affect disease range and severity. This may suggest that, in spite of the distinct patterns of plexin expression,[8] other plexins can sometimes partially compensate for loss of PLXNB2 or PLXNA1 to allow developmental processes vital to life to proceed. However, to prove such an effect would require the study of a large cohort of cases, and as yet no specific variants have been implicated in phenotype variation in either of these syndromes. Given that this cohort consists of only eight individuals with six different biallelic genotypes, the power to detect any genotype-phenotype correlation in this study is very limited.

Alternatively, the *PLXNB2* variants and/or genotypes in the families described herein may each have unique effects on PLXNB2 function. Variants could cause partial loss-of-function through hypomorphic alleles, with expression of the normal transcript reduced but not abolished,[40], or may act as "gain-of-function" alleles, creating a protein with altered or enhanced function or inappropriate persistence within the cell. Consistent with this

hypothesis, the affected cousins in family 1 and affected siblings in family 5, each having the

same genotype, have remarkably similar phenotypes (Table 1).

The missense variants identified in this study affect both the extracellular (p.(Ile805Phe) and

p.(Asp750Asn)) and intracellular portions of the protein (p.(Gly1537Ser)), with no particular

region or domain specifically affected by the variants identified in this study. In order to

assess the effects of the variants on protein structure, we attempted to model the changes

using I-TASSER-MTD.[32] However, due to the lack of homologous structures available for

the majority of the PLXNB2 protein, including the regions affected by the variants, the

accuracy of the modelling is likely to be low. The variants identified herein could be altering

the binding of PLXNB2 to semaphorins via the sema domain, the subsequent

homodimerization of PLXNB2 upon binding semaphorin, catalytic activity via the GAP

domain and interaction with GTPases, or the ability to interact with other proteins and to

effect other types of downstream signalling.[41] Further investigation will be required to

better understand the effects of specific variants and the basis of variation on phenotype in

the condition caused by *PLXNB2* variants.

Heterozygous carriers of the variants identified in these families appear in general to be

unaffected by disease, although one fetus (Family 3 II:2) did have facial clefting and carried a

heterozygous nonsense *PLXNB2* variant (c.750C>A, p.(Cys250*)). It is unknown whether

this fetus would have developed other clinical features similar to their sibling. It is possible

that another variant could be partially or entirely responsible for this particular phenotype

rather than the *PLXNB2* variant. In individuals with *PLXNA1* pathogenic variants, disease has

been observed with both biallelic and particular de novo heterozygous variants.[23]

The core phenotypes observed for the individuals carrying pathogenic *PLXNB2* variants reflect negative impacts on the development or function of the cochlea and ameloblasts. This led us to consider whether other plexins are expressed in the cochlea and inner enamel epithelium at a similar time to PLXNB2, or whether PLXNB2 is expressed alone in these tissues, excluding the possibility of partial compensation by a closely related protein. Our investigation of the expression of *Plxnb2* transcripts within the dental tissues of murine embryos gave similar results to the expression patterns previously detailed by Perala et al.[8] Their analysis of the expression of plexins in murine embryos using in situ hybridization revealed that *Plxnb2* transcripts are present in the brain, retina, cochlea and tooth bud at E14.[8] *Plxnb2* is the only plexin to be expressed at this time-point in all four tissues, although the expression of *Plxnd1* was not examined in this study and the expression of other plexins, most notably *Plxnb1*, do overlap that of *Plxnb2* in many tissues. *Plxnb2*, *Plxna2* and *Plxna3* transcripts are all expressed in the cochlea, although the relative levels of expression of each were not determined. Similarly, *Plxnb1* and *Plxnb2* transcripts were detected at relatively high levels at E14 in the oral epithelium and tooth bud, but other plexins were also detected at lower levels of expression. *Plxnb2* was shown to be expressed at relatively high levels in the inner enamel epithelium at E15, with expression sustained until at least E16. These findings suggest that PLXNA2, PLXNA3 and PLXNB1 are co-expressed in affected tissues at relevant timepoints, but are unlikely to be able to fully compensate for the loss of PLXNB2.

The variable phenotype and extremely low population frequency of the variants reported in this study may be the reasons why pathogenic variants in *PLXNB2* have not previously been reported as causing syndromic disease in humans. We suggest that the tooth enamel phenotype, AI, is a consistent, but potentially easily missed feature that flags this genetic diagnosis. AI is a heterogeneous group of genetic conditions characterized by a deficit in

enamel quantity and/or quality affecting all teeth of both dentitions.[42-44] It can present as

an isolated disease or can be part of more complex and diverse syndromes affecting other

tissues and organs. Once formed and following tooth eruption, enamel has no capacity for

cellular repair. Accordingly, AI provides a clear and persistent marker of abnormal

development that is recognisable at an early age. However, due to the presence of neuronal

deficits such as hearing loss and intellectual disability, AI might easily be overlooked or

dismissed as dental caries due to suboptimal diet and/or poor dental hygiene. A differential

diagnosis of AI as opposed to fluorosis or molar incisor hypomineralisation is also a possible

confounding issue. Diagnosis of AI may therefore require a specialist paediatric dental

professional. The disconnection between dental and general healthcare also presents barriers

to diagnosis and has been flagged as problematic previously in the differential diagnosis of

Usher and Heimler syndromes.[45]

The families presented have AI characterised by variable abnormalities of enamel volume

(hypoplasia) and mineralisation (hypomineralised). This is consistent with the understanding

of PLXNB2 function. Future laboratory investigation of enamel from affected individuals

will give insight into the characteristics of the disruption to enamel rod morphology and

mineralisation. It is unclear if the other dental morphological changes reported, including

conical lateral incisors, missing lateral incisors, flattened occlusal surfaces and mild

taurodontism are consequent to PLXNB2 functional disruption or are coincidence in these

families. As further families are described, the core features of the dental phenotype will

become clearer and will be an important clinical indicator to consider *PLXNB2* further.

In conclusion, we identify biallelic pathogenic variants in *PLXNB2* as a cause of a new

autosomal recessively inherited, phenotypically diverse syndrome including AI and SNHL as

core symptoms, with intellectual disability, ocular disease, ear developmental abnormalities and lymphoedema also present in multiple cases.

Data Availability

Raw sequencing data supporting the findings of this study are available from the corresponding author upon request.

Acknowledgements

Competing Interests

The authors declare no competing interests.

Funding

Collaborator Statement

The UK Inherited Retinal Dystrophy Consortium (UKIRDC) includes, in addition to listed authors CELS, RLT, GCB, and CFI, Martin McKibbin, Manir Ali, Carmel Toomes, Stuart Ingram, Forbes Manson, Panagiotis Sergouniotis, Gavin Arno, Alison J Hardcastle, Andrew Webster, Nikolas Pontikos, Michael Cheetham, Alessia Fiorentino, Susan Downes, Jing Yu, Stephanie Halford, Suzanne Broadgate and Veronica van Heyningen.

The Genomes England Research Consortium (GERC) includes Ambrose JC[1], Arumugam P[1], Bevers R[1], Bleda M[1], Boardman-Pretty F[1,2], Boustred CR[1], Brittain H[1], Brown MA[1], Caulfield MJ[1,2], Chan GC[1], Giess A[1], Griffin JN[1], Hamblin A[1], Henderson S[1,2], Hubbard

TJP[1], Jackson R[1], Jones LJ[1,2], Kasperaviciute D[1,2], Kayikci M[1], Kousathanas A[1], Lahnstein L[1], Lakey A[1], Leigh SEA[1], Leong IUS[1], Lopez FJ[1], Maleady-Crowe F[1], McEntagart M[1], Minneci F[1], Mitchell J[1], Moutsianas L[1,2], Mueller M[1,2], Murugaesu N[1], Need AC[1,2], O'Donovan P[1], Odhams CA[1], Patch C[1,2], Perez-Gil D[1], Pereira MB[1], Pullinger J[1], Rahim T[1], Rendon A[1], Rogers T[1], Savage K[1], Sawant K[1], Scott RH[1], Siddiq A[1], Sieghart A[1], Smith SC[1], Sosinsky A[1,2], Stuckey A[1], Tanguy M[1], Taylor Tavares AL[1], Thomas ERA[1,2], Thompson SR[1], Tucci A[1,2], Welland MJ[1], Williams E[1], Witkowska K[1,2], Wood SM[1,2], Zarowiecki M[1]

1. Genomics England, London, UK

2. William Harvey Research Institute, Queen Mary University of London, London. EC1M 6BQ, UK.

Contributorship Statement

Conceptualization: CFI, AJM, ABZ; Data Curation: CELS, VL, UKIRDC, GERC, RGF, GN, AR; Formal Analysis: CELS, VL, UH, SB, RLT, RGF, JAM, CMW; Funding Acquisition: CFI, AJM, ABZ; Investigation: CELS, VL, UH, SB, RLT, JCS, JP, UKIRDC, GERC, SBW, RGF, JAM, SA-B, GCB, SM, AJM, ABZ; Methodology: CELS, VL, UH, SB, RLT, JAP, UKIRDC, GERC, SBW, RGF, JAM, CFI, AJM, ABZ; Project administration: CELS, CFI, AJM, ABZ; Resources: GERC; Supervision: CELS, JAP, CMW, CFI, AJM; Validation: CELS, VL, UH, RLT, RGF; Visualization: CELS, VL, UH, SB, GN, AR, SBW, RGF, SM, AJM, ABZ; Writing-original draft: CELS, VL; Writing-review & editing: CELS, CMW, CFI, AJM, ABZ. All authors reviewed and commented critically on the final draft.

Ethics Declaration

This study underwent ethical review by the National Research Ethics Service Committee for Yorkshire & the Humber – South Yorkshire REC ref 13/YH/0028 IRAS project ID 82448

and underwent local ethical approval at other centers involved in this study. The 100,000

Genomes Project is covered by REC reference approval 14/EE/1112.

References

1 Perälä N, Sariola H and Immonen T. More than nervous: the emerging roles of plexins. *Differentiation* 2012;83:77-91. doi:10.1016/j.diff.2011.08.001

2 Alto LT and Terman JR. Semaphorins and their Signaling Mechanisms. *Methods Mol Biol* 2017;1493:1-25. doi:10.1007/978-1-4939-6448-2_1

3 Maestrini E, Tamagnone L, Longati P*, et al.* A family of transmembrane proteins with homology to the MET-hepatocyte growth factor receptor. *Proc Natl Acad Sci U S A* 1996;93:674-8. doi:10.1073/pnas.93.2.674

4 Tamagnone L, Artigiani S, Chen H*, et al.* Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. *Cell* 1999;99:71-80. doi:10.1016/s0092-8674(00)80063-x

5 Alves CH, Sanz AS, Park B*, et al.* Loss of CRB2 in the mouse retina mimics human retinitis pigmentosa due to mutations in the CRB1 gene. *Hum Mol Genet* 2013;22:35-50. doi:10.1093/hmg/dds398

6 Swiercz JM, Kuner R, Behrens J*, et al.* Plexin-B1 directly interacts with PDZ-RhoGEF/LARG to regulate RhoA and growth cone morphology. *Neuron* 2002;35:51-63. doi:10.1016/s0896-6273(02)00750-x

7 Perrot V, Vazquez-Prado J and Gutkind JS. Plexin B regulates Rho through the guanine nucleotide exchange factors leukemia-associated Rho GEF (LARG) and PDZ-RhoGEF. *J Biol Chem* 2002;277:43115-20. doi:10.1074/jbc.M206005200

8 Perälä NM, Immonen T and Sariola H. The expression of plexins during mouse embryogenesis. *Gene Expr Patterns* 2005;5:355-62. doi:10.1016/j.modgep.2004.10.001

9 Worzfeld T, Püschel AW, Offermanns S, *et al.* Plexin-B family members demonstrate non-redundant expression patterns in the developing mouse nervous system: an anatomical basis for morphogenetic effects of Sema4D during development. *Eur J Neurosci* 2004;19:2622-32. doi:10.1111/j.0953-816X.2004.03401.x

10 Luukko K and Kettunen P. Integration of tooth morphogenesis and innervation by local tissue interactions, signaling networks, and semaphorin 3A. *Cell Adh Migr* 2016;10:618-26. doi:10.1080/19336918.2016.1216746

11 Zhang L, Han Y, Chen Q, *et al.* Sema4D-plexin-B1 signaling in recruiting dental stem cells for vascular stabilization on a microfluidic platform. *Lab Chip* 2022;22:4632-44. doi:10.1039/d2lc00632d

12 Friedel RH, Kerjan G, Rayburn H, *et al.* Plexin-B2 controls the development of cerebellar granule cells. *J Neurosci* 2007;27:3921-32. doi:10.1523/jneurosci.4710-06.2007

13 Deng S, Hirschberg A, Worzfeld T, *et al.* Plexin-B2, but not Plexin-B1, critically modulates neuronal migration and patterning of the developing nervous system in vivo. *J Neurosci* 2007;27:6333-47. doi:10.1523/jneurosci.5381-06.2007

14 Zhao N, Ruan M, Koestler DC, *et al.* Epigenome-wide scan identifies differentially methylated regions for lung cancer using pre-diagnostic peripheral blood. *Epigenetics* 2022;17:460-72. doi:10.1080/15592294.2021.1923615

15 Liu H and Zhao H. Prognosis related miRNAs, DNA methylation, and epigenetic interactions in lung adenocarcinoma. *Neoplasma* 2019;66:487-93. doi:10.4149/neo_2018_181029N805

16 Lin L, Wang Y, Bian S, *et al.* A circular RNA derived from PLXNB2 as a valuable predictor of the prognosis of patients with acute myeloid leukaemia. *J Transl Med* 2021;19:123. doi:10.1186/s12967-021-02793-7

17 Saez-Atienzar S, Bandres-Ciga S, Langston RG, *et al.* Genetic analysis of amyotrophic

lateral sclerosis identifies contributing pathways and cell types. *Sci Adv* 2021;7.

doi:10.1126/sciadv.abd9036

18 Huang Y, Tejero R, Lee VK, *et al.* Plexin-B2 facilitates glioblastoma infiltration by

modulating cell biomechanics. *Commun Biol* 2021;4:145. doi:10.1038/s42003-021-01667-4

19 Yin J, Chun CA, Zavadenko NN, *et al.* Next Generation Sequencing of 134 Children with

Autism Spectrum Disorder and Regression. *Genes* 2020;11:853. 10.3390/genes11080853.

20 Hemida AS, Mareae AH, Elbasiony ASA, *et al.* Plexin-B2 in psoriasis; a clinical and

immunohistochemical study. *J Immunoassay Immunochem* 2020;41:718-28.

doi:10.1080/15321819.2020.1741385

21 Wang X, Shi W, Zhao S, *et al.* Whole exome sequencing in unexplained recurrent

miscarriage families identified novel pathogenic genetic causes of euploid miscarriage. *Hum

Reprod* 2023;38:1003-18. doi:10.1093/humrep/dead039

22 Guimier A, de Pontual L, Braddock SR, *et al.* Biallelic alterations in PLXND1 cause

common arterial trunk and other cardiac malformations in humans. *Hum Mol Genet*

2023;32:353-6. doi:10.1093/hmg/ddac084

23 Dworschak GC, Punetha J, Kalanithy JC, *et al.* Biallelic and monoallelic variants in

PLXNA1 are implicated in a novel neurodevelopmental disorder with variable cerebral and

eye anomalies. *Genet Med* 2021;23:1715-25. doi:10.1038/s41436-021-01196-9

24 Karczewski KJ, Francioli LC, Tiao G, *et al.* The mutational constraint spectrum quantified

from variation in 141,456 humans. *Nature* 2020;581:434-43. doi:10.1038/s41586-020-2308-7

25 Hebsgaard SM, Korning PG, Tolstrup N, *et al.* Splice site prediction in Arabidopsis

thaliana pre-mRNA by combining local and global sequence information. *Nucleic Acids Res*

1996;24:3439-52. doi:10.1093/nar/24.17.3439

26 Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF*, et al.* Predicting Splicing

from Primary Sequence with Deep Learning. *Cell* 2019;176:535-48.e24.

doi:10.1016/j.cell.2018.12.015

27 Rentzsch P, Schubach M, Shendure J*, et al.* CADD-Splice-improving genome-wide

variant effect prediction using deep learning-derived splice scores. *Genome Med* 2021;13:31.

doi:10.1186/s13073-021-00835-9

28 Ioannidis NM, Rothstein JH, Pejaver V*, et al.* REVEL: An Ensemble Method for

Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 2016;99:877-85.

doi:10.1016/j.ajhg.2016.08.016

29 Adzhubei IA, Schmidt S, Peshkin L*, et al.* A method and server for predicting damaging

missense mutations. *Nat Methods* 2010;7:248-9. doi:10.1038/nmeth0410-248

30 Kumar P, Henikoff S and Ng PC. Predicting the effects of coding non-synonymous

variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073-81.

doi:10.1038/nprot.2009.86

31 Plagnol V, Curtis J, Epstein M*, et al.* A robust model for read count data in exome

sequencing experiments and implications for copy number variant calling. *Bioinformatics*

2012;28:2747-54. doi:10.1093/bioinformatics/bts526

32 Zhou X, Zheng W, Li Y*, et al.* I-TASSER-MTD: a deep-learning-based platform for

multi-domain protein structure and function prediction. *Nat Protoc* 2022;17:2326-53.

doi:10.1038/s41596-022-00728-0

33 Pettersen EF, Goddard TD, Huang CC*, et al.* UCSF Chimera--a visualization system for

exploratory research and analysis. *J Comput Chem* 2004;25:1605-12. doi:10.1002/jcc.20084

34 Reis LM, Tyler RC, Muheisen S*, et al.* Whole exome sequencing in dominant cataract

identifies a new causative factor, CRYBA2, and a variety of novel alleles in known genes.

Hum Genet 2013;132:761-70. doi:10.1007/s00439-013-1289-0

35 Smedley D, Smith KR, Martin A, *et al.* 100,000 Genomes Pilot on Rare-Disease

Diagnosis in Health Care - Preliminary Report. *N Engl J Med* 2021;385:1868-80.

doi:10.1056/NEJMoa2035790

36 Chang YF, Imam JS and Wilkinson MF. The nonsense-mediated decay RNA surveillance

pathway. *Annu Rev Biochem* 2007;76:51-74.

doi:10.1146/annurev.biochem.76.050106.093909

37 Sobreira N, Schiettecatte F, Valle D, *et al.* GeneMatcher: a matching tool for connecting

investigators with an interest in the same gene. *Hum Mutat* 2015;36:928-30.

doi:10.1002/humu.22844

38 Melé M, Ferreira PG, Reverter F, *et al.* Human genomics. The human transcriptome

across tissues and individuals. *Science* 2015;348:660-5. doi:10.1126/science.aaa0355

39 Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in

60,706 humans. *Nature* 2016;536:285-91. doi:10.1038/nature19057

40 Bauwens M, Garanto A, Sangermano R, *et al.* ABCA4-associated disease as a model for

missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory,

structural, and recurrent hypomorphic variants. *Genet Med* 2019;21:1761-71.

doi:10.1038/s41436-018-0420-y

41 Pascoe HG, Wang Y and Zhang X. Structural mechanisms of plexin signaling. *Prog

Biophys Mol Biol* 2015;118:161-8. doi:10.1016/j.pbiomolbio.2015.03.006

42 Morr T. Amelogenesis imperfecta: More than just an enamel problem. *J Esthet Restor

Dent* 2023;35:745-57. doi:10.1111/jerd.13063

43 Dong J, Ruan W and Duan X. Molecular-based phenotype variations in amelogenesis

imperfecta. *Oral Dis* 2023, in press. doi:10.1111/odi.14599

44 Smith CEL, Poulter JA, Antanaviciute A, *et al.* Amelogenesis Imperfecta; Genes,

Proteins, and Pathways. *Front Physiol* 2017;8:435. doi:10.3389/fphys.2017.00435

45 Smith CE, Poulter JA, Levin AV*, et al.* Spectrum of PEX1 and PEX6 variants in Heimler

syndrome. *Eur J Hum Genet* 2016;24:1565-71. doi:10.1038/ejhg.2016.62

Figure legends

Figure 1. Pedigrees, Sanger sequencing and schematic diagram of the PLXNB2 protein. **A.** DNA was available for all labelled individuals on each pedigree. Arrows indicate the individuals whose DNA was exome or genome sequenced. Affected status is as reported by the families for individuals for which DNA was not available. Sanger sequencing traces showing the segregation of each variant with disease in each family, except for family 5, for which this is shown in Supplementary Figure 9. The schematic diagram shows B. the *PLXNB2* transcript (ENST00000359337.9, NM_012401.4; 6409 bp) and C. the PLXNB2 protein (ENSP00000352288.4, NP_036533.2; 1838 amino acids), with the positions marked for the pathogenic variants identified in this study.

Figure 2.

Clinical images that are illustrative of the shared and variable clinical features for affected individuals. **A.** The anterior clinical photograph showing primary teeth, the posterior teeth have changes consistent with loss of enamel (arrowheads) due to fracturing, whereas the enamel of the anterior teeth is developmentally thin and optically abnormal (Family 3 II:1). **B.** Panoramic radiograph of the adult permanent dentition illustrates that the enamel is more radiodense than the supporting dentine, but with a variably reduced enamel volume and irregular occlusal cusp morphology (arrowhead) consistent with AI characterised by enamel that is hypomineralised with variable hypoplasia (Family 4 II:2). **C.** SNHL was typical, with provision of hearing aids in childhood (Family 4 II:2). Blind-end skin tracts involving the skin of the ears or adjacent tissues (arrowhead) were observed in at least 3/6 families. **D.** Lower limb lymphoedema was observed in two families (Families 4 and 5, with Family 4 illustrated). **E.** Fundus autofluorescence images (Family 3 II:1) illustrate a developmental

macular abnormality with pale fundus and attenuated blood vessels. The individual also has

high myopia, nystagmus and microcornea.

Figure 3. Analysis of mouse *Plxnb2* transcript distribution by *in situ* hybridization. Selected

sections illustrating *Plxnb2* expression features in the developing skull bone, sensorial organs

and viscera are shown in the left-side panels (A–F), whereas right-side panels focus on

incisor (G, I and K) and molar (H, J and L) tooth development. Developmental stages and

section planes are: E14.5 frontal (A, B, G and H), E16.5 sagittal (C, D, E, I and J); E19.5

sagittal (F, K and L) sections. Am, ameloblasts; Co, cochlea; DP, dental papilla; EL,

epithelial loop; Gu, gubernaculum; IDE, inner dental epithelium; In, intestine; Ki, Kidney;

LI, lower incisor; Lu, lung; Od, odontoblasts; ODE, outer dental epithelium; OE, Olfactory

epithelium; Re, Retina; SB, skull bone; SR, stellate reticulum. Scale bars: 25 µm (J); 40 µm

(H and I); 50 µm (C and K); 60 µm (F and H), 80 µm (A D, E and L); 150 µm (B and G).

Figure 1. Pedigrees, Sanger sequencing and schematic diagram of the PLXNB2 protein. A. DNA was available for all labelled individuals on each pedigree. Arrows indicate the individuals whose DNA was exome or genome sequenced. Affected status is as reported by the families for individuals for which DNA was not available. Sanger sequencing traces showing the segregation of each variant with disease in each family, except for family 5, for which this is shown in Supplementary Figure 9. The schematic diagram shows B. the *PLXNB2* transcript (ENST00000359337.9, NM_012401.4; 6409 bp) and C. the PLXNB2 protein (ENSP00000352288.4, NP_036533.2; 1838 amino acids), with the positions marked for the pathogenic variants identified in this study.

190x245mm (96 x 96 DPI)

Figure 2. Clinical images that are illustrative of the shared and variable clinical features for affected individuals. A. The anterior clinical photograph showing primary teeth, the posterior teeth have changes consistent with loss of enamel (arrowheads) due to fracturing, whereas the enamel of the anterior teeth is developmentally thin and optically abnormal (Family 3 II:1). B. Panoramic radiograph of the adult permanent dentition illustrates that the enamel is more radiodense than the supporting dentine, but with a variably reduced enamel volume and irregular occlusal cusp morphology (arrowhead) consistent with AI characterised by enamel that is hypomineralised with variable hypoplasia (Family 4 II:2). C. SNHL was typical, with provision of hearing aids in childhood (Family 4 II:2). Blind-end skin tracts involving the skin of the ears or adjacent tissues (arrowhead) were observed in at least 3/6 families. D. Lower limb lymphoedema was observed in two families (Families 4 and 5, with Family 4 illustrated). E. Fundus autofluorescence images (Family 3 II:1) illustrate a developmental macular abnormality with pale fundus and attenuated blood vessels. The individual also has high myopia, nystagmus and microcornea.

190x245mm (233 x 233 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 3. Analysis of mouse *Plxnb2* transcript distribution by in situ hybridization. Selected sections illustrating *Plxnb2* expression features in the developing skull bone, sensorial organs and viscera are shown in the left-side panels (A–F), whereas right-side panels focus on incisor (G, I and K) and molar (H, J and L) tooth development. Developmental stages and section planes are: E14.5 frontal (A, B, G and H), E16.5 sagittal (C, D, E, I and J); E19.5 sagittal (F, K and L) sections. Am, ameloblasts; Co, cochlea; DP, dental papilla; EL, epithelial loop; Gu, gubernaculum; IDE, inner dental epithelium; In, intestine; Ki, Kidney; LI, lower incisor; Lu, lung; Od, odontoblasts; ODE, outer dental epithelium; OE, Olfactory epithelium; Re, Retina; SB, skull bone; SR, stellate reticulum. Scale bars: 25 µm (J); 40 µm (H and I); 50 µm (C and K); 60 µm (F and H), 80 µm (A D, E and L); 150 µm (B and G).

431x260mm (150 x 150 DPI)

Biallelic variants in Plexin B2 (*PLXNB2*) cause amelogenesis imperfecta, hearing loss and intellectual disability.

Claire E L Smith[1]*, Virginie Laugel-Haushalter[2]*, Ummey Hany[1], Sunayna Best[1,3], Rachel L Taylor[4,5,6], James A. Poulter[1], The UK Inherited Retinal Disease Consortium, Genomics England Research Consortium, Saskia B Wortmann[7,8], René G Feichtinger[7], Johannes A Mayr[7], Suhaila Al Bahlani[9], Georgios Nikolopoulos[10], Alice Rigby[1,11], Graeme C Black[4,5], Christopher M. Watson[1,12], Sahar Mansour[13,14], Chris F Inglehearn[1], Alan J Mighell[11]* and Agnès Bloch-Zupan[2,15,16]*

[1]Leeds Institute of Medical Research, St James's University Hospital, University of Leeds, Beckett Street, Leeds. LS9 7TF. UK.
[2]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258, CNRS-UMR7104, Université de Strasbourg, Illkirch, France.
[3]Yorkshire Regional Genetics Service, Leeds Teaching Hospitals NHS Trust, Leeds. LS7 4SA. UK.
[4]Division of Evolution and Genomic Sciences, Manchester Academic Health Science Centre, Faculty of Biology, Medicine and Health, School of Biological Sciences, University of Manchester, Manchester, M13 9NT. UK.
[5]Manchester Centre for Genomic Medicine, St. Mary's Hospital, Manchester University NHS Foundation Trust, Manchester. M13 9WL. UK.
[6]EMQN CIC, Pencroft Way, Manchester Science Park, Manchester. M15 6SE. UK.
[7]University Children's Hospital, Salzburger Landesklinken (SALK) and Paracelsus Medical University (PMU), Müllner Hauptstrasse 48, 5020 Salzburg, Austria.
[8]Amalia Children's Hospital, Radboud University Medical Center, Geert Grooteplein Zuid 10, 6525 GA Nijmegen, the Netherlands.
[9]Dental & O.M.F.S Clinic, Al Nahdha Hospital, Ministry of Health, Muscat. HGQJ+GV6. Oman.
[10]Institute for Fundamental Biomedical Research, BSRC Alexander Fleming, 16672 Vari, Attica, Greece.
[11]School of Dentistry, University of Leeds. LS2 9JT. UK.
[12]North East and Yorkshire Genomic Laboratory Hub, Central Lab, St. James's University Hospital, Leeds. LS9 7TF. UK.
[13]Lymphovascular Research Unit, Molecular and Clinical Sciences Research Institute, University of London St George's, London. SW17 0QT. UK.
[14]SW Thames Regional Genetics Service, St George's University Hospitals NHS Foundation Trust, London. SW17 0QT. UK.
[15]Faculté de Chirurgie Dentaire, Université de Strasbourg, 8 Rue Sainte Elisabeth, F-67000 Strasbourg, France.
[16]Centre de référence des maladies rares orales et dentaires O-Rares, Filière Santé Maladies rares TETE COU, European Reference Network CRANIO, Pôle de Médecine et Chirurgie Bucco-dentaires, Hôpital Civil, Hôpitaux Universitaires de Strasbourg (HUS), 1 place l'hôpital, 67000 Strasbourg, France.

* These authors contributed equally to the study.
Correspondence to:
Dr Claire E L Smith email C.E.L.Smith@leeds.ac.uk
Telephone (+44) 113 343 8445

Supplementary Materials and Methods.

Sequencing and analysis

Since individuals were recruited and sequenced by different institutions, the sequencing and filtering methods used vary and are detailed below.

Family 1

Regions of homozygosity were identified using the GeneChip® Human Mapping 250K *Nsp* assay (Affymetrix). Results were analysed using Affymetrix software. WES was performed by IntegraGen (Evry, France; https://integragen.com). WES and bioinformatic analysis followed the protocol previously described.[1] Briefly exons were captured using the SureSelect Human All Exon V5 + UTR 75 Mb reagent (Agilent Technologies, USA) and sequenced using an Illumina HiSeq 2000 sequencer (Illumina, USA). The Burrows-Wheeler Aligner (v7.12) was used to align reads to the Human genome reference sequence GRCh37. Variant calling was carried out using the HaplotypeCaller module of Genome Analysis Toolkit (GATK, v.3.4.36). Copy number variants were called using CANOES.[2] The variants were filtered to exclude variants with an allele frequency of more than 1% in public variation databases-including the 1,000 Genomes,[3] the Genome Aggregation Database (gnomAD) (v.2.2.1).[4] Variants in the 5' or 3' UTR and variants with intronic locations and no prediction of local splice effect, as well as synonymous variants without prediction of local splice effect were also filtered. Splice effects were predicted using MaxEntScan,[5] NNSplice[6] and SpliceSitFinder.[7] Genes with homozygous and compound heterozygous variants consistent with a recessive mode of inheritance were prioritised for investigation.

Family 2

Three micrograms of genomic DNA were processed according to the SureSelect XT Library Prep protocol (Agilent Technologies, USA). SureSelect Human All Exon V5 (Agilent Technologies) was used as the exome capture reagent. Sequencing was performed using a 100 bp paired-end protocol on an Illumina HiSeq 2500 sequencer (Illumina, USA). The resulting fastq files were aligned to the human reference genome (GRCh37) using Burrows-Wheeler Aligner (BWA, v.0.7.12).[8] The alignment was processed according to GATK best practice. Exome depth was used for CNV analysis according to the developers' guidelines.[9]

Variants were filtered to exclude all changes other than missense, frameshift or stop variants, exonic insertion/deletions or variants located at splice consensus sites. Synonymous variants within the splice consensus region and predicted to affect splicing were also retained. Variants present in the gnomAD (v2.2.1)[4] were excluded if present at a frequency of 1% or higher. Variants were also filtered based on the mode of inheritance. In families known to be consanguineous, homozygous variants were prioritised. Variant lists were also filtered to exclude population specific high frequency variants and platform artefacts by removing variants also present in exomes of individuals of the same ethnicity without dental disease sequenced using the same reagents and platform.

Family 3

One microgram of genomic DNA was processed according to the Agilent SureSelect XT Library Prep protocol (Agilent Technologies). SureSelect Human All Exon V6 (Agilent Technologies) was used as the exome capture reagent. Whole exome libraries were sequenced with a 150 bp paired-end protocol on an Illumina Hi-Seq 3000 sequencer. Alignment, variant calling and filtering were as previously described for Family 2.

Family 4

Investigation of the 100,000 genomes rare disease dataset[10] was carried out using the Gene-Variant Workflow script (available from https://research-help.genomicsengland.co.uk/display/GERE/Gene-Variant+Workflow). This identified all *PLXNB2* variants from all available genome VCF files and annotated them using Ensembl Variant Effect Predictor. A custom Python script called filter_gene_variant_workflow.py (available from https://github.com/sunaynabest/filter_100K_gene_variant_workflow) was used to exclude common variants using the following criteria: 100K major allele frequency (MAF) ≥0.002; gnomAD allele frequency (AF) ≥0.002 and variants called in non-canonical transcripts. Variants were prioritised if they were also marked deleterious by SIFT. Individuals with biallelic high impact variants (variants annotated by VEP as 'high impact' include stop_gained, stop_lost, start_lost, splice_acceptor_variant, splice_donor_variant, frameshift_variant, transcript_ablation, transcript_amplification) were prioritised for study. Variants were excluded if they were only identified in unaffected relatives and not identified in probands. Variant segregation was confirmed by examining the accompanying genome data from parents.

Family 5

50 ng of genomic DNA was processed according to the Human Comprehensive Exome kit (Twist Bioscience) protocol. The exome library was sequenced using an Illumina P3 300 cycle kit on an Illumina NextSeq2000 instrument.

After checking the quality of the raw data by FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adaptor trimming by Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), sequence alignments were generated against the reference genome (hg19/GRCh37) using the BWA (v0.7.12-r1.39).[8] SAMtools was used to convert SAM files to BAM files and PCR duplicates were removed by Picard (v2.5.0), to produce sorted BAM files.

Variants were filtered to exclude all changes other than missense, frameshift or stop variants, exonic small insertion/deletions or variants located at splice consensus sites. Synonymous variants within the splice consensus region and predicted to affect splicing were also retained. Identified variants were excluded if the CADD (v1.6) score was <15 or minor allele frequency (MAF) was >0.01. The filtered variant list was examined for candidate pathogenic variants in known or potential candidate AI or sensorineural hearing loss genes.

The sorted BAM files were used for custom CNV detection using a pipeline based on the R package ExomeDepth (v1.1.12) using default parameters.[9] A custom reference panel was created using BAM files from another 47 samples sequenced in the same run. Common CNVs[11] were filtered out and CNVs calls prioritised by Bayes factor (the $\log_{10}$ of the likelihood ratio of data for the CNV call divided by that of the normal copy number call).

Soft clipped reads were visualised in Integrated Genomics Viewer to identify CNV breakpoints.

Family 6

Whole exome libraries were prepared from genomic DNA using the SureSelect60Mbv6 (Agilent) reagent using a 100 bp paired-end protocol on a HiSeq 4000 platform (Illumina). Reads were aligned to GRCh37/hg19 using the BWA (v.0.5.87.5).[8] Detection of genetic variation was performed using PINDEL (v 0.2.4t),[12] and ExomeDepth (v 1.0.0).[9] The

cut-off for biallelic inheritance was set to <0.01 allele frequency. The size of reference entries was >27,000 exomes in the in-house database at the time of analysis. Only remaining variants present at <0.01 allele frequency in gnomAD (v2.2.1) were retained.

Missense, nonsense, stoploss, splice, splice consensus, frameshift, and indel variants were included in standard evaluation. Synonymous variants outside of the splice consensus sequence, 5-UTR, 3-UTR, non-coding, miRNA, intronic, intergenic, and regulatory variants were only considered if previously published as pathogenic relevant or if indicated by a distinctive phenotype.[13] For variant interpretation, CADD, Polyphen2, and SIFT scores were used for prioritization. Sanger sequencing for all families was performed with the BigDye Terminator v3.1 kit (Life Technologies) according to the manufacturer's instructions and resolved on an ABI3130xl sequencer (Life Technologies). Results were analysed with SeqScape 2.5 (Life Technologies).

Mouse tissue preparation
Mouse embryos/fetuses were collected at E14.5, E16.5, and on the day of birth (hereafter referred to as E19.5), after matings between C57BL6 mice. For E14.5 and older samples, the whole body was embedded in OCT 4583 medium (Tissue-TEK, Sakura) and frozen on the surface of dry ice. E12.5 embryos were fixed overnight in 4% paraformaldehyde (pH 7.5, w/v) in PBS, cryoprotected by overnight incubation in 20% sucrose (w/v) in PBS, and cryoembedded as described above. Cryosections (Leica CM3050S cryostat) at 10 µm were collected on Superfrost plus slides and stored at –80 °C until hybridization. E14.5 samples were sectioned in a frontal plane, whereas other stages were sectioned sagittally.

Probe Synthesis
The *Plxnb2* probe was synthesized from PCR-generated DNA templates kindly provided by the EURExpress consortium (http://www.eurexpress.org). The template sequence is provided in the supplementary information. DIG-labeled antisense riboprobe was transcribed in vitro by incubation for 2 h at 37 ° C using 1 µg of the PCR product, 20 U SP6 RNA polymerase, 5X transcription buffer (Promega), 10 X DIG RNA labelling mix (Roche), 0.5 M DTT, 20 U RNAse inhibitor (Roche) in a 20 µl volume. The reaction was stopped with 2 µl EDTA (0.2 M, pH 8), and RNA was precipitated with 1 µl yeast tRNA (10 mg/ml), 2.5 µl LiCl (4 M) and 75 µl absolute ethanol, followed by an incubation for 30 min at –80 ° C and centrifugation at 12,000 rpm (30 min at 4° C). The pellet was washed with 70% ethanol and recentrifuged. The supernatant was discarded and the pellet was allowed to dry. The probe was then diluted in 20 µl sterile H_2O. The quality of the probe was verified by electrophoresis in a 1% agarose gel. If no smear was observed and the size was as expected, the probe was considered to be ready for use. The quantity of RNA was evaluated by a Nanodrop spectrophotometer (ND-1000, Labtech) and adjusted to 150 ng/ µl in hybridization buffer, then stored at –20 °C until use.

In situ Hybridization
Slides were allowed to thaw to room temperature (RT) for 2 h. Then they were post-fixed on ice in 4% paraformaldehyde (diluted in PBS) for 10 min and rinsed in PBS. The hybridization buffer was composed of 50% deionized formamide, 10% dextran sulfate, 1mg/ml yeast tRNA, 1 X Denhardt's solution, and 1 X salt solution (0.195 M NaCl, 5 mM Tris pH 7.2, 5 mM $NaH_2PO_4.H_2O$, 5 mM $Na_2HPO_4.12H_2O$, 5 mM EDTA pH 8). The probe was diluted in hybridization buffer at a concentration of 1 µg/ml. The probe mix was denatured by a 10 min incubation at 70 °C and placed on ice. 100 µl was applied on each slide and covered by a coverslip. This was allowed to hybridize overnight at 65 °C in

humidified chambers. The slides were then washed twice for 30 min at 65 °C in 1X standard saline citrate (SSC), 50% formamide, 0.1% Tween-20, and twice for 30 min at RT in MABT buffer (1 X MAB (maleic acid buffer): 0.5 M maleic acid (Roche), 0.75 M NaCl pH 7.5, 0.1% Tween-20). Probe detection was performed using antibodies and reagents from Roche. Slides were incubated for 1 hr at RT with a blocking solution (20% goat serum, 2% blocking reagent in MABT). The anti-DIG antibody was diluted 1: 2,500 in blocking solution, and 200 µl was added to each slide, which were covered by Parafilm and incubated overnight at 4 ° C. Slides were washed 5 times in MABT for 20 min and then twice for 10 min in NTMT buffer (100 mM NaCl, 100 mM Tris-HCl pH 9.5, 50 mM $MgCl_2.6H_2O$, 0.1% Tween-20). An aliquot of 200 µl of freshly prepared staining solution (3.5 µl nitro-blue tetrazolium chloride (Roche), 3.5 µl 5-bromo-4-chloro-3'indolylphosphate p-toluidine salt (Roche) in NTMT buffer) was placed on each slide, covered by a Parafilm and incubated overnight in the dark at RT. The staining solution was changed every day and when signal was optimal, the slides were rinsed twice for 5 min in NTMT buffer. The slides were further rinsed with PBS and water, allowed to dry overnight, and mounted in Coverquick 2000 mounting medium (Labonord).

Target (GRCh37)	Forward primer (5′)	Reverse primer (3′)	Target (bp)	Primer Identifier (Use)
chr22:50722118-50722591	TACGGCAAGAATATCGACAGCAAG	AAATTGGACCCCAGGATGGTGA	474	Exon 14 (Family 1)
chr22:25600810-25601544	TGGGAAGATTAAACCGGCATCTGG	TGAGGCAGGAGAATGGCATGAAC	735	*CRYBB3* exon 5 (Family 1)
chr22:50722428-50722719	GAGGCTGCCCTTGTGACC	AGAGCTGCCGTCAGTGGT	292	Exon 13 (Family 2)
chr22:50721764-50722004	ACTTGGAAGGTGAACTGGACA	CTGTCAGCTGGGTCAGTACC	241	Exon 16 (Family 4)
chr22:50719090-50719291	TGTGGATGAACTGTAGGGGC	GCGACAAGGACGTGATGATC	202	Exon 24 (Family 4)
chr22:50715038-50715798	GTCCCTGGTGGTCTCCTTAC	CAGCACACGAGCAGAGAAAC	761	F5-R7 (Family 5)
chr22:50714866- 50715798	AAGCAGGAGCAGCTCATACT	CAGCACACGAGCAGAGAAAC	933	F6-R7 (Family 5)
chr22:50716913-50717351	GATCATTGACCAGGTGTACCG	CACCTTGGACAGGATGAGG	439	Exon 29 (Family 6)

Supplementary Table 1. Primers used to verify the variants identified in Families 1, 2, 4-6. Primers are for regions of *PLXNB2* unless otherwise stated. Coordinates are for the GRCh37 human reference genome.

Supplementary Figure 1. Family 1: Clinical images of affected individuals. Oral photographs and panoramic radiographs are shown for II:3 and II:6 which illustrate changes over time. The dentitions of both affected individuals are characterised by variable degrees of enamel hypoplasia AI with only a thin layer of enamel present in some places. Crown morphology is atypical in some teeth with striking flattening of normal cusp morphology of the permanent molar teeth evident on radiographs prior to eruption. The radiographic appearances of the dental pulp spaces are within expected limits other than mild taurodontism in some permanent molar teeth. The root morphology is within expected limits. Individual II:6 is shown with bilateral hearing aids.
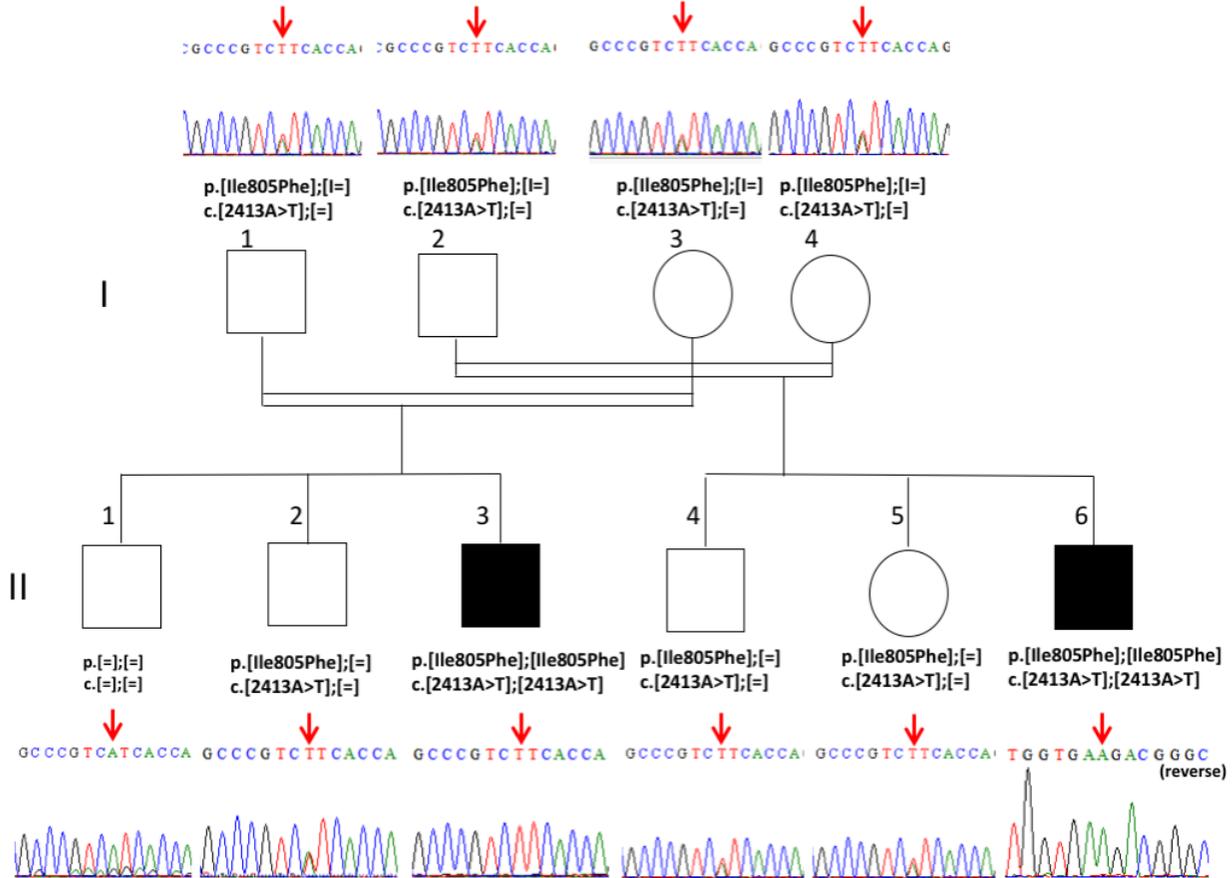
Supplementary Figure 2. Segregation of the *PLXNB2* variant c.2413A>T, p.(Ile805Phe) identified by exome sequencing of II:3 and II:6.

Family carrying variant and allelic status	Variant (GRCh37)	Transcript (Exon) ENST00000359337.9 / NM_012401.4	Protein	CADD v1.6 GRCh37	REVEL	Polyphen-2 (Hum Var)	SIFT	gnomAD v2.1.1; dbSNP155
1 homozygous	g.50722270T>A	c.2413A>T 14	p.(Ile805Phe)	25.2	0.649	0.97 predicted damaging	0 deleterious	Not present in either
2 homozygous	g.50722576C>T	c.2248G>A 13	p.(Asp750Asn)	23.6	0.159	0.177 benign	0.01 deleterious	Not present; rs1332617247
3 compound heterozygous	g.50728264G>T	c.750C>A 3	p.(Cys250*)	36	N/A	N/A	N/A	Not present in either
3 compound heterozygous	g.50720613C>T	c.3117G>A 19	p.(Thr1039=)	22.7	N/A	N/A	N/A	Not present; rs1234372437
4 compound heterozygous	g.50721841delA	c.2606del 16	p.(Phe869Serfs*45)	17.61	N/A	N/A	N/A	4.024 x10^{-6} 1/248518; not present
4 compound heterozygous	g.50719182_50719186 delAAAG	c.3982_3986del 24	p.(Phe1328Hisfs*65)	34	N/A	N/A	N/A	Not present in either
5 homozygous	g.50715085_50715672 del	c.5197-337_5310del 34-35	p.(Asp1733_Arg1779del)$^{\blacklozenge}$	N/A	N/A	N/A	N/A	N/A
6 homozygous	g.50717063C>T	c.4609G>A 29	p.(Gly1537Ser)	29.1	0.297	0.973 probably damaging	0 deleterious	Not present in either

Supplementary Table 2. Variant interpretation and population frequency data of the *PLXNB2* variants identified in the individuals included in this study. Various pathogenicity prediction scores are included to aid variant interpretation, including Combined Analysis Dependent Depletion (CADD),[14] REVEL,[15] Polyphen-2[16] and SIFT.[17] The population frequency of each variant in the gnomAD database[4] and its presence or absence in dbSNP155 is also shown. The Genomics England GRCh37 cohort was also checked for these variants, but none were present. Analysis by Franklin (https://franklin.genoox.com), showed all variants were classified as VUS (data not shown). Variants are based on genome build GRCh37 and *PLXNB2* transcript ENST00000359337.9, NM_012401.4 and PLXNB2 protein ENSP00000352288.4, NP_036533.2. $^{\blacklozenge}$Note that splice prediction software predicts the skipping of all of exon 35, despite only partial deletion.

Journal of Medical Genetics

Asp750

```
Zebrafish     NVTFYIKDKDTDRKIDSTLRVVLYNCSVRREDCSLCKNAHKNDSCVWCDT---TKSCIYR 818
Chicken       KLKLSMR--TGKVKIDSKLMVTLYNCSYGRSDCSLCLAAHEDYKCVWCEEDGKPSKCVYE 804
Mouse         PLHLYVK--SFGKNIDSKLQVTLYNCSFGRSDCSLCLAADPAYRCVWCRG---QNRCVYE 793
Human         PLHLYVK--SYGKNIDSKLHVTLYNCSFGRSDCSLCRAANPDYRCAWCGG---QSRCVYE 791
Gorilla       PLHLYVK--SYGKNIDSKLHVTLYNCSFGRSDCSLCRAANPDYRCAWCGG---QSRCVYE 791
Red deer      PLHLYVK--SSGKNVDSRLQVTLYNCSFGRSDCSLCLAADPAYRCVWCSG---QSRCVYE 795
Dog           PLHLYVK--SYGKNVDSKLQVTLYNCSFGRSDCSLCLAADPAYKCVWCAG---QGRCVYE 793
Cat           PLHLYVK--SYGKNVDSKLQVTLYNCSFGRSDCSLCLAADPVYKCVWCGG---QSRCVYE 793
Giant panda   PLHLYVK--SYDKNVDSKLQVTLYNCSFGRSDCSLCLAADPAYKCVWCAG---QSGCVYE 793
Pig           PLHLYVK--SYGKNIDSRLQVTLYNCSFGRSDCSLCLAADPAYKCVWCSG---QSRCVYE 793
Narwhal       PLHLYVK--SYGKNIDSRLQVTLYNCSFGRSDCSLCLAADPVYRCVWCSG---QSRCVYA 793
Baleen whale  PLHLYVK--SYGKNIDSRLQVTLYNCSFGRSDCSLCLAADPAYQCVWCSG---QSGCVYA 793
                 :  :  ::          ::** * *.***** *.***** *.      *.**          *.*
```

Ile805

```
Zebrafish     SLCDNE-LHQCPPPKITDIVPRYGPMNGRISVTIKGSNLGIKKEDIKKITVAKVPCIHQP 877
Chicken       KLCRAPPTNTCPHPEITRIEPKTGPLDGGILLTIMGSNLGIKAEDVKNITVADKECVFKE 864
Mouse         ALCSNV-TSECPPPVITRIQPETGPLGGGILVTIHGSNLGVKADDVKKITVAGQNCAFEP 852
Human         ALCN-T-TSECPPPVITRIQPETGPLGGGIRITILGSNLGVQAGDIQRISVAGRNCSFQP 849
Gorilla       ALCN-T-TSECPPPVITRIQPETGPLGGGIRITILGSNLGVQAGDIQRISVAGRNCSFQP 849
Red deer      ALCSNA-TSECPPPVVTRIQPETGPLGGGIRITILGSNLGVRADDVKRVTVAGQNCAFEP 854
Dog           ALCSNT-TSECPPPVITRIHPETGPLGGGIRITILGSNLGVTADDVKRVTVAGQNCAFEP 852
Cat           ALCSNA-TSECPPPVITRIHPETGPLGGGIRITILGSNLGVRADDVKRITVAEQNCVFQP 852
Giant panda   ALCSNT-TSECPPPVITRIHPETGPLGGGIRITIVGSNLGVRAEDVRGVTVAGQNCVFEP 852
Pig           ALCSNA-TSECPPPVITRIHPETGPLGGGIHVTILGSNLGVKADDVTRVTVAGRNCTFEP 852
Narwhal       ALCGNA-TSECPPPVITRIQPETGPLGGGIRVTILGSNLGVRADDVKRVTVAGRSCAFEP 852
Baleen whale  ALCGHA-TSECPPPVITRIQPETGPLGGGIRITILGSDLGVRADDVKRVTVAGQNCAFEP 852
                **        ** *    :* * *. **:.* * :** **:**:      *:    ::**    *  .:
```

Supplementary Figure 3. Multiple sequence alignment of PLXNB2 sequences from a variety of species to show the evolutionary conservation of the mutated residues suggested to be disease causing. The evolutionary conservation for residues Asp750 (Family 2), Ile805 (Family 1) and Gly1537 (Family 6) and surrounding residues is shown. Disease causing residues are highlighted by a red box. Human (*Homo sapiens*) NP_001363798.1; mouse (*Mus musculus*); gorilla (*Gorilla gorilla gorilla*) XP_030863195.1; pig (*Sus scrofa*) XP_020946968.1; dog (*Canis lupus familiaris*) XP_038535376.1; cat (*Felis catus*) XP_044918511.1; red deer (*Cervus elaphus*) XP_043737305.1; giant panda (*Ailuropoda melanoleuca*) XP_034499687.1; chicken (*Gallus gallus*) XP_046764781.1; narwhal (*Monodon monoceros*) XP_029061558.1; baleen whales (*Balaenoptera musculus*) XP_036721225.1; zebrafish (*Danio rerio*) NP_001155072.1.
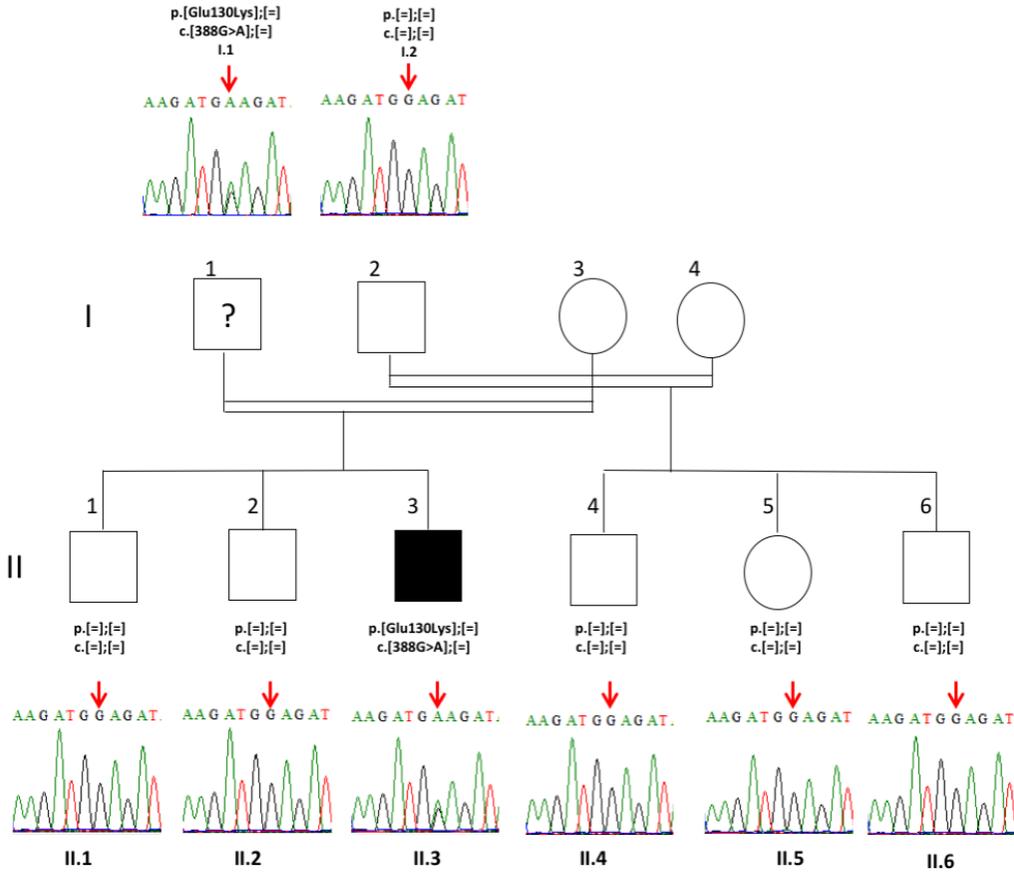
Supplementary Figure 4. Segregation of the *CRYBB3* variant c.388G>A, p.(Glu130Lys) identified by exome sequencing of II:3. The variant was inherited from his father (I:1). It is unknown whether I:1 has cataracts.

Family	Other genetic findings	Reference sequences
1	chr22:25601247G>A, Homozygous *CRYBB3* variant c.388G>A, p.(Glu130Lys). Did not segregate with disease for all family members, but was likely responsible for cataract phenotype in some individuals (Supplementary Figure 4).	MANE Select transcript NM_004076.5 (ENST00000215855.7) RefSeq protein NP_004067.1 (ENSP00000215855.2)
2	Chr21:47544826G>A, Homozygous *COL6A2* variant c.1762G>A:p.(Gly588Ser) [3 hets in gnomAD] segregated with disease for available family members CADDv1.6 score 27.1, in dbSNP (rs139488626), in ClinVar as a benign variant / variant of uncertain significance. However, the published COL6A2 disease phenotype is myopathy which was not observed in the affected individual examined. Homozygous variants with CADD>15 were also observed in *AQP1*/*FAM188B* (one variant in multiple transcripts) and *TRAK1*, but variants in these genes have no relevant known disease association and the encoded proteins have no obvious link with symptoms observed in patients, so they were not investigated further.	MANE Select transcript NM_0041849.4 (ENST00000300527.9) RefSeq protein NP_001840.4 (ENSP00000300527.4)
3	No other relevant variants identified. Known and candidate retinal disease genes were specifically checked as this individual was enrolled into a study on inherited retinal disease. The *PLXNB2* variant was scored highest by pathogenicity prediction software after variant filtering, with no other variants with comparable scores.	
4	Genomes England data had undergone tiering. No other likely causative variants for the phenotypes that the individual was entered into the study under were apparent in Tiers 1-3. The top hit variants for Exomiser were compound heterozygous variants in *FLNB* (comp het, c.2773G>T, p.(Gly925Cys) [0.1% in gnomAD] and c.124C>T, p.(Arg42*) [absent in gnomAD]). The only other known disease gene with biallelic variants was *PKD1* (comp het, c.11800G>A, p.(Gly3934Arg) [1 het in gnomAD] and c.872C>T, p.(Ser291Leu) [absent from gnomAD but only present in ENST00000567946, not the MANE Select one]). For both genes, their reported disease phenotype does not fit with the disease phenotype of the affected individual.	*FLNB* MANE Select transcript NM_001457.4 (ENST00000295956.9) FLNB RefSeq protein NP_001448.2 (ENSP00000295956.5) *PKD1* MANE Select transcript NM_001009944.3 (ENST00000262304.9) PKD1 RefSeq protein NP_001009944.3 (ENSP00000262304.4)
5	Family 5 was identified as having a similar phenotype by the same clinician treating family 4. Homozygous variants with CADD>15 were identified in *DENNDH6B*, *RHBDL1*, *TSC2*, *RIMS1*, *GPBP1*, *PKD1*, *SMPD3* and *COX3*. No other rare variants in genes for which variants are known to cause aspects of the phenotype observed in this patient or in candidate genes were identified. The phenotypes expected for these genes, where already known, either do not match the mode of disease inheritance or the phenotype of the affected individual. Parents were noted to be unaffected by disease.	N/A

Family	Other genetic findings	Reference sequences
6	Family 6 was identified through Genematcher, where labs report variants in genes believed to be causative for disease but for which only have one family identified. This flagged an individual with biallelic *PLXNB2* variants with an overlapping disease phenotype to the other cases. The homozygous variant identified in *PLXNB2* was scored highest by pathogenicity prediction software after variant filtering, with no other homozygous variants with comparable scores.	N/A

Supplementary Table 3. Details of other genetic findings in affected individuals after variant filtering.

Supplementary Figure 5. Family 2: Clinical images of IV:2 are consistent with hypomineralised AI.
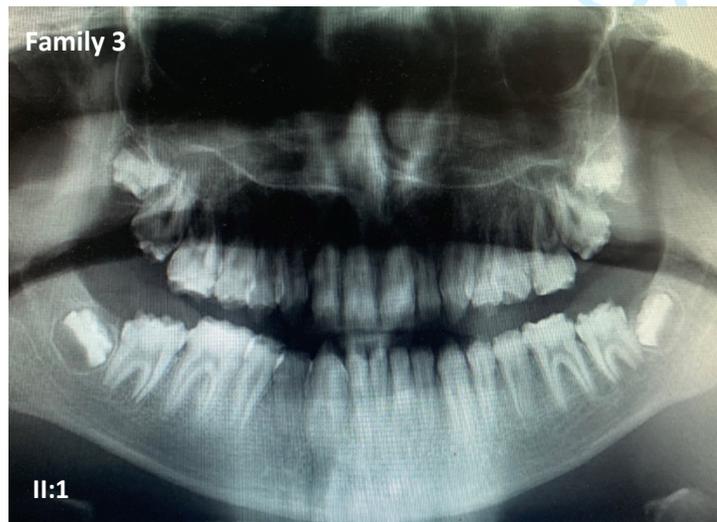
Supplementary Figure 6. Family 3: **II:1** The dental panoramic tomograph (OPT) captures the abnormal dental enamel formation with variable enamel hypoplasia and crown morphology that is especially evident in the unerupted teeth (arrowhead) where the cervical enamel is a more normal morphology compared to the occlusal enamel, which is markedly hypoplastic. These observations are consistent with those for the OPT presented for Family 4 II:1.

Genotype	pos 5'->3' phase strand	confidence	5' exon intron 3'
Wildtype	721 0 +	0.82	GCCCATGACG^GTCAGTCCTG
Mutant	ND		

Supplementary Table 4. NetGene2 (v2.4.2) donor splice site predictions for the effect of the variant identified in Family 2. A section spanning the flanking introns and exons was analysed (chr22:50720258-50721332, based on genome build GRCh37, reverse complement), selected results are shown.   Wildtype and mutant *PLXNB2* (g.50720613C>T) were investigated. ND not detected.

Supplementary Figure 7. SpliceAI predictions for the effect of the variant identified in Family 2. SpliceAI gave the variant a delta score of 0.67 for splice donor loss, which is above the threshold for alterations of splicing of 0.5.
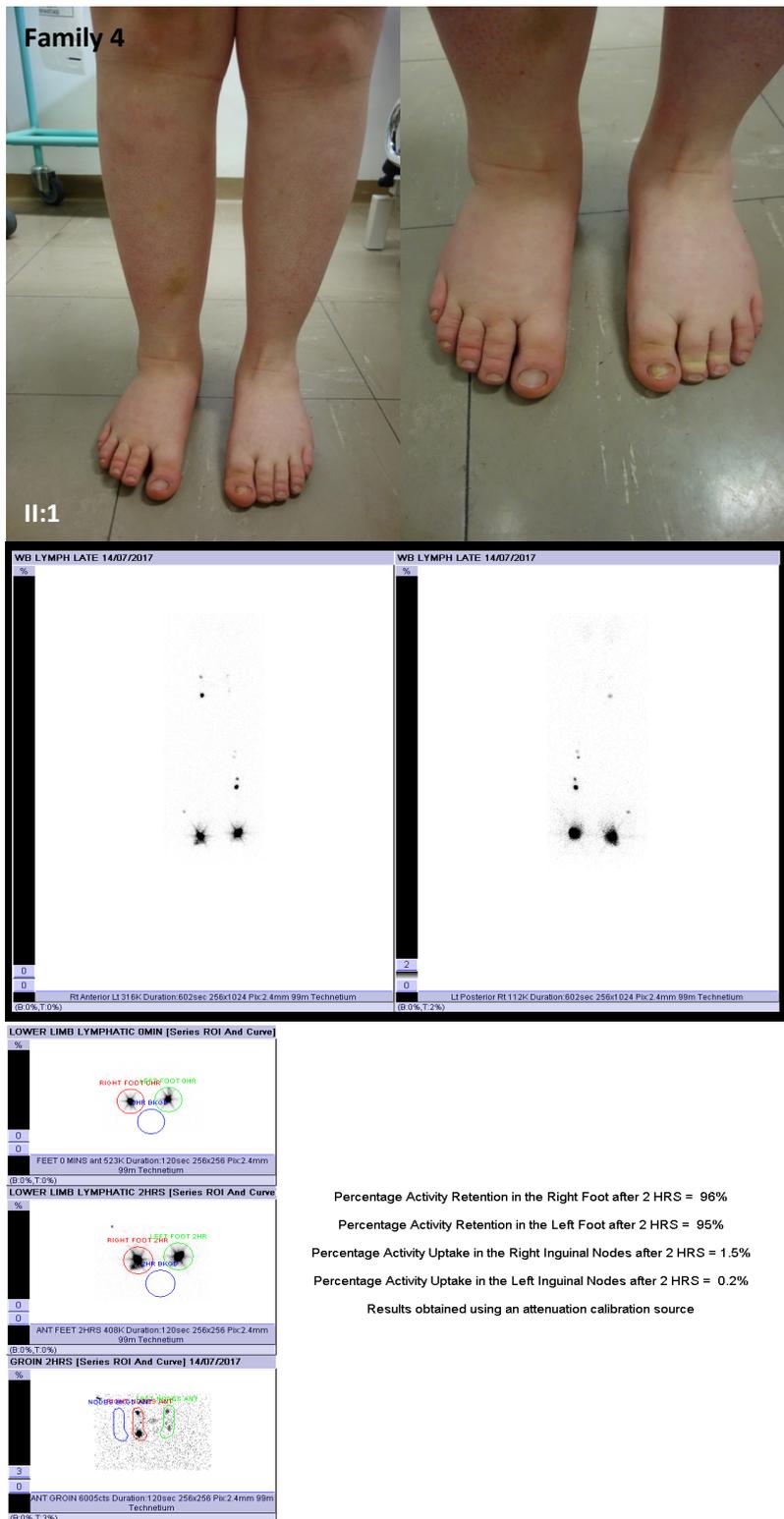
Supplementary Figure 8. Family 4: Bilateral lower limb lymphoedema in Family 4 II:1. Images show bilateral lower limb lymphoedema. Bilateral lower limb lymphoscintigraphy imaging at 2 hours after injecting radioactive [99]technetium into toe web spaces is also shown. Imaging demonstrates poor uptake into the inguinal lymph nodes with few nodes visualised. The uptake into the inguinal lymph nodes at 2 hours was 1.5% on the right and 0.2% on the left, in contrast, normal uptake is >12%. The appearance of popliteal lymph nodes on the left suggesting deep rerouting of the lymphatic tracts.
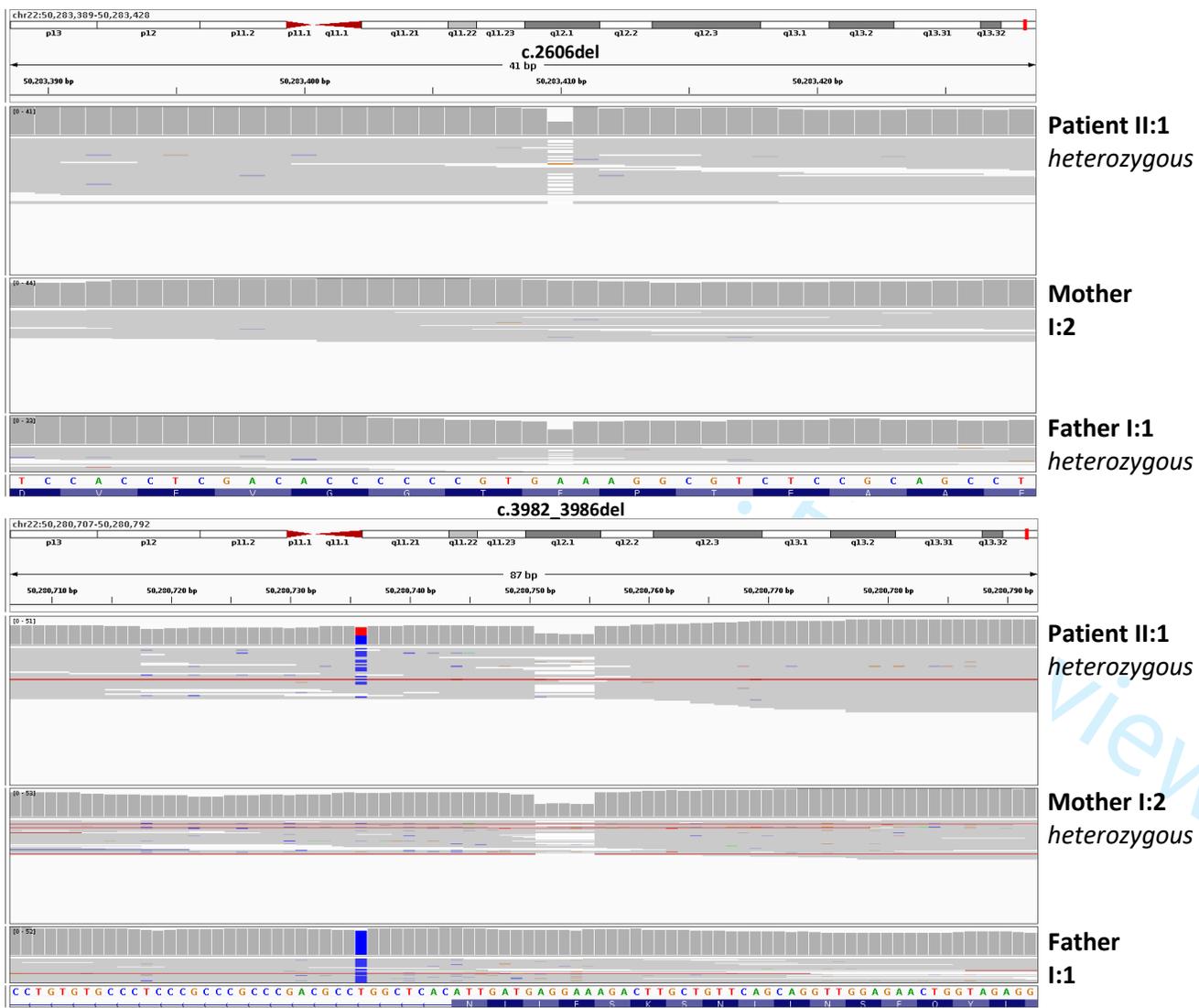
Supplementary Figure 9. IGV images of the variants positions (GRCh38) for individuals from Family 4. IGV traces show c.2606del (upper panel) and c.3982_3986del (lower panel) showing the reduction in read depth relative to the surrounding positions and in comparison to the sequencing of other family members.
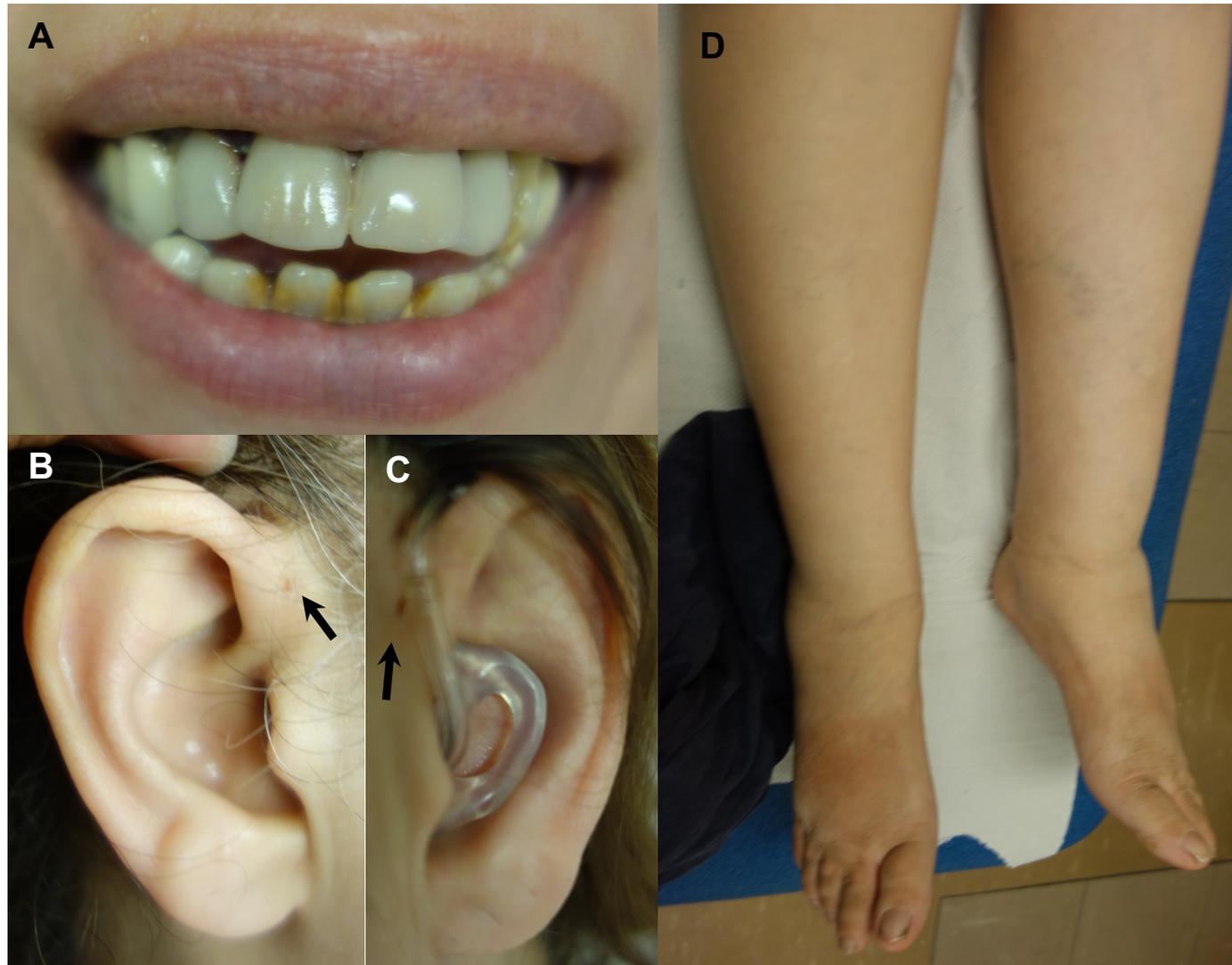
Supplementary Figure 10. Clinical images of Family 5 II:1. A The permanent dentition has been extensively restored and there are calculus deposits on the lower anterior teeth. B and C Periauricular pits are present (blind ended fistulas; arrows). The individual uses a hearing aid. D Bilateral lymphoedema of the lower limbs.
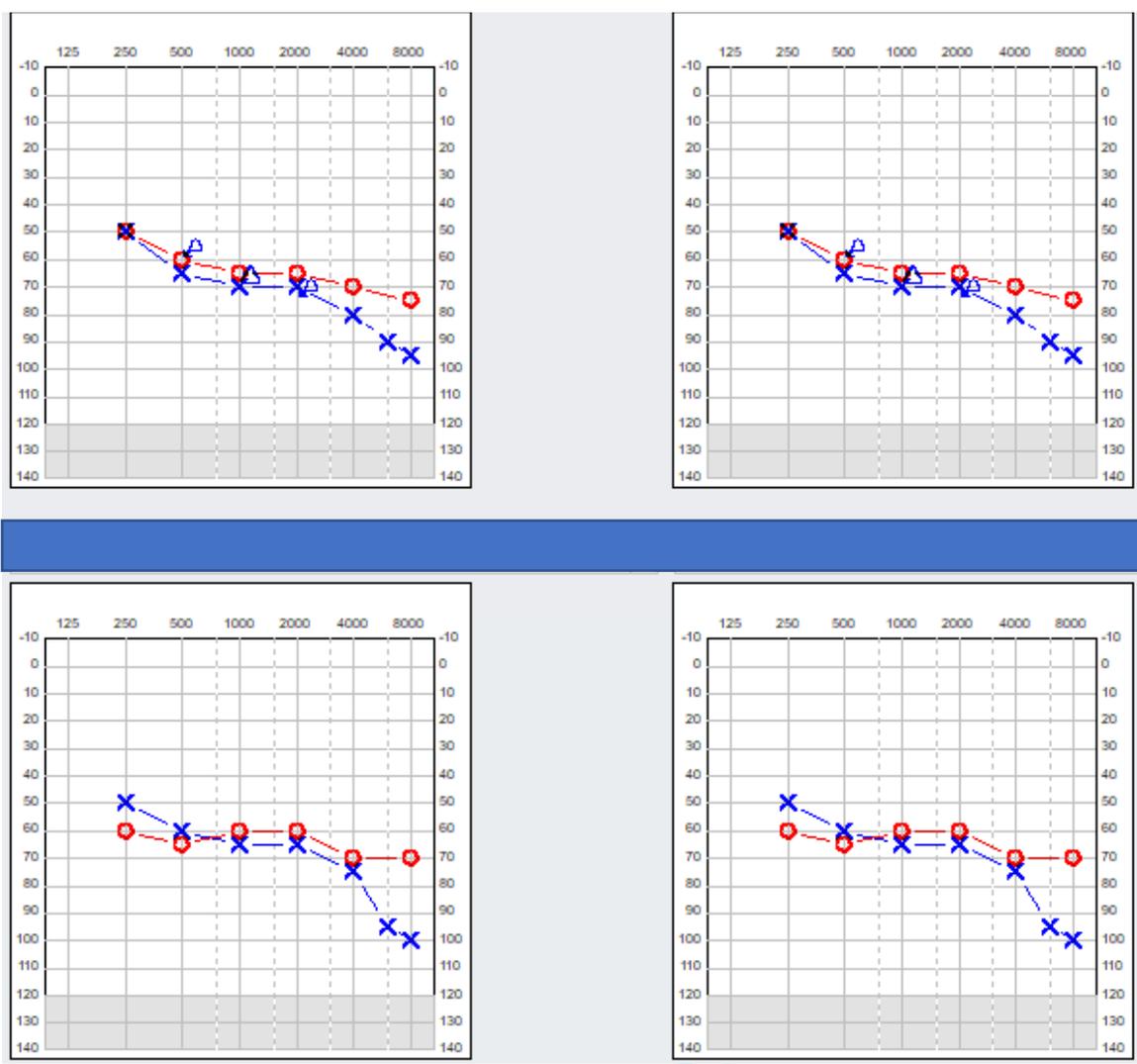
Supplementary Figure 11. Auditory exam history of Family 5 II:1. Results for the left ear (blue) and right ear (red) with headphones are shown for sound frequencies 250-8000 Hz. Results for tests carried out in the past year (top left), one year earlier (top right) and five years earlier are included (both tests shown at the bottom).
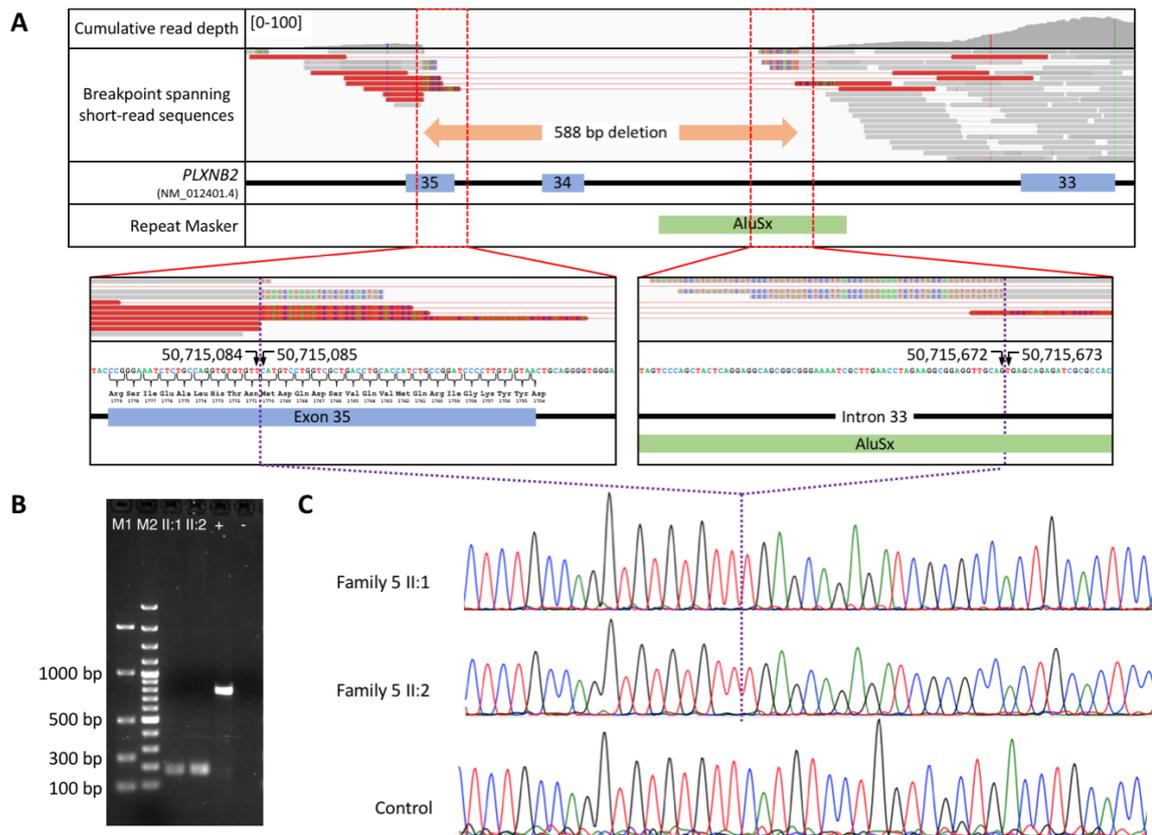
Supplementary Figure 12. Characterisation of the *PLXNB2* deletion in Family 5 by whole exome and Sanger sequencing. A Diagram showing the results of short read data analysis. Soft clipped reads revealed the breakpoints to be at chr22:50715085 and chr22:50715672 (based on genome build GRCh37), resulting in the complete deletion of exon 34 and a partial deletion of exon 35, c.5197-337_5310del (NM_012401.4), predicted to result in an in-frame deletion p.(Asp1733_Met1770del) (NP_036533.2), however splicing prediction predicts that the remainder of exon 35 will be skipped, resulting in p.(Asp1733_Arg1779del). B Gel image showing products resulting from PCR amplification using primers F5-R7 (Supplementary Table 1). DNA from both II:1 and II:2 but not a control individual produces an approx. 175 bp product when a product of 760 bp is expected. Legend: M1: EasyLadderI (Bioline), M2: 100 bp ladder (Fermentas), + control DNA, - water. C Sanger sequencing, using F6-R7 (Supplemental Table 1). confirmed these breakpoint coordinates, revealed that the same deletion was also present in the affected sibling and showed that there were no other alterations around the breakpoints.

Genotype	pos 5'->3' phase strand	confidence	5' exon intron 3'
Wildtype	158 2 +	0.95	TCTGCCTCAG^CTTACCGCTC
Mutant	ND		

Supplementary Table 5. NetGene2 (v2.4.2) acceptor splice site predictions for the effect of the deletion identified in Family 5. A section spanning the flanking introns and exons was analysed (chr22:50,713,408-50,716,325, reverse complement), selected results are shown. Wildtype and mutant *PLXNB2* (g.50715085_50715672del) were investigated. No alternative splice acceptor site was detected. ND not detected.

Supplementary Figure 13. Family 6: Examination of the dentition of III:1 was limited by patient compliance, but extensive post-eruptive enamel changes were clear, with dental caries as a contributory factor.
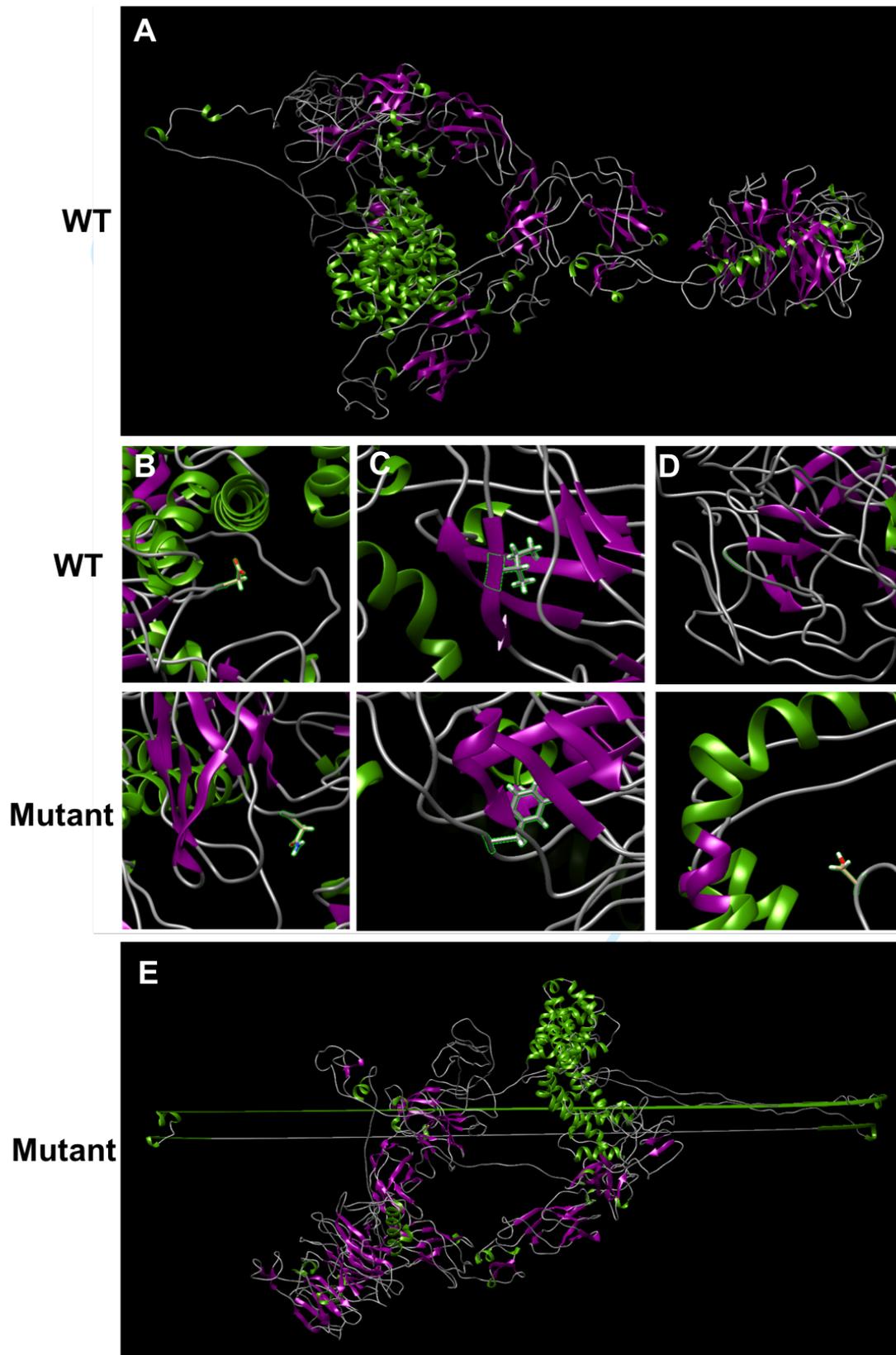
Supplementary Figure 14. Predicted tertiary structure of PLXNB2 using the I-Tasser-MTD algorithm. A Wild-type (WT) protein structure. B Comparison of WT with Asp750Asn mutant structure. C Comparison of WT with Ile805Phe mutant structure. D Comparison of WT with Gly1537Ser mutant structure. E. Mutant p.(Asp1733_Arg1779del) protein structure.

CAAAGCAGATGACGTCAAGAAGATAACTGTGGCTGGCCAGAACTGTGCCTTTGA
ACCAAGAGGGTACTCCGTATCCACCCGGATTGTGTGTGCAATTGAGGCTTCGGAG
ATGCCCTTCACAGGAGGCATTGAGGTGGATGTTAATGGAAAACTCGGCCATTCA
CCGCCACACGTCCAGTTCACTTATCAACAACCCCAGCCTCTCAGTGTGGAGCCAC
GACAGGGGCCACAGGCAGGTGGCACCACATTGACCATCAATGGCACTCACCTGG
ACACAGGCTCCAAGGAGGATGTGCGGGTGACACTCAATGACGTCCCTTGTGAAG
TGACAAAGTTTGGAGCACAGCTACAGTGTGTCACAGGTCAACAGTTGGCTCCAG
GCCAGGTGACACTAGAAATCTACTATGGGGGCTCCAGAGTGCCCAGCCCCGGCA
TCTCTTTCACCTACTGCGAGAACCCCATGATACGAGCCTTTGAGCCATTGAGAAG
CTTTGTCAGTGGTGGCCGGAGCATCAACGTTACTGGCCAGGGCTTCAGCCTCATC
CAGAAGTTTGCCATGGTTGTCATCGCTGAGCCCTTGCGGTCCTGGAGGCGGCGGC
GGCGGGAGGCTGGAGCCCTGGAGCGTGTGACGGTCGAGGGCATGGAGTACGTGT
TCTACAACGACACCAAGGTCGTCTTCTTGTCTCCTGCTGTCCCCGAAGAGCCCGA
GGCTTACAACCTCACCGTGCTGATA

Supplementary Figure 15. Murine *Plxnb2* template sequence (733 bp) used for in situ hybridization.

Supplementary Figure 16. Bulk tissue gene expression for *PLXNB2* from GTEx portal.[18]

References

1 Laugel-Haushalter V, Bär S, Schaefer E, *et al.* A New SLC10A7 Homozygous Missense

Mutation Responsible for a Milder Phenotype of Skeletal Dysplasia With Amelogenesis

Imperfecta. *Front Genet* 2019;10:504. doi:10.3389/fgene.2019.00504

2 Backenroth D, Homsy J, Murillo LR, *et al.* CANOES: detecting rare copy number variants

from whole exome sequencing data. *Nucleic Acids Res* 2014;42:e97. doi:10.1093/nar/gku345

3 Abecasis GR, Auton A, Brooks LD, *et al.* An integrated map of genetic variation from

1,092 human genomes. *Nature* 2012;491:56-65. doi:10.1038/nature11632

4 Karczewski KJ, Francioli LC, Tiao G, *et al.* The mutational constraint spectrum quantified

from variation in 141,456 humans. *Nature* 2020;581:434-43. doi:10.1038/s41586-020-2308-7

5 Yeo G and Burge CB. Maximum entropy modeling of short sequence motifs with

applications to RNA splicing signals. *J Comput Biol* 2004;11:377-94.

doi:10.1089/1066527041410418

6 Reese MG, Eeckman FH, Kulp D, *et al.* Improved splice site detection in Genie. *J Comput

Biol* 1997;4:311-23. doi:10.1089/cmb.1997.4.311

7 Shapiro MB and Senapathy P. RNA splice junctions of different classes of eukaryotes:

sequence statistics and functional implications in gene expression. *Nucleic Acids Res*

1987;15:7155-74.

8 Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler

transform. *Bioinformatics* 2009;25:1754-60. doi:10.1093/bioinformatics/btp324

9 Plagnol V, Curtis J, Epstein M, *et al.* A robust model for read count data in exome

sequencing experiments and implications for copy number variant calling. *Bioinformatics*

2012;28:2747-54. doi:10.1093/bioinformatics/bts526

10 Turnbull C, Scott RH, Thomas E, *et al.* The 100 000 Genomes Project: bringing whole

genome sequencing to the NHS. *Bmj* 2018;361:k1687. doi:10.1136/bmj.k1687

11 Conrad DF, Pinto D, Redon R*, et al.* Origins and functional impact of copy number variation in the human genome. *Nature* 2010;464:704-12. doi:10.1038/nature08516

12 Ye K, Guo L, Yang X*, et al.* Split-Read Indel and Structural Variant Calling Using PINDEL. *Methods Mol Biol* 2018;1833:95-105. doi:10.1007/978-1-4939-8666-8_7

13 El-Gazzar A, Mayr JA, Voraberger B*, et al.* A novel cryptic splice site mutation in COL1A2 as a cause of osteogenesis imperfecta. *Bone Rep* 2021;15:101110. doi:10.1016/j.bonr.2021.101110

14 Rentzsch P, Schubach M, Shendure J*, et al.* CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med* 2021;13:31. doi:10.1186/s13073-021-00835-9

15 Ioannidis NM, Rothstein JH, Pejaver V*, et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 2016;99:877-85. doi:10.1016/j.ajhg.2016.08.016

16 Adzhubei IA, Schmidt S, Peshkin L*, et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248-9. doi:10.1038/nmeth0410-248

17 Ng PC and Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812-4.

18 Melé M, Ferreira PG, Reverter F*, et al.* Human genomics. The human transcriptome across tissues and individuals. *Science* 2015;348:660-5. doi:10.1126/science.aaa0355